METAPROTEOMICS OF MARINE THIOTROPHIC SYMBIOSES

—

ANALYSIS OF HOST-MICROBE INTERACTIONS AND SYMBIONT PHYSIOLOGIES WITH OPTIMIZED METHODS

I n a u g u r a l d i s s e r t a t i o n

zur

Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Universität Greifswald

vorgelegt von

Tjorven Hinzke

Greifswald, April 2020

Dekan:	Prof. Dr. Gerald Kerth
1. Gutachter:	Prof. Dr. Thomas Schweder
2. Gutachter:	Prof. Dr. Silvia Bulgheresi
3. Gutachter:	Prof. Dr. François Lallier
Tag der Promotion:	14.09.2020

To see a World in a Grain of Sand

And a Heaven in a Wild Flower

Hold Infinity in the palm of your hand

And Eternity in an hour

William Blake, *Auguries of Innocence*

CONTENTS

Abstract

Symbiotic interactions are a key element of biological systems. One powerful strategy to gain insight into these interactions, and into biological systems in general, is the analysis of proteins expressed *in situ* using metaproteomics. In this thesis, host-microbe interactions in two mutualistic associations between chemosynthetic sulfur-oxidizing endosymbionts and marine invertebrates, the deep-sea tubeworm *Riftia pachyptila* and the shallow-water clam *Codakia orbicularis*, were studied by adapted and optimized metaproteomics methods.

The *Riftia* symbiosis, which inhabits hydrothermal vents in the deep sea, and in which the host completely depends on its symbiont for nutrition, has fascinated researchers for about four decades. Yet, the interaction mechanisms between both partners have been understudied so far. Additionally, while different aspects of the host's biology have been described, a comprehensive analysis has been lacking. Moreover, although only one symbiont 16S rRNA phylotype is present in *Riftia*, the symbiont population of the same host expresses proteins of various redundant or opposed metabolic pathways at the same time. As the symbionts also exhibit a wide variety in size and shape, symbionts of different size might have dissimilar physiological functions, which remained as of now to be elucidated. In this thesis, we addressed both, the host-symbiont interaction mechanisms, and physiological roles of symbiont subpopulations. A comprehensive *Riftia* host and symbiont protein database was generated as prerequisite for metaproteomics studies by *de novo* sequencing the host's transcriptome and combining it with existing symbiont protein databases. This database was then used for metaproteomics comparisons of symbiont-containing and symbiont-free *Riftia* tissues, to gain insights into host-symbiont interactions on the protein level. The impact of energy availability on host-symbiont interactions was studied by comparing specimens with stored sulfur (i.e., high energy availability) with specimens in which sulfur storages were depleted. We employed optimized liquid chromatography peptide separation to increase metaproteome coverage. With this analysis, we identified proteins and mechanisms likely involved in maintaining the symbiosis, under varying environmental conditions. We unraveled key interaction mechanisms, i.e.: (i) the host likely digests its symbionts using abundant digestive enzymes, and, at the same time, (ii) a considerable part of the worm's proteome is involved in creating stable internal conditions, thus maintaining the symbiont population. Furthermore, (iii) the symbionts probably employ

eukaryote-like proteins to communicate with the host. (iv) Under conditions of restricted energy availability, the host apparently increases digestion pressure on the symbiotic population to sustain itself.

Riftia symbionts of different size apparently have dissimilar metabolic roles, as revealed in this thesis. We enriched symbionts of different sizes using gradient centrifugation. These enrichments were subjected to protein extraction using a protocol optimized for the small sample amount available. Metaproteomics analysis included a gel-based workflow and evaluation of the complex dataset with machine learning techniques. Based on our metaproteomics study, we propose that *Riftia* symbionts of different cell size correspond to dissimilar physiological differentiation stages. Smaller cells are apparently engaged in cell differentiation and host interactions. Larger cells, on the other hand, seem to be more involved in synthesis of various organic compounds. Supposedly, in large symbionts endoreduplication cycles lead to polyploidy. Our results indicate that the *Riftia* symbiont employs a large part of its metabolic repertoire at the same time in the stable host environment.

The symbiont of the shallow-water clam *Codakia orbicularis*, which, like the *Riftia* symbiont, relies on reduced sulfur compounds as energy source and fixes inorganic carbon, is, unexpectedly, also able to fix atmospheric nitrogen, as shown by metaproteomic, genomic and biochemical analysis. Potentially, this benefits the host, as *Codakia* digests its symbiont and might thus supplement its diet with organic nitrogen fixed by the symbionts in addition to organic carbon in its nitrogen-poor seagrass habitat.

Zusammenfassung

Symbiotische Interaktionen sind ein Schlüsselelement biologischer Systeme. Eine leistungsfähige Möglichkeit, Erkenntnisse über diese Interaktionen und auch über biologische Systeme im Allgemeinen zu gewinnen, ist die Analyse von *in situ* exprimierten Proteinen mittels Metaproteomics. In dieser Arbeit wurden mittels adaptierter und optimierter Metaproteomics-Methoden Wirt-Mikroorganismen-Interaktionen in zwei mutualistischen Assoziationen zwischen chemosynthetischen schwefeloxidierenden Endosymbionten und marinen Invertebraten untersucht, dem Tiefsee-Röhrenwurm *Riftia pachyptila* und der Flaschwassermuschel *Codakia orbicularis*.

Die *Riftia*-Symbiose, die Tiefsee-Hydrothermalquellen bewohnt, und in welcher der Wirt komplett abhängig von seinem Symbionten als einziger Nahrungsquelle ist, fasziniert Forscher seit über vier Jahrzehnten. Dennoch wurden Interaktionsmechanismen zwischen beiden Partnern der Symbiose bislang nur wenig untersucht. Zudem wurden einzelne Aspekte der Wirtsbiologie beschrieben, eine umfassende Analyse gab es bislang jedoch nicht. Darüber hinaus besteht die *Riftia*-Symbiontenpopulation zwar nur aus einem einzelnen 16S rRNA-Phylotypen, aber die Symbionten eines Wirtstieres exprimieren Proteins verschiedener redundanter oder gegenläufiger Stoffwechselwege zur selben Zeit. Da die Symbionten zudem in Größe und Form variieren, ist es möglich, dass Symbionten unterschiedlicher Größe verschiedene physiologische Funktionen haben. Dies wurde bislang nicht untersucht.

In dieser Arbeit wurden sowohl Mechanismen der Wirt-Symbionten-Interaktion untersucht, als auch physiologische Rollen unterschiedlicher Symbionten-Subpopulationen. Als Grundlage für Metaproteom-Studien wurde eine kombinierte Wirts- und Symbionten-Proteindatenbank von *Riftia* erstellt. Dafür wurde das Wirtstranskriptom *de novo* sequenziert und mit existierenden Symbionten-Proteindatenbanken kombiniert. Die resultierende Datenbank wurde genutzt, um die Metaproteome von symbionten-haltigen *Riftia*-Wirtsgeweben mit symbiontenfreien Wirtsgeweben zu vergleichen, um Erkenntnisse über Wirt-Symbionten-Interaktionen auf Proteinebene zu gewinnen. Der Einfluss von Energieverfügbarkeit auf diese Interaktionen wurde untersucht, indem Tiere mit gespeichertem Schwefel (entsprechend hoher Energieverfügbarkeit) mit solchen verglichen wurden, in denen die Schwefelspeicher geleert waren. Um die Metaproteom-Abdeckung zu erhöhen, wurde eine optimierte Peptidseparation mittels Flüssigchromatographie genutzt. Auf diese

Weise wurden Proteine und Mechanismen identifiziert, die wahrscheinlich am Erhalt der Symbiose unter unterschiedlichen Umweltbedingungen beteiligt sind. Es wurden Schlüsselmechanismen der Wirt-Symbionten-Interaktion identifiziert: (i) Der Wirt verdaut wahrscheinlich seine Symbionten mittels abundanter Verdauungsenzyme, und gleichzeitig (ii) ist anscheinend ein großer Teil der vom Wirt exprimierten Proteine daran beteiligt, stabile interne Bedingungen zu schaffen und damit die Symbiontenpopulation zu erhalten. (iii) Die Symbionten nutzen mutmaßlich eukaryoten-artige Proteine zur Kommunikation mit dem Wirt. (iv) Bei reduzierter Energieverfügbarkeit erhöht der Wirt vermutlich die Verdauung von Symbionten, um sich selbst zu erhalten.

Weiters wurde in dieser Arbeit gezeigt, dass *Riftia*-Symbionten unterschiedlicher Größe sich wahrscheinlich in ihren physiologischen Funktionen unterscheiden. Um die Physiologie von Symbionten unterschiedlicher Größe zu untersuchen, wurden diese mittels Gradientenzentrifugation angereichert. Aus den Anreicherungen wurden Proteine unter Nutzung eines für geringe Probenmengen optimierten Protokolls extrahiert. Die Metaproteom-Analyse wurde mittels eines gelbasierten Verfahrens durchgeführt. Der resultierende komplexe Datensatz wurde mit Techniken des maschinellen Lernens ausgewertet. Nach den Ergebnissen dieser Metaproteom-Studie handelt es sich bei Symbionten unterschiedlicher Größe wahrscheinlich um physiologisch verschiedene Zelldiffenzierungs-Stadien. Kleinere Zellen teilen sich vermutlich aktiv und etablieren Interaktionen mit dem Wirt. Größere Zellen hingegen sind anscheinend primär aktiv in der Synthese verschiedener organischer Verbindungen. In größeren Symbionten scheint zudem Endoreduplikation stattzufinden, die zu Polyploidie führt. Den Ergebnissen dieser Studie zufolge nutzt der *Riftia*-Symbiont wahrscheinlich simultan einen Großteil seines metabolischen Potentials in der stabilen Wirtsumgebung.

Der Symbiont der Flachwasser-Muschel *Codakia orbicularis* nutzt wie der *Riftia*-Symbiont reduzierte Schwefelverbindungen als Energiequelle und fixiert anorganischen Kohlenstoff. Darüber hinaus ist der *Codakia*-Symbiont unerwarteterweise in der Lage, atmosphärischen Stickstoff zu fixieren. Dies wurde mittels metaproteomischer, genomischer und biochemischer Analysen gezeigt. Die Stickstoff-Fixierung durch den Symbionten könnte ein Vorteil für den Wirt der in einer stickstoffarmen Umgebung lebenden *Codakia*-Symbiose sein, da der *C. orbicularis*-Wirt seine Symbionten verdaut und somit zusätzlich zu organischem Kohlenstoff auch von den Symbionten fixierten organischen Stickstoff erhalten könnte.

1 Introduction

1.1. Symbiosis as key element of biological systems

Eukaryote-microbe interactions are omnipresent (McFall-Ngai, 2008; Gilbert et al., 2012; McFall-Ngai et al., 2013; Douglas, 2014a; Bang et al., 2018). The significance of such interactions, or symbioses, for higher organisms is exemplified in the endosymbiont theory for the origin of eukaryotic cell organelles from symbiotic organisms, as put forward in Lynn Margulis' seminal manuscript (Sagan, 1967). Beneficial symbioses (i.e., mutualistic associations) with prokaryotes can allow eukaryotes to cope with otherwise uninhabitable conditions, and enable them to utilize resources which would be inaccessible without the respective microbes. In this thesis, the term "symbiosis" will be used in this meaning *sensu strictu*, i.e., for mutualistic associations.

Still during the 1980s, intricate animal-host symbioses have been considered a rare phenomenon (McFall-Ngai, 2014). The discovery of chemosynthetic symbioses, which were first described in the early 1980s for the hydrothermal vent tubeworm *Riftia pachyptila* (here *Riftia*, Figure 1), was therefore truly an epiphany (Cavanaugh et al., 1981; Felbeck, 1981). Here, the symbiont uses inorganic reduced sulfur compounds to generate energy (i.e., the symbiont is thiotrophic) and fixes carbon dioxide using two different pathways (Cavanaugh et al., 1981; Markert et al., 2007).

Figure 1: Schematic depiction of *Riftia pachyptila* tubeworms.

After that, chemosynthetic associations have also been found in many other organisms, including more easily accessible examples like shallow-water invertebrates (e.g., *Codakia* clams, Berg and Alatalo, 1984; *Olavius* oligochaetes, Woyke et al., 2006).

From those early years onwards, research on eukaryote-microbe symbioses has increasingly shown the immense impact of these associations: In insect-bacteria symbioses, the symbionts nurture the

host and protect it against pathogens and parasites (Douglas, 2014b). The squid-vibrio symbiosis, where each partner can be cultured separately, has enabled insight into initial host colonization by bacteria, as well as into symbiosis development and maintenance (McFall-Ngai, 2014). Macroalgae can form nutritional endo- and episymbioses with bacteria with various degrees of complexity and ensuing functional differences (Hollants et al., 2013). Our own associated microbes have vast influence on our health, which is a topic of intense research (Cho and Blaser, 2012).

We are now addressing major open questions, e.g.: Which molecular mechanisms form the basis of host-symbiont interactions? Which evolutionary impact do symbionts have on their host? How do these associations adapt to extreme habitats? How do environmental changes influence symbioses? (McFall-Ngai et al., 2013; Bang et al., 2018). The advance of -omics technologies, which allow for the generation of comprehensive data of biological systems without the need to culture them, enables us to elucidate these fundamental aspects of symbiotic interactions in various habitats.

1.2. METAPROTEOMICS – A POWERFUL TOOL TO STUDY BIOLOGICAL COMMUNITIES

The term "metaproteomics", originally defined as the study of the proteins expressed by an environmental microbial community at a certain time (Wilmes and Bond, 2004), is not unambiguous. While it is often used as a synonym for "community proteomics" or "environmental proteomics", some have suggested to reserve the term "metaproteomics" for highly complex systems, in which assigning of proteins to species is often not possible (VerBerkmoes et al., 2009; Schneider and Riedel, 2010). In this thesis, we will use the term "metaproteomics" in the broad meaning, i.e., to refer to protein identification and quantification in microbial communities (Kleiner, 2019), and also include the microbiota's host in it. Metaproteomics offers an ideal method to understand the functions of microbial symbionts and their interactions with their eukaryote host: Metaproteomic tools can be applied to study communities *in situ*, without the need for cultivation – which is often not possible, especially not separately for all partners of a symbiosis. As expressed proteins determine cellular functions, insights into organismal activities and interactions can be gained using metaproteomics.

While not even two decades ago a metaproteomics study with 2D polyacrylamide gel electrophoresis (PAGE) identified a total of three proteins (Wilmes and Bond, 2004), rapid technological and methodological developments now allow to identify and quantify hundreds to thousands of proteins, e.g., in lichen (Schneider et al., 2011), plankton (Sowell et al., 2011), marine invertebrate symbioses (Markert et al., 2011;

Wippler et al., 2016; Ponnudurai et al., 2017), hydrothermal vent communities (Pjevac et al., 2018) and the human gut (Xiong et al., 2015b). Metaproteomics applications are increasingly gaining momentum, as exemplified by the number of publications which include the term "metaproteo*" in the title or abstract on PubMed (Figure 2). Ongoing developments are expected to broaden the applicability and significance of metaproteomics methods even more (see Kleiner, 2019 for an overview).

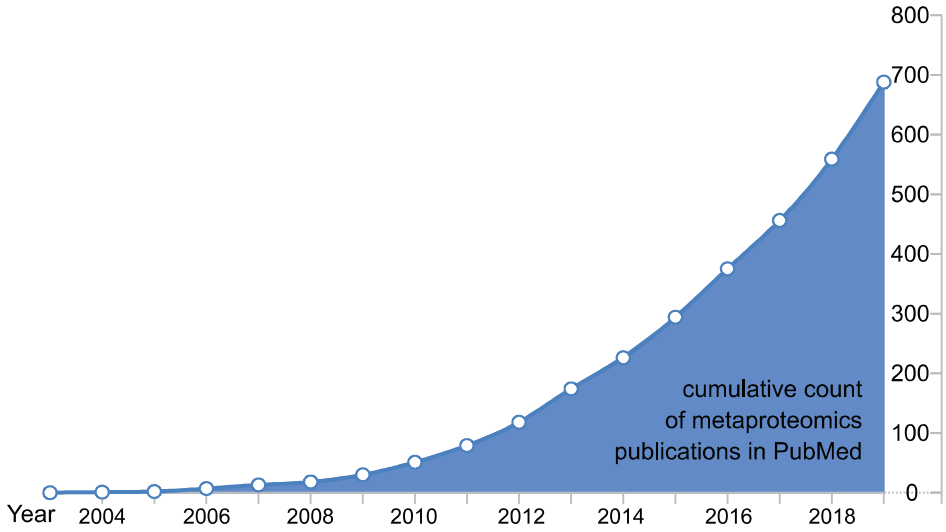

Figure 2: Cumulative count of publications in PubMed containing the term "metaproteo*" in the title or abstract. Note that this figure includes only studies in which the title or abstract contains the term "metaproteo*". Other metaproteomics studies, which use metaproteomics methods, but do not mention the term in the title or abstract, are not included. Also excluded are metaproteomics studies which are referred to, e.g., as "community proteomics" or "environmental proteomics".

Many applications of metaproteomics rely on protein digestion into peptides, followed by peptide separation by liquid chromatography (LC) and mass spectrometry (MS) based identification via generation of fragment ions (shotgun or bottom-up approach, Zhang et al., 2013). For protein identification, raw mass spectrometry files, containing the mass spectra from peptide fragmentation, can be searched against a database containing the proteins which could theoretically be present in the sample and from which theoretical mass spectra are generated *in silico*. Various search algorithms have been developed to tackle this highly complex challenge (Verheggen et al., 2017).

Metaproteomics techniques enable not only to identify proteins, but also to quantify them. In bottom-up (meta)proteomics, proteins can be quantified using stable isotope label-based or label-free approaches. Quantification can be absolute or relative, i.e., estimate the concentrations of proteins present in the same sample or the amount as compared to other samples or other proteins in the same sample (Blein-Nicolas and Zivy,

2016; Lindemann et al., 2017; Ankney et al., 2018). Label-free quantification has the advantage of being less time-consuming and less expensive as compared to label-based approaches (Ankney et al., 2018). Label-free quantification can be based on counting fragment ion spectra assigned to a protein (spectral counting), or on the intensity of peptide precursor ions of a protein as calculated by the area under the curve or the peak height for the respective peptide (Nahnsen et al., 2013; Blein-Nicolas and Zivy, 2016; Ankney et al., 2018). Intensity-based quantification is far more computationally intensive than spectral counting and necessitates a trade-off between identification and quantification, both of which need MS scanning time. This is especially problematic for metaproteomics, due to the comparatively high sample complexity. Spectral counting, on the other hand, is less sensitive with regard to small abundance changes (Nahnsen et al., 2013; Blein-Nicolas and Zivy, 2016; Ankney et al., 2018). However, the results of both approaches are comparably related to mRNA gene expression data (Ning et al., 2012).
Obtaining high and reliable numbers of identified and quantified proteins during metaproteomics applications faces several challenges, concerning: (i) sample complexity and diversity, (ii) protein extraction (iii) generation of a database that fits the sample in question and (iv) assigning peptides to proteins, especially if proteins are highly similar, known as the protein inference problem (Nesvizhskii and Aebersold, 2005; VerBerkmoes et al., 2009; Muth et al., 2013; Wang et al., 2014; Wilmes et al., 2015; Xiong et al., 2015a; Kleiner et al., 2017; Timmins-Schiffman et al., 2017; Verheggen et al., 2017). Additionally, stringent experimental design and statistical evaluation of generated data are crucial to obtain meaningful biological insights.

1.3. SCOPE OF THIS THESIS

This thesis focuses on metaproteomics analyses of marine chemosynthetic symbioses. It addresses current key methodological challenges presented by metaproteomics applications and uses state-of-the-art metaproteomics methods to unravel host-symbiont interactions and physiological capabilities of chemosynthetic symbionts in the deep-sea tube worm *Riftia pachyptila* and the shallow-water clam *Codakia orbicularis*.

(i) Method development for metaproteomics studies included testing and adaptation of gradient centrifugation methods to enrich bacterial symbiont subpopulations of different sizes. Furthermore, various 1D and 2D LC separation methods were tested and compared for their applicability for metaproteomics studies. With regard to metaproteomics data evaluation,

several statistical analysis approaches were applied to different biological questions, and a method to calculate stable carbon isotope ratios from metaproteomics data was evaluated.

(ii) Marine thiotrophic symbioses were studied using adapted and optimized metaproteomics methods. Special attention was paid to host-symbiont interactions in the *Riftia* symbiosis, to throw light on the molecular mechanisms which form the basis of this intriguing system. In addition, the effect of varying environmental conditions on these interactions was analyzed. Moreover, physiological differences of *Riftia* symbiont subpopulations were in the focus of this thesis. Additionally, the physiology of the symbiont of the lucinid clam *Codakia orbicularis* was examined, with special focus on its ability to fix inorganic nitrogen.

2 METHOD DEVELOPMENT AND ADAPTATION FOR METAPROTEOMICS

Metaproteomics analyses have paved the way for comprehensive analysis of the functions of biological systems, including symbiotic associations between pro- and eukaryotes. For these analyses, a sufficiently high metaproteome coverage is crucial. Achieving such a coverage depends on addressing several key challenges of metaproteomics, including dealing with the high sample diversity and complexity, optimization of protein extraction, database generation, protein identification and data analysis (Nesvizhskii and Aebersold, 2005; VerBerkmoes et al., 2009; Muth et al., 2013; Wang et al., 2014; Wilmes et al., 2015; Xiong et al., 2015a; Kleiner et al., 2017; Timmins-Schiffman et al., 2017; Verheggen et al., 2017). In this thesis, we addressed the challenges of sample complexity (Hinzke et al., 2019b – Publication II, Hinzke et al., 2018 – Publication VI), as well as data analysis, including statistical evaluation (Sokolov et al., 2019 – Publication III, Kleiner et al., 2018 – Publication IV, König et al., 2016 – Publication V). We employed adapted methods for our analysis of the *Riftia* symbiosis (Hinzke et al., 2019a – Publication I, Hinzke et al., submitted – Publication VII).

2.1 DECREASING (META)PROTEOMICS SAMPLE COMPLEXITY

Increasing protein identification necessitates reduction of sample complexity. Sample complexity can already be reduced at the level of the biological sample itself, or at the level of extracted proteins or peptide mixtures.

One possibility to decrease sample complexity at the level of the biological sample is to enrich the components of interest. Centrifugation-based methods allow to separate host and symbiont populations in symbiotic associations and to enrich different (symbiont) cell types (Publication VI). We thoroughly tested different methods for centrifugation-based cell enrichment, and presented comprehensive protocols for i) differential pelleting, which separates particles based on size and to some degree on density, and ii) density gradient centrifugation methods. The latter includes rate-zonal density gradient centrifugation, where particles with similar densities are separated in a size- and shape-dependent fashion, and isopycnic density gradient centrifugation, where particles are separated depending on their density. In the scope of this thesis, we employed rate-zonal

density gradient centrifugation to enrich *Riftia* symbionts of different sizes (Publication VII).

As another means of reducing sample complexity, protein or peptide mixtures can be separated before MS analysis by chromatographic methods. We used a fully characterized mock community to test over 20 different 1D and 2D LC peptide separation methods (Publication II). For 2D methods, peptide separation was on-line, i.e., peptides were injected into the mass spectrometer directly after separation. We also compared these methods to gel-based protein separation (GeLC). We used very long analytical columns (50 and 75 cm) and long LC gradients of up to 12 h.

These tests resulted in a comprehensive comparison of the relative capabilities and drawbacks of the different methods and in recommendations for their use for different research questions (Figure 3).

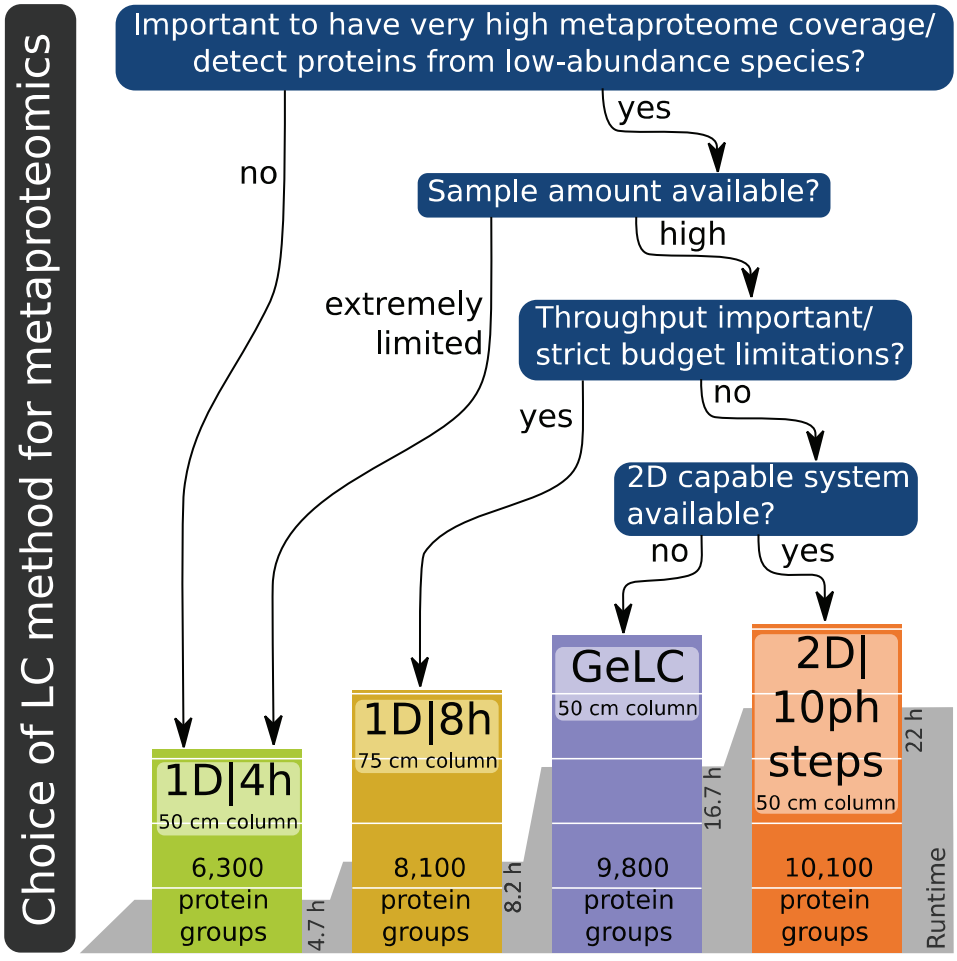

Figure 3: Recommendations for the selection of liquid chromatography (LC) methods for metaproteomics experiments. Figure adapted from Publication II.

While shorter gradients in a 1D approach are faster and well suited for systems with lower complexity, longer methods using 2D separation lead to identification also of low-abundance proteins, while at the same time being more expensive. The performance of the GeLC method in terms of number of identified proteins is comparable to that of 2D LC separation, while sample preparation is far less time consuming for 2D LC separation. Based on this method comparison, we selected the most appropriate method for our analysis of host-symbiont interactions in the *Riftia* symbiosis (Publication I).

2.2 ANALYSIS OF METAPROTEOMICS DATA – STATISTICAL EVALUATION AND STABLE CARBON ISOTOPE RATIO CALCULATION

Metaproteomics analysis generates huge amounts of data which have to be analyzed and interpreted. Statistical evaluation of measured protein abundances is often crucial to draw meaningful conclusions from comparative (meta)proteomics experiments. Another valuable information, which up to now was not straightforward to obtain for metaproteomics samples, is the stable carbon isotope ratio ($\delta^{13}C$) of different taxa in the sample.

For statistical evaluation, data needs to be normalized, and subsequently subjected to an appropriate statistical test to find proteins with significant abundance changes across samples or conditions. Data normalization is necessary to be able to compare values across samples and to adjust the data distribution to the statistical analysis. A widely used approach is the calculation of normalized spectral abundance factors (NSAF, Zybailov et al., 2006), where spectral count data are normalized by the sum of spectral counts per sample and by protein length. Another possibility for normalization, which has first been introduced for transcriptomics data, but can also be used for proteomics, is referred to as relative log expression (RLE) normalization (https://www.rdocumentation.org/packages/edgeR/versions/3.14.0/topics/calcNormFactors): here, size factors can be derived from calculating a pseudo-reference sample, e.g., the geometric mean across samples, and then calculating the median of the ratios of the samples to the reference sample (Anders and Huber, 2010). This method appeared to be more robust than NSAF calculation when comparing samples of very dissimilar composition and was therefore used for our analysis of the *Riftia* symbiosis (Publications I, VII) and of oyster mitochondrial proteomes in reaction to hypoxia-reoxygenation stress (Publication III).

After normalization, data can be sorted together by clustering to find common patterns. Clustering techniques can therefore be used to structure metaproteomics data by finding

proteins with similar abundance trends. One such clustering technique is implemented in the Short Time-series Expression Miner (STEM): Here, small ordered gene expression datasets, e.g., time series, are clustered. STEM pre-defines model expression profiles, which represent the expression patterns possible for the analyzed dataset. Genes are then assigned to the best-fitting model expression profile (Ernst et al., 2005; Ernst and Bar-Joseph, 2006). We used this approach for our analysis of protein abundance patterns across enrichments of *Riftia* symbionts with different cell size (Publication VII).

To find proteins with statistically significant abundance changes between samples or conditions, a classical t-test is widely applied. However, when applied to NSAF data, it can result in large numbers of false-positive results in samples of very different composition (e.g., if many more proteins have been detected in one of the samples than in the other), or false-negatives, if differences between samples are small, but consistent (personal observation, unpublished results).

Another promising tool for statistical analysis of metaproteomics data are random forests (Figure 4). This machine-learning technique uses decision trees to sort together similar data. Random forests can be used to find variables (proteins) which are important to distinguish groups, i.e., which change consistently in abundance between groups (Degenhardt et al., 2017). Random forests can detect differentially abundant proteins even if differences are small, as long as they are consistent. Random forests are robust against noise, hard to overfit and suitable for datasets with many more variables than samples. Additionally, measures of variable importance can be directly generated (Díaz-Uriarte and de Andrés, 2006). These properties make random forests well suited for metaproteomics data.

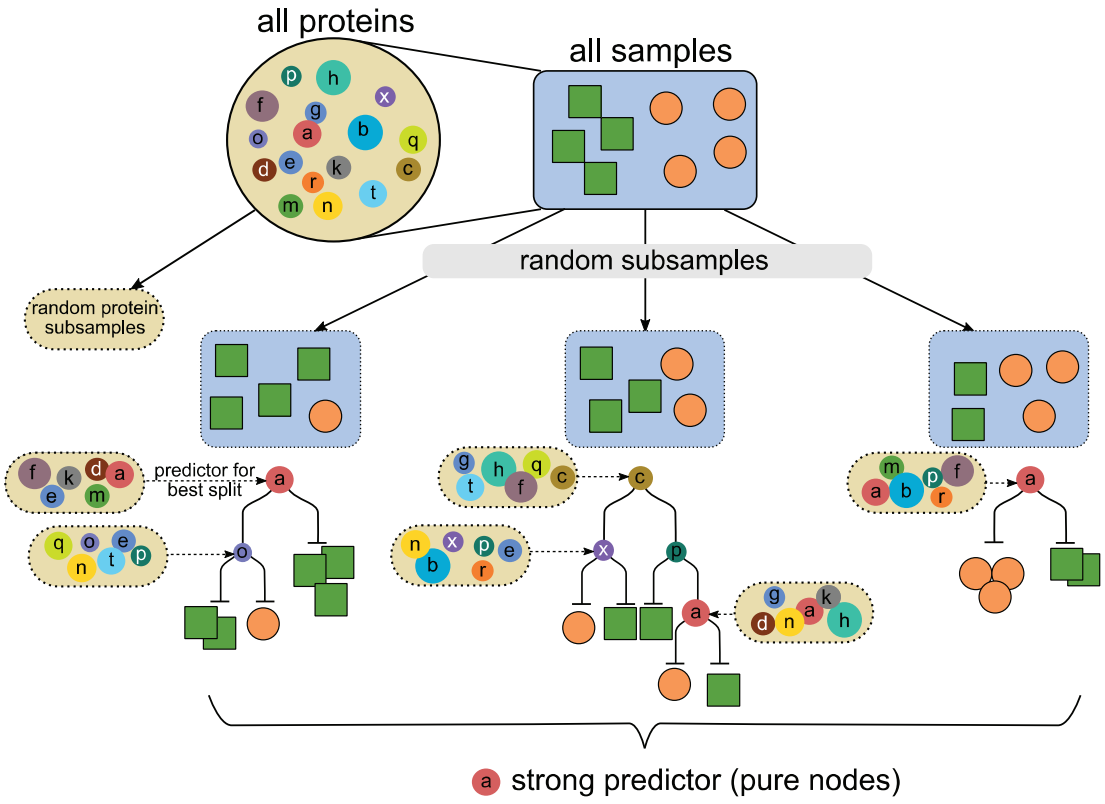

Figure 4: Principle of random forest analysis, shown for (meta)proteomics data. Replicate samples from different sample types (e.g., two different incubation conditions) are indicated by green squares (=sample type 1), and by orange circles (=sample type 2). Proteins detected in any sample are shown as circles with small letters (different colors are for visual clarity). During random forest analysis, random subsamples are taken from all available samples. Then, a random subsample of all detected proteins is chosen and the protein whose abundance best separates the subsample into sample types is found (=predictor for best split). The procedure is repeated until nodes are pure, i.e., contain only one sample type. In the example above, protein "a" is a good predictor for the sample type, as splitting by the abundance of protein "a" results in pure nodes. This would mean that "a" is a protein with consistent abundance changes between the sample types.

Random forest analysis was used to detect differentially abundant proteins in oyster (*Crassostrea gigas*) mitochondrial proteomes and phosphoproteomes during hypoxia and reoxygenation stress (Publication III). We gained insights into changes in the proteome, which include proteins involved in oxidative phosphorylation, β-oxidation, ion homeostasis, mitochondrial quality control, redox balance, as well as mitochondrial substrate uptake. Reversible protein phosphorylation seems to have a regulatory function during recovery of *C. gigas* after hypoxia treatment. As at the same time mitochondrial function appears to be largely preserved during treatment, these changes could lead to mitochondrial resilience against hypoxia-reoxygenation stress in *C. gigas*. For the comparative analysis of physiological functions of *Riftia* symbiont cells with differing sizes, we used samples enriched in these cell sizes (Publication VII; see below).

Results from random forests and STEM clustering enabled us to define functional differences between symbiont subpopulations in this challenging dataset.

Stable carbon isotope ratios ($\delta^{13}C$) can be used to infer diets and carbon assimilation pathways of organisms (DeNiro and Epstein, 1978; Pearson, 2010). While previous methods for $\delta^{13}C$ analysis were not suitable to metaproteomics samples due to the long time needed for analysis or their limited ability to distinguish between taxa, MS raw data can now be used directly to calculate $\delta^{13}C$ values (Publication IV). This method was termed "direct protein-SIF" (SIF: stable isotope fingerprints, i.e. stable isotope ratios) and a software package, Calis-p (The CALgary approach to ISotopes in Proteomics), was made available to the public. We used this for our *Riftia* data to calculate symbiont and host $\delta^{13}C$ values (Publication I) and thus to gain insights into nutrient transfer modes from symbiont to host. For the Calis-p analysis, samples are measured together with a reference sample, the $\delta^{13}C$ value of which is known, e.g., human hair. Peptide spectrum matches (PSMs) and raw MS data are used as input for Calis-p. Calis-p extracts isotopic peaks of PSMs and sums them up. SIFs are calculated by fitting theoretical isotope peak distributions for a range of $\delta^{13}C$ values to the experimental distribution. After Calis-p computation, SIFs can be averaged per taxon and the $\delta^{13}C$ value of the reference material can be used to correct for isotope fractionation during MS analysis to obtain the final $\delta^{13}C$ values per taxon.

3 Host-microbe interactions and symbiont physiological capabilities in marine thiotrophic symbioses

Riftia pachyptila, the conspicuous giant deep sea tubeworm, was the first organism in which a symbiosis with chemosynthetic bacteria was discovered (Cavanaugh et al., 1981; Felbeck, 1981). This system continued to fascinate researchers in the following decades (e.g., Felbeck and Turner, 1995; Girguis et al., 2000; Markert et al., 2007, 2011; Robidart et al., 2008, 2011; Gardebrecht et al., 2012; Scott et al., 2012; see Stewart and Cavanaugh, 2005 for a review). The tubeworm lacks a digestive system and therefore completely depends on its bacterial endosymbiotic partner, named "*Candidatus* Endoriftia persephone" (here Endoriftia), which is housed in the so-called trophosome, for nutrition (Cavanaugh et al., 1981; Hand, 1987; Distel et al., 1988; Robidart et al., 2008). The symbionts, on the other hand, do apparently have a free-living stage (Harmer et al., 2008) and infect each worm generation anew (Nussbaumer et al., 2006). The symbionts belong to one 16S rRNA phylotype of thiotrophic gammaproteobacteria (Felbeck, 1981; Robidart et al., 2008; Polzin et al., 2019). Their oxidation of reduced sulfur compounds to sulfate drives the high growth rates of this association: The worm grows very fast (up to >85 cm in tube length per year (Lutz et al., 1994)) and can reach a length of up to 1.5 m (Jones, 1981); its cell division rates have been compared to those of tumors and healing wounds (Pflugfelder et al., 2009). Inside the worms, the symbionts exhibit distinct morphological differences, encompassing small, dividing rods and cocci and larger, but non-dividing cocci. These different cell types have been speculated to be part of a common cell cycle (Bright and Sorgo, 2003). Modern -omics methods have made the metabolism of the uncultivable symbiont easier accessible, using its metagenome as basis for metagenomics and metaproteomics analyses (Markert et al., 2007, 2011; Robidart et al., 2008, 2011; Gardebrecht et al., 2012). Research of the host's metabolism and especially of host-symbiont interactions on the protein level, on the other hand, has been hampered by the lack of sequence data for the host. Therefore, in this thesis a comprehensive protein database of both, host and symbiont proteins, was generated. For that, the host's transcriptome was sequenced *de novo* and combined with the symbiont's proteome sequences (Publication I). This database was used as foundation for metaproteomics studies of host-symbiont interactions (Publication I), as well as for differential analysis of physiological roles of *Riftia* symbionts differing in cell size (Publication VII). Metaproteomes were generated and analyzed based on adapted and

optimized protocols for enriching cell subpopulations of differing size (Publication VI), as well as for LC-MS/MS analysis (Publication II), and for statistical evaluation (Publication III).
Thiotrophic symbioses exist not only in the deep sea, where chemoautotrophy forms the basis of food chains, but also in shallow-water environments in the presence of reduced sulfur compounds. Here, metaproteomics analyses, including absolute protein quantification, underpinned the unexpected finding that the thiotrophic *Codakia orbicularis* symbiont is capable to fix atmospheric nitrogen (Publication V).

3.1 HOST-MICROBE INTERACTIONS UNRAVELED IN THE DEEP-SEA *RIFTIA* SYMBIOSIS

The *Riftia* tubeworm host provides the symbionts with all compounds necessary for chemosynthesis, and the symbionts, in turn, efficiently supply the host with organic matter for fast growth. This intricate interdependency, together with the low complexity (only two partners) and high efficiency of the association, make the *Riftia* symbiosis an intriguing system to study eukaryote-microbe interactions in detail.
We compared metaproteomes of symbiont-free and symbiont-containing *Riftia* tissues to unravel host-symbiont interactions on the protein level. To assess the influence of energy availability on these interactions, we analyzed both sulfur-rich (energy-rich) and sulfur-depleted specimens (Publication I). As a prerequisite for these analyses, we generated a combined host and symbiont protein database, based on existing symbiont protein sequences and a *de novo* generated host transcriptome. Techniques for mRNA extraction were thoroughly adapted and optimized. Tissue disruption via bead-beating in TRIzol (Thermo Fisher Scientific) and subsequent homogenization with QIAShredder (Qiagen) columns proved to be critical to obtain high-quality mRNA samples from the different *Riftia* tissue types (plume, body wall, vestimentum, trophosome; see Figure 1). For metaproteome analysis, we used a high-end mass spectrometer, the Orbitrap Q Exactive (Thermo Fisher Scientific), and a 4 h 1D LC gradient for optimized peptide separation (Publication II). We used the metaproteomics data to calculate $\delta^{13}C$ values as described in Publication IV (see above).
Based on this fully replicated dataset and a stringent experimental design, we identified proteins with significant abundance differences between tissue types. We used these data to characterize key interactions in the *Riftia* symbiosis, which can be grouped in three main complexes: (i) transfer of metabolites between host and symbiont, (ii) *Riftia* host measures to maintain its symbiont population and (iii) symbiont strategies to thrive inside the host. Our comparison of *Riftia* specimens from different energy regimes allowed us to identify sulfur availability-dependent changes of these symbiotic interactions. The main *Riftia* host-symbiont interactions are summarized in Figure 5.

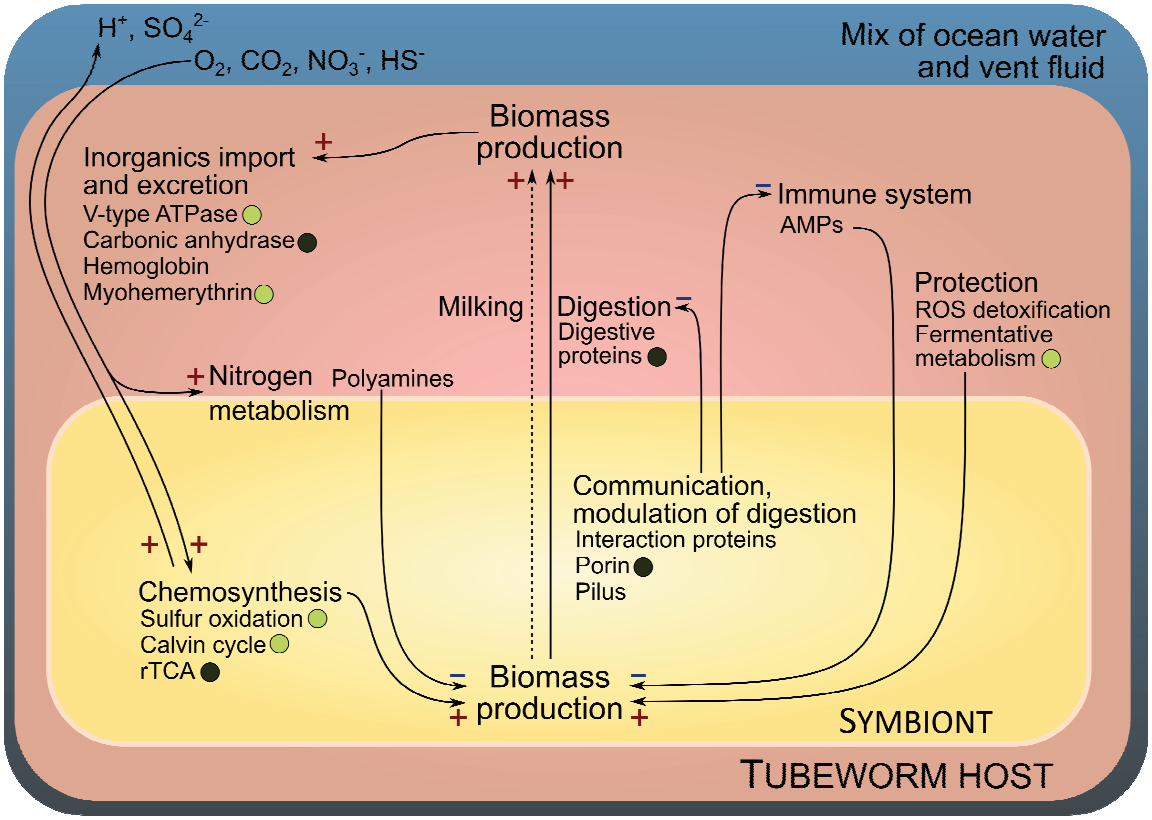

Figure 5: Schematic overview of the main interactions in the *Riftia pachyptila* symbiosis. Interactions which are likely stimulating are marked with a "plus" sign, whereas potentially inhibiting interactions are marked with a "minus" sign. When present, circles indicate that the respective proteins are more abundant in sulfur-rich (energy-rich) *Riftia* specimens (light circles) or in sulfur-depleted (energy-depleted) specimens (dark circles). The dashed arrow refers to the putative transfer of small organic molecules, also referred to as "milking". Figure adapted from Publication I. AMPs: antimicrobial peptides, ROS: reactive oxygen species, rTCA: reverse tricaboxylic acid cycle.

Our metaproteomics comparison revealed that *Riftia* likely digests its symbionts to gain organic matter, whereas the active secretion of organic compounds by the symbionts ("milking") apparently plays a minor role – if any – in the nutrition of the *Riftia* host. Digestion apparently increases during sulfur starvation, possibly because sulfur-starved symbionts possess a lower nutritional value for the host. Our results suggest that symbionts might be digested in a process resembling endosome maturation, and point to a role of the protease cathepsin B in cell death in *Riftia* trophosome. This is in contrast to previous microscopy-based studies, which suggested an autophagic or apoptotic cell death in *Riftia* trophosomes (Pflugfelder et al., 2009).

Riftia not only uses its symbionts for its own nutrition, but also provides ideal conditions for their chemosynthetic activities. A large part of the identified *Riftia* host proteins seems to be involved in transport of inorganic substrates for chemosynthesis to the symbionts. This indicates a high metabolic effort of the host to sustain its symbionts. Interestingly, the host

apparently possesses different O_2-transporting proteins: Hemoglobins make up about one third of all host trophosome and one fifth of all host plume proteins according to our analysis. They transport both O_2 and sulfide to the symbionts (Bailly and Vinogradov, 2005; Flores et al., 2005; Hourdez and Weber, 2005). In addition, *Riftia* also possesses myohemerythrin as another O_2-transporting protein. *Riftia* myohemerythrin might, in addition to hemoglobin, be involved in O_2 transfer to the symbionts and additionally have other, tissue-specific functions, as we detected different myohemerythrin isoforms in different *Riftia* tissues. Additionally, the host might generate hypoxic conditions in the trophosome, which would also benefit its microaerophilic symbionts. All of these mechanisms likely respond to the availability of the respective compounds in the environment, buffering the symbionts against changes in environmental conditions.

While the *Riftia* host thus gains nutrition and maintains the symbiont population, the symbionts themselves likely interact with the host employing eukaryote-like proteins to ensure their survival inside the host cells. These eukaryote-like proteins might especially interact with the host's cytoskeleton and thereby inhibit endosome maturation. Various proteins similar to that of the *Riftia* symbiont have also been found in pathogens, suggesting that the *Riftia* symbiont uses a strategy resembling that of pathogens to establish an intracellular lifestyle in the host.

3.2 PHYSIOLOGICAL ROLES OF *RIFTIA* SYMBIONTS SUBPOPULATIONS

The *Riftia* trophosome has an intricate structure: The highly vascularized tissue consists of very small lobules (approximately 200-500 µm in diameter), each of which features a central efferent blood vessel, which is surrounded by several layers of symbiont-containing bacteriocytes. The peripheral lobule layer consists of symbiont-free host cells (Figure 6; Bright and Sorgo, 2003).

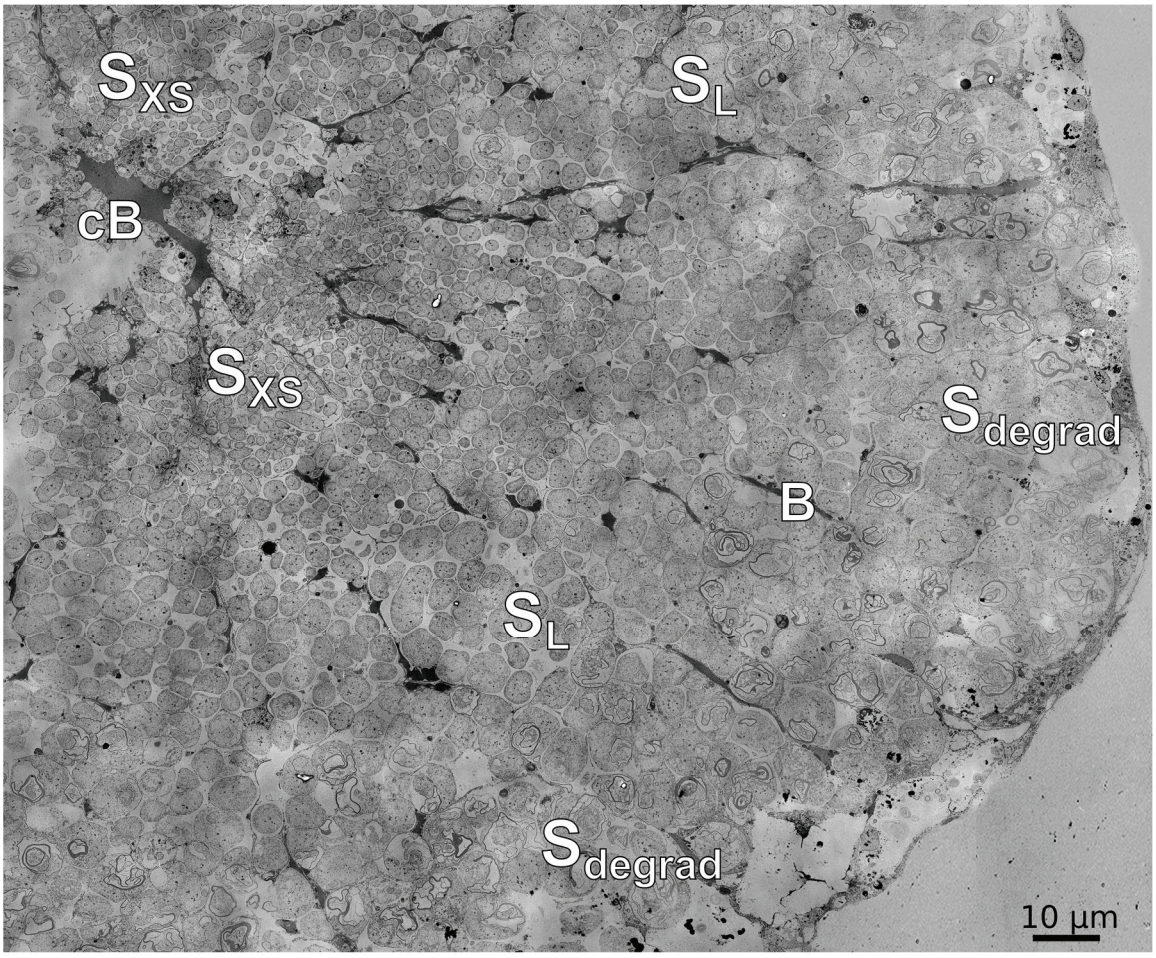

Figure 6: Electron micrograph of a *Riftia* trophosome lobule cross section. Bacteriocytes with very small (S_{XS}) symbionts are located near to the central blood vessel (cB). Symbionts towards the periphery are larger (S_L: large symbionts) and are digested, together with their bacteriocytes, in the outermost symbiont-containing lobule zone (S_{degrad}). The bacteriocyte zone also contains smaller blood vessels (B). The image was assembled from individual transmission electron micrographs. Brightness and contrast were adapted for visual clarity. Image adapted from Publication VII.

Despite belonging to the same 16S rRNA phylotype (Polzin et al., 2019), the symbiont population of the same host animal expresses proteins of redundant or even opposing metabolic pathways simultaneously, e.g., proteins of the reverse tricarboxylic acid (rTCA) cycle and the Calvin cycle for CO_2 fixation, or proteins involved in glycogen generation and degradation (Markert et al., 2007, 2011; Gardebrecht et al., 2012, Publication I). As previous metaproteomics analyses of the symbiont's metabolism targeted the population as a whole (Markert et al., 2007, 2011; Gardebrecht et al., 2012), it was unclear whether these pathways were active in the same cell or rather in metabolically different subpopulations.

Riftia symbionts are not only metabolically, but also morphologically diverse: Endoriftia in central host cells are small and mainly rod-shaped, whereas towards the lobule periphery, the symbiont cells become larger and coccoid (Bright and Sorgo, 2003, Figure 6). In addition to

that, microscopy-based studies also showed differences between symbionts of different size in cell division and carbon incooporation rates, as well as in glycogen and sulfur storage (Bright et al., 2000; Sorgo et al., 2002; Bright and Sorgo, 2003; Pflugfelder et al., 2005). We therefore hypothesized that symbionts of different size have dissimilar metabolic profiles.

To test our hypothesis, we enriched Endoriftia of different sizes to elucidate their metabolic function in the *Riftia* symbiosis (Publication VII). For enrichment, we employed rate-zonal density gradient centrifugation, i.e., we separated the symbionts based on their sizes (Publication VI). We subjected the resulting fractions to catalyzed reporter deposition-fluorescence in situ hybridization (CARD-FISH) to verify the enrichments. Cell counting and measurement was done using a custom Fiji (Schindelin et al., 2012) macro. For further metaproteomics analysis, we used fractions enriched in symbionts of four size ranges, representing a quartile split of the counted and measured symbiont population in un-enriched homogenate (i.e., 25% of all measured symbiont cells were assigned to each of the ranges defined). These ranges were very small (≥2–<3.9 µm), small (≥3.9–<5.3 µm), medium (≥5.3–<6.8 µm) and large (≥6.8–20 µm) cells. As available sample amounts were very low, we extensively tested protein extraction as well as protein and peptide fractionation techniques, resulting in a publicly available protocol for optimized protein yield from small biological samples (Hinzke and Markert, 2017). We then subjected enriched fractions representing the four different symbiont size classes to individual metaproteomics analysis using our optimized protein extraction method and GeLC with subsequent MS analysis. For statistical evaluation of the resulting dataset, we employed random forests and abundance profile clustering to find proteins with significant abundance differences across different size classes (see 2.2).

Our analysis revealed that, indeed, symbionts of different size apparently are functionally dissimilar. We propose a physiological differentiation process of Endoriftia during symbiosis: While smaller cells supposedly represent the actively dividing subpopulation, larger cells seem to be mainly active in CO_2 fixation and organic material production (Figure 7).

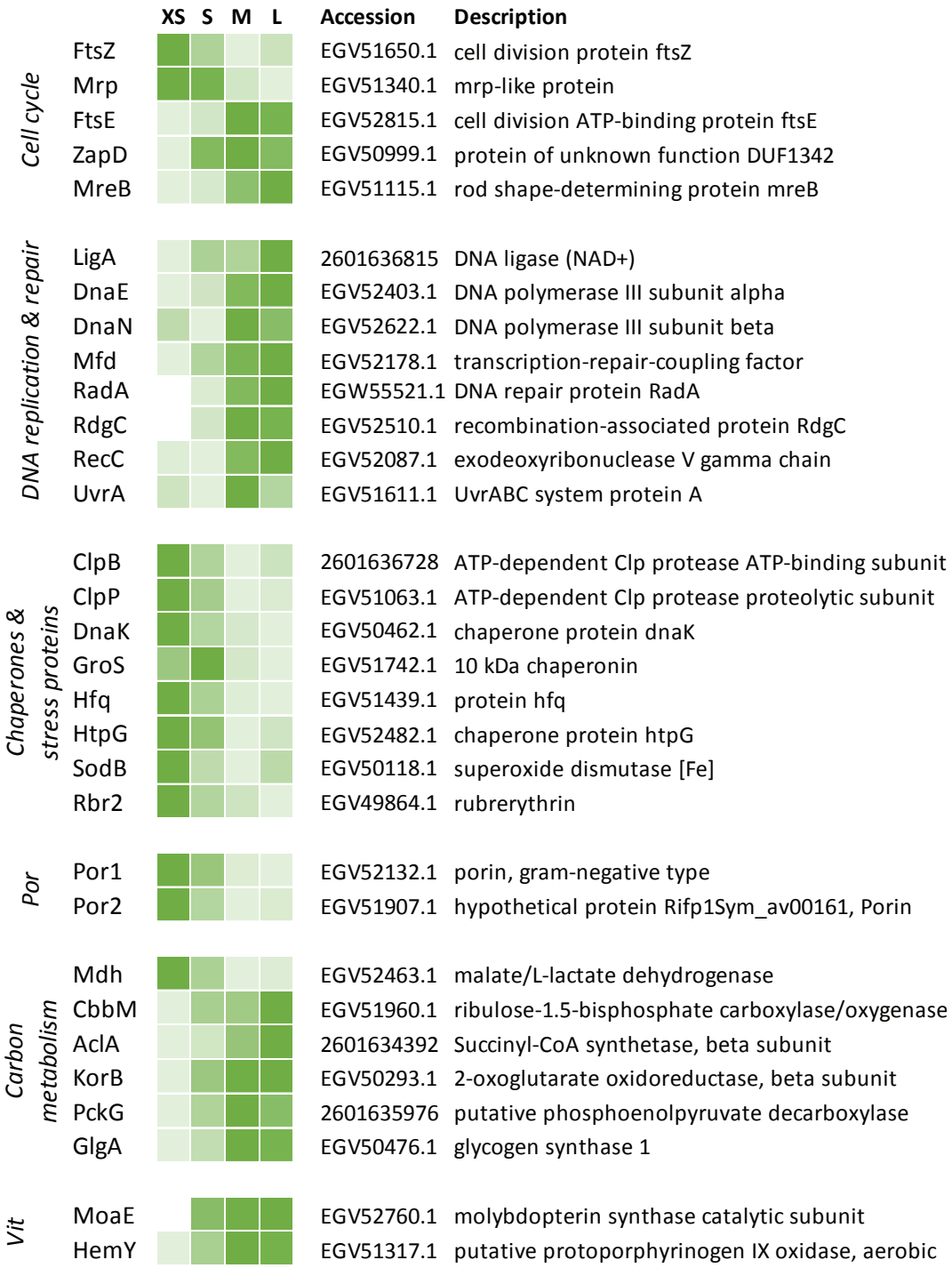

Figure 7: Heatmap showing protein abundance trends across fractions enriched in very small (XS) to large (L) *Riftia* symbiont cells. Shown are selected proteins for symbionts enriched from sulfur-rich *Riftia* specimens with significantly different abundances (consistent differences across replicates according to random forest or abundance profile clustering analysis). Light green indicates the lowest protein abundance across the four fractions, dark green indicates the highest protein abundance. White indicates that the protein was not detected in the respective fraction. Please note that comparisons are possible for the same protein across fractions, but not between proteins. DUF: domain of unknown function, Vit: vitamin and cofactor metabolism. Figure adapted from Publication VII.

In enrichments of smaller *Riftia* symbiont cells, Ftsz, as well as chaperones and reactive oxygen species (ROS) scavengers, and porins were more abundant than in enrichments of larger cells. As FtsZ is a key protein of bacterial cell division, its comparatively higher abundance in enrichments of smaller cells indicates that in these smaller cells, cell division has a higher priority than in larger cells. This is in accordance with the microscopy-based theory of Bright and Sorgo (2003), stating that small cells in the trophosome lobule center function as stem cells for the Endoriftia population, and that larger cells towards the lobule periphery cease to divide. The higher abundance of chaperones and ROS scavengers in enrichments of smaller Endoriftia suggests that these smaller Endoriftia experience higher stress levels than larger symbionts. Such putatively higher stress levels in small symbionts might result, e.g., from ROS produced by the host. This would be similar to other animal and plant cells, which produce ROS as defense against bacteria (Heath, 2000; Lynch and Kuramitsu, 2000; D'Haeze and Holsters, 2004). As smaller symbionts are relatively loosely packed inside the bacteriocytes and have a higher surface-to-volume ratio than larger symbiont cells, this could make them more prone to host-produced ROS. A comparatively higher relevance of host-symbiont interactions in smaller Endoriftia cells was also suggested by the extremely high abundance of porins (over 1/10 of all detected proteins) in enrichments of small symbionts, whereas their abundance was about 3 times lower in enrichments of large symbiont cells. Porins, water-filled channels in the outer bacterial membrane (Fernández and Hancock, 2012), have been described to be involved in host interactions in *Vibrio splendidus*, where a porin is involved in host-symbiont recognition (Duperthuy et al., 2011) as well as in *Neisseria gonorrhoeae*, where a porin inhibits phagocytosis by human immune cells (Mosleh et al., 1998; Lorenzen et al., 2000). The *Riftia* symbiont porins might fulfill a similar role in ensuring survival inside the host bacteriocytes.

Enrichments of larger cells, on the other hand, exhibited relatively higher abundances of enzymes for CO_2 fixation and various biosynthesis pathways, including synthesis of carbohydrates, amino acids, lipids, and cofactors. Supposedly, larger *Riftia* symbiont cells thus engage relatively more in organic material synthesis than smaller cells. This might be the result of biochemical gradients in the lobules, caused by the blood flow from the lobule periphery (where the larger symbionts reside) towards the center (Felbeck and Turner, 1995). Differential availability of substrates might then lead to differential gene expression of smaller and larger symbionts. In addition to presumably having more substrate available for biosynthesis, larger symbionts can also likely invest relatively more cellular resources in organic material production than smaller symbionts, as smaller symbionts need to maintain

cell division. Moreover, based on our metaproteomics analysis results, showing high abundances of DNA ligase and DNA polymerase III in enrichments of large symbionts, we hypothesized that larger symbiont cells might undergo endoreduplication cycles of their DNA and thus become polyploid. This was backed up by flow cytometry and fluorescence activated cell sorting (FACS), indicating that large cells have an up to 10fold higher DNA content than small cells. A higher ploidy level could increase the metabolic activity and productivity of the larger *Riftia* symbiont, as suggested for plant cells (Kondorosi and Kondorosi, 2004). Relatively higher organic material production in larger symbionts would be beneficial for the *Riftia* host, as it digests large symbionts at the periphery of the trophosome lobules.

Taken together, we suggest that Endoriftia undergoes a process of physiological differentiation. This differentiation could, e.g., be caused by substrate gradients inside the trophosome lobules, or by specific host effectors. Overall, the constant provisioning of the symbionts by the host with all necessary nutrients might cause symbiont gene regulation to be less stringent as compared to free-living bacteria, leading to proteins of redundant pathways being expressed at the same time. This supposedly leads to high productivity, which could benefit the symbiosis as a whole.

3.3 THIOTROPHIC *CODAKIA ORBICULARIS* SYMBIONTS FIX NITROGEN

The *Codakia orbicularis* symbiosis, living in shallow water sediment amongst the rhizomes of *Thalassia testudinum* seagras (Jackson, 1973), represents another thiotrophic marine symbiosis with a single bacterial symbiont phylotype (Durand et al., 1996), which infects each host generation anew (Gros et al., 1996). Also like the *Riftia* symbiont, the *Codakia* symbiont is apparently restricted in its cell division during symbiosis (Caro et al., 2007) and digested by the host (Caro et al., 2009; König et al., 2015). In contrast to *Riftia* symbionts, which can be released after host death (Klose et al., 2015), *C. orbicularis* symbionts appear not to be able to exit their intracellular stage after death of the host, though (Brissac et al., 2009).

Metaproteomics analysis, including gel-free absolute protein quantification, combined with evaluation of genomics data, showed that the *C. orbicularis* symbiont contains the genes necessary for nitrogen fixation and expresses them under symbiotic conditions (Publication V). Nitrogenase activity was confirmed by assaying acetylene to ethylene reduction in host tissues and in isolated symbionts. Metaproteomics data also suggested that ammonium obtained by nitrogen fixation might be assimilated using the glutamine synthetase-glutamine oxoglutarate aminotransferase (GS-GOGAT) pathway. Based on absolute protein quantification, sulfide starvation (and thus energy starvation) apparently leads to inhibition

of nitrogen fixation by the symbiont, as significantly lower concentrations of nitrogenase subunits were present under these conditions. As symbiont rubrerythrin showed the same trend, it could be co-regulated with nitrogen fixation proteins and could be involved in protecting the symbiont, and especially its nitrogenase, against oxidative stress.

Digestion of *C. orbicularis* symbionts by their host could lead to symbiont-generated organic nitrogen being available for the host. This might pose an additional advantage of the *C. orbicularis* symbiont, besides that of sulfide detoxification and organic carbon generation, for the host, especially considering the low inorganic nitrogen concentrations in its seagrass bed habitat.

4 Outlook

Symbiosis research profoundly benefits from the insights into symbiosis functions under *in situ* conditions. These insights are possible by using advanced –omics, including metaproteomics, techniques.

This thesis used metaproteomics techniques to shed light on the role of the *Riftia* tubeworm host in this deep-sea symbiosis and on the physiological functions of *Riftia* symbiont subpopulations, and revealed the unexpected capability for nitrogen fixation in the symbiont of *Codakia orbicularis*. The core interactions in the *Riftia* symbiosis, including symbiont digestion by the host, a high host investment in symbiont maintenance, and putative communication of the symbiont with the host by means of eukaryote-like proteins, might be widespread in host-symbiont interactions in general. This hypothesis could be tested by studying and comparing a variety of symbiotic associations from different environments. Showing that one 16S rRNA phylotype of symbionts can consist of functionally dissimilar subgroups challenges the perception of a two-partner symbiosis as being low-complex and opens new avenues of research in comparable systems.

Moving forward, metaproteomics studies in general can use the extensive LC methods comparison undertaken in this thesis. The combined *Riftia* host and symbiont database facilitates further metaproteomics studies of this intriguing symbiosis, including proteins and mechanisms involved in the initial infection mechanism of *Riftia* by its symbiont, the reactions of the symbiosis to experimentally changed environmental parameters, as well as host adaptations to different habitats. Additionally, this thesis provides the basis for further functional analysis of key proteins in the *Riftia* symbiosis identified here. Protein-protein interaction studies could be used to unravel direct interactions between symbiont and host proteins. Moreover, first tests of heterologous gene expression of single symbiont proteins for further functional analysis already showed promising results.

5 Publications

Overview of publications discussed in this thesis and Author Contributions

Peer-reviewed manuscripts:

I **Hinzke, T.**, Kleiner, M., Breusing, C., Felbeck, H., Häsler, R., Sievert, S. M., Schlüter, R., Rosenstiel, P., Reusch, T. B. H., Schweder, T., Markert, S. Host-microbe interactions in the chemosynthetic *Riftia pachyptila* symbiosis. mBio 10, e02243-19 (2019). doi: 10.1128/mBio.02243-19

Author contributions: TH, SM, and MK designed experiments. TH prepared and analyzed samples for metaproteomics with input from MK, compiled the metaproteomics database with input from MK and SM, performed statistical analyses, prepared samples for RNA sequencing with input from CB, prepared figures, and wrote the manuscript. TS was involved in project coordination. TR, RH, and PR coordinated transcriptome sequencing. HF helped with sampling. SMS obtained funding for the research cruises and coordinated sampling as chief scientist. CB assembled and annotated transcriptomic data. All authors contributed to the final manuscript.

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

Author contributions: MK conceived the study, obtained and created bacterial stocks for mock communities. MK and TH designed the experiments. TH and R-AH prepared samples for mass spectrometry. TH measured samples, analyzed the data, and wrote the manuscript with input from MK. AK did mass spectrometry measurements. MS and AK revised the manuscript. All authors read and approved the final manuscript.

III Sokolov, E. P., Markert, S., **Hinzke, T.**, Hirschfeld, C., Becher, D., Ponsuksili, S., Sokolova, I. M. Effects of hypoxia-reoxygenation stress on mitochondrial proteome and bioenergetics of the hypoxia-tolerant marine bivalve *Crassostrea gigas*. J Proteomics 194, 99-111 (2019). doi: 10.1016/j.jprot.2018.12.009

Author contributions: IMS, EPS and SP conceived the study and secured funding. IMS and EPS wrote the manuscript with input from all coauthors. SM prepared proteomics samples. DB coordinated proteomics analysis. CH performed proteomics analysis. TH performed statistical analyses of proteomics data.

IV Kleiner, M., Dong, X., **Hinzke, T.**, Wippler, J., Thorson, E., Mayer, B., Strous, M. Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. PNAS 115, E5576-E5584 (2018). doi: 10.1073/pnas.1722325115

Author contributions: MK, conceived study, obtained and created bacterial stocks for pure culture and mock community experiments, designed experiments, generated mass spectrometric data, developed and tested the algorithms together with MS and XD, wrote the paper with input from all co-authors; XD, designed and implemented the algorithm to extract isotopic patterns from raw data, prepared downloadable software package and online wiki page on sourceforge, and revised manuscript; TH, prepared samples for IRMS analyses and revised manuscript; JW, collected and processed *O. algarvensis* samples for metaproteomics; ET, prepared proteomics samples, carried out experiments and operated mass spectrometer, revised the manuscript; BM, provided IRMS analyses and revised manuscript; MS, conceived study, developed and coded the algorithm for estimating $\delta^{13}C$ values, revised manuscript.

V König, S., Gros, O., Heiden, S. E., **Hinzke, T.**, Thürmer, A., Poehlein, A., Meyer, S., Vatin, M., Mbéguié-A-Mbéguié, D., Tocny, J., Ponnudurai, R., Daniel, R., Becher, D., Schweder, T., Markert, S. Nitrogen fixation in a chemoautotrophic lucinid symbiosis. Nat. Microbiol. 2, 16193 (2016). doi: 10.1038/nmicrobiol.2016.193

Author contributions: SK and OG planned and performed the experiments with *C. orbicularis*. Sediment chemistry analyses were conducted by JT, western blots and immunohistochemistry by MV, and acetylene reduction measurements by OG and DM-M. Proteome analyses were planned by SK, TS, SMa and DB, and conducted by SK, SMe and DB. AT, AP and RD performed the genome sequencing and submitted the sequences to NCBI. SK and SEH assembled and annotated the genome. SEH performed the phylogenetic analyses and gene cluster comparisons. OG, TS and SMa coordinated and supervised the laboratory work. SK and SMa analysed the proteome data. SMa wrote the manuscript. TH was involved in manuscript writing, figure generation and, together with RP, in MS data evaluation.

Book chapter:

VI **Hinzke, T.**, Kleiner, M., Markert, S. Centrifugation-based enrichment of bacterial cell populations for metaproteomic studies on bacteria-invertebrate symbioses. In Becher, D., (ed), Methods in Molecular Biology: Microbial Proteomics, Humana Press, New York (2018). doi: 10.1007/978-1-4939-8695-8_22

Author contributions: TH analyzed the data and wrote the manuscript with input from all coauthors.

Submitted manuscript:

VII **Hinzke, T.**, Kleiner, M., Meister, M., Schlüter, R., Hentschker, C., Pané-Farré, J., Hildebrandt, P., Felbeck, H., Sievert, S. M., Bonn, F., Völker, U., Becher, D., Schweder, T., Markert, S. Bacterial symbiont subpopulations have different roles in a deep-sea symbiosis. *Submitted*, preprint available at bioRxiv, doi: 10.1101/2020.04.08.032177

Author contributions: SM conceived the study, SM, TH, MK designed the experiments, TH and HF took samples. TH prepared samples for CARD-FISH and metaproteomics analysis, analyzed data, conducted statistical analyses and prepared figures. TH and SM wrote the manuscript with input from all coauthors. TS was involved in project coordination, SMS obtained funding for

the research cruises and coordinated sampling as chief scientist. CH and FB performed MS measurements of metaproteomics samples, DB coordinated MS measurements. MM and JP-F contributed to fluorescence microscopy and RS conducted electron microscopy analyses. PH performed flow cytometry analyses, and UV coordinated flow cytometry measurements.

PUBLICATION I

The article “Host-microbe interactions in the chemosynthetic *Riftia pachyptila* symbiosis” elucidates host-symbiont interaction mechanisms in the deep-sea tubeworm *Riftia pachyptila* on the protein level.

Supplementary data is provided in the CD.

AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE
Host-Microbe Biology

Host-Microbe Interactions in the Chemosynthetic *Riftia pachyptila* Symbiosis

Tjorven Hinzke,[a,b,c] Manuel Kleiner,[c,d] Corinna Breusing,[e] Horst Felbeck,[f] Robert Häsler,[g] Stefan M. Sievert,[h] Rabea Schlüter,[i] Philip Rosenstiel,[g] Thorsten B. H. Reusch,[j] Thomas Schweder,[a,b] Stephanie Markert[a,b]

[a]Institute of Marine Biotechnology e.V., Greifswald, Germany
[b]Institute of Pharmacy, Department of Pharmaceutical Biotechnology, University of Greifswald, Greifswald, Germany
[c]Energy Bioengineering Group, University of Calgary, Calgary, Canada
[d]Department of Plant & Microbial Biology, North Carolina State University, Raleigh, North Carolina, USA
[e]Monterey Bay Aquarium Research Institute, Moss Landing, California, USA
[f]Scripps Institution of Oceanography, University of California San Diego, San Diego, California, USA
[g]Institute of Clinical Molecular Biology (IKMB), Kiel University, Kiel, Germany
[h]Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, USA
[i]Imaging Center of the Department of Biology, University of Greifswald, Greifswald, Germany
[j]Marine Evolutionary Ecology, GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

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 served as a basis for comparative metaproteomic analyses of symbiont-containing versus symbiont-free tissues, both under energy-rich and energy-limited conditions. Our results suggest that metabolic interactions include nutrient allocation from symbiont to host by symbiont digestion and substrate transfer to the symbiont by abundant host proteins. We furthermore propose 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.

IMPORTANCE All animals are associated with microorganisms; hence, host-microbe interactions are of fundamental importance for life on earth. However, we know little about the molecular basis of these interactions. Therefore, we studied the deep-sea *Riftia pachyptila* symbiosis, a model association in which the tubeworm host is associated with only one phylotype of endosymbiotic bacteria and completely depends on this sulfur-oxidizing symbiont for nutrition. Using a metaproteomics approach, we identified both metabolic interaction processes, such as substrate transfer between the two partners, and interactions that serve to maintain the symbiotic balance, e.g., host efforts to control the symbiont population or symbiont strategies to modulate these host efforts. We suggest that these interactions are essential principles of mutualistic animal-microbe associations.

KEYWORDS host-microbe interactions, symbiosis, holobiont, chemosynthesis, hydrothermal vents, metaproteomics

All animals are associated with microorganisms (1–3), and consequently, mutualistic bacterium-animal symbioses play critical roles in the physiology, ecology, and evolution of animals, thereby shaping life on our planet. Many of these mutualistic

Citation Hinzke T, Kleiner M, Breusing C, Felbeck H, Häsler R, Sievert SM, Schlüter R, Rosenstiel P, Reusch TBH, Schweder T, Markert S. 2019. Host-microbe interactions in the chemosynthetic *Riftia pachyptila* symbiosis. mBio 10:e02243-19. https://doi.org/10.1128/mBio.02243-19.

Invited Editor Daniel Distel, Northeastern University

Editor Edward G. Ruby, University of Hawaii at Manoa

Address correspondence to Tjorven Hinzke, tjorven.hinzke@outlook.com, or Stephanie Markert, stephanie.markert@uni-greifswald.de.

Received 28 August 2019
Accepted 15 November 2019
Published 17 December 2019

symbioses are based on nutritional benefits for both partners. Symbionts supply their host with nutrients otherwise lacking in the host's diet, while the host in turn provides the symbionts with metabolites, shelter, and optimal growth conditions (4). To establish and stably maintain their alliance, the partners have to interact on the molecular level. The host's immune system needs to control the symbiont population without erasing it altogether (5), for example, by restricting the symbionts to certain organs and/or by downregulating its own immune response (reviewed in reference 6). 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 used by both symbionts and pathogens for interactions with the host (4, 7–9).

In many animals, host-microbe interactions are difficult to assess due to the high number of microbes potentially involved and the presence of long- and short-term associations, which are hard to distinguish (9). Therefore, low-complexity models are important to identify and characterize interaction mechanisms (10). 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 a few symbiont types, which enable the eukaryotes to prevail in otherwise hostile environments (11). 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*), which thrives around deep-sea hydrothermal vents of the East Pacific (12, 13). 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 its chemosynthetic endosymbiont (12–15). The host in turn provides the endosymbiont with all necessary inorganic compounds for chemosynthesis (16). This association is remarkably productive: *Riftia* grows extraordinarily fast (>85-cm increase in tube length per year [17]) and reaches body lengths of up to 1.5 m (18).

The uncultured gammaproteobacterial *Riftia* symbiont, a single 16S rRNA phylotype tentatively named "*Candidatus* Endoriftia persephone" (19–21), densely populates bacteriocytes in the host trophosome, a specialized organ that fills most of the worm's body cavity (14). The bacteria oxidize inorganic reduced compounds, such as hydrogen sulfide, to generate energy for carbon fixation (13, 22–26). The symbiont can store elemental sulfur, an intermediate of sulfide oxidation, in sulfur globules (27). Trophosome tissue containing large amounts of stored sulfur has a light yellowish color. During sulfide limitation, i.e., when energy availability is restricted due to low environmental sulfide concentrations, stored sulfur is consumed and the trophosome appears much darker (27–29). Thus, the energetic status of the symbiosis can 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 (for examples, see references 24 and 29–32; see references 16 and 33 for reviews). The symbiont's metabolism has been studied in detail as well (16), in particular by means of metagenomics and metaproteomics (19, 25, 34, 35). However, 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 (19, 34), until 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 the dynamics of host-symbiont interactions under high- and low-energy conditions. Our analysis sheds light on metabolite

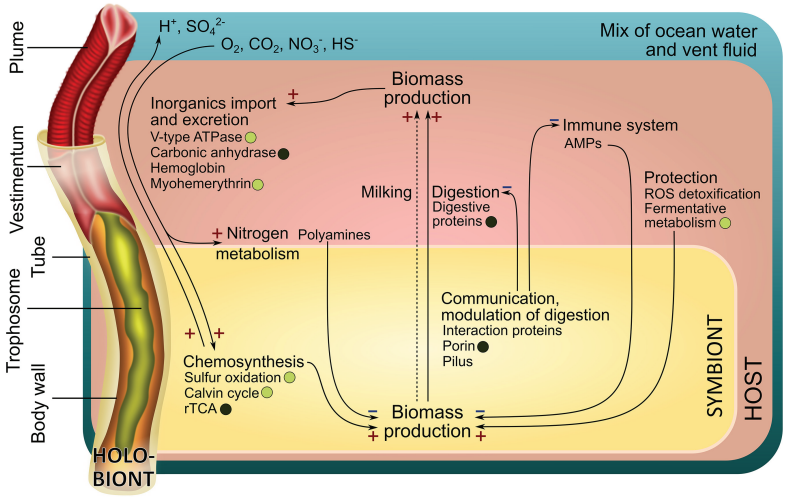

FIG 1 Main interactions in the *Riftia* symbiosis. "HOST" refers to processes in *Riftia* host tissues, while "SYMBIONT" refers to processes in the bacterial endosymbiont. A plus sign indicates presumably stimulating interactions, and a minus sign indicates presumably inhibiting interactions. For example, host efforts that protect the symbiont population from oxidative stress, i.e., ROS detoxification and fermentative metabolism (on the right), can promote symbiont biomass production (+). In contrast, host immune system-related proteins and antimicrobial peptides (AMPs) may inhibit symbiont biomass production (−). Circles, where present, indicate that the respective proteins are more abundant in S-rich (energy-rich) specimens (light circles) or S-depleted (energy-limited) specimens (dark circles). The dashed arrow indicates putative transfer of small organic compounds "Milking"; see Text S1, section 3).

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 Materials and Methods). Our metaproteomic analysis included comparisons between symbiont-containing and symbiont-free tissues of specimens with light and dark trophosomes. As trophosome color and bacterial sulfur content are directly correlated (27, 28), samples from specimens with light and dark trophosomes will here be referred to as sulfur-rich (S-rich) and sulfur-depleted (S-depleted) samples, respectively. A fully replicated data set and stringent study 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 Text S1, section 1, and Fig. S1 in the supplemental material. We identified numerous molecular interaction processes (Fig. 1), 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) sulfur availability affects symbiotic interactions in *Riftia*. Beyond the results presented here, our comprehensive metaproteome data sets and our newly established transcriptome-based *Riftia* host database (all available from the PRIDE archive; see below) also provide a valuable resource for future *Riftia* studies and microbe-eukaryote symbiosis research in general.

Metabolite exchange between host and symbiont. (i) *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

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[a]: S-rich troph	Sig in[a]: S-depl troph	Secreted/membrane[b]
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	x	x	
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-coenzyme A carboxylase beta chain, mitochondrial	x	o	
Host_DN41664_c1_g5_i6::g.166806	Peroxisomal bifunctional enzyme	o	x	

[a]Sig, Significance (x, significant; o, nonsignificant; false discovery rate, 0.05); troph, trophosome; S-depl, S depleted.
[b]Subcellular localization (M, membrane-associated; S, secreted) was predicted using Phobius, TMHMM, and SignalP. Possibly M or S indicates localization prediction based on one tool only.

as in glycolysis and fatty acid beta oxidation. Twenty-two 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., higher organism-specific normalized spectral abundance factor values, or %orgNSAF) in the symbiont-bearing trophosome than in other tissues, both in S-rich and S-depleted specimens (Fig. 2). Many of the protein degradation-related proteins contain signal peptides and thus are likely either contained in lysosomes or secreted into the symbiont-containing vesicles to digest the symbiont cells (Table 1 and Table S1a).

Our findings are in accordance with previous biochemical, autoradiographic, and microscopic studies, which suggested symbiont digestion in the *Riftia* trophosome (14, 36–38). Moreover, abundant degradative enzymes and symbiont digestion appear to be common in other mutualistic symbioses as well, including deep-sea mussels (39, 40), shallow-water clams (41, 42), and the gutless oligochaete *Olavius algarvensis* (43, 44).

Our metaproteome analysis suggests that symbiont digestion in *Riftia* involves maturation of symbiont-containing host vesicles in a process resembling the maturation of endosomes. Endosomes form after endocytosis of extracellular compounds and mature from early to late endosomes, which ultimately fuse with lysosomes (45). The endosome-associated proteins Rab5 and Rab7 showed significantly higher abundances in trophosome samples than in other host tissues (Table S1a). Rab5 and Rab7 localize to early and late stages, respectively, of endosomes and autophagosomes and are markers for these recycling-related organelles (45–47). The idea of symbiont degradation via an endosome-like maturation process in *Riftia* is additionally supported by our transmission electron microscopy (TEM) images of *Riftia* bacteriocytes (Fig. 3), which

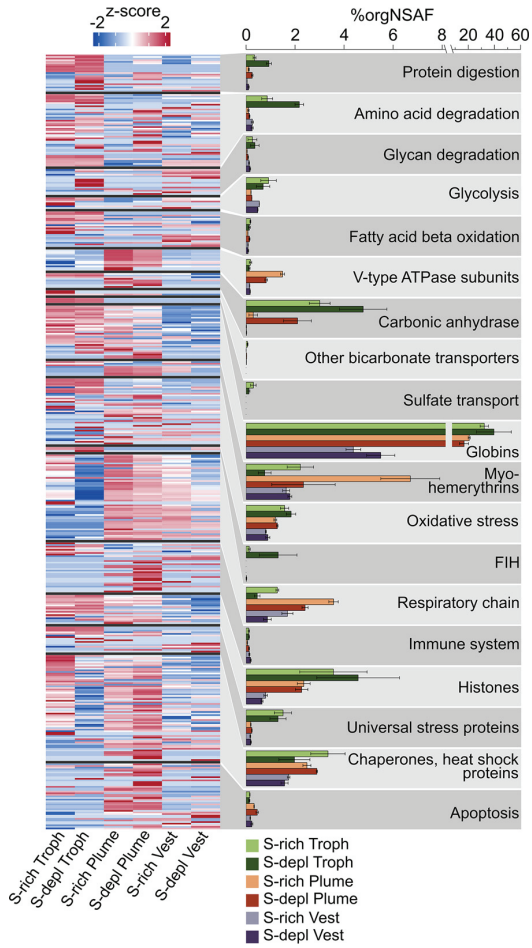

FIG 2 Functional groups of selected *Riftia* host proteins and their relative abundances in tissue samples. The heatmap shows log-normalized, centered, and scaled protein abundances. The bar chart shows summed abundances in %orgNSAF (percent normalized spectral abundance factor per organism, i.e., of all host proteins) of all 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. Vest, vestimentum. Troph, trophosome. For a list of all identified proteins and their abundances, see Table S1a. (Categories presented in this figure are labeled with X in Table S1a in the column labeled Figure 2. The table can be filtered for these categories.)

showed multilamellar bodies. These myelin-like structures can form in endosomes (48) and also during autophagic digestion and have, therefore, previously been attributed to autophagy in the *Riftia* trophosome (37). However, our results suggest that autophagy plays a less prominent role in symbiont digestion, as we detected only two autophagy-related proteins (Table S1b) in the trophosome metaproteome.

Moreover, only 12 of 41 detected apoptosis-related *Riftia* proteins were identified in the trophosome, mostly with similar or significantly lower abundances than in other tissues, and caspases, the main apoptotic effectors, were not detected at all on the protein level in trophosome samples (see also Text S1, section 2). These results suggest that bacteriocyte cell death, which follows after symbiont digestion, probably does not involve apoptosis. This contradicts previous observations (37) but is in line with microscopic results, which did not indicate apoptosis in the trophosome (49). We therefore suggest that an alternative, nonapoptotic cell death mechanism exists in *Riftia* trophosomes. A nonapoptotic, nonautophagic cell death mechanism was recently

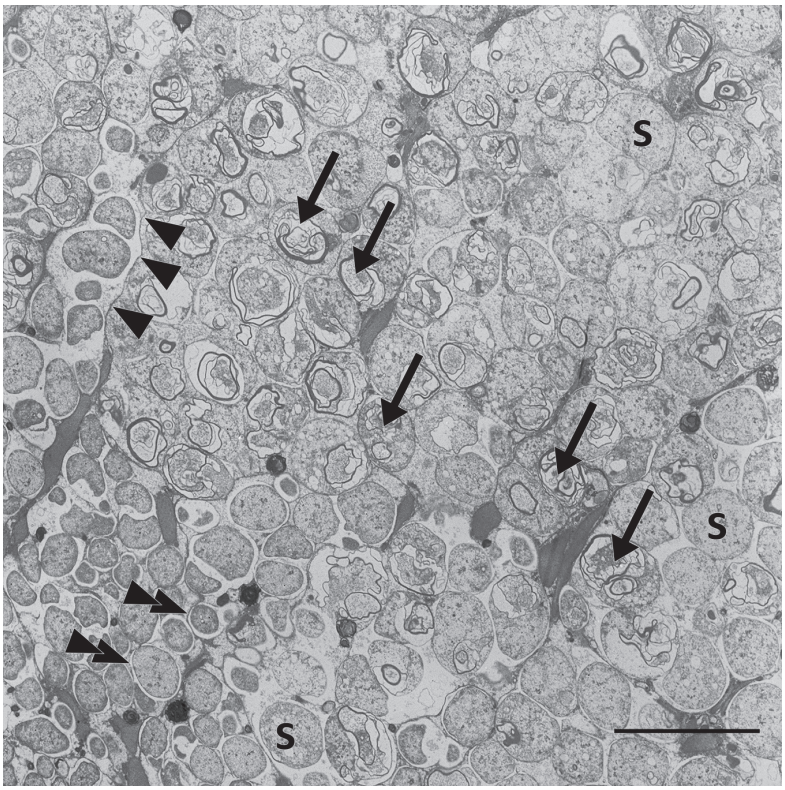

FIG 3 Transmission electron micrograph of a *Riftia* trophosome tissue section. Within the lobular trophosome tissue, this section shows the median and peripheral zones of an individual lobule with host bacteriocytes containing intracellular coccoid symbionts (S) located in dedicated vesicles (arrowheads, bacteriocyte membrane; double arrowheads, vesicle membrane). While the lower left area of the image shows mostly intact symbiont cells, arrows in the central area point to symbiont cells in the state of digestion by the host, where cell degradation is indicated by the presence of lamellar bodies. Image brightness and contrast were adjusted for visual clarity. Scale bar, 10 μm.

described in pea aphid bacteriocytes (50). In the aphids, the proposed mechanism involved hypervacuolation of host bacteriocytes, which, however, was not observed in the *Riftia* trophosome. Also, in cancer cells, a caspase-independent (nonapoptotic) cell death mechanism was described, which involves the lysosomal protease cathepsin B (51), a putative regulator of lysosome production and autophagic activity (52). As cathepsin B was significantly more abundant in trophosome than in other *Riftia* tissues, we speculate that this protease, among other degradative enzymes, is involved in controlled cell death in the *Riftia* trophosome.

Besides symbiont digestion, a second mode of nutrient transfer, the release of small organic carbon compounds by intact symbionts (termed "milking"), was suggested to be present in *Riftia* (36, 53). Our calculated $\delta^{13}C$ ratios might support this hypothesis (Text S1, section 3). However, as we did not detect dedicated symbiont exporters for organic acids or sugars on the proteome level, nutrient transfer by milking is probably less relevant for overall host nutrition than symbiont digestion.

(ii) *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 (Fig. 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 proteome in the plume (i.e., the worm's symbiont-free gas exchange organ;

Fig. 2), with extracellular hemoglobins being particularly abundant (in sum, 32 to 40 %orgNSAF in trophosome and 17 to 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 (54–56). We detected several of these subunits, including isoforms that are (to our knowledge) hitherto undescribed (Table S1a). *Riftia's* extracellular hemoglobins have been shown to bind both O_2 and sulfide (56, 57; reviewed in references 58 and 59). Consequently, abundant hemoglobins in the highly vascularized plume would ensure efficient uptake of these compounds for transport to the symbionts (see Text S1, section 4, for more details on sulfur metabolism in the host). Moreover, reversible O_2 and sulfide binding to abundant hemoglobins in the trophosome not only provides the bacteria with chemosynthetic substrates and prevents spontaneous sulfide oxidation but also protects the symbionts from oxygen (60). As suggested previously (23), *Riftia* symbionts are microaerophilic, i.e., sensitive to high oxygen levels. This idea is corroborated by the presence of several ROS-scavenging enzymes (superoxide dismutase, alkyl hydroperoxide reductase, and rubrerythrin) and cytochrome *c* oxidase cbb_3 subunits in the symbiont metaproteome. cbb_3 has a high affinity for oxygen and participates in microaerobic respiration (61). In addition to extracellular hemoglobins, we identified four low-abundance (0.002 to 0.084 %orgNSAF) globins that are probably intracellular and might store O_2 (Text S1, section 5).

Besides hemoglobins, myohemerythrins were detected in all tissues, with particularly high abundances of 6.7 %orgNSAF in S-rich plumes. With their comparatively high oxygen-binding capacity (62), hemerythrins could facilitate oxygen uptake from the environment into the plume and are possibly also involved in O_2 storage and intracellular transport in *Riftia*. Moreover, the abundance distribution of the nine detected myohemerythrins suggests a tissue-specific function (Text S1, section 6).

V-type ATPase subunits were found with highest total abundances of up to 1.5 %orgNSAF in *Riftia* plumes (Fig. 2), and almost all of the detected subunits were significantly more abundant or exclusively detected in the plumes. V-type ATPases have a pivotal function in regulating internal pH and CO_2 uptake (63) and thus in symbiont provisioning. The high energy demand of V-type ATPase-dependent pH regulation could be met via a relatively higher respiration activity in the plume, as indicated by comparatively higher total abundances of respiratory chain proteins (Fig. 2), ATP synthase, and mitochondrial ribosomal proteins in this tissue. Additionally, carbonic anhydrase (CA), another important enzyme for CO_2 uptake, was detected in all tissues. While we observed tissue-specific abundance patterns of individual CAs (Text S1, section 7 and Fig. S4), overall CA abundance was highest in the trophosome (Fig. 2). CA facilitates CO_2 diffusion into the plume by converting it to HCO_3^- (63, 64) 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 are membrane bound (Text S1, section 7) and could thus facilitate CO_2 diffusion into bacteriocytes by converting HCO_3^- to CO_2 on the bacteriocyte cell surface (65, 66). Transport of HCO_3^- to the bacteriocytes could be mediated by sodium bicarbonate exchangers, which we identified in trophosome and plume samples (Table S1a).

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 (67). We did identify host phosphoenolpyruvate carboxykinase and pyruvate carboxylase, which could be involved in this process (Text S1, section 8).

(iii) *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 likely also with ammonium produced by bacterial nitrate reduction (Fig. 4 and Text S1, section 9). 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 (68) and produce polyamines (69). In contrast to those studies, we found the multifunctional CAD protein (which combines the three enzyme func-

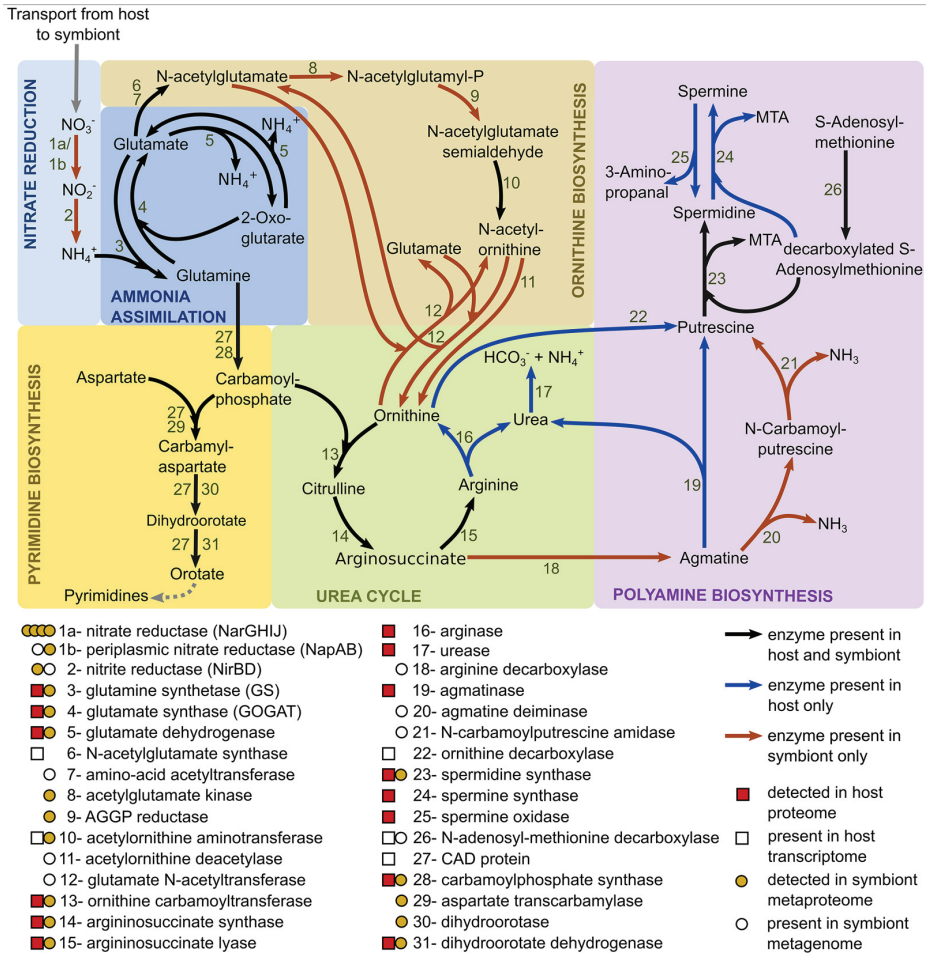

FIG 4 Main nitrogen metabolic pathways in *Riftia* symbiosis. AGGP reductase, *N*-acetyl-gamma-glutamyl-phosphate reductase; CAD protein, multifunctional carbamoyl-phosphate synthetase 2, aspartate transcarbamoylase, and dihydroorotase protein; MTA, 5′-methylthioadenosine. Note that the symbiont might also be capable of nitrate respiration (25, 60), which is not depicted here.

tions 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 host's metatranscriptome and also detected several of the respective proteins in the host's metaproteome (Fig. 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 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 (Text S1, section 10). 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 (Text S1, section 9). These results suggest 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.

Host strategies of symbiont maintenance. (i) *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, and 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 than in other tissues (Fig. 2 and Text S1, section 11). *Riftia*'s ROS-detoxifying enzymes probably protect not only the host but also the microaerophilic symbiont against ROS. Upregulation of host proteins involved in ROS detoxification was previously shown in the *Wolbachia* symbiosis (70, 71). Additionally, malate dehydrogenase was highly abundant in trophosomes. This enzyme is regularly observed in different invertebrates under anaerobic conditions (72) and is involved in maintaining redox balance during anaerobiosis (73). Therefore, the host might generate hypoxic conditions in the trophosome, as also indicated by the overall lower abundance of host respiratory chain proteins in trophosome than in 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 (23, 60) and with the suggestion of fermentative metabolism under hypoxic and even normoxic conditions in *Riftia* based on biochemical results (74). 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 also decrease the competition between the *Riftia* host and its symbionts for oxygen.

(ii) The *Riftia* immune system might be involved in symbiont population control. We detected several proteins that potentially are 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 (75). 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 (76). 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, PRPs accumulate during defense responses against pathogens (reviewed in reference 77). PRPs have also been described in nematodes (78) and humans (79), 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 (Table S1a). Besides being crucial for DNA interactions, histones and histone-derived peptides can have antimicrobial effects (80–82). A BLASTP search of the detected *Riftia* histones against the antimicrobial peptide (AMP) database APD3 (83) gave hits for four of the *Riftia* histones (Table S1c), stimulating the speculation that these histones 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 actively dividing stem cells into growing but nondividing larger cells (49). As various AMPs were shown to inhibit cell division or septum formation and to cause filamentous cell morphologies (reviewed in reference 84), we speculate that *Riftia* AMPs inhibit cell division as well, e.g., via interaction with the symbiont protein GroEL. Interaction between a host AMP and a symbiont GroEL has been proposed to lead to cell elongation

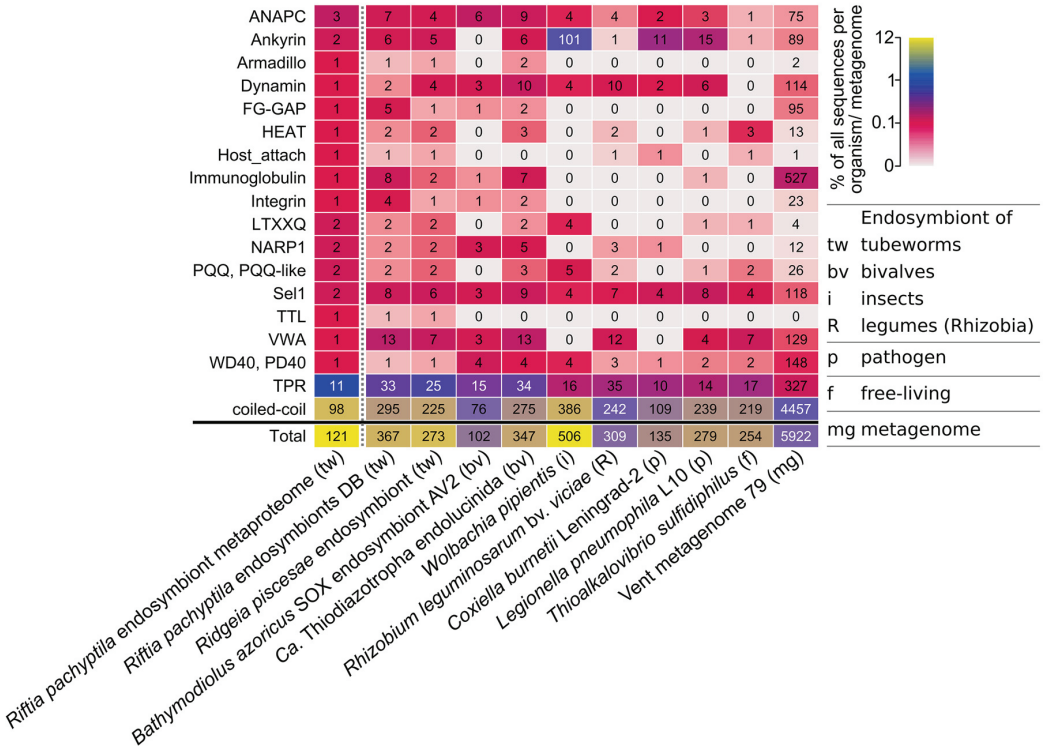

FIG 5 Selected domains with eukaryote-like structures and with putative functions in symbiont-host interactions in the *Riftia* symbiont and in selected other organisms and metagenomes. Color scale shows the percentage of genes/proteins containing the respective domain relative to all gene/protein sequences in this organism or metagenome. Numbers indicate the total number of genes/proteins containing the respective domain. For an overview of all analyzed organisms and domains, see Text S1, Fig. S5. For details on the organisms and communities, see Table S1d. The vent metagenome was sampled from hydrothermal vent fluid at a diffuse-flow vent site (Crab Spa) (137), which also houses *Riftia*. For further information about the selected protein groups, see Table S2. *Riftia pachyptila* endosymbiont metaproteome refers to the *Riftia* symbiont proteins detected in this study.

of bacterial weevil symbionts (85). A role of histones and histone-derived peptides in immune system responses has been described or suggested in various other organisms, including catfish (80), Komodo dragons (86), toads (81), and humans (82).

Beyond these and a few other individual immunity-related proteins, we did not observe an overall higher abundance of host immune system proteins (such as lysozyme, complement system proteins, or peptidoglycan recognition proteins) in the trophosome than in symbiont-free tissues. This indicates that the host immune system does not play a major role in controlling symbiont population size. More likely, symbiont population control might to a large part be a result of digestion of symbionts (a "mowing" process), which effectively prevents the symbionts from escaping their compartments and/or overgrowing the host. Nevertheless, the immune system might be involved in phage protection and symbiont recognition during establishment of the symbiosis (Text S1, section 12).

Symbiont persistence mechanisms. (i) Eukaryote-like protein structures in the symbiont might be involved in host communication. The metagenome of the *Riftia* symbiont "*Ca.* E. persephone" encodes several protein groups with possible roles in symbiont-host interactions, including eukaryote-like protein (ELP) structures, as revealed by our SMART analysis (Table S2). We detected more than 100 of these symbiont proteins in the trophosome samples (Fig. 5), which points to a symbiosis-relevant function.

Among the ELPs detected in the symbiont metaproteome were two ankyrin repeat-containing proteins, which contain a signal peptide and are therefore likely secreted

(predicted by Phobius, http://phobius.sbc.su.se/). Ankyrin repeats were found to mediate protein-protein interactions (87). In the sponge *Cymbastela concentrica*, symbiont ankyrins were proposed to interact with the eukaryote's phagocytosis system: the symbiont ankyrins were heterologously expressed in *Escherichia coli* and led to inhibition of phagocytosis by amoebae (88). Likewise, a secreted *Legionella pneumophila* ankyrin protein apparently interferes with host endosome maturation (89). The "*Ca.* E. persephone" ankyrin repeat-containing proteins therefore could directly interact with host proteins as well, e.g., to modulate endosome maturation and thus to interfere with symbiont digestion by the host. Similarly, proteins with tetratricopeptide repeat (TPR)/Sel1 domains, which we also detected in the "*Ca.* E. persephone" metaproteome, have been shown to impact phagocytosis by amoebae (90).

The *Riftia* symbiont furthermore encodes eukaryote-like proteins of the tubulin-tyrosine ligase family (TTL proteins). These proteins posttranslationally modify tubulin and thus interact with the eukaryotic cytoskeleton (91). We found one TTL protein in the "*Ca.* E. persephone" metaproteome. Other protein groups that are involved in protein-protein interactions in eukaryotes, e.g., with cytoskeletal proteins, and that we detected in "*Ca.* E. persephone" include armadillo repeat proteins (92) and HEAT repeat-containing proteins (93). As several of the protein structures analyzed here are also found in other mutualistic symbionts and pathogens (Text S1, section 13, and Table S2), it is conceivable that parallels exist between interaction processes of mutualistic and pathogenic associations and that the *Riftia* symbiont employs a strategy similar to that of pathogens to communicate with its host on the molecular level.

(ii) Symbiont membrane proteins may export effector proteins into host cells and lead to strain adaptation. We detected various outer membrane-related proteins in the "*Ca.* E. persephone" proteome, including a porin (Sym_EGV52132.1), which was one of the most abundantly expressed symbiont proteins, and 12 type IV pilus (T4P) system proteins (PilQ, PilF, PilC, PilBTU, PilM, PilN, PilP, FimV, PilH, and PilY1). Five additional T4P structure proteins were encoded in the metagenome (*pilVWXE* and *pilO*). These proteins are in direct contact with the host cells and thus are likely involved in interactions between both symbiosis partners, including such processes that facilitate the symbiont's persistence inside the host cells.

The abundant symbiont porins could transport effector molecules, e.g., to modulate digestion by the host. A role of porins in effector transport during symbiosis has been hypothesized for *Vibrio fischeri* OmpU, a channel protein that is important for symbiont recognition by the squid host (94).

The T4P system is a complex structure, which in *Pseudomonas aeruginosa* comprises more than 40 proteins, including structural and regulatory proteins (95). It can have several functions in different species: adhesion, secretion, and natural transformation (95–98). As the "*Ca.* E. persephone" T4P system likely is not involved in adhesion to host cells during symbiosis (although it might be during the initial infection), it could participate in protein secretion and/or natural transformation. The *Riftia* symbiont's T4P system could export putative effector proteins (e.g., ankyrins and SET domain proteins; Text S1, sections 13 and 14) for host interactions. Interestingly, in the pathogen *Francisella tularensis* subsp. *novicida*, a T4P structure is involved in secretion of infection-moderating proteins (97).

Besides their putative function in effector protein export, symbiont membrane proteins may also lead to bacterial strain adaptation. The *Riftia* symbiont population is polyclonal, i.e., although there is only one 16S rRNA phylotype, this phylotype consists of several distinct strains (20). T4P system-mediated exchange of genetic material between different symbiont strains would add to this diversity in the symbiosis and might additionally enable exchange of symbiosis-related genes within the free-living "*Ca.* E. persephone" population. Natural transformation in symbionts has only recently been shown for *V. fischeri* in culture (99) and the earthworm symbiont *Verminephrobacter eiseniae*, which likely employs a T4P structure for DNA uptake (98). As microbial cell densities are comparatively high in eukaryote-prokaryote mutualisms, natural transformation in these systems might actually be more common than previously

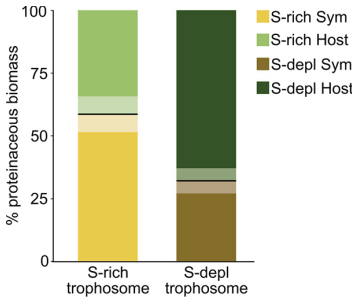

FIG 6 Percent proteinaceous biomass contributions of host and symbiont as calculated from the share of host and symbiont spectral counts in all spectral counts of the respective samples (127; see Materials and Methods for details). Boldface lines indicate the means, and semitransparent areas indicate standard error of the mean. Sym, symbiont; S-depl, S depleted.

recognized. While mostly only one to three symbiont cells are located in one host vesicle, individual vesicles with up to 14 symbiont cells have also been reported (49), which might allow for exchange of genetic material. The proposed DNA uptake by the *Riftia* symbiont may not only facilitate exchange between symbiont strains but also promote horizontal gene transfer between host and symbiont, e.g., of eukaryote-like proteins. This hypothesis, as well as the speculation that "*Ca.* E. persephone" is capable of conjugation (Text S1, section 14), certainly warrant further investigations.

S availability affects symbiotic interactions in *Riftia*. (i) S-depleted *Riftia* hosts digest more symbionts than S-rich specimens. We compared the metaproteomes of *Riftia* specimens with and without stored sulfur (i.e., energy-rich versus energy-depleted specimens; Text S1, Fig. S1) to examine how energy availability impacts symbiotic interactions (see Table S1f for total numbers of differentially abundant proteins). Metabolite transfer is apparently especially influenced by the energy regime: the host supposedly relies more on symbiont digestion in times of S shortage. Proteinaceous symbiont biomass was notably lower in S-depleted trophosomes (32%) than in S-rich trophosomes (58%) (Fig. 6). Simultaneously, overall abundances for several groups of host digestive enzymes were higher in S-starved trophosomes (Fig. 2), and a number of individual host proteins were significantly more abundant in these S-depleted samples, such as enzymes involved in protein digestion (including cathepsin B), amino acid degradation, the late endosome-related protein Rab7, and histones (Table S1a). One reason for this supposed increase in symbiont digestion in S-depleted trophosomes could be a lower nutritional value of the energy-depleted symbionts. S-depleted symbionts have lower abundances of enzymes involved in sulfur oxidation, probably due to lower S availability. Therefore, less energy might be available for biosynthesis under S depletion, rendering the symbiont less nutritious for the host. The animal would then, especially if S depletion is prolonged, have to rely on increased symbiont digestion in order to still satisfy its basal metabolic demands. Thus, S-depleted hosts may, despite increased symbiont digestion, have less energy available. This idea is supported by the observation that host proteins involved in the energy-generating glycolysis, tricarboxylic acid (TCA) cycle, respiratory chain, ATP synthesis, and biosynthetic pathways were less abundant in S-depleted trophosomes than in S-rich trophosomes. Concomitant with the postulated lower nutritional value of S-depleted symbionts, the Calvin cycle key enzyme RubisCO had an about 10-fold lower abundance in S-depleted symbionts. Abundance of the reverse TCA (rTCA) cycle key enzyme ATP citrate lyase (NCBI accession no. EGV51152.1), on the other hand, was slightly higher in S-depleted symbionts than in S-rich symbionts, albeit only 1.4-fold. Under S-depleted conditions, symbionts apparently rely relatively more on the rTCA cycle, which is more energy efficient than the Calvin cycle (35). The Calvin cycle could be used in addition to the rTCA cycle under favorable conditions to maximize carbon fixation. Moreover, symbiont enzymes involved in translation were overall more abundant in S-rich tro-

phosomes than in S-depleted trophosomes. Less protein biosynthesis in S-depleted symbionts would not only impact the nutritional value of these symbionts but also directly decrease the proteinaceous symbiont biomass. The reason for the lower proteinaceous biomass of symbionts in S-depleted trophosomes is, therefore, probably 2-fold: the host digests more symbionts and the symbionts produce less biomass than in energy-rich trophosomes.

These findings are in contrast to previous results (29), which showed no significant differences in autotrophic activity and symbiont abundance between *Riftia* specimens from high- versus low-sulfide habitats. Increased symbiont digestion may be a short-term adaptation to fluctuating environmental conditions, whereas under long-term low-S conditions the symbiosis might adapt by other means, e.g., by reduced growth rates. Decrease in symbiont abundance or total protein under energy-limiting conditions also has been noted in *Bathymodiolus* (100) and *Codakia orbicularis* bivalves (42) as well as in *O. algarvensis* oligochaetes (43). Thus, relying on the symbionts as a nutrient source also under unfavorable conditions appears to be a common symbiosis mechanism that would ensure survival of the host and a subset of the symbiont population, ultimately prolonging survival of the individual holobiont.

(ii) S availability influences CO_2 uptake, pH regulation, and O_2 regime in the *Riftia* host. S-depleted hosts seem to invest relatively more biosynthetic capacities in CO_2 uptake and less in pH regulation, and their trophosomes are supposedly less hypoxic than those of S-rich hosts (Text S1, sections 11 and 15). At the same time, S availability appeared to have little influence on non-symbiont-related processes in the host, as only very few (i.e., <10) individual proteins differed significantly in abundance between S-rich and S-depleted plume and vestimentum samples. This indicates that the host's metabolism is very well buffered against changes in environmental conditions.

(iii) Higher digestion pressure might result in symbiont countermeasures. A putative "*Ca.* E. persephone" dodecin was significantly more abundant in S-depleted *Riftia* specimens than in S-rich specimens. This protein might be involved in protecting the symbiont against oxygen and/or digestion stress (Text S1, section 14). A symbiont porin, which was also significantly more abundant in S-depleted specimens, might be involved in counteracting the supposedly higher digestion pressure (described above) (Text S1, section 14).

Conclusions. To fully understand the biology of organisms, it is crucial to study them together with their symbiotic partners as holobionts (101). 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 host's 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.

MATERIALS AND METHODS

Sampling. *Riftia* 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. Specimens were kept at 4°C in cold seawater until dissection, which was performed within 4 h after recovery. Only healthy-looking specimens were used. Sampling dates for all *Riftia* tissue samples for proteomics, transcriptomics, and transmission electron microscopy (TEM) are summarized in Table S1e in the supplemental material. Different specimens were used for proteomics, transcriptomics, and TEM. *Riftia* specimens were dissected onboard, and samples from four different organs (here referred to as tissues) were 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. As trophosome color is directly correlated to the tissue's elemental sulfur content (27, 28), we classified specimens as sulfur rich (S rich), S depleted, and medium S according to their trophosome color (yellow/light green, dark green/black, or medium green, respectively). To ensure comparability of the classifications, the same light source and dissection tray were used for all samples. Sulfur-rich and sulfur-depleted specimens were used for transcriptome sequencing and comparative metaproteomics, while specimens with medium sulfur content were only used for transcriptome sequencing (Table S1e).

Extraction of whole-tissue RNA. RNA was extracted from a total of 22 tissue samples from 9 specimens with high, medium, and low trophosome sulfur content (6× trophosome, 6× body wall, 5× plume, 5× vestimentum) (Fig. 1). Tissue samples were homogenized by bead beating with lysing matrix D (MP Biomedicals) in 1 ml TRIzol (Thermo Fisher Scientific; 3 times at 6.5 m/s for 30 s, with 3 min of cooling on ice between steps). After 5 min of acclimatization to room temperature, samples were applied onto QIAShredder columns (Qiagen) and centrifuged (16,000 × *g*, 3 min, 4°C). Afterwards, RNA was isolated from the aqueous flowthrough according to the TRIzol extraction protocol, with the modification that samples were centrifuged for 20 min at 12,000 × *g* and 4°C for phase separation. Ten micrograms of 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, including DNA removal with DNase (Qiagen). Quality of extracted RNA was assessed using NanoDrop (Thermo Fisher Scientific) and Bioanalyzer (Agilent) analyses.

Transcriptome sequencing and assembly. (i) Transcriptome sequencing. Transcriptome sequencing was performed employing the TruSeq stranded mRNA [poly(A)-based] library protocol (Illumina) on a HiSeq 4000 (Illumina) according to the manufacturer's guidelines.

(ii) Transcriptome assembly. High-throughput paired-end Illumina sequencing resulted in an average of about 26 million reads per end per library (minimum of 16,045,121 reads per end, maximum of 31,318,532 reads per end; 95% confidence interval, 1,673,590). After demultiplexing and quality checking of reads in FastQC v0.11.5 (102), we trimmed low-quality bases and adapters with Trimmomatic v0.32 (103) using the settings ILLUMINACLIP:AllAdapters.fa:2:30:10 SLIDINGWINDOW:4:20 and LEADING:5 TRAILING:5 HEADCROP:15 MINLEN:75. Although bacterial mRNA does not possess a poly(A) tail, previous research has shown that bacterial reads can still be present in poly(A)-enriched RNA-sequencing libraries (104). To filter out potential symbiont contaminations from our host transcriptomes, we used the Bowtie 2 v2.2.9 aligner (105) in very-sensitive mode to map the quality-filtered paired-end reads against the published genomes of the endosymbionts of *Riftia* (Riftia1, NCBI locus tag prefix RIFP1SYM; Riftia2, locus tag prefix RIFP2SYM) and *Tevnia jerichonana* (34). Unmapped paired-end reads were subsequently extracted using SAMtools v1.4.1 (106). Potential environmental sequence contaminations from sample handling were excluded with DeconSeq v0.4.3 (107), using coverage and identity thresholds of 0.90 and 0.95, respectively. The decontaminated host reads were normalized, pooled, and assembled with Trinity v2.3.2 (108). To optimize the transcriptome assembly, we performed four different assemblies with different parameters and input files: (i) only paired reads, (ii) paired and unpaired reads, (iii) only paired reads plus Jaccard-clip option (to reduce chimeras), and (iv) paired and unpaired reads plus Jaccard-clip option.

To assess the completeness of the different assemblies, we compared our transcriptomes to the BUSCO v2.0 eukaryote and metazoan orthologous data sets (109). Overall, the best results in terms of transcriptome completeness and quality were obtained by the assembly approach using paired and unpaired reads plus the Jaccard-clip option (complete BUSCO, 99.0%) (Table S3). This data set was used for all further analyses.

(iii) ORF prediction. TransDecoder v3.0.1 (110) was used to identify coding regions in the assembled transcripts. To improve open reading frame (ORF) prediction, we examined all candidate ORFs for homology to known proteins by searching the Swiss-Prot (http://www.uniprot.org) and Pfam (111) databases (downloaded 3 January 2017) with BLASTP (E value of 1e−05) (112) and HMMER3 (113), respectively. ORFs that were longer than 100 amino acids and/or had a database entry were retained. The FASTA headers of the TransDecoder output files were modified with a custom PERL script to include the BLASTP protein annotations.

Database generation. A common database for protein identification of *Riftia* host and symbiont was generated. To this end, host protein sequences were clustered at 95% identity with CD-HIT v. 4.6 (114). For symbiont sequences, the three proteomes of the *Riftia*1 symbiont (NCBI PRJNA60889, ID 60889), *Riftia*2 symbiont (NCBI PRJNA60891, JGI 2600255285), and *Tevnia* symbiont (NCBI PRJNA60887, ID 60887) (34) were used. *Riftia*1 was used as basis for clustering the symbiont protein sequences with CD-Hit-2D (114). Subsequently, the combined symbiont database was clustered at 95% identity. Identifier prefixes

were added to distinguish between host and symbiont sequences for Calis-p (115 and see below). Host and symbiont databases were concatenated, and the cRAP database containing common laboratory contaminants (116) was added. The final database contained 71,194 sequences (67,092 host and 3,986 symbiont protein sequences).

Proteomics sample preparation and analysis. For metaproteomics analysis, we used three biological replicates per tissue (trophosome, vestimentum, and plume) and condition (specimens with S-rich and S-depleted trophosomes), which resulted in a total of 18 samples. Tissues were disrupted by bead beating for 45 s at 6.0 m/s with lysing matrix D tubes (MP Biomedicals) in SDT buffer (4% [wt/vol] sodium dodecyl sulfate [SDS], 100 mM Tris-HCl, pH 7.6, 0.1 M dithiothreitol [DTT]), followed by heating to 95°C for 10 min. Tryptic peptides were generated following the FASP protocol of Wiśniewski et al. (117), with minor modifications as described by Hamann et al. (118). Peptide concentrations were determined with the Pierce Micro BCA (bicinchoninic acid) assay (Thermo Scientific Pierce) according to the manufacturer's instructions. The tryptic digest was desalted on-line during liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis.

All samples were analyzed by one-dimensional LC-MS/MS as described by Hinzke et al. (119), using 4-h gradients. Samples were analyzed in a randomized block design (120) and run in technical triplicates. Two technical replicate runs were acquired with a 50-cm analytical column, one with a 75-cm analytical column. To standardize the stable isotope fingerprinting (SIF) analysis (115), human hair was measured in technical duplicate alongside the *Riftia* samples in the replicate run using a 75-cm column.

Proteomics data evaluation. (i) Protein identification, quantification, and statistical analyses. For protein identification, MS/MS spectra of combined technical triplicate runs were searched against the combined host and symbiont database using the Sequest HT node in Proteome Discoverer version 2.0.0.802 (Thermo Fisher Scientific) as described in Kleiner et al. (115). For protein abundance estimates, normalized spectral abundance factors (NSAFs) (121) were calculated per sample and organism (%orgNSAF) (122). Statistical evaluation was performed based on spectral counts using the edgeR package (123) in R (124). The edgeR package uses an overdispersed Poisson model for analysis of count data. Overdispersion is moderated across proteins using empirical Bayes methods, and differentially abundant proteins are detected using an overdispersion-adapted analog to Fisher's exact test (123). We filtered for proteins with at least 10 spectral counts for host proteins and at least 5 spectral counts for symbiont proteins in at least three samples and employed a false discovery rate (FDR) of 0.05 to assign statistical significance to protein abundance differences. For graphical representation, heatmaps were generated with the R package ComplexHeatmaps (125) and intersection plots with the R package UpsetR (126). Protein biomasses of host and symbiont were calculated as described in Kleiner et al. (127). Spectral counts of all symbiont proteins and of all host proteins identified with at least two unique peptides were summed individually, and these two sums were divided by the sum of all spectral counts (host plus symbiont proteins with at least two unique peptides) and multiplied by 100 to give the percentage of proteinaceous biomass for host and symbiont.

$\delta^{13}C$ values of *Riftia* symbiont and host were calculated from mass spectrometry data with Calis-p (115) using one technical replicate LC-MS/MS run (75-cm analytical column). Human hair was used as a reference material.

(ii) Protein annotations, functional characterization, and categorization. Besides the annotations included in the database, proteins where further characterized using the online tools described in Table S4. Proteins were manually categorized into functional groups based on their annotations and on protein function information in the UniProt (128), NCBI (https://www.ncbi.nlm.nih.gov/), and InterPro (129) databases. We used the Transporter Automatic Annotation Pipeline (TransAAP) (http://www.membranetransport.org/transportDB2/TransAAP_login.html) of TransportDB2 (130) and TCDB (131) with gblast 2 (http://www.tcdb.org/labsoftware.php) to annotate transporters in the *Riftia*1 symbiont metagenome database. To detect potential antimicrobial peptides (AMPs) among the host proteins, we searched the detected host proteins against the antimicrobial peptide database APD3 (83) using BLASTP (112) in BLAST+ 2.7.1 (132). Results were filtered for identity of >75% and E value of <0.005. We screened the *Riftia* proteome for homologs of known autophagy-related *Drosophila melanogaster* proteins (as listed in reference 133) by BLAST searching (BLASTP [112] in BLAST+ 2.8.1 [132]) the *Riftia* host proteome against the respective *Drosophila* amino acid sequences (Table S1b).

(iii) SMART analysis of eukaryote-like and potential interaction domains. We used the SMART tool (134) to screen the *Riftia* symbiont protein database for proteins and domains that could be involved in symbiont-host interactions. Structures that 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 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-bacterium interactions in other organisms based on the literature. All annotations we included are given in Table S2. The organisms we used for comparison and their associated proteome accession numbers can be found in Table S1d. Proteins with structures that did not pass the threshold criterion in SMART were removed.

(iv) 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.

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. (135). 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 combined host and symbiont database have been deposited to the ProteomeXchange Consortium via the PRIDE (136) partner repository with the data set 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/bioproject/PRJNA534438/).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mBio.02243-19.

TEXT S1, PDF file, 1.9 MB.
TABLE S1, XLSX file, 1.8 MB.
TABLE S2, PDF file, 0.4 MB.
TABLE S3, PDF file, 0.3 MB.
TABLE S4, PDF file, 0.4 MB.

ACKNOWLEDGMENTS

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 U.S. 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. Marie-Katherin Zühlke provided helpful feedback on figure generation. 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 the Natural Sciences and Engineering Research Council of Canada NSERC through a Banting Fellowship (M.K.), the U.S. 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. was supported by a grant from the 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 DFG CRC1182 "Origin and Function of Metaorganisms," subprojects B2, Z3, and INF.

REFERENCES

1. Bosch TCG, McFall-Ngai MJ. 2011. Metaorganisms as the new frontier. Zoology 114:185–190. https://doi.org/10.1016/j.zool.2011.04.001.
2. Bang C, Dagan T, Deines P, Dubilier N, Duschl WJ, Fraune S, Hentschel U, Hirt H, Hülter N, Lachnit T, Picazo D, Pita L, Pogoreutz C, Rädecker N, Saad MM, Schmitz RA, Schulenburg H, Voolstra CR, Weiland-Bräuer N, Ziegler M, Bosch T. 2018. Metaorganisms in extreme environments: do microbes play a role in organismal adaptation? Zoology 127:1–19. https://doi.org/10.1016/j.zool.2018.02.004.
3. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF, Hentschel U, King N, Kjelleberg S, Knoll AH, Kremer N, Mazmanian SK, Metcalf JL, Nealson K, Pierce NE, Rawls JF, Reid A, Ruby EG, Rumpho M, Sanders JG, Tautz D, Wernegreen JJ. 2013. Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci U S A 110:3229–3236. https://doi.org/10.1073/pnas.1218525110.
4. Moya A, Peretó J, Gil R, Latorre A. 2008. Learning how to live together: genomic insights into prokaryote–animal symbioses. Nat Rev Genet 9:218–229. https://doi.org/10.1038/nrg2319.
5. Feldhaar H, Gross R. 2009. Insects as hosts for mutualistic bacteria. Int J Med Microbiol 299:1–8. https://doi.org/10.1016/j.ijmm.2008.05.010.
6. Nyholm SV, Graf J. 2012. Knowing your friends: invertebrate innate immunity fosters beneficial bacterial symbioses. Nat Rev Microbiol 10:815–827. https://doi.org/10.1038/nrmicro2894.
7. Hentschel U, Steinert M, Hacker J. 2000. Common molecular mechanisms of symbiosis and pathogenesis. Trends Microbiol 8:226–231. https://doi.org/10.1016/s0966-842x(00)01758-3.

8. Dale C, Moran NA. 2006. Molecular interactions between bacterial symbionts and their hosts. Cell 126:453–465. https://doi.org/10.1016/j.cell.2006.07.014.
9. McFall-Ngai M. 2008. Are biologists in "future shock"? Symbiosis integrates biology across domains. Nat Rev Microbiol 6:789–792. https://doi.org/10.1038/nrmicro1982.
10. Webster NS. 2014. Cooperation, communication, and co-evolution: grand challenges in microbial symbiosis research. Front Microbiol 5:164. https://doi.org/10.3389/fmicb.2014.00164.
11. Dubilier N, Bergin C, Lott C. 2008. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. Nat Rev Microbiol 6:725–740. https://doi.org/10.1038/nrmicro1992.
12. Felbeck H. 1981. Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila* Jones (Vestimentifera). Science 213:336–338. https://doi.org/10.1126/science.213.4505.336.
13. Cavanaugh CM, Gardiner SL, Jones ML, Jannasch HW, Waterbury JB. 1981. Prokaryotic cells in the hydrothermal vent tube worm *Riftia pachyptila* Jones: possible chemoautotrophic symbionts. Science 213:340–342. https://doi.org/10.1126/science.213.4505.340.
14. Hand SC. 1987. Trophosome ultrastructure and the characterization of isolated bacteriocytes from invertebrate-sulfur bacteria symbioses. Biol Bull 173:260–276. https://doi.org/10.2307/1541878.
15. Felbeck H, Childress JJ, Somero GN. 1981. Calvin-Benson cycle and sulphide oxidation enzymes in animals from sulphide-rich habitats. Nature 293:291–293. https://doi.org/10.1038/293291a0.
16. Stewart FJ, Cavanaugh CM. 2005. Symbiosis of thioautotrophic bacteria with *Riftia pachyptila*, p 197–225. *In* Overmann J (ed), Molecular basis of symbiosis. Progress in molecular and subcellular biology, vol 41. Springer, Berlin, Germany.
17. Lutz RA, Shank TM, Fornari DJ, Haymon RM, Lilley MD, Von Damm KL, Desbruyeres D. 1994. Rapid growth at deep-sea vents. Nature 371:663–664. https://doi.org/10.1038/371663a0.
18. Jones ML. 1981. *Riftia pachyptila* Jones: observations on the vestimentiferan worm from the Galápagos Rift. Science 213:333–336. https://doi.org/10.1126/science.213.4505.333.
19. Robidart JC, Bench SR, Feldman RA, Novoradovsky A, Podell SB, Gaasterland T, Allen EE, Felbeck H. 2008. Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed through metagenomics. Environ Microbiol 10:727–737. https://doi.org/10.1111/j.1462-2920.2007.01496.x.
20. Polzin J, Arevalo P, Nussbaumer T, Polz MF, Bright M. 2019. Polyclonal symbiont populations in hydrothermal vent tubeworms and the environment. Proc Biol Sci 286:20181281. https://doi.org/10.1098/rspb.2018.1281.
21. Distel DL, Lane DJ, Olsen GJ, Giovannoni SJ, Pace B, Pace NR, Stahl DA, Felbeck H. 1988. Sulfur-oxidizing bacterial endosymbionts: analysis of phylogeny and specificity by 16S rRNA sequences. J Bacteriol 170:2506–2510. https://doi.org/10.1128/jb.170.6.2506-2510.1988.
22. Dover CLV. 2000. The ecology of deep-sea hydrothermal vents. Princeton University Press, Princeton, NJ.
23. Fisher CR, Childress JJ, Minnich E. 1989. Autotrophic carbon fixation by the chemoautotrophic symbionts of *Riftia pachyptila*. Biol Bull 177:372–385. https://doi.org/10.2307/1541597.
24. Robidart JC, Roque A, Song P, Girguis PR. 2011. Linking hydrothermal geochemistry to organismal physiology: physiological versatility in *Riftia pachyptila* from sedimented and basalt-hosted vents. PLoS One 6:e21692. https://doi.org/10.1371/journal.pone.0021692.
25. Markert S, Gardebrecht A, Felbeck H, Sievert SM, Klose J, Becher D, Albrecht D, Thürmer A, Daniel R, Kleiner M, Hecker M, Schweder T. 2011. Status quo in physiological proteomics of the uncultured *Riftia pachyptila* endosymbiont. Proteomics 11:3106–3117. https://doi.org/10.1002/pmic.201100059.
26. Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R, Hourdez S, Girguis PR, Wankel SD, Barbe V, Pelletier E, Fink D, Borowski C, Bach W, Dubilier N. 2011. Hydrogen is an energy source for hydrothermal vent symbioses. Nature 476:176–180. https://doi.org/10.1038/nature10325.
27. Pflugfelder B, Fisher CR, Bright M. 2005. The color of the trophosome: elemental sulfur distribution in the endosymbionts of *Riftia pachyptila* (Vestimentifera; Siboglinidae). Mar Biol 146:895–901. https://doi.org/10.1007/s00227-004-1500-x.
28. Wilmot DB, Jr, Vetter RD. 1990. The bacterial symbiont from the hydrothermal vent tubeworm *Riftia pachyptila* is a sulfide specialist. Mar Biol 106:273–283. https://doi.org/10.1007/BF01314811.
29. Scott KM, Boller AJ, Dobrinski KP, Le Bris N. 2012. Response of hydrothermal vent vestimentiferan *Riftia pachyptila* to differences in habitat chemistry. Mar Biol 159:435–442. https://doi.org/10.1007/s00227-011-1821-5.
30. Liu H, Wang H, Cai S, Zhang H. 2017. A novel ω3-desaturase in the deep sea giant tubeworm *Riftia pachyptila*. Mar Biotechnol 19:345–350. https://doi.org/10.1007/s10126-017-9753-9.
31. Drozdov AL, Galkin SV. 2012. Morphology of gametes and insemination in the vestimentiferan *Riftia pachyptila*. OJMS 02:96–102. https://doi.org/10.4236/ojms.2012.23013.
32. Sanchez S, Hourdez S, Lallier FH. 2007. Identification of proteins involved in the functioning of *Riftia pachyptila* symbiosis by Subtractive Suppression Hybridization. BMC Genomics 8:337. https://doi.org/10.1186/1471-2164-8-337.
33. Bright M, Lallier FH. 2010. The biology of vestimentiferan tubeworms. Oceanogr Mar Biol 48:213–266. https://doi.org/10.1201/EBK1439821169-c4.
34. Gardebrecht A, Markert S, Sievert SM, Felbeck H, Thürmer A, Albrecht D, Wollherr A, Kabisch J, Le Bris N, Lehmann R, Daniel R, Liesegang H, Hecker M, Schweder T. 2012. Physiological homogeneity among the endosymbionts of *Riftia pachyptila* and *Tevnia jerichonana* revealed by proteogenomics. ISME J 6:766–776. https://doi.org/10.1038/ismej.2011.137.
35. Markert S, Arndt C, Felbeck H, Becher D, Sievert SM, Hügler M, Albrecht D, Robidart J, Bench S, Feldman RA, Hecker M, Schweder T. 2007. Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila*. Science 315:247–250. https://doi.org/10.1126/science.1132913.
36. Bright M, Keckeis H, Fisher CR. 2000. An autoradiographic examination of carbon fixation, transfer and utilization in the *Riftia pachyptila* symbiosis. Mar Biol 136:621–632. https://doi.org/10.1007/s002270050722.
37. Pflugfelder B, Cary SC, Bright M. 2009. Dynamics of cell proliferation and apoptosis reflect different life strategies in hydrothermal vent and cold seep vestimentiferan tubeworms. Cell Tissue Res 337:149–165. https://doi.org/10.1007/s00441-009-0811-0.
38. Boetius A, Felbeck H. 1995. Digestive enzymes in marine invertebrates from hydrothermal vents and other reducing environments. Mar Biol 122:105–113. https://doi.org/10.1007/BF00349283.
39. Ponnudurai R, Kleiner M, Sayavedra L, Petersen JM, Moche M, Otto A, Becher D, Takeuchi T, Satoh N, Dubilier N, Schweder T, Markert S. 2017. Metabolic and physiological interdependencies in the *Bathymodiolus azoricus* symbiosis. ISME J 11:463–477. https://doi.org/10.1038/ismej.2016.124.
40. Streams ME, Fisher CR, Fiala-Médioni A. 1997. Methanotrophic symbiont location and fate of carbon incorporated from methane in a hydrocarbon seep mussel. Mar Biol 129:465–476. https://doi.org/10.1007/s002270050187.
41. König S, Le Guyader H, 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. https://doi.org/10.1002/jemt.22458.
42. Caro A, Got P, Bouvy M, Troussellier M, 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. https://doi.org/10.1128/AEM.02659-08.
43. Wippler J, Kleiner M, Lott C, Gruhl A, Abraham PE, Giannone RJ, Young JC, Hettich RL, Dubilier N. 2016. Transcriptomic and proteomic insights into innate immunity and adaptations to a symbiotic lifestyle in the gutless marine worm *Olavius algarvensis*. BMC Genomics 17:942. https://doi.org/10.1186/s12864-016-3293-y.
44. Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO, Boffelli D, Anderson IJ, Barry KW, Shapiro HJ, Szeto E, Kyrpides NC, Mussmann M, Amann R, Bergin C, Ruehland C, Rubin EM, Dubilier N. 2006. Symbiosis insights through metagenomic analysis of a microbial consortium. Nature 443:950–955. https://doi.org/10.1038/nature05192.
45. Hyttinen JMT, Niittykoski M, Salminen A, Kaarniranta K. 2013. Maturation of autophagosomes and endosomes: a key role for Rab7. Biochim Biophys Acta 1833:503–510. https://doi.org/10.1016/j.bbamcr.2012.11.018.
46. Chavrier P, Parton R, Hauri H, Simons K, Zerial M. 1990. Localization of low molecular weight GTP binding proteins to exocytic and endocytic compartments. Cell 62:317–329. https://doi.org/10.1016/0092-8674(90)90369-p.
47. Vieira OV, Botelho RJ, Grinstein S. 2002. Phagosome maturation: aging gracefully. Biochem J 366:689–704. https://doi.org/10.1042/BJ20020691.

48. Marchetti A, Mercanti V, Cornillon S, Alibaud L, Charette SJ, Cosson P. 2004. Formation of multivesicular endosomes in *Dictyostelium*. J Cell Sci 117:6053–6059. https://doi.org/10.1242/jcs.01524.
49. Bright M, Sorgo A. 2003. Ultrastructural reinvestigation of the trophosome in adults of *Riftia pachyptila* (Annelida, Siboglinidae). Invertebr Biol 122:345–366. https://doi.org/10.1111/j.1744-7410.2003.tb00099.x.
50. Simonet P, Gaget K, Balmand S, Ribeiro Lopes M, Parisot N, Buhler K, Duport G, Vulsteke V, Febvay G, Heddi A, Charles H, Callaerts P, Calevro F. 2018. Bacteriocyte cell death in the pea aphid/*Buchnera* symbiotic system. Proc Natl Acad Sci U S A 115:E1819–E1828. https://doi.org/10.1073/pnas.1720237115.
51. Bröker LE, Huisman C, Span SW, Rodriguez A, Kruyt FA, 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. https://doi.org/10.1158/0008-5472.can-03-3060.
52. Man SM, Kanneganti T-D. 2016. Regulation of lysosomal dynamics and autophagy by CTSB/cathepsin B. Autophagy 12:2504–2505. https://doi.org/10.1080/15548627.2016.1239679.
53. Felbeck H, Jarchow J. 1998. Carbon release from purified chemoautotrophic bacterial symbionts of the hydrothermal vent tubeworm *Riftia pachyptila*. Physiol Zool 71:294–302. https://doi.org/10.1086/515931.
54. Zal F, Lallier FH, Green BN, Vinogradov SN, Toulmond A. 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. https://doi.org/10.1074/jbc.271.15.8875.
55. Zal F, Leize E, Lallier FH, Toulmond A, Van Dorsselaer A, Childress J. 1998. S-sulfohemoglobin and disulfide exchange: the mechanisms of sulfide binding by *Riftia pachyptila* hemoglobins. Proc Natl Acad Sci U S A 95:8997–9002. https://doi.org/10.1073/pnas.95.15.8997.
56. Flores JF, Fisher CR, Carney SL, Green BN, Freytag JK, Schaeffer SW, Royer WE. 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. https://doi.org/10.1073/pnas.0407455102.
57. Arp AJ, Childress JJ, Vetter RD. 1987. The sulphide-binding protein in the blood of the vestimentiferan tube-worm, *Riftia pachyptila*, is the extracellular haemoglobin. J Exp Biol 128:139–158.
58. Bailly X, Vinogradov S. 2005. The sulfide binding function of annelid hemoglobins: relic of an old biosystem? J Inorg Biochem 99:142–150. https://doi.org/10.1016/j.jinorgbio.2004.10.012.
59. Hourdez S, Weber RE. 2005. Molecular and functional adaptations in deep-sea hemoglobins. J Inorg Biochem 99:130–141. https://doi.org/10.1016/j.jinorgbio.2004.09.017.
60. Hentschel U, Felbeck H. 1993. Nitrate respiration in the hydrothermal vent tubeworm *Riftia pachyptila*. Nature 366:338–340. https://doi.org/10.1038/366338a0.
61. Pitcher RS, Brittain T, Watmough NJ. 2002. Cytochrome cbb_3 oxidase and bacterial microaerobic metabolism. Biochem Soc Trans 30:653–658. https://doi.org/10.1042/bst0300653.
62. Mangum CP. 1992. Physiological function of the hemerythrins, p 173–192. *In* Mangum CP (ed), Blood and tissue oxygen carriers. Advances in comparative and environmental physiology, vol 13. Springer, Berlin, Germany.
63. De Cian M-C, Andersen AC, Bailly X, Lallier FH. 2003. Expression and localization of carbonic anhydrase and ATPases in the symbiotic tubeworm *Riftia pachyptila*. J Exp Biol 206:399–409. https://doi.org/10.1242/jeb.00074.
64. Goffredi SK, Girguis PR, Childress JJ, Desaulniers NT. 1999. Physiological functioning of carbonic anhydrase in the hydrothermal vent tubeworm *Riftia pachyptila*. Biol Bull 196:257–264. https://doi.org/10.2307/1542950.
65. Sanchez S, Andersen AC, Hourdez S, Lallier FH. 2007. Identification, sequencing, and localization of a new carbonic anhydrase transcript from the hydrothermal vent tubeworm *Riftia pachyptila*. FEBS J 274:5311–5324. https://doi.org/10.1111/j.1742-4658.2007.06050.x.
66. De Cian M-C, Andersen AC, Toullec J-Y, Biegala I, Caprais J-C, Shillito B, Lallier FH. 2003. 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. https://doi.org/10.1007/s00227-002-0931-5.
67. Felbeck H. 1985. CO_2 fixation in the hydrothermal vent tube worm *Riftia pachyptila* (Jones). Physiol Zool 58:272–281. https://doi.org/10.1086/physzool.58.3.30155998.
68. Minic Z, Simon V, Penverne B, Gaill F, Hervé G. 2001. Contribution of the bacterial endosymbiont to the biosynthesis of pyrimidine nucleotides in the deep-sea tube worm *Riftia pachyptila*. J Biol Chem 276:23777–23784. https://doi.org/10.1074/jbc.M102249200.
69. Minic Z, Hervé G. 2003. Arginine metabolism in the deep sea tube worm *Riftia pachyptila* and its bacterial endosymbiont. J Biol Chem 278:40527–40533. https://doi.org/10.1074/jbc.M307835200.
70. Brennan LJ, Keddie BA, Braig HR, Harris HL. 2008. The endosymbiont *Wolbachia pipientis* induces the expression of host antioxidant proteins in an *Aedes albopictus* cell line. PLoS One 3:e2083. https://doi.org/10.1371/journal.pone.0002083.
71. Zug R, 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. https://doi.org/10.3389/fmicb.2015.01201.
72. Hourdez S, Lallier FH. 2007. Adaptations to hypoxia in hydrothermal-vent and cold-seep invertebrates. Rev Environ Sci Biotechnol 6:143–159. https://doi.org/10.1007/s11157-006-9110-3.
73. Fields JHA, Quinn JF. 1981. Some theoretical considerations on cytosolic redox balance during anaerobiosis in marine invertebrates. J Theor Biol 88:35–45. https://doi.org/10.1016/0022-5193(81)90327-1.
74. Arndt C, Schiedek D, Felbeck H. 1998. Metabolic responses of the hydrothermal vent tube worm *Riftia pachyptila* to severe hypoxia. Mar Ecol Prog Ser 174:151–158. https://doi.org/10.3354/meps174151.
75. Elsbach P, Weiss J. 1998. Role of the bactericidal/permeability-increasing protein in host defence. Curr Opin Immunol 10:45–49. https://doi.org/10.1016/s0952-7915(98)80030-7.
76. Chen F, Krasity BC, Peyer SM, Koehler S, Ruby EG, Zhang X, McFall-Ngai MJ. 2017. Bactericidal permeability-increasing proteins shape host-microbe interactions. mBio 8:e00040-17. https://doi.org/10.1128/mBio.00040-17.
77. Van Loon CL, Van Strien EA. 1999. The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. Physiol Mol Plant Pathol 55:85–97. https://doi.org/10.1006/pmpp.1999.0213.
78. Asojo OA, Goud G, Dhar K, Loukas A, Zhan B, Deumic V, Liu S, Borgstahl GEO, Hotez PJ. 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. https://doi.org/10.1016/j.jmb.2004.12.023.
79. Eberle HB, Serrano RL, Füllekrug J, Schlosser A, Lehmann WD, Lottspeich F, Kaloyanova D, Wieland FT, Helms JB. 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.
80. Park IY, Park CB, Kim MS, Kim SC. 1998. Parasin I, an antimicrobial peptide derived from histone H2A in the cat fish, *Parasilurus asotus*. FEBS Lett 437:258–262. https://doi.org/10.1016/S0014-5793(98)01238-1.
81. Cho JH, Sung BH, Kim SC. 2009. Buforins: histone H2A-derived antimicrobial peptides from toad stomach. Biochim Biophys Acta 1788:1564–1569. https://doi.org/10.1016/j.bbamem.2008.10.025.
82. Rose F, Bailey K, Keyte JW, Chan WC, Greenwood D, Mahida YR. 1998. Potential role of epithelial cell-derived histone H1 proteins in innate antimicrobial defense in the human gastrointestinal tract. Infect Immun 66:3255–3263.
83. Wang G, Li X, Wang Z. 2016. APD3: the antimicrobial peptide database as a tool for research and education. Nucleic Acids Res 44:D1087–D1093. https://doi.org/10.1093/nar/gkv1278.
84. Brogden KA. 2005. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat Rev Microbiol 3:238–250. https://doi.org/10.1038/nrmicro1098.
85. Login FH, Balmand S, Vallier A, Vincent-Monégat C, Vigneron A, Weiss-Gayet M, Rochat D, Heddi A. 2011. Antimicrobial peptides keep insect endosymbionts under control. Science 334:362–365. https://doi.org/10.1126/science.1209728.
86. Bishop BM, Juba ML, Russo PS, Devine M, Barksdale SM, Scott S, Settlage R, Michalak P, Gupta K, Vliet K, Schnur JM, van Hoek ML. 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 16:1470–1482. https://doi.org/10.1021/acs.jproteome.6b00857.
87. Li J, Mahajan A, Tsai M-D. 2006. Ankyrin repeat: a unique motif mediating

protein-protein interactions. Biochemistry 45:15168–15178. https://doi.org/10.1021/bi062188q.
88. Nguyen M, Liu M, Thomas T. 2014. Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. Mol Ecol 23:1635–1645. https://doi.org/10.1111/mec.12384.
89. Pan X, Lührmann A, Satoh A, Laskowski-Arce M, Roy CR. 2008. Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. Science 320:1651–1654. https://doi.org/10.1126/science.1158160.
90. Reynolds D, Thomas T. 2016. Evolution and function of eukaryotic-like proteins from sponge symbionts. Mol Ecol 25:5242–5253. https://doi.org/10.1111/mec.13812.
91. Prota AE, Magiera MM, Kuijpers M, Bargsten K, Frey D, Wieser M, Jaussi R, Hoogenraad CC, Kammerer RA, Janke C, Steinmetz MO. 2013. Structural basis of tubulin tyrosination by tubulin tyrosine ligase. J Cell Biol 200:259–270. https://doi.org/10.1083/jcb.201211017.
92. Coates JC. 2003. Armadillo repeat proteins: beyond the animal kingdom. Trends Cell Biol 13:463–471. https://doi.org/10.1016/s0962-8924(03)00167-3.
93. Yoshimura SH, Hirano T. 2016. HEAT repeats–versatile arrays of amphiphilic helices working in crowded environments? J Cell Sci 129:3963–3970. https://doi.org/10.1242/jcs.185710.
94. Nyholm SV, Stewart JJ, Ruby EG, McFall-Ngai MJ. 2009. Recognition between symbiotic *Vibrio fischeri* and the haemocytes of *Euprymna scolopes*. Environ Microbiol 11:483–493. https://doi.org/10.1111/j.1462-2920.2008.01788.x.
95. Leighton TL, Buensuceso RNC, Howell PL, Burrows LL. 2015. Biogenesis of *Pseudomonas aeruginosa* type IV pili and regulation of their function. Environ Microbiol 17:4148–4163. https://doi.org/10.1111/1462-2920.12849.
96. Stone BJ, Kwaik Y. 1999. Natural competence for DNA transformation by *Legionella pneumophila* and its association with expression of type IV pili. J Bacteriol 181:1395–1402.
97. Hager AJ, Bolton DL, Pelletier MR, Brittnacher MJ, Gallagher LA, Kaul R, Skerrett SJ, Miller SI, Guina T. 2006. Type IV pili-mediated secretion modulates *Francisella virulence*. Mol Microbiol 62:227–237. https://doi.org/10.1111/j.1365-2958.2006.05365.x.
98. Davidson SK, Dulla GF, Go RA, Stahl DA, Pinel N. 2014. Earthworm symbiont *Verminephrobacter eiseniae* mediates natural transformation within host egg capsules using type IV pili. Front Microbiol 5:546. https://doi.org/10.3389/fmicb.2014.00546.
99. Pollack-Berti A, Wollenberg MS, Ruby EG. 2010. Natural transformation of *Vibrio fischeri* requires tfoX and tfoY. Environ Microbiol 12:2302–2311. https://doi.org/10.1111/j.1462-2920.2010.02250.x.
100. Stewart FJ, Newton ILG, Cavanaugh CM. 2005. Chemosynthetic endosymbioses: adaptations to oxic-anoxic interfaces. Trends Microbiol 13:439–448. https://doi.org/10.1016/j.tim.2005.07.007.
101. Gilbert SF, Sapp J, Tauber AI. 2012. A symbiotic view of life: we have never been individuals. Q Rev Biol 87:325–341. https://doi.org/10.1086/668166.
102. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
103. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
104. Egas C, Pinheiro M, Gomes P, Barroso C, Bettencourt R. 2012. The transcriptome of *Bathymodiolus azoricus* gill reveals expression of genes from endosymbionts and free-living deep-sea bacteria. Mar Drugs 10:1765–1783. https://doi.org/10.3390/md10081765.
105. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
106. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
107. Schmieder R, Edwards R. 2011. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS One 6:e17288. https://doi.org/10.1371/journal.pone.0017288.
108. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren B, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652. https://doi.org/10.1038/nbt.1883.
109. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
110. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512. https://doi.org/10.1038/nprot.2013.084.
111. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279–D285. https://doi.org/10.1093/nar/gkv1344.
112. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
113. Eddy SR. 2009. A new generation of homology search tools based on probabilistic inference. Genome Inform 23:205–211. https://doi.org/10.1142/9781848165632_0019.
114. Huang Y, Niu B, Gao Y, Fu L, Li W. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26:680–682. https://doi.org/10.1093/bioinformatics/btq003.
115. Kleiner M, Dong X, Hinzke T, Wippler J, Thorson E, Mayer B, Strous M. 2018. Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. Proc Natl Acad Sci U S A 115:E5576–E5584. https://doi.org/10.1073/pnas.1722325115.
116. The Global Proteome Machine Organization. 2012. The global proteome machine: cRAP protein sequences. https://www.thegpm.org/crap/.
117. Wiśniewski JR, Zougman A, Nagaraj N, Mann M. 2009. Universal sample preparation method for proteome analysis. Nat Methods 6:359–362. https://doi.org/10.1038/nmeth.1322.
118. Hamann E, Gruber-Vodicka H, Kleiner M, Tegetmeyer HE, Riedel D, Littmann S, Chen J, Milucka J, Viehweger B, Becker KW, Dong X, Stairs CW, Hinrichs K-U, Brown MW, Roger AJ, Strous M. 2016. Environmental Breviatea harbour mutualistic *Arcobacter* epibionts. Nature 534:254–258. https://doi.org/10.1038/nature18297.
119. Hinzke T, Kouris A, Hughes R-A, Strous M, Kleiner M. 2019. More is not always better: evaluation of 1D and 2D-LC-MS/MS methods for metaproteomics. Front Microbiol 10:238. https://doi.org/10.3389/fmicb.2019.00238.
120. Oberg AL, Vitek O. 2009. Statistical design of quantitative mass spectrometry-based proteomic profiling experiments. J Proteome Res 8:2144–2156. https://doi.org/10.1021/pr8010099.
121. Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, Washburn MP. 2006. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. J Proteome Res 5:2339–2347. https://doi.org/10.1021/pr060161n.
122. Mueller RS, Denef VJ, Kalnejais LH, Suttle KB, Thomas BC, Wilmes P, Smith RL, Nordstrom DK, McCleskey RB, Shah MB, VerBerkmoes NC, Hettich RL, Banfield JF. 2010. Ecological distribution and population physiology defined by proteomics in a natural microbial community. Mol Syst Biol 6:374. https://doi.org/10.1038/msb.2010.30.
123. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140. https://doi.org/10.1093/bioinformatics/btp616.
124. R Core Team. 2017. R: a language and environment for statistical computing. http://www.R-project.org/. R Foundation for Statistical Computing, Vienna, Austria.
125. Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32:2847–2849. https://doi.org/10.1093/bioinformatics/btw313.
126. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. 2014. UpSet: visualization of intersecting sets. IEEE Trans Vis Comput Graph 20:1983–1992. https://doi.org/10.1109/TVCG.2014.2346248.
127. Kleiner M, Thorson E, Sharp CE, Dong X, Liu D, Li C, Strous M. 2017. Assessing species biomass contributions in microbial communities via metaproteomics. Nat Commun 8:1558. https://doi.org/10.1038/s41467-017-01544-x.

128. The UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Res 45:D158–D169. https://doi.org/10.1093/nar/gkw1099.
129. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang H-Y, Dosztányi Z, El-Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H, Huang X, Letunic I, Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka G, Orengo CA, Park Y, Pesseat S, Piovesan D, Potter SC, Rawlings ND, Redaschi N, Richardson L, Rivoire C, Sangrador-Vegas A, Sigrist C, Sillitoe I, Smithers B, Squizzato S, Sutton G, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Xenarios I, Yeh L-S, Young S-Y, Mitchell AL. 2017. InterPro in 2017–beyond protein family and domain annotations. Nucleic Acids Res 45:D190–D199. https://doi.org/10.1093/nar/gkw1107.
130. Elbourne LDH, Tetu SG, Hassan KA, Paulsen IT. 2017. TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. Nucleic Acids Res 45:D320–D324. https://doi.org/10.1093/nar/gkw1068.
131. Saier MH, Jr, Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G. 2016. The transporter classification database (TCDB): recent advances. Nucleic Acids Res 44:D372–D379. https://doi.org/10.1093/nar/gkv1103.
132. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.
133. Chang Y-Y, Neufeld TP. 2010. Autophagy takes flight in Drosophila. FEBS Lett 584:1342–1349. https://doi.org/10.1016/j.febslet.2010.01.006.
134. Letunic I, Bork P. 2018. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res 46:D493–D496. https://doi.org/10.1093/nar/gkx922.
135. Hammerschmidt S, Wolff S, Hocke A, Rosseau S, Müller E, Rohde M. 2005. Illustration of pneumococcal polysaccharide capsule during adherence and invasion of epithelial cells. Infect Immun 73:4653–4667. https://doi.org/10.1128/IAI.73.8.4653-4667.2005.
136. Vizcaíno JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, Xu Q-W, Wang R, Hermjakob H. 2016. 2016 update of the PRIDE database and its related tools. Nucleic Acids Res 44:D447–D456. https://doi.org/10.1093/nar/gkv1145.
137. Labonté JM, Pachiadaki M, Fergusson E, McNichol J, Grosche A, Gulmann LK, Vetriani C, Sievert SM, Stepanauskas R. 2019. Single cell genomics-based analysis of gene content and expression of prophages in a diffuse-flow deep-sea hydrothermal system. Front Microbiol 10:1262. https://doi.org/10.3389/fmicb.2019.01262.

PUBLICATION II

The publication “More is not always better: Evaluation of 1D and 2D-LC-MS/MS methods for metaproteomics” provides an in-depth evaluation and comparison of 1D and 2D liquid chromatography separation methods for metaproteomics.

Supplementary data is provided in the CD.

frontiers in Microbiology

METHODS
published: 14 February 2019
doi: 10.3389/fmicb.2019.00238

More Is Not Always Better: Evaluation of 1D and 2D-LC-MS/MS Methods for Metaproteomics

Tjorven Hinzke[1,2,3], Angela Kouris[1], Rebecca-Ayme Hughes[4], Marc Strous[1] and Manuel Kleiner[1,4]**

[1] Department of Geoscience, University of Calgary, Calgary, AB, Canada, [2] Institute of Pharmacy, Department of Pharmaceutical Biotechnology, University of Greifswald, Greifswald, Germany, [3] Institute of Marine Biotechnology e.V., Greifswald, Germany, [4] Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC, United States

Metaproteomics, the study of protein expression in microbial communities, is a versatile tool for environmental microbiology. Achieving sufficiently high metaproteome coverage to obtain a comprehensive picture of the activities and interactions in microbial communities is one of the current challenges in metaproteomics. An essential step to maximize the number of identified proteins is peptide separation via liquid chromatography (LC) prior to mass spectrometry (MS). Thorough optimization and comparison of LC methods for metaproteomics are, however, currently lacking. Here, we present an extensive development and test of different 1D and 2D-LC approaches for metaproteomic peptide separations. We used fully characterized mock community samples to evaluate metaproteomic approaches with very long analytical columns (50 and 75 cm) and long gradients (up to 12 h). We assessed a total of over 20 different 1D and 2D-LC approaches in terms of number of protein groups and unique peptides identified, peptide spectrum matches (PSMs) generated, the ability to detect proteins of low-abundance species, the effect of technical replicate runs on protein identifications and method reproducibility. We show here that, while 1D-LC approaches are faster and easier to set up and lead to more identifications per minute of runtime, 2D-LC approaches allow for a higher overall number of identifications with up to >10,000 protein groups identified. We also compared the 1D and 2D-LC approaches to a standard GeLC workflow, in which proteins are pre-fractionated via gel electrophoresis. This method yielded results comparable to the 2D-LC approaches, however with the drawback of a much increased sample preparation time. Based on our results, we provide recommendations on how to choose the best LC approach for metaproteomics experiments, depending on the study aims.

Keywords: microbiota, microbiome, mock community, method evaluation, microbial ecology, Q Exactive, liquid chromatography, GeLC

OPEN ACCESS

Edited by:
Martin G. Klotz, Washington State University, United States

Reviewed by:
Florian-Alexander Herbst, Aalborg University, Denmark
Dirk Benndorf, Otto-von-Guericke-Universität Magdeburg, Germany

****Correspondence:***
Tjorven Hinzke
tjorven.hinzke@outlook.com
Manuel Kleiner
manuel_kleiner@ncsu.edu

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *14 September 2018*
Accepted: *28 January 2019*
Published: *14 February 2019*

Citation:
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

Abbreviations: 1D-LC-MS/MS, one-dimensional liquid chromatography coupled to tandem mass spectrometry; 2D-LC-MS/MS, two-dimensional liquid chromatography coupled to tandem mass spectrometry; ABC, ammonium bicarbonate; APS, ammonium persulfate; DTT, dithiothreitol; FASP, filter-aided sample preparation; GeLC-MS/MS, 1D gel electrophoresis coupled to tandem mass spectrometry; HPLC, high performance liquid chromatography; LC, liquid chromatography; MS, mass spectrometry; MS/MS, tandem mass spectrometry; NSAF, normalized spectral abundance factor; pI, isoelectric point; PSM, peptide spectrum match; protein-SIF, protein stable isotope fingerprinting; RP, reversed-phase; SCX, strong cation exchange; SDS, sodium dodecyl sulfate; SIP, stable isotope probing; TEMED, tetramethylethylenediamine.

INTRODUCTION

Metaproteomics, the analysis of expressed proteins in a microbial community (Wilmes and Bond, 2004), is a powerful tool which has enabled new insights into the role of microorganisms in a variety of environments. Metaproteomics not only provides information about gene expression of uncultured microorganisms, it can also be used for direct activity measurements with stable isotope probing (SIP) approaches (Jehmlich et al., 2016), the determination of carbon sources of individual species in microbial communities using stable isotope fingerprinting (protein-SIF) (Kleiner et al., 2018) and the assessment of community structure based on proteinaceous biomass (Kleiner et al., 2017). Numerous systems have been studied using metaproteomics, including acid mine drainage biofilms (Ram et al., 2005; Belnap et al., 2010), lichen (Schneider et al., 2011), marine symbioses (Markert et al., 2011; Gardebrecht et al., 2012; Kleiner et al., 2012; Wippler et al., 2016; Ponnudurai et al., 2017), plankton (Sowell et al., 2011), biogas plant communities (Heyer et al., 2016), and the human gut (Xiong et al., 2015a; Xiao et al., 2017). In the most common metaproteomic workflow, the protein mixture which is extracted from an environmental sample is digested into peptides. This peptide mixture is then analyzed in a mass spectrometer. There are numerous challenges associated with this workflow that must be addressed in order to obtain the largest possible number of identified and quantified proteins. The foremost challenge is high sample diversity and complexity. Others include finding a method for effective and unbiased cell lysis and protein extraction, generation of a database matching the sample, and correct assignment of peptides to proteins, especially for highly similar proteins from the same or different organisms (the protein inference problem) (Nesvizhskii and Aebersold, 2005; Muth et al., 2013; Wang et al., 2014; Xiong et al., 2015a; Kleiner et al., 2017). For two comprehensive reviews that provide more details on these challenges and how they can be addressed see VerBerkmoes et al. (2009) and Wilmes et al. (2015).

Here, we address the challenge posed by sample complexity, which translates into the large number of proteins or, ultimately, peptides that need to be analyzed. To increase the number of identifiable peptides and thereby proteins, it is crucial to reduce the sample complexity prior to measurement in the mass spectrometer (Taylor et al., 2009; Mostovenko et al., 2013; Weston et al., 2013). The sample complexity can be reduced by separating either proteins or peptides based on differing chemical properties. On-line separation is especially convenient, as the peptides are injected directly into the mass spectrometer after separation. In addition, on-line separations decrease bias and require less time, effort and material than manual off-line separations (Magdeldin et al., 2014; Camerini and Mauri, 2015; Richard et al., 2017). Frequently, one on-line peptide separation step is accomplished by reversed-phase liquid chromatography (RP-LC) of the peptides (Hsieh et al., 2013; Weston et al., 2013). To increase the separation – and thus the metaproteome coverage – an off-line sample separation can be used before the RP-LC step. The pre-separation of proteins by 1D SDS gel electrophoresis, or GeLC, is sometimes used in metaproteomics (Schneider et al., 2011; Stokke et al., 2012; König et al., 2016; Ponnudurai et al., 2017). Also possible is an off-line pre-separation of peptides (Keiblinger et al., 2012; Zheng et al., 2015). Alternatively to these off-line separations, a second on-line separation step can be used. For this additional on-line separation, a strong cation exchange (SCX) column can be added upstream of the RP column, resulting in 2D-LC separation (Taylor et al., 2009).

While multi-dimensional separation is currently the state-of-the-art approach for metaproteomics (Hettich et al., 2012), recent advances in column technology have made 1D-LC without additional separations a possible competitive alternative to multidimensional separations. For standard, single-organism proteomics it has been shown that longer columns enable extended 1D-LC runs, which achieve high resolution and at the same time use less sample material and take less MS runtime than multi-dimensional approaches (Thakur et al., 2011; Nagaraj et al., 2012; Pirmoradian et al., 2013).

One important difference between single-organism proteomics and multi-species metaproteomics is that, while in single-organism proteomics a fairly comprehensive set of peptides can be identified with a limited amount of MS run time, this is not the case in metaproteomics, where several orders of magnitude more peptides are present in a sample. This means that some proteomics methods cannot be immediately transferred to metaproteomics. For example, new methods such as BoxCar (Meier et al., 2018) have recently been developed, which allow to largely increase the number of quantified peptides in single-organism proteomics. This increased number of quantified peptides, however, can only be achieved by increasing the amount of time spent for the acquisition of MS spectra. Thus, less time is available for the MS/MS scans needed for peptide identification. BoxCar relies on the transfer of identifications between LC-MS/MS runs, as well as on spectral libraries to overcome the loss in MS/MS scans. Such a transfer of identifications between sample runs is, however, not feasible in metaproteomics, where each sample may contain different species and thus proteins. This makes the acquisition of maximal MS/MS scan numbers for peptide identifications indispensable in metaproteomics.

In this study, we compared the performance of 1D and on-line 2D-LC separation methods for metaproteomics. Additionally, we compared the performance of these on-line separations with the off-line GeLC method. We used a defined mock community, consisting of 32 strains and species, for this method comparison. We assessed the number of identified protein groups, unique peptides and peptide spectrum matches (PSMs), the detection of low-abundance species and the reproducibility of the methods. We show that, while 2D-LC methods can lead to more total identifications, they are not necessarily the best choice for metaproteomic studies. 1D-LC methods with long columns and gradients offer a suitable alternative in terms of throughput and ease of use. GeLC can be an alternative to 2D-LC, if no 2D-capable system is available. This data enables us to provide recommendations to researchers in the field of metaproteomics for the mass spectrometric experimental design based on their study goals.

MATERIALS AND METHODS

Mock Community Generation

We used a defined mock community sample to test and compare different LC separation methods. The mock community corresponded to the UNEVEN community described by Kleiner et al. (2017). This mock community was assembled to cover a large taxonomic range and was not meant to resemble any specific metaproteome. Additionally, the aim was to address potential challenges for metaproteomic approaches, such as (1) large differences in species abundances in terms of proteinaceous biomass in the community, which challenges the dynamic range of the approaches, (2) different cell types influencing extraction efficiencies and (3) difficulties with protein inference due to highly similar strains from one species and multiple species from the same genus. The mock community contains a total of 32 different microbial species and strains of all three domains of life as well as viruses. The species and strains are present in different abundances in the UNEVEN community, similar to what can be expected for a natural microbial community. The community was assembled from pure cultures as described by Kleiner et al. (2017) and the detailed composition can be found in **Supplementary Tables S3** and **S6**.

Peptide Sample Preparation

LC-MS/MS (On-Line Separation)

Cells in mock community samples were disrupted by bead-beating (6.0 m/s, 45 s) in lysing matrix B tubes (MP Biomedicals) in SDT lysis buffer [4% (w/v) SDS, 100 mM Tris-HCl pH 7.6, 0.1 M DTT] followed by heating to 95°C for 10 min. Tryptic digests of protein extracts were prepared according to the filter-aided sample preparation (FASP) protocol described by Wiśniewski et al. (2009). Peptides were desalted with Sep-Pak C18 Plus Light Cartridges (Waters). Acetonitrile from the peptide elution step was exchanged for 0.1% formic acid (v/v) using a centrifugal vacuum concentrator. The desalting step was necessary to enable binding of peptides to the SCX column during sample loading for the 2D-LC methods. Peptide concentrations were determined using the Pierce Micro BCA assay (Thermo Scientific Pierce) according to manufacturer's instructions. The peptide mixture was aliquoted and frozen at −80°C. For mass spectrometric analyses, fresh aliquots were regularly thawed and formic acid concentration increased to 0.2% (v/v). All aliquots used here were prepared at the same time from the same peptide mixture to eliminate between-sample variation.

GeLC-MS/MS (Gel-Based Separation)

To compare the on-line LC separation of peptides with the 1D pre-separation of proteins on an SDS gel, followed by 1D-LC separation (GeLC approach), aliquots of the UNEVEN mock community were prepared for GeLC as similarly as possible to the FASP approach above. Briefly, we disrupted cells with the same method as for the on-line separations, with the modification that no DTT was used in the lysis buffer. After centrifugation for 5 min at 21.000 × *g*, protein concentrations in the supernatant were determined using the Pierce BCA assay (Thermo Scientific Pierce) according to the manufacturer's instructions for the enhanced protocol. 30 μg of protein were mixed with loading buffer (4x Laemmli Sample Buffer, Bio-Rad; containing 50 mM DTT) and separated on 12% polyacrylamide gels [12% (v/v) acrylamide, 0.38 M Tris pH 8.8, 0.1% (w/v) SDS, 0.1% (w/v) APS, 0.04% (v/v) TEMED]. Gels were fixed for 15 min [40% (v/v) ethanol, 10% (v/v) glacial acetic acid], washed twice in deionized water and stained with QC Colloidal Coomassie Stain (Bio-Rad) for 2 h. After overnight destaining, each gel lane was cut into 10 equal-sized pieces. In-gel digestion of proteins and peptide elution was done as described previously by Eymann et al. (2004) with several small modifications. In brief, gel pieces were washed three times (15 min, 900 rpm shaking, 37°C) in 300 μl destaining solution (200 mM ABC in 50% v/v acetonitrile) and dried for 30 min in a vacuum centrifuge. Gel pieces were rehydrated for 30 min in an aqueous 2 ng μl^{-1} trypsin solution (sequencing grade modified trypsin, Thermo Scientific Pierce). Trypsin solution not absorbed by the gel pieces was removed and the gel pieces were incubated at 37°C overnight for digestion. Peptides were eluted by adding 40 μl of distilled water and applying ultrasound for 15 min in a sonication bath. The supernatant was transferred to HPLC vials for analysis. For two biological replicates of the UNEVEN mock community (U1 and U2), peptides from two gel lanes were combined to increase loading amount.

MS Analysis

1D-LC-MS/MS and GeLC-MS/MS

For 1D-LC-MS/MS and GeLC-MS/MS analysis, an UltiMateTM 3000 RSLCnano Liquid Chromatograph (Thermo Fisher Scientific) with two 10-port valves in the column oven was used to load the respective amount of peptide mixture (**Table 1**) with loading solvent A (2% acetonitrile, 0.05% trifluoroacetic acid) onto a 5 mm, 300 μm ID C18 Acclaim® PepMap100 pre-column (Thermo Fisher Scientific) at a flow rate of 5 $\mu l\ min^{-1}$. The pre-column was then switched into line with the analytical column, which was either a 50 cm × 75 μm analytical EASY-Spray column packed with PepMap RSLC C18, 2 μm material (Thermo Fisher Scientific), heated to 45°C (1D and GeLC), or a 75 cm × 75 μm analytical column with the same packing material (Thermo Fisher Scientific), heated to 60°C (1D-LC only). An EASY-Spray source connected the analytical column to a Q Exactive Plus hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific). Elution and separation of peptides on the analytical column was achieved at a flow rate of 225 $nl\ min^{-1}$ using a gradient of eluent A (0.1% formic acid) and eluent B (0.1% formic acid, 80% acetonitrile) with the times as specified in **Supplementary Table S1**. Eluting peptides were ionized using electrospray ionization. Two wash runs with injection of 20 μl acetonitrile and 99% eluent B and one blank run were done between samples to reduce carryover. Data was acquired in the Q Exactive Plus as described by Petersen et al. (2016).

2D-LC-MS/MS

For the 2D-LC-MS/MS experiments, the same LC as for the 1D experiments was used. The respective amount of peptide mixture (**Table 1**) was loaded with loading solvent B (2% acetonitrile, 0.5% formic acid) onto a 10 cm, 300 μm ID Poros

TABLE 1 | Overview of LC methods developed and tested in this study.

Method	Method overview[a]	Peptide loaded [μg]	Runtime [h][b]
2D-LC (all with 50 cm analytical column)			
2D|3salt$_{4.5}$	bt; 5, 100, 2000 mM (each 300 min)	4.5	20
2D|3salt$_{9}$	bt; 5, 100, 2000 mM (each 300 min)	9	20
2D|11salt$_{4.5}$	bt; 1, 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000 mM (each 120 min)	4.5	24
2D|11salt$_{9}$	bt; 1, 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000 mM (each 120 min)	9	24
2D|10pH_G1	bt; pH 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 7.0, 8.0 (each 120 min)	4.5	22
2D|10pH_G2$_{4.5}$	bt (300 min); pH 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 7.0, 8.0 (each 120 min)	4.5	25
2D|10pH_G2$_{9}$	bt (300 min); pH 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 7.0, 8.0 (each 120 min)	9	25
2D|5pH+2salt	bt (228 min); 1 mM (228 min); pH 3.0 (120 min), 4.0 (105 min), 4.5 (90 min), 5.0 (90 min), 6.0 (79 min); 2000 mM (60 min)	4.5	16.7
2D|6pH+1salt_G1	bt (228 min); pH 2.5 (228 min), 3.5 (120 min), 4.5 (90 min), 5.0 (90 min), 6.0 (79 min), 8.0 (60 min); 2000 mM (60 min)	4.5	15.9
2D|6pH+1salt_G2	bt (466 min); pH 2.5 (228 min), 3.5 (120 min), 4.5 (90 min), 5.0 (90 min), 6.0 (79 min), 8.0 (60 min); 2000 mM (60 min)	4.5	19.9
2D|8pH+1salt_G1$_{4.5}$	bt (300 min); pH 2.5 (120 min), 3.0 (120 min), 3.5 (120 min), 4.0 (105 min), 4.5 (105 min), 5.0 (105 min), 6.0 (105 min), 8.0 (80 min); 2000 mM (80 min)	4.5	20.7
2D|8pH+1salt_G1$_{9}$	bt (300 min); pH 2.5 (120 min), 3.0 (120 min), 3.5 (120 min), 4.0 (105 min), 4.5 (105 min), 5.0 (105 min), 6.0 (105 min), 8.0 (80 min); 2000 mM (80 min)	9	20.7
2D|8pH+1salt_G2	bt (228 min); pH 2.5 (228 min), 3.0 (120 min), 3.5 (120 min), 4.0 (105 min), 4.5 (90 min), 5.0 (90 min), 6.0 (79 min), 8.0 (60 min); 2000 mM (60 min)	4.5	19.7
2D|8pH+4salt$_{4.5}$	bt (300 min); 1 mM (120 min); pH 2.5 (120 min), 3.0 (120 min), 3.5 (120 min), 4.0 (105 min), 4.5 (105 min), 5.0 (105 min), 6.0 (105 min), 8.0 (80 min); 500 mM (80 min), 1000 mM (80 min), 2000 mM (80 min)	4.5	25.3
2D|8pH+4salt$_{9}$	bt (300 min); 1 mM (120 min); pH 2.5 (120 min), 3.0 (120 min), 3.5 (120 min), 4.0 (105 min), 4.5 (105 min), 5.0 (105 min), 6.0 (105 min), 8.0 (80 min); 500 mM (80 min); 1000 mM (80 min); 2000 mM (80 min)	9	25.3
2D|10pH+1salt	bt (228 min); pH 2.5 (228 min), 3.0 (120 min), 3.5 (120 min), 4.0 (105 min), 4.5 (90 min), 5.0 (90 min), 5.5 (90 min), 6.0 (79 min), 7.0 (79 min), 8.0 (60 min), 2000 mM (60 min)	4.5	22.5
2D|11salt[c]	bt; 1, 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000 mM (each 120 min)	11	24
1D-LC			
50 cm analytical column			
1D|8h_50$_{1.6}$	460min_Pre20 (460 min plus 20 min pre-equilibration)	1.6	8
1D|8h_50$_{2}$	460min_Pre20 (460 min plus 20 min pre-equilibration)	2	8
1D|8h_50$_{2.5}$	460min_Pre30 (460 min plus 30 min pre-equilibration to also work with the 75 cm column)	2.5	8.2
1D|12h_50	720min_Pre30 (720 min plus 30 min pre-equilibration to also work with the 75 cm column)	2.5	12.5
1D|4h[c]	260min_Pre20 (260 min plus 20 min pre-equilibration)	0.8–2	4.7
1D|8h[c]	460min_Pre20 (460 min plus 20 min pre-equilibration)	2	8
75 cm analytical column			
1D|8h_75	460min_Pre30 (460 min plus 30 min pre-equilibration)	2.5	8.2
1D|12h_75	720min_Pre30 (720 min plus 30 min pre-equilibration)	2.5	12.5
GeLC (50 cm analytical column)			
GeLC	10x (Load, 90 min)	(30–60)[d]	16.7

[a] *For description of the gradients, see* ***Supplementary Table S1****.*
bt, breakthrough; i.e., peptides that do not sufficiently bind to the SCX column and are directly eluted onto the C18 RP column before any salt or pH steps.
Concentrations correspond to NaCl concentrations in loading buffer B.
pH values correspond to commercial pH plugs (CTIBiphase buffers, Column Technology, Inc.).
[b] *Method runtime without intermediary times for washing etc., but with pre-equilibration for 1D methods.*
[c] *This method was run on four biological replicates as described by Kleiner et al. (2017).*
[d] *Protein concentration was determined before loading samples on the gel.*

10 S SCX column (Thermo Fisher Scientific) at a flow rate of 5 μl min^{-1}. The specific plumbing scheme used in the RSLCnano corresponded to the standard set up recommended by the manufacturer for on-line 2D salt plug separations[1]. During loading, the C18 pre-column (see above) was in-line downstream of the SCX column to capture peptides that did not bind to the SCX column (breakthrough). After loading, the C18 pre-column was switched in-line with the 50 cm × 75 μm analytical column (same as for 1D) and the breakthrough was separated using a gradient of eluent A and eluent B (**Supplementary Table S1**). Subsequently, elution of peptides from the SCX

[1]https://tools.thermofisher.com/content/sfs/manuals/87961-MAN-4820-4103-UltiMate3000-RSLCnano-Applicationsmanual-V1R3.pdf

to the C18 pre-column (same as for 1D-LC) took place by injection of 20 µl of salt plugs (different concentrations of NaCl in loading buffer B) or pH plugs (CTIBiphase buffers, Column Technology, Inc.) from the autosampler. The salt concentrations and pH steps were dependent on the method used (**Table 1**). The C18 pre-column was then again switched in-line with the analytical column and peptides separated with gradients of eluent A and B (**Supplementary Table S1**). Two washes of the SCX column (injection of 20 µl 4 M NaCl in loading solvent B, 100% eluent B), one RP column wash (injection of 20 µl acetonitrile, 99% eluent B) and one blank run were done between samples to reduce carryover. Data acquisition in the mass spectrometer was done as described by Petersen et al. (2016).

Protein Identification and Quantification

The same database as in Kleiner et al. (2017) was used, containing all protein sequences from the reference genomes of the organisms present in the mock community (**Supplementary Tables S3, S6**). The database also included the cRAP protein sequence database with common laboratory contaminants[2]. The database contained 123,100 protein sequences and is available from the PRIDE (Vizcaíno et al., 2016) repository (PXD008017).

The raw MS files were searched against this database using the MaxQuant software version 1.5.8.3 (Cox and Mann, 2008; Tyanova et al., 2016). MaxQuant identifies protein groups, which are defined as the set of all proteins that cannot be separated based on their detected peptides (Cox and Mann, 2008). At least one unique peptide was required for protein group identification, in addition to an FDR of 1% on the peptide and on the protein group level. All other parameters were left at default values.

For MaxQuant analyses of several runs together, the "Second peptides" and the "Match between runs" options were disabled. Where needed, we used the normalized spectral abundance factor (NSAF) (Zybailov et al., 2006) for quantification. To quantify the number of proteins identified per species and to calculate protein isoelectric point (pI) distributions, size distributions and NSAFs, we used the first protein in the Majority Protein ID assignment, if different proteins were present in the protein group.

Data Evaluation

For data evaluation, the proteinGroups.txt output from the MaxQuant searches was filtered as follows: protein groups with zero MS/MS (i.e., only identified by co-eluting 'second peptides'), or those identified by modification site only as well as reverse hits were removed. We exclusively defined protein groups which passed these criteria as being identified and used them for further analyses. Missing values were replaced by zero for calculations. For overlap calculations, a custom Venn diagram tool from the University of Gent was used[3]. Protein pIs were calculated using the package seqinR (Charif and Lobry, 2007) in R (R Core Team, 2018).

[2]http://thegpm.org/crap/

[3]http://bioinformatics.psb.ugent.be/webtools/Venn/

RESULTS AND DISCUSSION

Method Development

We tested different 1D and 2D-LC separation methods to find which ones are the best for metaproteomics. For the 1D-LC separations, we used C18 RP-LC with acetonitrile-based gradients. For 2D-LC, we used SCX LC with step-wise salt (Taylor et al., 2009) or pH bumps (Dai et al., 2005) as the first dimension and C18 RP-LC as the second dimension. The salt and pH bumps were injected directly from the autosampler, similar to what has been described by Taylor et al. (2009). Our system had separate SCX and C18 pre-columns and a switching valve that connected the C18 pre-column to the analytical column, while Taylor et al. (2009) used a biphasic pre-column connected to the analytical column in a vented-column setup.

We compared different gradient lengths and shapes for both 1D and 2D-LC as well as combinations of salt and pH steps for the 2D-LC (**Table 1**). We chose the different gradient and salt/pH bumps to achieve a roughly equal distribution of eluting peptides over the gradients and between steps.

Additionally, we compared the on-line 1D and 2D-separations to a GeLC approach, for which we used 12% SDS gels and 1D C18 RP-LC.

How to Get the Most Proteins?

We achieved the highest number of identified protein groups [i.e., sets of proteins which cannot be separated based on their peptides by MaxQuant (Cox and Mann, 2008)] with a 2D-LC method, 2D|10pH_G1 (**Figure 1A**). This method had 10 pH plugs, each eluted with a 120 min gradient. With 2D|10pH_G1 we identified a mean of 10,148 protein groups, based on 35,127 unique peptides.

Interestingly, two of the 1D methods performed better than several 2D methods. With the best 1D method (1D|12h_75) we identified 8,261 protein groups, corresponding to 31,283 unique peptides. With a 75 cm column, we increased the number of identifications for an 8 h as well as for a 12 h 1D run as compared to the 50 cm column (**Supplementary Table S2**).

The total number of identified unique peptides correlated well with the number of protein groups for all methods (**Figure 1A**). However, some methods provided more unique peptides per protein group than others. The best method in this regard was 1D|12h_75 with 3.8 unique peptides/protein group (**Supplementary Table S2**). This higher number of unique peptides per protein group increases the identification and quantification confidence of the protein groups identified in 1D|12h_75 as compared to the other methods (Mallick and Kuster, 2010).

The percentage of protein groups with one unique peptide and one peptide total (i.e., the unique peptide is the only peptide found) and only one PSM was mostly higher for the 2D methods (**Supplementary Table S2**), likely because in the 2D methods peptides of less abundant proteins were also targeted for identification (in addition to higher-abundant ones detected by 1D methods as well) and these peptides of low-abundant proteins have a lower likelihood of being detected. The datasets for which four biological replicates were measured generally had a higher

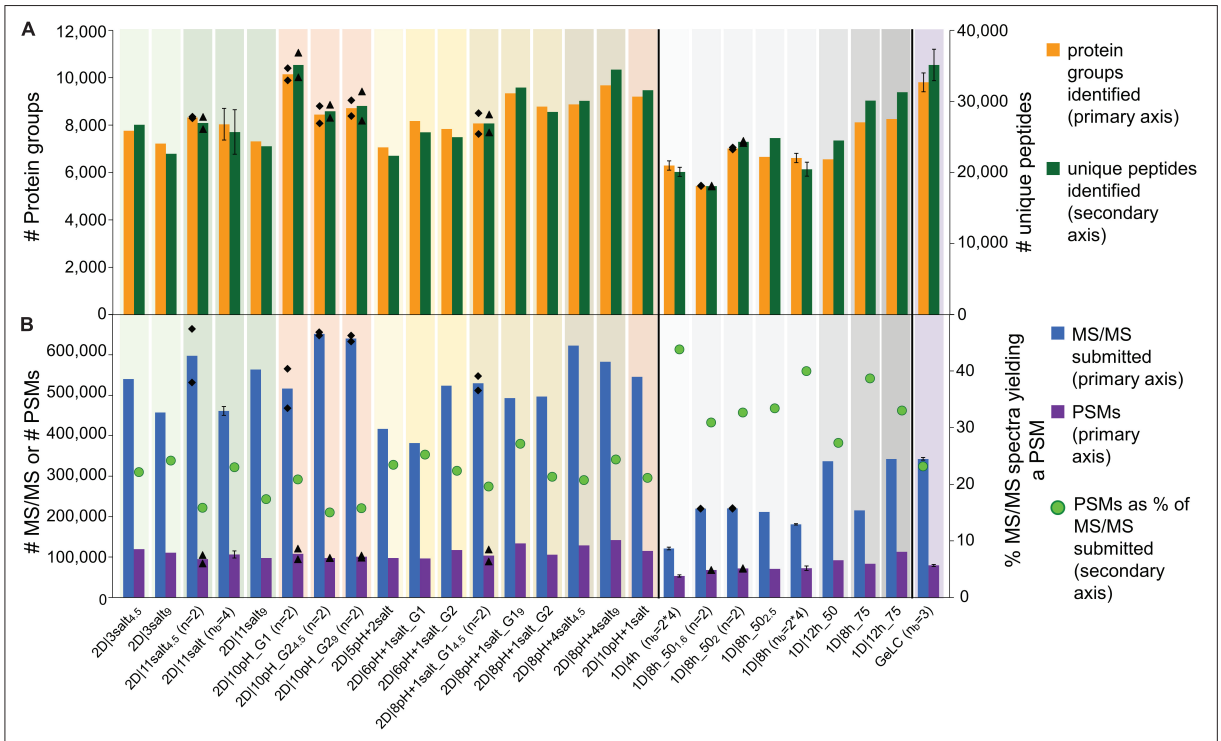

FIGURE 1 | Comparison of data amount obtained with different 1D and 2D-LC-MS/MS and the GeLC method. **(A)** Number of protein groups and unique peptides (secondary axis) identified by MaxQuant. **(B)** Number of submitted MS/MS, PSMs and the percentage of PSMs of submitted MS/MS (secondary axis). For $n = 2$, the means of two independent runs with the same method (bars) and the two individual values (black diamonds) are shown. For $n_b = 3$, $n_b = 4$ and $n_b = 2*4$, the means of three ($n_b = 3$) or four biological replicates ($n_b = 4$), or four biological replicates, measured in technical duplicate ($n_b = 2*4$), are shown. Error bars indicate standard deviations.

proportion of protein groups with one (unique) peptide or one PSM, which might be due to the additional variance introduced by biological replication.

In terms of time efficiency, i.e., protein groups identified per minute of gradient runtime (see also Köcher et al., 2011), the 1D|4h method outperformed every other method, with 22.5 protein groups identified per min. In comparison, the best 2D method in this regard (2D|6pH+1salt_G1) only reached about 8.6 protein groups per min and the second best 1D method (1D|8h_75) identified 16.6 proteins per min (**Supplementary Table S2**).

We also tested the effect of loading different amounts of digested sample on the identification numbers. The sample amount has to be high enough to allow for sufficient ion intensities for peptide identifications, while not so high as to overload the column and cause ion suppression. The amount of sample needed thus depends on the run length: longer runs need more sample to achieve high enough ion intensities over the whole run. We did see the different effects of loading not enough and loading too much sample: in some cases, the number of identified protein groups increased when we used more sample material (especially when we loaded 2 µg instead of 1.6 µg as measured by BCA assay for the 1D|8h_50 methods). For some methods, we identified fewer protein groups after loading more material. While 2 µg of peptide are sufficient for a 1D 8 h run, for 2D runs at least 4 µg should be used, as we still see an increase in identifications for 9 µg as compared to 4 µg for several 2D methods.

Overall, we observed good performance of 1D-LC separation approaches with a long gradient and long analytical column for metaproteomics. When comparing the number of identified protein groups between the 4 h 1D method and the 8 h 1D method run on the same column length (50 cm) and with at least 2 µg peptide loaded for the 8 h runs, we found that the 4 and 8 h runs both performed equally. The 4 h method only identified on average 6.2% less protein groups as compared to the 8 h methods (**Supplementary Figure S1** and **Supplementary Table S2**).

In the following, we will compare the data obtained in this study with existing literature data. We would like to add a word of caution though, since to our knowledge LC-MS/MS methods for metaproteomics have never been assessed in this depth before and both LC instruments and mass spectrometers have undergone major improvements over the last few years. Comparisons with numbers reported in the literature can thus only give a very rough impression of the comparative method

performance and must be interpreted with care. Similar to our study, where we identified up to over 8,000 proteins with 1D runs using 50 and 75 cm columns, other authors reported good performance of 1D methods using pure cultures or simpler communities. Nagaraj et al. (2012) identified, on average, over 3,900 yeast proteins by using a 4 h LC gradient with a 50 cm column. Richards et al. (2015) describe a method to identify up to 4,000 yeast proteins in 70 min using a 30 cm column. Likewise, Pirmoradian et al. (2013) detected over 4,800 protein groups in a human cell line with a 4 h run and a 50 cm column. In bacterial symbiont-containing gill tissue of a lucinid clam, Petersen et al. (2016) identified up to 1,400 proteins of the symbiont alone using a 50 cm column and a 460 min gradient. Another study of symbiont-containing mussel gills with the same column and gradient lengths led to over 7,700 identified proteins of host and symbionts (Rubin-Blum et al., 2017).

In contrast to our best 2D runs, which resulted in over 10,000 identified proteins, numbers reported in the literature for metaproteomes analyzed with 2D setups are lower. For example, a 2D-LC analysis of a symbiosis consisting of a marine oligochaete with at least four different bacterial species yielded a total of 4,355 protein groups (Wippler et al., 2016). In fermented fish microbiota, a total of 2,175 proteins were identified from different subsamples after 2D-LC (Ji et al., 2017). Two studies of human infant gut microbiota used 2D-LC to identify up to 1,264 protein groups (Xiong et al., 2015b) or up to 4,031 proteins (Young et al., 2015).

More MS/MS – More Identifications?

The number of MS/MS spectra and the number of PSMs generated from these MS/MS spectra are important parameters to assess metaproteomics data quality: the identification of protein groups is based on the MS/MS spectra that match to peptides in a database comparison and thus generate PSMs (Zhang et al., 2013). Besides being the building blocks of protein identifications, PSMs can also be used to quantify proteins via spectral counting based methods, e.g., with NSAFs (Zybailov et al., 2006).

The matching of MS/MS spectra to peptides in the database depends on the quality of the database, i.e., how well the database fits to the actual sample (Timmins-Schiffman et al., 2017), and on the quality of the MS/MS spectra themselves. As we used a mock community with a known composition of sequenced organisms, we were able to remove confounding effects of database quality. We however want to emphasize the importance of using a database that actually fits the sample in question.

The 1D-LC methods generally produced less MS/MS spectra than the 2D-LC methods, due to the shorter time available to the mass spectrometer to acquire spectra in the 1D runs (Hsieh et al., 2013). The proportion of identified MS/MS spectra (PSMs), on the other hand, was higher for the 1D runs than for the 2D runs (**Figure 1B** and **Supplementary Table S2**). The likely reason for this is that in the 2D runs not only the higher abundant peptides were sampled (as during the 1D runs), but also more spectra of low abundant peptides were acquired during 2D runs due to the better separation, as indicated by more identified proteins of low-abundance species in 2D runs. Those low-abundance peptides also include non-tryptic peptides which cannot be identified with the search engine settings used, leading to a lower proportion of identified MS/MS spectra in 2D runs, while the absolute number of PSMs is still higher in 2D runs, e.g., twice as high in the 2D|10pH runs as compared to a 4 h 1D run (**Supplementary Figure S1**).

We also evaluated our LC methods in terms of absolute number of PSMs (**Figure 1B**). The total number of PSMs is generally higher for 2D than for 1D-LC methods, with the exception of 1D|12h_75. Interestingly, 2D|10pH_G1, which identified the most protein groups and unique peptides, led to a comparatively low number of about 108,000 PSMs. This indicates the higher separation efficiency of the 2D|10pH_G1 method, as new peptides are targeted more often for sequencing than in the other 2D methods.

How Beneficial Are Run Repetitions?

One potential way to increase metaproteome coverage is to repeat a run once or multiple times. In order to evaluate the effect of repeating runs, we used increasing numbers of the 1D|8h_50 runs as one experiment in MaxQuant

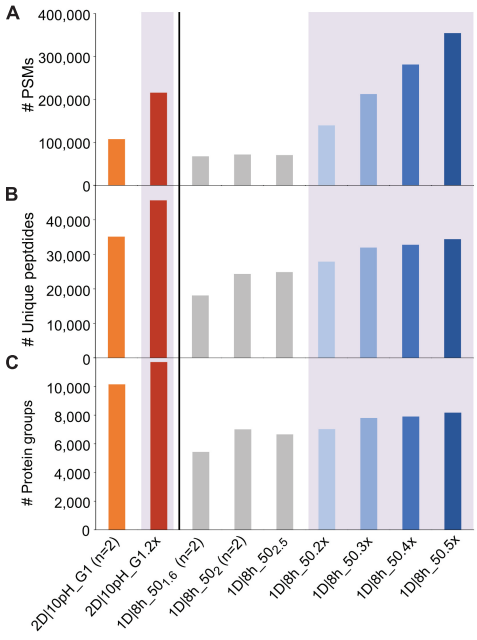

FIGURE 2 | Number of PSMs, unique peptides and protein groups for replicate runs of 2D|10pH_G1 and repetitions of 1D|8h_50 runs. **(A)** Number of PSMs, **(B)** number of identified unique peptides, **(C)** number of identified protein groups. For (n = 2), the mean value of two independent runs is shown. Results for searches combining runs have a purple colored background and are labeled with the number of runs searched together in MaxQuant, followed by "x" (e.g., 2D|10pH_G1 (n = 2) means that two runs have been searched separately and the mean is shown, whereas 2D|10pH_G1.2x means that both replicates have been run in MaxQuant as one experiment and the output of the combined search is depicted).

searches. Additionally, we compared the effect of running our best 2D method in terms of protein groups identified (2D|10pH_G1) once versus in technical replicate (**Figure 2** and **Supplementary Table S3**).

Each additional run of the 1D|8h_50 runs increased the number of PSMs. The total number of unique peptide and protein group identifications also increased, albeit not as much as the number of PSMs. Similarly, the replicate run 2D|10pH_G1.2x led to double the amount of PSMs compared to a single run, while the number of unique peptides increased by 29.9%, and the number of identified protein groups by 15.3%. The percentage of protein groups with one unique peptide, one peptide total or one PSM total decreased with more runs (**Supplementary Table S3**). Overall, there is a decreasing benefit of increasing run numbers for the protein group identifications, which has also been observed by others (Thakur et al., 2011). The reason for this is that during replicate runs, foremost peptides of the same (more abundant) protein groups are sequenced. Therefore, run repetition increases the amount of spectra for proteins already identified in the first run and, to a lesser extent, metaproteome coverage.

We also analyzed how a single long 2D run competes against repeated shorter 1D runs. While in three combined runs of 1D|8h_50 we identified about 7,800 protein groups, the 2D|10pH_G1 alone led to over 10,000 protein groups, even though three 1D|8h_50 runs together have about the same total runtime as a single 2D|10pH_G1 run (24 vs. 22 h). Köcher et al. (2011), too, identified more protein groups in HeLa digests with a single long run than with replicate shorter runs, using 1D-LC with a long analytical column. The number of PSMs, on the other hand, was almost double for the three 1D|8h_50 runs (over 212,000 as compared to about 108,000 for 2D|10pH_G1). Thus, while the 2D|10pH_G1 led to more protein groups overall, the 1D|8h_50 runs taken together contained on average more information per protein.

How Reproducible Are the LC Methods?

We assessed the reproducibility of selected methods in terms of total per species protein quantification within and between methods and overlap of identified protein groups within methods. For this, we used the biological and technical replicate runs of the UNEVEN mock community from Kleiner et al. (2017). In this study, 1D methods (1D|4h, 1D|8h) and a 2D method (2D|11salt) were used to measure four independently generated biological replicates of the UNEVEN mock community. In addition to this, for the 1D methods all samples were measured in technical duplicates.

To assess how reproducible the methods are in terms of quantification of individual species' proteins, we summed the relative abundance of proteins of each species for all runs (**Figure 3**). We found that the mean and standard deviations of biological replicates are comparable between technical replicates of the same method and between different methods.

Additionally, we compared the method reproducibility between technical and biological replicates for the total metaproteome coverage. For this, we calculated the overlap between identified protein groups of technical and biological replicates (**Supplementary Table S5**). Technical replicates of the 1D methods had an overlap between 82 and 86%. These

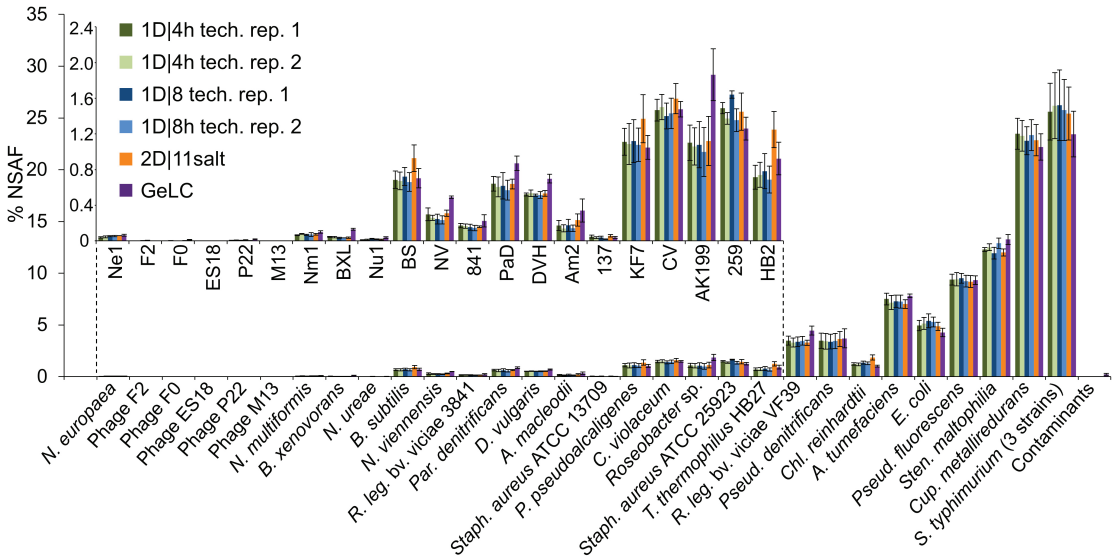

FIGURE 3 | % NSAFs summed up for each organism in the mock community. Each bar represents the mean of four biological replicates analyzed with 1D-LC or 2D-LC and three biological replicates analyzed with GeLC-MS/MS. For the 1D-LC-MS/MS methods, technical replicate (tech. rep.) datasets are presented as separate bars. Error bars indicate standard deviations. Small insert: % NSAF for the species with lower abundance. Data for this figure in tabular format can be found in **Supplementary Table S4**.

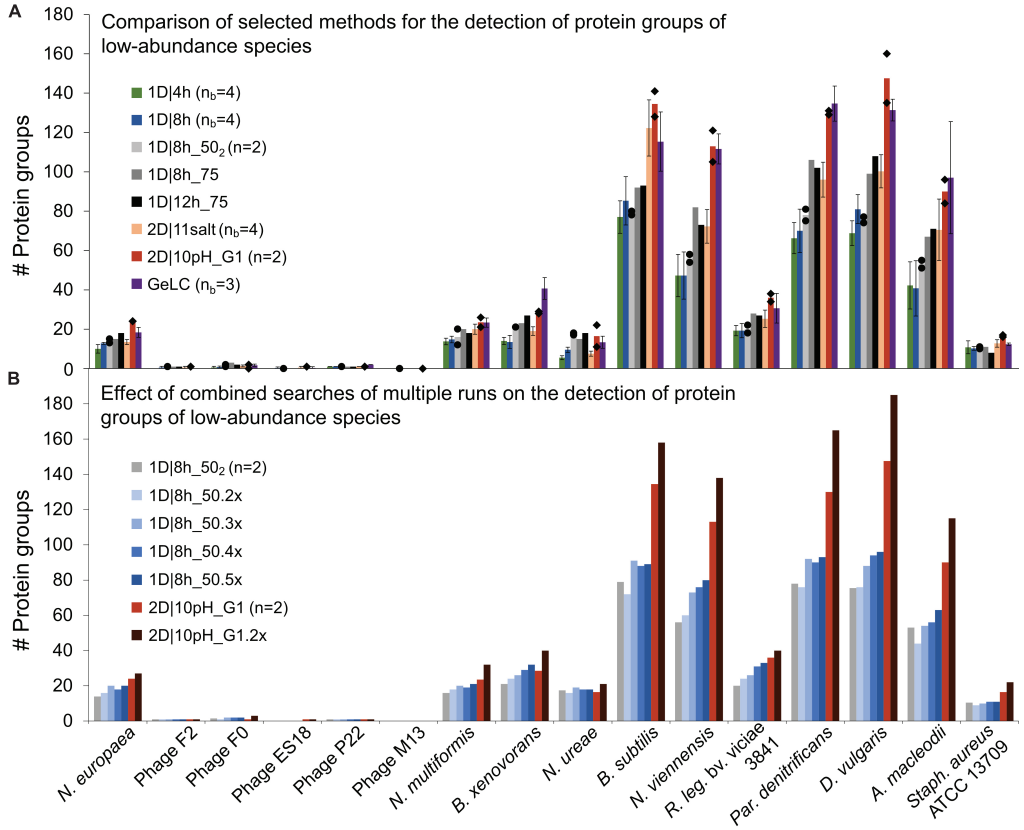

FIGURE 4 | Number of protein groups of low-abundance organisms in the mock community identified by different LC methods. **(A)** Comparison of selected 1D and 2D-LC methods and the GeLC method. Shown are means (bars), individual values (black diamonds) for $n = 2$ and standard deviations for $n_b = 3$ and $n_b = 4$. **(B)** Results for searches combining multiple runs. The number of runs combined for searches in MaxQuant is indicated by the number of runs followed by "x." n, number of technical replicates; n_b, number of biological replicates.

values are similar to previously published overlap values for technical replicates in proteomic experiments with yeast (i.e., a less complex sample than our mock community): Richards et al. (2015) reported an 83% overlap of protein identifications in five technical replicates of 1 h 1D runs. Nagaraj et al. (2012) noted a 92% overlap of six technical replicates of 4 h 1D runs on the protein level. While Nagaraj et al. (2012) enabled the "Match between runs" option in MaxQuant, which should increase the overlap between runs, we deliberately disabled this option such that only peptides identified via MS/MS sequencing during a run in our analysis were included.

On average between 74 and 76% of the protein groups identified in one run were detected in all biological replicates, regardless of the method (1D or 2D) used (**Supplementary Table S5**). A lower overlap between biological replicates as compared to technical replicates is to be expected, since biological replication introduces more variation. Moreover, two technical replicates, but four biological replicates (each of which represents an additional source of variance) were used. In summary, in terms of reproducibility, 1D and 2D methods perform comparably.

How to Acquire Data for Low-Abundance Organisms?

An important aspect of metaproteomics is the analysis depth – how do we identify proteins of as many community members as possible, including low-abundance ones, that despite their low abundance can be important members of the system (Podar et al., 2007; Hajishengallis et al., 2011; Baldrian et al., 2012)?

To determine which LC method is best suited to detect low-abundance organisms, we analyzed the datasets for which four biological replicates were measured (1D|4h, 1D|8h, and 2D|11salt), as well as the 1D|8h_50$_2$, 1D|8h_75, 1D|12h_75, and 2D|10pH runs. We especially considered the 16 organisms that had <1% total protein abundance in the mock community (**Figure 4A** and **Supplementary Table S6** for all organisms).

Surprisingly, the 1D|4h and 1D|8h_50 methods performed similarly in detecting proteins of low-abundance species.

For several organisms, the 2D|10pH_G1 method clearly outperformed the other methods. For the viruses (F2, F0, ES18, P22, and M13), which consist of only very few major proteins, there was almost no difference between methods. Aside from the viruses, which have a very small genome, we did not see a correlation between genome size (which roughly translates into the theoretically observable number of proteins) and number of proteins identified (**Supplementary Tables S3, S6**).

For the lowest-abundance bacterium (*Nitrosomonas europaea*, 0.082% total protein abundance in the mock community), at least eight protein groups were detected by every run considered and with 2D|10pH_G1, we detected 24 protein groups containing this organism. The protein groups which contain only *N. europaea* proteins, for example, housekeeping proteins such as a 50S ribosomal protein (Ne1_Q82VV4), allow us to confidently deduce the presence of this species in the sample. Additionally, some of the detected proteins such as ammonia monooxygenase (Ne1_Q04508), which was identified even in 1D|4h runs, would, when analyzing an uncharacterized community, provide a hint that this organism is an ammonia-oxidizing chemoautotroph. Nitrosocyanin (Ne1_Q820S6) and cytochrome c-552 (Ne1_P95339) were also identified already in several 1D|4h as well as in at least one 2D|10pH_G1 run. Nitrosocyanin is a red copper protein with a potential central role in metabolism of ammonia-oxidizing bacteria, while cytochrome c-552 is part of the electron transport chain (Arciero et al., 2002; Zorz et al., 2018). In both 2D|10pH_G1 runs, phosphoenolpyruvate synthase (Ne1_Q81ZR7), phosphoenolpyruvate carboxylase (Ne1_Q82WS3) and phosphoglycerate kinase (Ne1_Q82XE8), all involved in carbohydrate metabolism, were also identified.

Besides more peptide separation, run replication could also increase the data from low-abundance community members. To test whether run repetitions could indeed lead to higher analysis depth, we analyzed repeated runs of 1D|8h_50 and 2D|10pH_G1 (**Figure 4B** and **Supplementary Table S3**). The number of identified protein groups increased for most organisms with more runs. For the viruses there was very little-to-no change in protein identification numbers, again likely due to the limited total number of abundant proteins in these organisms.

A single 2D|10pH_G1 run mostly provides more protein groups for low-abundance organisms than three repetitions of a 1D|8h_50 run, while both have roughly the same runtime (22 vs. 24 h). Thus, for detecting low-abundance community members (or low-abundance proteins) a 2D-LC separation is better suited than repeated 1D-LC runs. Since low-abundance proteins are generally less reliably identified than high-abundance ones (Zubarev, 2013), replicate 2D runs would be even better to generate more data from low-abundance proteins or species, if those are important for the study question.

How Do the On-Line LC-MS/MS Approaches Compare to the GeLC-MS/MS Approach?

We compared the performance of a standard GeLC approach, which has been used in different metaproteomic studies (Schneider et al., 2011; Stokke et al., 2012; König et al., 2016; Ponnudurai et al., 2017), to the on-line-only approaches that we tested. The total number of protein groups and unique peptide identifications was comparable between the GeLC approach and our best performing 2D-LC-MS/MS approach (2D|10pH_G1) and thus higher than our 1D-LC-MS/MS-only approaches. A similar performance difference was recently observed in a study of biogas plant communities, where a GeLC method performed much better than a 120 min 1D-LC-MS/MS method (Wenzel et al., 2018). The GeLC method also performed comparably to the 2D methods in terms of run reproducibility (**Figure 3**) and identification of protein groups of low-abundance organisms (**Figure 4**).

The number of MS/MS generated by the GeLC approach was in the range of the 1D|12h runs. In contrast, the % identified spectra was much lower than in the 1D approaches and comparable to that of the 2D methods (**Figure 1**). Taken together, this indicates that the GeLC approach leads to the acquisition of a lower number of "redundant" spectra from the same peptides, as shown by the highest unique peptides to PSM ratio among all tested methods.

One potential concern with GeLC approaches is that no SDS gel electrophoresis method exists that works equally well for all protein sizes (Rabilloud et al., 2009). Therefore, it can be expected that biases, especially concerning very small and large proteins, are introduced by this step. Considering this, we checked our data for size and, additionally, pI biases in the identified proteins. We noted a small bias of the GeLC method against smaller proteins and proteins with a very low or high pI (**Supplementary Table S7**). For example, whereas in all on-line approaches proteins with a length of <150 amino acids made up

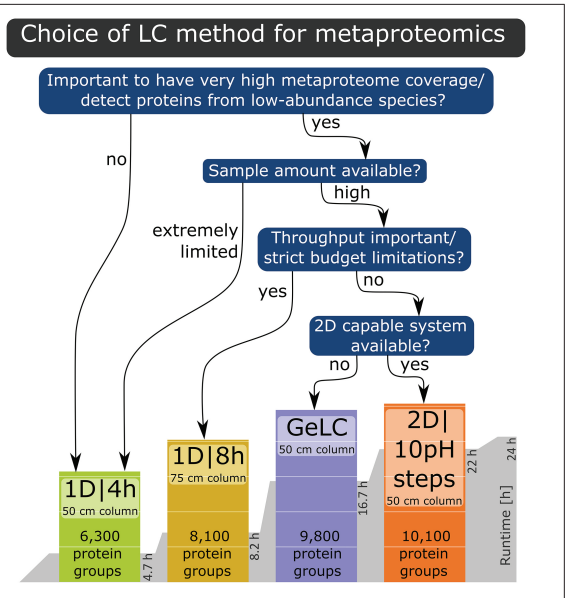

FIGURE 5 | Decision tree for the LC method to choose for a specific experiment, based on the results of our study.

15–17% of all proteins, in the GeLC identifications this protein size fraction only made up 11.7%.

While 2D on-line separations and the GeLC approach overall performed equally well, the amount of sample preparation work and the possible throughput is not comparable between the methods: with the FASP approach, peptides for analysis can be generated in 1–1.5 days, and hands-on time is mainly limited to adding reagents to the filter units. For the GeLC approach, at least 2.5 days are necessary, with substantial hands-on time during gel slicing and washing. Even more important is the fact that there is a large difference in the number of samples that can be prepared in parallel. With FASP, which is used for preparation of peptides for the on-line only approaches, up to 44 samples can be prepared in parallel with a high-capacity centrifuge rotor. The number of reaction tubes that fit into the rotor is the main limitation for parallelization during FASP preparations. During the GeLC approach, on the other hand, each sample is split in at least 10 subsamples, meaning that preparing more than 3–4 samples in parallel is hardly feasible. This translates into an at least 10-fold higher sample preparation throughput for FASP with subsequent on-line separation than for the GeLC-MS/MS approach.

CONCLUSION

In this study, we tested 1D and 2D-LC methods for metaproteomics. We included different gradient lengths, peptide loads, analytical column lengths (50 and 75 cm), 2D separations (salt and pH bumps), numbers of run repetitions, as well as reproducibility estimates with biological and technical replicates. We also compared these on-line separation methods with the off-line GeLC method. A graphical summary of our main findings can be found in **Supplementary Figure S1**.

We demonstrated that, when using a 50 cm column, an increase from 4 to 8 h 1D run time leads to only a small gain in identified protein groups, whereas a further increase in run length to 12 h does not improve the number of identified protein groups. This makes the 1D 4 h run the most time-effective choice (**Figure 1** and **Supplementary Figure S1**).

Peptide load has a strong influence on the number of identified protein groups: up to a certain threshold, loading more peptide increases the number of identifications, after which the number of identifications decreases again (**Figure 1**).

When switching from a 50 cm to a 75 cm column for 8 and 12 h 1D runs, we found a large gain in identification numbers (**Figure 1** and **Supplementary Figure S1**).

For the 2D runs, we found the best performance for the method using 10 pH bumps, each with a RP runtime of 120 min. This method outperforms all other methods for total number of identified protein groups (**Figure 1**) and the detection of protein groups of low-abundance species (**Figure 4**).

Repeating runs several times does not lead to a large gain in metaproteome coverage, but rather increases the data amount for already identified proteins (**Figure 2**).

When comparing the reproducibility of 1D versus 2D LC runs of biological replicate samples and technical replicate runs, we found that both approaches perform equally well (**Figure 3** and **Supplementary Table S5**).

The GeLC approach performs equally well as the 2D method with 10 pH bumps (**Figure 1**), but is extremely limited in throughput as compared to the on-line methods.

Based on our findings, we provide the following recommendations for LC method selection (**Figure 5**): if the goal is a very high metaproteome coverage or the detection of proteins from low-abundance species, then either a 1D|8h gradient with a 75 cm column, the GeLC approach or a 2D LC run should be used, depending on restrictions of sample quantity, equipment availability and throughput as well as budget considerations. If the sample amount is extremely limited or the characterization of proteins from low-abundance community members is not the primary goal, then a 1D|4h run with a 50 cm column is well-suited for metaproteomics experiments.

DATA AVAILABILITY

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al., 2016) partner repository with the dataset identifier PXD008017. In addition to the datasets generated in this study, we also re-analyzed the UNEVEN eight 4 h, eight 8 h and four 2D-LC-MS/MS datasets generated by Kleiner et al. (2017). The corresponding raw data has been deposited to PRIDE with the dataset identifier PXD006118.

AUTHOR CONTRIBUTIONS

MK conceived the study, obtained and created bacterial stocks for mock communities. MK and TH designed the experiments. TH and R-AH prepared samples for mass spectrometry. TH measured samples, analyzed the data, and wrote the manuscript with input from MK. AK did mass spectrometry measurements. MS and AK revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was supported by the Western Canadian Microbiome Center, the Government of Canada through Genome Canada, the Government of Alberta through Genome Alberta, Genome Prairie, Research Manitoba and Genome Quebec, the Canada First Research Excellence Fund, the Campus Alberta Innovation Program (MS), the Canadian Foundation for Innovation (MS), the Natural Sciences and Engineering Research Council of Canada NSERC through a Banting fellowship (MK) and a discovery grand (MS), the NC State Chancellor's Faculty Excellence Program Cluster on Microbiomes and Complex Microbial Communities (MK), the German Research Foundation DFG, grant MA 6346/2-1 (TH) and the German Academic Exchange Service DAAD (TH).

ACKNOWLEDGMENTS

We are grateful to Jianwei Chen, Emmo Hamann, Marc Mußmann, Jessica Kozlowski, Sean Booth, Jessica Duong, Johanna Voordouw, Kenneth Sanderson, Joong-Jae Kim, Joenel Alcantara, Anupama P. Halmillawewa, Michael F. Hynes, and Heidi Gibson for donations of cultures for the mock communities, to Erin Thorson and Christine Sharp for the help with mock community generation and to Jackie K. Zorz and Maryam Ataeian for assistance with MS measurements. We are also thankful to the two reviewers for thoughtful feedback that helped to improve this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.00238/full#supplementary-material

REFERENCES

Arciero, D. M., Pierce, B. S., Hendrich, M. P., and Hooper, A. B. (2002). Nitrosocyanin, a red cupredoxin-like protein from *Nitrosomonas europaea*. *Biochemistry* 41, 1703–1709. doi: 10.1021/bi015908w

Baldrian, P., Kolařík, M., Štursová, M., Kopecký, J., Valášková, V., Větrovský, T., et al. (2012). Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME J.* 6, 248–258. doi: 10.1038/ismej.2011.95

Belnap, C. P., Pan, C., VerBerkmoes, N. C., Power, M. E., Samatova, N. F., Carver, R. L., et al. (2010). Cultivation and quantitative proteomic analyses of acidophilic microbial communities. *ISME J.* 4, 520–530. doi: 10.1038/ismej.2009.139

Camerini, S., and Mauri, P. (2015). The role of protein and peptide separation before mass spectrometry analysis in clinical proteomics. *J. Chromatogr. A* 1381, 1–12. doi: 10.1016/j.chroma.2014.12.035

Charif, D., and Lobry, J. R. (2007). "SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis," in *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations*, eds U. Bastolla, M. Porto, H. E. Roman, and M. Vendruscolo (Berlin: Springer Verlag), 207–232. doi: 10.1007/978-3-540-35306-5_10

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372. doi: 10.1038/nbt.1511

Dai, J., Shieh, C. H., Sheng, Q.-H., Zhou, H., and Zeng, R. (2005). Proteomic analysis with integrated multiple dimensional liquid chromatography/mass spectrometry based on elution of ion exchange column using pH steps. *Anal. Chem.* 77, 5793–5799. doi: 10.1021/ac050251w

Eymann, C., Dreisbach, A., Albrecht, D., Bernhardt, J., Becher, D., Gentner, S., et al. (2004). A comprehensive proteome map of growing *Bacillus subtilis* cells. *Proteomics* 4, 2849–2876. doi: 10.1002/pmic.200400907

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

Hajishengallis, G., Liang, S., Payne, M. A., Hashim, A., Jotwani, R., Eskan, M. A., et al. (2011). Low-abundance biofilm species orchestrates inflammatory periodontal disease through the commensal microbiota and complement. *Cell Host Microbe* 10, 497–506. doi: 10.1016/j.chom.2011.10.006

Hettich, R. L., Sharma, R., Chourey, K., and Giannone, R. J. (2012). Microbial metaproteomics: identifying the repertoire of proteins that microorganisms use to compete and cooperate in complex environmental communities. *Curr. Opin. Microbiol.* 15, 373–380. doi: 10.1016/j.mib.2012.04.008

Heyer, R., Benndorf, D., Kohrs, F., De Vrieze, J., Boon, N., Hoffmann, M., et al. (2016). Proteotyping of biogas plant microbiomes separates biogas plants according to process temperature and reactor type. *Biotechnol. Biofuels* 9:155. doi: 10.1186/s13068-016-0572-4

Hsieh, E. J., Bereman, M. S., Durand, S., Valaskovic, G. A., and MacCoss, M. J. (2013). Effects of column and gradient lengths on peak capacity and peptide identification in nanoflow LC-MS/MS of complex proteomic samples. *J. Am. Soc. Mass Spectrom.* 24, 148–153. doi: 10.1007/s13361-012-0508-6

Jehmlich, N., Vogt, C., Lünsmann, V., Richnow, H. H., and von Bergen, M. (2016). Protein-SIP in environmental studies. *Curr. Opin. Biotechnol.* 41, 26–33. doi: 10.1016/j.copbio.2016.04.010

Ji, C., Zhang, J., Lin, X., Han, J., Dong, X., Yang, S., et al. (2017). Metaproteomic analysis of microbiota in the fermented fish, *Siniperca chuatsi*. *LWT Food Sci. Technol.* 80, 479–484. doi: 10.1016/j.lwt.2017.03.022

Keiblinger, K. M., Wilhartitz, I. C., Schneider, T., Roschitzki, B., Schmid, E., Eberl, L., et al. (2012). Soil metaproteomics – Comparative evaluation of protein extraction protocols. *Soil Biol. Biochem.* 54, 14–24. doi: 10.1016/j.soilbio.2012.05.014

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

Kleiner, M., Wentrup, C., Lott, C., Teeling, H., Wetzel, S., Young, J., et al. (2012). Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc. Natl. Acad. Sci. U.S.A.* 109, E1173–E1182. doi: 10.1073/pnas.1121198109

Köcher, T., Swart, R., and Mechtler, K. (2011). Ultra-high-pressure RPLC hyphenated to an LTQ-Orbitrap Velos reveals a linear relation between peak capacity and number of identified peptides. *Anal. Chem.* 83, 2699–2704. doi: 10.1021/ac103243t

König, S., Gros, O., Heiden, S. E., Hinzke, T., Thürmer, A., Poehlein, A., et al. (2016). Nitrogen fixation in a chemoautotrophic lucinid symbiosis. *Nat. Microbiol.* 2:16193. doi: 10.1038/nmicrobiol.2016.193

Magdeldin, S., Moresco, J. J., Yamamoto, T., and Yates, J. R. III (2014). Off-line multidimensional liquid chromatography and auto sampling result in sample loss in LC/LC-MS/MS. *J. Proteome Res.* 13, 3826–3836. doi: 10.1021/pr500530e

Mallick, P., and Kuster, B. (2010). Proteomics: a pragmatic perspective. *Nat. Biotechnol.* 28, 695–709. doi: 10.1038/nbt.1658

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

Meier, F., Geyer, P. E., Virreira Winter, S., Cox, J., and Mann, M. (2018). BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nat. Methods* 15, 440–448. doi: 10.1038/s41592-018-0003-5

Mostovenko, E., Hassan, C., Rattke, J., Deelder, A. M., van Veelen, P. A., and Palmblad, M. (2013). Comparison of peptide and protein fractionation methods in proteomics. *EuPA Open Proteom.* 1, 30–37. doi: 10.1016/j.euprot.2013.09.001

Muth, T., Benndorf, D., Reichl, U., Rapp, E., and Martens, L. (2013). Searching for a needle in a stack of needles: challenges in metaproteomics data analysis. *Mol. BioSyst.* 9, 578–585. doi: 10.1039/C2MB25415H

Nagaraj, N., Kulak, N. A., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., et al. (2012). System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Mol. Cell. Proteomics* 11:M111.013722. doi: 10.1074/mcp.M111.013722

Nesvizhskii, A. I., and Aebersold, R. (2005). Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteomics* 4, 1419–1440. doi: 10.1074/mcp.R500012-MCP200

Petersen, J. M., Kemper, A., Gruber-Vodicka, H., Cardini, U., van der Geest, M., Kleiner, M., et al. (2016). Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation. *Nat. Microbiol.* 2:16195. doi: 10.1038/nmicrobiol.2016.195

Pirmoradian, M., Budamgunta, H., Chingin, K., Zhang, B., Astorga-Wells, J., and Zubarev, R. A. (2013). Rapid and deep human proteome analysis by single-dimension shotgun proteomics. *Mol. Cell. Proteomics* 12, 3330–3338. doi: 10.1074/mcp.O113.028787

Podar, M., Abulencia, C. B., Walcher, M., Hutchison, D., Zengler, K., Garcia, J. A., et al. (2007). Targeted access to the genomes of low-abundance organisms in complex microbial communities. *Appl. Environ. Microbiol.* 73, 3205–3214. doi: 10.1128/AEM.02985-06

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

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at: https://www.R-project.org/

Rabilloud, T., Vaezzadeh, A. R., Potier, N., Lelong, C., Leize-Wagner, E., and Chevallet, M. (2009). Power and limitations of electrophoretic separations in proteomics strategies. *Mass Spectrom. Rev.* 28, 816–843. doi: 10.1002/mas.20204

Ram, R. J., VerBerkmoes, N. C., Thelen, M. P., Tyson, G. W., Baker, B. J., Blake, R. C. II, et al. (2005). Community proteomics of a natural microbial biofilm. *Science* 308, 1915–1920. doi: 10.1126/science

Richard, V. R., Domanski, D., Percy, A. J., and Borchers, C. H. (2017). An online 2D-reversed-phase – Reversed-phase chromatographic method for sensitive and robust plasma protein quantitation. *J. Proteomics* 168, 28–36. doi: 10.1016/j.jprot.2017.07.018

Richards, A. L., Hebert, A. S., Ulbrich, A., Bailey, D. J., Coughlin, E. E., Westphall, M. S., et al. (2015). One-hour proteome analysis in yeast. *Nat. Protoc.* 10, 701–714. doi: 10.1038/nprot.2015.040

Rubin-Blum, M., Antony, C. P., Borowski, C., Sayavedra, L., Pape, T., Sahling, H., et al. (2017). Short-chain alkanes fuel mussel and sponge *Cycloclasticus* symbionts from deep-sea gas and oil seeps. *Nat. Microbiol.* 2:17093. doi: 10.1038/nmicrobiol.2017.93

Schneider, T., Schmid, E., de Castro, J. V. Jr., Cardinale, M., Eberl, L., Grube, M., et al. (2011). Structure and function of the symbiosis partners of the lung lichen (*Lobaria pulmonaria* L. Hoffm.) analyzed by metaproteomics. *Proteomics* 11, 2752–2756. doi: 10.1002/pmic.201000679

Sowell, S. M., Abraham, P. E., Shah, M., VerBerkmoes, N. C., Smith, D. P., Barofsky, D. F., et al. (2011). Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. *ISME J.* 5, 856–865. doi: 10.1038/ismej.2010.168

Stokke, R., Roalkvam, I., Lanzen, A., Haflidason, H., and Steen, I. H. (2012). Integrated metagenomic and metaproteomic analyses of an ANME-1-dominated community in marine cold seep sediments. *Environ. Microbiol.* 14, 1333–1346. doi: 10.1111/j.1462-2920.2012.02716.x

Taylor, P., Nielsen, P. A., Trelle, M. B., Hørning, O. B., Andersen, M. B., Vorm, O., et al. (2009). Automated 2D peptide separation on a 1D nano-LC-MS system. *J. Proteome Res.* 8, 1610–1616. doi: 10.1021/pr800986c

Thakur, S. S., Geiger, T., Chatterjee, B., Bandilla, P., Fröhlich, F., Cox, J., et al. (2011). Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Mol. Cell. Proteomics* 10:M110.003699. doi: 10.1074/mcp.M110.003699

Timmins-Schiffman, E., May, D. H., Mikan, M., Riffle, M., Frazar, C., Harvey, H. R., et al. (2017). Critical decisions in metaproteomics: achieving high confidence protein annotations in a sea of unknowns. *ISME J.* 11, 309–314. doi: 10.1038/ismej.2016.132

Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* 11, 2301–2319. doi: 10.1038/nprot.2016.136

VerBerkmoes, N. C., Denef, V. J., Hettich, R. L., and Banfield, J. F. (2009). Functional analysis of natural microbial consortia using community proteomics. *Nat. Rev. Microbiol.* 7, 196–205. doi: 10.1038/nrmicro2080

Vizcaíno, J. A., Csordas, A., del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., et al. (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 44, D447–D456. doi: 10.1093/nar/gkv1145

Wang, D.-Z., Xie, Z.-X., and Zhang, S.-F. (2014). Marine metaproteomics: current status and future directions. *J. Proteomics* 97, 27–35. doi: 10.1016/j.jprot.2013.08.024

Wenzel, L., Heyer, R., Schallert, K., Löser, L., Wünschiers, R., Reichl, U., et al. (2018). SDS-PAGE fractionation to increase metaproteomic insight into the taxonomic and functional composition of microbial communities for biogas plant samples. *Eng. Life Sci.* 18, 498–509. doi: 10.1002/elsc.201800062

Weston, L. A., Bauer, K. M., and Hummon, A. B. (2013). Comparison of bottom-up proteomic approaches for LC-MS analysis of complex proteomes. *Anal. Methods* 5, 4615–4621. doi: 10.1039/c3ay40853a

Wilmes, P., and Bond, P. L. (2004). The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environ. Microbiol.* 6, 911–920. doi: 10.1111/j.1462-2920.2004.00687.x

Wilmes, P., Heintz-Buschart, A., and Bond, P. L. (2015). A decade of metaproteomics: where we stand and what the future holds. *Proteomics* 15, 3409–3417. doi: 10.1002/pmic.201500183

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

Xiao, M., Yang, J., Feng, Y., Zhu, Y., Chai, X., and Wang, Y. (2017). Metaproteomic strategies and applications for gut microbial research. *Appl. Microbiol. Biotechnol.* 101, 3077–3088. doi: 10.1007/s00253-017-8215-7

Xiong, W., Abraham, P. E., Li, Z., Pan, C., and Hettich, R. L. (2015a). Microbial metaproteomics for characterizing the range of metabolic functions and activities of human gut microbiota. *Proteomics* 15, 3424–3438. doi: 10.1002/pmic.201400571

Xiong, W., Giannone, R. J., Morowitz, M. J., Banfield, J. F., and Hettich, R. L. (2015b). Development of an enhanced metaproteomic approach for deepening the microbiome characterization of the human infant gut. *J. Proteome Res.* 14, 133–141. doi: 10.1021/pr500936p

Young, J. C., Pan, C., Adams, R. M., Brooks, B., Banfield, J. F., Morowitz, M. J., et al. (2015). Metaproteomics reveals functional shifts in microbial and human proteins during a preterm infant gut colonization case. *Proteomics* 15, 3463–3473. doi: 10.1002/pmic.201400563

Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C., and Yates, J. R. III (2013). Protein analysis by shotgun/bottom-up proteomics. *Chem. Rev.* 113, 2343–2394. doi: 10.1021/cr3003533

Zheng, Q., Lin, B., Wang, Y., Zhang, Q., He, X., Yang, P., et al. (2015). Proteomic and high-throughput analysis of protein expression and microbial diversity of microbes from 30- and 300-year pit muds of Chinese Luzhou-flavor liquor. *Food Res. Int.* 75, 305–314. doi: 10.1016/j.foodres.2015.06.029

Zorz, J. K., Kozlowski, J. A., Stein, L. Y., Strous, M., and Kleiner, M. (2018). Comparative proteomics of three species of ammonia-oxidizing bacteria. *Front. Microbiol.* 9:938. doi: 10.3389/fmicb.2018.00938

Zubarev, R. A. (2013). The challenge of the proteome dynamic range and its implications for in-depth proteomics. *Proteomics* 13, 723–726. doi: 10.1002/pmic.201200451

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

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

PUBLICATION III

The reaction of the *Crassostrea gigas* mitochondrial proteome and mitochondrial respiration during hypoxia-reoxygenation stress is described in the article "Effects of hypoxia-reoxygenation stress on mitochondrial proteome and bioenergetics of the hypoxia-tolerant marine bivalve *Crassostrea gigas*"

Journal of Proteomics 194 (2019) 99–111

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

Effects of hypoxia-reoxygenation stress on mitochondrial proteome and bioenergetics of the hypoxia-tolerant marine bivalve *Crassostrea gigas*

Eugene P. Sokolov[a,*], Stephanie Markert[b,c], Tjorven Hinzke[b,c], Claudia Hirschfeld[d], Dörte Becher[d], Siriluck Ponsuksili[e], Inna M. Sokolova[f,g]

[a] Leibniz Institute for Baltic Sea Research, Leibniz Science Campus Phosphorus Research Rostock, Warnemünde, Germany
[b] Institute of Marine Biotechnology e. V., 17489 Greifswald, Germany
[c] Institute of Pharmacy, University of Greifswald, Germany
[d] Institute of Microbiology, University of Greifswald, Germany
[e] Leibniz Institute for Farm Animal Biology (FBN), Institute for Genome Biology, Functional Genome Analysis Research Unit, Dummerstorf, Germany
[f] Department of Marine Biology, University of Rostock, Rostock, Germany
[g] Department of Maritime Systems, Interdisciplinary Faculty, University of Rostock, Rostock, Germany

A B S T R A C T

Mitochondria are key intracellular targets of hypoxia-reoxygenation (H/R) stress due to their central role in generation of ATP and reactive oxygen species (ROS). Intertidal oysters *Crassostrea gigas* are adapted to frequent H/R cycles and maintain aerobic function despite frequent oxygen fluctuations. To gain insight into the molecular mechanisms of H/R tolerance, we assessed changes in mitochondrial respiration and (phospho)proteome of *C. gigas* during hypoxia and recovery. Oyster mitochondria maintained OXPHOS capacity despite a decline in cytochrome *c* oxidase activity during H/R stress. Rearrangements of the mitochondrial proteome during H/R stress involved upregulation of mitochondrial electron transport system and iron-binding proteins, and suppression of the pathways that channel electrons to ubiquinone, possibly as a mechanism to limit ROS production. H/R stress led to upregulation of a mitophagic activator PGAM5 and dephosphorylation of metalloendopeptidase OMA1, indicating stimulation of mitochondrial quality control mechanisms. Changes in abundance and phosphorylation levels of key proteins involved in mitochondrial protein homeostasis indicate suppression of protein synthesis during hypoxia, likely as an energy-saving mechanism, and its subsequent reactivation during reoxygenation. Thus, shifts in the mitochondrial (phospho-)proteome might play an important role in H/R stress resistance of oysters ensuring mitochondrial integrity and function during oxygen fluctuations.

Significance: Hypoxia-reoxygenation (H/R) stress elicits shifts in proteome and phosphoproteome of mitochondria in a hypoxia-tolerant model bivalve, oyster *Crassostrea gigas*, upregulating electron transport system, limiting electron flow to ubiquinone and activating mitochondrial quality control and protein homeostasis mechanisms. These findings provide insights into the potential role of proteomic shifts in adaptive response to H/R stress and serve as an important benchmark to understand the mechanisms of mitochondrial sensitivity to hypoxia and reoxygenation.

1. Intoduction

Oxygen (O_2) variability is a major stressor for aerobic organisms leading to energy deficiency during hypoxia (low O_2 levels) and oxidative injury during reoxygenation [1,2]. Extensive research in humans and mammalian models showed that mitochondrial damage is the main cause of cellular injury and organ failure during hypoxia-reoxygenation (H/R) stress such as caused by infarction, stroke, organ transplantation, or sleep apnea [3–6]. Hypoxia causes severe energy limitations and impairs activity of ATP-dependent ion pumps resulting in disturbances of cellular ion homeostasis [7,8]. A common strategy to overcome the energy limitation in hypoxia-tolerant organisms involves metabolic rate depression (a concomitant suppression of ATP producing and consuming pathways) combined with ion channel arrest to suppress passive ion movements thereby minimizing disturbances to ion homeostasis [8–10]. Notably, although mitochondrial activity is suppressed during severe hypoxia, mitochondria of hypoxia-tolerant organisms (such as fish or mollusks) undergo functional rearrangements that increase antioxidant defense and enhance mitochondrial respiratory capacity potentially facilitating post-hypoxic recovery [1,11–14].

Paradoxically, mitochondria of hypoxia-sensitive species (such as terrestrial mammals and hypoxia-sensitive invertebrates) sustain major damage when oxygen flux to the tissue is reinstated [3,11,15–18]. The mitochondrial injury during reoxygenation involves elevated production of reactive oxygen species (ROS), Ca^{2+} overload, damage to the mitochondrial proteins and membranes, and release of pro-apoptotic factors [3,11,15–18]. These alterations suppress mitochondrial capacity for oxidative phosphorylation (OXPHOS), reinforce energy deficiency and ultimately result in cell death [1,3,19]. In hypoxia-sensitive organisms even a single episode of H/R is highly damaging to

* Corresponding author.
E-mail address: evgeny.sokolov@io-warnemuende.de (E.P. Sokolov).

https://doi.org/10.1016/j.jprot.2018.12.009
Received 19 October 2018; Received in revised form 3 December 2018; Accepted 10 December 2018
Available online 12 December 2018

mitochondria and potentially fatal to the organism [3,11,15–18]. In contrast, hypoxia-tolerant intertidal organisms (such as marine bivalves and fish) withstand frequent periods of H/R stress and maintain high mitochondrial aerobic capacity throughout H/R cycles [1,11,13,14]. Furthermore, in some highly tolerant species such as hard shell clams (*Mercenaria mercenaria*) or oysters (*Crassostrea virginica*), mitochondrial capacity increases during reoxygenation, as reflected in elevated activity of the electron transport (ETS) and oxidative phosphorylation (OXPHOS) systems [11,20]. To date, the molecular mechanisms underlying this remarkable ability of hypoxia-tolerant intertidal organisms to maintain mitochondrial integrity and functional capacity and avoid mitochondrial injury during hypoxia and reoxygenation are not fully understood and require further investigation.

Proteome rearrangements are a common response to environmental stress [21] including hypoxia and reoxygenation [22–25]. Studies in mammalian models showed that H/R stress affects the abundance of proteins involved in aerobic respiration and antioxidant defense [26,27], induces expression of alternative isoforms of cytochrome *c* oxidase (COX) [28,29] and results in post-translational modifications of proteins involved in cell signaling, glycolysis, ion transport and protein synthesis [30–32]. H/R stress also has impact on the proteome of marine organisms. Hypoxic exposure affected abundance of mitochondrial proteins (including subunits of ATP synthase and citrate synthase) in the annelid *Hydroides elegans* [23] and the shrimp *Fenneropenaeus chinensis* [24], and cyclic hypoxia simulating tidal cycles upregulated proteins involved in anaerobic ATP production and antioxidant defense in the intertidal bivalve *Geukensia demissa* [22]. The effects of H/R stress on the phosphoproteome have not been extensively studied, but targeted investigations of key regulatory enzymes of energy metabolism (including hexokinase, pyruvate kinase, phosphofructokinase, and pyruvate dehydrogenase) [33–39], ion pumps (such as Na^+/K^+ ATPase) [40] and transcription factors [11,41] showed that reversible phosphorylation of these proteins plays an essential role in regulation of metabolic fluxes during hypoxic exposures. To date, investigations on proteome plasticity in response to hypoxia or H/R stress have almost exclusively focused on the total cellular proteome [21–24,42], and organelle-specific proteome rearrangements during H/R stress have not been extensively analyzed. Due to the critical regulatory, signaling and bioenergetics role of mitochondria [1,43,44], understanding shifts in the mitochondrial proteome and phosphoproteome during H/R stress can shed important new light on the mechanisms of metabolic regulation and cellular homeostasis during oxygen fluctuations and identify the potential pathways involved in tolerance to intermittent hypoxia.

The Pacific oyster *Crassostrea gigas* is a common inhabitant of intertidal zones and a key aquaculture species around the world. Originating from the eastern Pacific, this species has been introduced in aquaculture on all five continents and established wild populations in many coastal areas outside its native range serving as an ecosystem engineer in intertidal and estuarine ecosystems [45,46]. Oysters are eurybiont species with a broad tolerance to abiotic stressors [47] including oxygen fluctuations [48]. Earlier studies in *C. gigas* and a closely related species *Crassostrea virginica* showed the importance of mitochondrial adjustments in adaptive and stress responses to intermittent hypoxia [12,14,20,49], similar to other hypoxia-tolerant intertidal bivalves [11,12]. This makes oysters an excellent model species to investigate the molecular mechanisms of H/R-induced mitochondrial adjustments. The aim of the present study was to assess the effects of intermittent hypoxia (24 h of hypoxia followed by 1 h of recovery) on mitochondrial functions of *C. gigas* and to investigate the alterations of the mitochondrial proteome and phosphoproteome during H/R stress. Such a discovery-oriented approach is essential for generating novel hypotheses concerning the potentially adaptive pathways that are involved in functional reorganization of oyster mitochondria during oxygen fluctuations and underlie their remarkable tolerance to frequent H/R stress.

2. Materials and methods

2.1. Chemicals

All chemicals were purchased from Sigma Aldrich (Munich, Germany), Fisher Scientific (Schwerte, Germany), or Carl Roth (Karlsruhe, Germany) and were of the analytical grade or higher.

2.2. Animal collection and care

Adult Pacific oysters *C. gigas* (shell length 10–15 cm) were collected from the island of Sylt in German Wadden Sea, and transported within 24 h of collection to the University of Rostock. Oysters were kept at 15 ± 0.5 °C and salinity 30 ± 1 (practical salinity units) in tanks with aerated natural Baltic Sea water adjusted to 30 salinity with Instant Ocean sea salt (Aquarium Systems, Sarrebourg, France). Oysters were fed ad libitum with a commercial algal blend (DT's Live Marine Phytoplankton, CoralSands, Wiesbaden, Germany) according to the manufacturer's recommendations.

Oysters were exposed to hypoxia in chambers with tightly closed lids filled with approximately 0.7 l of seawater and bubbled with nitrogen until oxygen concentration dropped below ~1%. Oxygen content of seawater was measured using a fiber optic oxygen sensor connected to a FireStingO_2 Optical Oxygen Meter (PyroScience GmbH, Aachen, Germany). The chamber was submerged in the aquarium during exposure to maintain temperature (15 ± 0.5 °C). After 24 h of exposure mollusks were placed on ice and processed immediately for mitochondrial isolation. For reoxygenation, a subset of oysters exposed to 24 h of hypoxia was placed into a fully aerated aquarium for 1 h prior to mitochondrial isolation. Mitochondria were isolated from gill tissues since this is the main organ of gas exchange in oysters as well as a target for H/R stress due to the large surface area and direct contact with the ambient water [50]. Each mitochondrial isolate was obtained from a single oyster. In each experimental group (control, hypoxia and reoxygenation), 6 oysters were used for measurements of mitochondrial respiration, 3 oysters – for total mitochondrial proteome analyses and 5 oysters – for analyses of mitochondrial phosphoproteome.

2.3. Mitochondrial isolation

Mitochondria were isolated from gill tissues of oysters as described earlier [51]. Briefly, 1–2 g of gills were homogenized in an ice-cold buffer containing 100 mmol l^{-1} sucrose, 200 mmol l^{-1} KCl, 100 mmol l^{-1} NaCl, 8 mmol l^{-1} ethylene glycol-bis(2-aminoethylether)-N,N,N′,N′-tetraacetic acid (EGTA), 50 μg l^{-1} aprotinin, 1 mmol l^{-1} phenylmethylsulfonyl fluoride (PMSF), and 30 mmol l^{-1} 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES), pH 7.5 using several passes of a Potter–Elvenhjem homogenizer at 200 r.p.m. The homogenate was centrifuged at 4 °C and 2000 *g* for 4 min to remove cell debris, and the supernatant was centrifuged at 4 °C and 8500 *g* for 8 min to obtain a mitochondrial pellet. The mitochondria were cleaned up two times in the isolation media, collected each time by brief centrifugation (8500 *g* for 5 min) and resuspended in ice-cold assay media containing 440 mmol l^{-1} sucrose, 130 mmol l^{-1} KCl, 10 mmol l^{-1} NaCl, 30 mmol l^{-1} HEPES, 10 mmol l^{-1} glucose 1 mmol l^{-1} $MgCl_2$, 10 mmol l^{-1} KH2PO4 and 1% BSA, pH 7.2. Resulting mitochondrial suspensions were split into two aliquots for high resolution respirometry and proteome analysis, respectively. It is worth noting that mitochondrial isolates used in this study consisted of mitochondria-enriched crude isolates rather than purified mitochondrial fractions. Mitochondrial purification by centrifugation on density gradients [52,53] results in considerable and potentially selective loss of certain mitochondrial fractions that might bias the results. Therefore, in the present study we analyzed the proteome of crude enriched mitochondrial fractions as recommended for comprehensive mitoproteome analyses [54], and manually curated the proteomic data to include only

mitochondrial proteins.

2.4. High resolution respirometry

To test the effects of H/R stress on mitochondrial function of *C. gigas*, we measured the mitochondrial oxygen consumption and mitochondria membrane potential (Δψ) under the fully oxygenated conditions [11,55]. This approach allows determining the maximum capacity of the respective mitochondrial pathways in the absence of substrate or oxygen limitation and is commonly used in mitochondrial physiology [56]. Oxygen consumption of isolated mitochondria was measured using an Oxygraph 2 k high resolution respirometer (Oroboros, Innsbruck, Austria) at 15 °C. Two-point calibration for 0% and 100% air saturation was conducted. The following substrate-uncoupler-inhibitor titration (SUIT) protocol was used: 5 mmol l^{-1} pyruvate with 2 mmol l^{-1} malate to stimulate Complex I (NADH-linked) respiration; 10 mmol l^{-1} succinate to stimulate the electron flow through Complex II; 2.5 mmol l^{-1} ADP to achieve state 3 (ADP-stimulated) respiration; 2.5 μmol l^{-1} oligomycin to inhibit mitochondrial F_O, F_1-ATPase and achieve state 4 respiration indicative of the proton leak; 7.5 mmol l^{-1} carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP) to uncouple the mitochondrial ETS; 1 μmol l^{-1} rotenone to inhibit electron flux through Complex I; 2.5 μmol l^{-1} antimycin A to inhibit electron flux through Complex III; 0.5 mmol l^{-1} *N,N,N′,N′*-tetramethyl-*p*-phenylenediamine (TMPD) and 2 mmol l^{-1} ascorbate to stimulate activity of Complex IV (cytochrome *c* oxidase, or COX), and 20 mmol l^{-1} KCN to inhibit COX. Pilot experiments with a stepwise titration of oyster mitochondria with CCCP (0.5–10 mmol l^{-1}) showed that 7.5–10 mmol l^{-1} CCCP fully uncoupled the mitochondria without inhibiting ETS activity (data not shown). Thus, we used 7.5 mmol l^{-1} CCCP in our assays. Residual oxygen consumption after the addition of antimycin A was ~5–7% of the CCCP uncoupled respiration, and residual COX activity after addition of KCN was < 10% of COX activity, and did not differ between the experimental treatments (data not shown). Mitochondrial respiratory states and control indices were calculated as described elsewhere [56,57]. Briefly, oxidative phosphorylation (OXPHOS) flux was determined from the rate of ATP-stimulated mitochondrial respiration reflective of the ATP synthesis capacity; the rate of proton leak (LEAK) was measured as the oxygen consumption of resting mitochondria in the absence of ADP and in the presence of saturating concentrations of substrates (pyruvate + malate or succinate); maximum activity of the electron transport system (ETS) was assessed as the oxygen consumption of mitochondria with saturating concentrations of substrates in the presence of CCCP to collapse mitochondrial membrane potential. Respiratory control ratio (RCR) was calculated as the ratio of OXPHOS to LEAK flux as described elsewhere [57] and used as an index of mitochondrial coupling.

Mitochondrial membrane potential (Δψ) was determined simultaneously to the oxygen signal using a tetraphenylphosphonium (TPP^+)-selective electrode module of the Oxygraph-2k high resolution respirometer (Oroboros, Innsbruck, Austria), as TPP^+ accumulates into the mitochondria in a Δψ-dependent manner [58]. The TPP^+ electrode was calibrated before each measurement using stepwise additions of TPP^+ (0–6 mM). To calculate Δψ we used the Nernst equation based on extramitochondrial and intramitochondrial TPP^+ concentrations corrected for non-specific TPP+ binding as described elsewhere [11,55]. Corrections for the nonspecific binding of TTP were conducted after fully collapsing Δψ with 1 mM of KCN [59]. All mitochondrial assays were completed within 60 min after the isolation. After the assays, a 50 μL aliquot of each mitochondrial suspension was mixed 1:20 vol:vol with ultrapure water and lysed by three rapid freeze-thaw cycles. Protein concentrations were measured using a Bio-Rad protein assay (Bio-Rad, Hercules, CA, USA), and corrected for the BSA content of the assay media. Mitochondrial respiration rates were expressed as μmol O_2 min^{-1} g^{-1} protein, and mitochondrial membrane potential in mV.

2.5. Sample preparation for LC-MS/MS analysis: Total proteins

Mass spectrometric analyses were performed after exposure of *C. gigas* to hypoxia (H), hypoxia followed by reoxygenation (R), and normoxia (control, C). Two sets of *C. gigas* proteins were analyzed in this study: 1) the total mitochondrial proteome (all proteins extracted from isolated mitochondria), and 2) the enriched mitochondrial phosphoproteome (including phosphorylated mitochondrial proteins)

Total mitochondrial proteome samples were analyzed in three biological replicates ($N = 3$) for each condition (H, R and C). To extract the total proteome, mitochondrial suspensions were subjected to sonication (2 × 25 s; Sonoplus HD2070 sonication probe, Bandelin, Berlin, Germany) to disrupt the mitochondrial membranes. Protein concentration in the lysates was determined according to Bradford [60] using Roti-Nanoquant reagent (Carl Roth, Karlsruhe, Germany). For lysates of the total mitochondrial samples, 30 μg of protein were loaded onto precast SDS mini gels (4–20% TGX, BioRad) and separated by 1D PAGE at 150 V for 50 min. Protein bands were fixed in 40% ethanol, 10% acetic acid and stained with Coomassie Brilliant Blue (CBB) for 12 h. Gels were washed in double-distilled water to remove excess CBB. Protein-containing gel lanes were excised and each sample lane was divided into 10 equal-sized subsamples, which were transferred to individual microcentrifuge tubes. Gel pieces were repeatedly washed (200 mM NH_4HCO_3, 30% acetonitrile, 37 °C, 30 min) until destained and dried in a vacuum centrifuge before in-gel digestion with 2 μg/μl trypsin solution (sequencing grade modified trypsin, Promega) at 37 °C for 16 h. Peptides were eluted from the gel pieces in double-distilled water in an ultrasonic bath (15 min), concentrated to 10 μl and frozen at −80 °C until MS analysis in an LTQ Orbitrap Classic mass spectrometer.

2.6. Sample preparation for LC-MS/MS analysis: Phosphoproteins

Phosphorylated proteins were isolated from mitochondrial isolates (N = 5) by means of affinity chromatography using Pro-Q® Diamond Phosphoprotein Enrichment Kit (Invitrogen, Carlsbad, CA, USA) according to manufacturer's protocol for non-denatured proteins. This method enables efficient, nonradioactive isolation of phosphoproteins from complex cellular extracts using phosphoprotein-binding resin to capture both native and denatured phosphoproteins. Briefly, 40–60 mg of mitochondrial suspension was lysed on ice in presence of protease and phosphatase inhibitors (Sigma, St.Louis, MO, USA) and diluted to achieve ~0.1–0.2 mg of protein per ml^{-1}. The diluted lysate was applied to a column containing 1 ml of resin, and the flow-through was saved for analysis. The columns were washed several times followed by elution of the bound proteins. Eluate was concentrated with a centrifuge concentrator (Vivaspin) to 50–70 μl and the elution buffer was replaced with several rounds of centrifugation with 25 mM Tris, pH 7.5, 0.25% 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate (CHAPS). Concentration of resulting protein was measured according to Bradford using a Bio-Rad protein assay (Bio-Rad, Hercules, CA, USA).

Mitochondrial phosphoprotein-enriched samples were analyzed in five biological replicates per condition. Enriched phosphoprotein extracts (containing approximately 50–100 μg protein) were concentrated to 15 μl and mixed with 5 μl sample buffer (200 mM Tris-HCl pH 6.8, 50 mM EDTA pH 8.0, 40% glycerol, 8% SDS, 0.08% bromophenol blue, 1 mM dithiothreitol (DTT)). Proteins were denatured at 90 °C for 10 min and extracts were loaded onto 10% SDS mini gels for 1D PAGE. Electrophoresis was stopped after 40 min at 50 V, when the samples had spread out across approximately 10 mm of the gel. This step served to purify the samples and to remove CHAPS buffer before mass spectrometric analysis. Gels were fixed in 40% ethanol, 10% acetic acid, stained with CBB (3–6 h) and washed in double-distilled water. Protein-containing gel areas were excised, transferred to microcentrifuge tubes (without separation into subsamples), and dried in a vacuum centrifuge. Subsequently, proteins in the gel pieces were subjected to

reduction in 20 mM DTT (30 min, 55 °C, 800 rpm), washing in 50 mM NH_4HCO_3 (2 min, room temperature) and alkylation in 40 mM iodoacetamide (30 min, room temperature, in the dark). After another washing step in 50 mM NH_4HCO_3 (2 min, room temperature), gel pieces were dehydrated in 100% acetonitrile (HPLC grade, 2 min, room temperature) and dried in a vacuum centrifuge again, before digestion with 4 μg/μl trypsin solution (37 °C, 16 h). Peptides were eluted as described above and peptide extracts were concentrated to approximately 40 μl before determination of peptide concentrations using a Nanodrop micro-volume photometer. Peptide purification using StageTips was performed as described elsewhere [61]. Concentrations of the peptide extracts were adjusted to 2 μg/μl with double-distilled water and samples were frozen at −80 °C until mass spectrometric analysis in an LTQ Orbitrap Velos mass spectrometer.

2.7. LC-MS/MS analysis

Purified peptides of enriched phosphoproteins were analyzed by reversed phase liquid chromatography (LC) electrospray ionization (ESI) MS/MS using an LTQ Orbitrap Velos (Thermo Fisher Scientific, Waltham, USA). In-house self-packed nano-LC columns (100 μm × 20 cm) containing C18 material (ReproSil-Pur 120, 3 μm, Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) were used to perform LC with an EASY-nLC II system (Thermo Fisher Scientific, Waltham, USA). The peptides were loaded with solvent A (0.1% acetic acid (*v*/v)). Subsequently, the peptides were eluted by a non-linear binary gradient of 166 min from 5% to 99% solvent B (0.1% acetic acid (v/v), 99.9% acetonitrile (v/v)) in solvent A at a constant flow rate of 300 nl min^{-1}. After injection into the mass spectrometer, a full scan was recorded in the Orbitrap (*m/z* 300–2000) with a resolution of 30,000 at 400 m/z. The twenty most abundant precursor ions were consecutively isolated in the linear ion trap and fragmented via collision-induced dissociation (CID). Unassigned charge states as well as singly charged ions were rejected and the lock mass correction was enabled. Selected ions were fragmented with 35% normalized collision energy (NCE) and analyzed in data-dependent MS/MS in the linear trap quadrupole (LTQ).

Peptides obtained from total mitochondrial protein samples were analyzed in an LTQ Orbitrap Classic coupled to an EASY-nLC II system (Thermo Fisher Scientific, Waltham, USA). Peptides were loaded on in-house self-packed nano-LC columns as described before and eluted by a non-linear binary gradient of 77 min from 5% to 99% solvent B (0.1% acetic acid (*v*/v), 99.9% acetonitrile (v/v)) in solvent A at a constant flow rate of 300 nl min^{-1}. In the Orbitrap (*m/z* 300–2000) a full scan was recorded with a resolution of 30,000 at 400 m/z. The five most abundant precursor ions were consecutively isolated in the ion trap and fragmented via collision-induced dissociation (CID). Unassigned charge states and singly charged ions were rejected, the lock mass correction was enabled. Selected ions were fragmented with 35% NCE and analyzed in data-dependent MS/MS in the LTQ.

2.8. MS data analysis

MS/MS spectra were extracted and analyzed using the Sorcerer-Sequest software (Sorcerer v3.5, Sage-N Research; Sequest version v.27, rev. 11, Thermo-Finnigan, Thermo Fisher Scientific, Germany) as described elsewhere [62]. We used a combined target-decoy database containing all proteins encoded in the *C. gigas* genome (NCBI project accession PRJNA276446), 42 common laboratory contaminants, and the reversed amino acid sequences of all these proteins as decoys to determine false discovery rates (FDR; total number of ORFs in database: 78,206). MS/MS-based peptide and protein identifications were evaluated with Scaffold version 4.8.3 (http://www.proteomesoftware.com). Protein level FDR and peptide level FDR were set to 1%. A minimum of 2 peptides was required for protein identification. For semi-quantitative analysis and comparison of relative protein abundances between samples, normalized spectral abundance factors (NSAF) were calculated [63] using total spectrum counts. The mass spectrometry proteomics data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository [64] with the dataset identifier PXD011365 and https://doi.org/10.6019/PXD011365. Proteome and phosphoproteome data were manually curated to include only the mitochondrial proteins in further analyses.

2.9. Statistics

Data on the functional traits of mitochondria (oxygen consumption $\dot{M}_{O2}$ and membrane potential $\Delta\psi$) were tested for the normal distribution and homogeneity of variances using Shapiro-Wilk and Levine test, respectively. The effects of hypoxia and reoxygenation on mitochondrial $\dot{M}_{O2}$ and $\Delta\psi$ were tested using ANOVA with oxygen regime as a fixed factor with three levels (control, 24 h of hypoxia and 1 h of reoxygenation). Post-hoc analyses were conducted using Fisher's least significant differences (LSD) tests using planned contrasts between the groups of interest. The statistical analyses were conducted using IBM® SPSS® Statistics ver. 22.0.0.0 (IBM Corp., Armonk, NY, USA) and GraphPad Prism ver 7.02 (GraphPad Software Inc., La Jolla, CA, USA) software. All differences were considered significant if the probability of Type II error was < 0.05.

2.10. Random Forest analysis

Normalized abundances of all mitochondrial proteins and enriched phosphoproteins were analyzed using the package randomForest [65] implemented in R [66] to find proteins with statistically relevant abundance differences between the three conditions (C, H, R). Random forests are a machine learning technique. For building the forest, decision trees are calculated on bootstrapped samples. For every split in the tree, the best split variable from a random variable subset is used. The final classification in the forest is the majority vote of the trees [67]. As a criterion for variable importance, the mean decrease in accuracy calculates the difference of the prediction error for the original and permuted observations for each variable [65]. Random forests are well suited for proteomics (as for other –omics) data, as they are noise-robust, can be used for datasets with many more variables than samples, are not prone to overfitting and directly return measures of variable importance [68].

Ten forests were trained for both datasets, with 12,500 trees per forest for the dataset containing all mitochondrial proteins and 21,000 trees per forest for the phosphoproteome dataset. Mean decrease of accuracy (unscaled) was averaged per protein over all forests (i.e. mean of the means) and minimal and maximal mean decrease in accuracy considered across all forests was assigned to each protein. Mitochondrial proteins that had a minimal mean decrease in accuracy > 0 over all forests, i.e. that were predicted to be relevant for sample class differentiation, were chosen for further analysis.

3. Results

Overall, proteomics analysis of total and phosphoprotein-enriched mitochondrial fractions identified 422 mitochondrial proteins. In the total mitochondrial fractions (N = 3), 234 mitochondrial proteins were found, with 43 unique proteins not found in the phosphoprotein-enriched fraction. In the phosphoprotein-enriched fractions (N = 5), 379 mitochondrial proteins were indentified, with 188 unique proteins that were not detected in the total protein fraction.

3.1. Total abundance of mitochondrial proteins

Out of 234 mitochondrial proteins identified in the total mitochondrial fractions, 49 proteins were identified by the random forest analysis as distinguishing different treatment groups (i.e. having a

Table 1
Differentially abundant mitochondrial proteins in the gills of control (C), hypoxia- (H) and reoxygenation-exposed (R) *C. gigas*.

Only the proteins identified by the random forest analysis as contributing to distinguishing between two and more of the experimental groups are shown. The last two columns of the table indicate relative change of the abundance of the respective protein compared to its abundance in the control group (taken as 100%). ~ - change in abundance < 50%; single arrow – change between 1.5 and 2-fold; double arrows – change > 2-fold compared with the control. Changes larger than 50% compared with the control levels are highlighted in bold.

Accession #	MW (kDa)	Identified Proteins	Separates groups	H	R	H, % control	R, % control
Upregulated in H and R							
XP_011441201	**38**	**NADH dehydrogenase (ubiquinone) complex I, assembly factor 6-like**	**C–H**	↑↑	↑↑	**474**	**308**
XP_011413812	**19**	**cytochrome *c* oxidase subunit 5B, mitochondrial**	**C–H**	↑	↑	**156**	**142**
XP_019924659	**54**	**probable D-lactate dehydrogenase, mitochondrial isoform X1**	**C–H, H–R**	↑↑	↑	**466**	**176**
XP_011456349	**70**	**acyl-CoA dehydrogenase family member 9, mitochondrial**	**C–R, H–R**	~	↑↑	**133**	**238**
XP_011428335	**32**	**serine/threonine-protein phosphatase PGAM5, mitochondrial isoform X5**	**C–H**	↑↑	↑↑	**433**	**331**
XP_011429335	**47**	**ferrochelatase, mitochondrial**	**H–R**	↑↑	~	**238**	**104**
XP_011428121	**78**	**calcium-binding mitochondrial carrier protein Aralar2 isoform X2**	**H–R**	~	↑↑	**123**	**203**
XP_011429655	**35**	**persulfide dioxygenase ETHE1, mitochondrial**	**C–R, H–R**	~	↑	**127**	**169**
XP_011451520	**66**	**glycerol-3-phosphate dehydrogenase, mitochondrial isoform X2**	**C–R**	↑	↑↑	**177**	**211**
XP_011428031	38	calcium uniporter protein, mitochondrial	H–R	~	~	111	101
XP_011450475	57	aldehyde dehydrogenase, mitochondrial	C–R, H–R	~	~	102	126
Downregulated in H and R							
XP_011432897	**20**	**succinate dehydrogenase cytochrome *b*560 subunit, mitochondrial isoform X1**	**H–R**	~	↓↓	**85**	**35**
XP_011441687	**32**	**mitochondrial dicarboxylate carrier**	**C–H, H–R**	↓↓	~	**49**	**83**
XP_011416632	**18**	**mitochondrial carrier homolog 1-like**	**C–H, H–R**	↓↓	~	**39**	**91**
XP_011425087	**40**	**isocitrate dehydrogenase [NAD] subunit α, mitochondrial isoform X2**	**C–R**	↓↓	↓↓	**42**	**43**
XP_011418526	**48**	**glutaryl-CoA dehydrogenase, mitochondrial**	**C–R, H–R**	~	↓	**88**	**56**
XP_011450365	**47**	**short-chain specific acyl-CoA dehydrogenase, mitochondrial**	**C–R**	~	↓	**80**	**54**
XP_011442388	**50**	**isovaleryl-CoA dehydrogenase, mitochondrial isoform X1**	**C–R**	~	↓	**98**	**65**
XP_011449935	**33**	**protein SCO1 homolog, mitochondrial**	**C–R, H–R**	↓	↓↓	**56**	**37**
XP_011455482	**20**	**peroxiredoxin-5, mitochondrial**	**C–R, H–R**	↓	↓	**63**	**57**
XP_011456199	**52**	**calcium uptake protein 3, mitochondrial-like**	**C–R**	↓	↓↓	**57**	**42**
XP_019921452	**47**	**kynurenine/alpha-aminoadipate aminotransferase, mitochondrial**	**C–R, H–R**	~	↓↓	**82**	**48**
XP_011433638	54	cytochrome *b*-c1 complex subunit 2, mitochondrial	C–H, H–R	~	~	74	91
XP_011412259	14	cytochrome *b*-c1 complex subunit 7	C–H	~	~	76	90
XP_011437055	33	ADP, ATP carrier protein	H–R	~	~	69	96
NP_001292269	51	isocitrate dehydrogenase [NADP], mitochondrial-like	C–R	~	~	78	67
XP_011456884	60	glutamate dehydrogenase, mitochondrial	C–H, H–R	~	~	70	87
XP_011413584	47	aspartate aminotransferase, mitochondrial	C–H	~	~	69	67
XP_011428695	59	propionyl-CoA carboxylase beta chain, mitochondrial	C–H, H–R	~	~	80	96
XP_011434348	46	2-amino-3-ketobutyrate coenzyme A ligase, mitochondrial	C–H, C–R, H–R	~	~	79	71
XP_011456348	53	mitochondrial-processing peptidase subunit beta	C–H, C–R	~	~	77	69
Upregulated in H, downregulated in R							
XP_011448969	**18**	**ATP synthase subunit delta, mitochondrial**	**H–R**	↑↑	~	**211**	**83**
XP_011436137	**36**	**ubiquinone biosynthesis protein COQ9, mitochondrial-like isoform X2**	**C–R, H–R**	~	↓↓	**111**	**0**
XP_011421875	**53**	**mitochondrial import inner membrane translocase subunit TIM44**	**H–R**	↑↑	~	**217**	**73**
XP_011427846	**43**	**enoyl-CoA delta isomerase 2, mitochondrial isoform X1**	**H–R**	~	↓	**112**	**65**
XP_011430333	**27**	**GTP:AMP phosphotransferase AK3, mitochondrial-like**	**H–R**	↑↑	↓	**227**	**51**
XP_011445975	32	succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial-like	H–R	~	~	110	89
XP_011421683	37	2,4-dienoyl-CoA reductase, mitochondrial	C–H, H–R	~	~	136	95
Downregulated in H, upregulated in R							
XP_011431701	**16**	**NADH dehydrogenase [ubiquinone] iron-sulfur protein 6, mitochondrial**	**H–R**	~	↑↑	**84**	**332**
XP_011449111	**24**	**peroxiredoxin-6**	**C–H, H–R**	↓↓	~	**47**	**100**
XP_011433038	**20**	**frataxin, mitochondrial-like isoform X2**	**C–R, H–R**	~	↑	**91**	**198**
XP_011434533	**75**	**carnitine O-palmitoyltransferase 2, mitochondrial-like**	**C–R, H–R**	↓	↑	**54**	**199**
XP_011450919	**68**	**very long-chain specific acyl-CoA dehydrogenase, mitochondrial**	**H–R**	↓	~	**50**	**114**
XP_019919459	**10**	**mitochondrial import inner membrane translocase subunit Tim9**	**C–H, H–R**	↓↓	~	**0**	**108**
XP_011450157	35	tricarboxylate transport protein, mitochondrial	C–R	~	~	100	137
XP_011454174	35	mitochondrial 2-oxoglutarate/malate carrier protein	H–R	~	~	68	124
XP_019930702	64	delta-1-pyrroline-5-carboxylate dehydrogenase, mitochondrial	H–R	~	~	88	110
XP_011439166	36	bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase, mitochondrial-like	H–R	~	~	90	138
XP_011439505	57	probable methylmalonate-semialdehyde dehydrogenase [acylating], mitochondrial	C–H, H–R	~	~	79	116

minimal mean decrease in accuracy over all forests of > 0) (Table 1). Most of these differentially abundant proteins showed large changes in mean abundance (by 2-fold or more) in hypoxia or reoxygenation compared with the respective controls. However, some mitochondrial proteins identified in the random forest analysis as distinguishing different experimental groups showed relatively small changes in abundance (by ~10–40%) in response to H/R stress (Table 1). This indicates that these changes, albeit small, were consistent in different biological replicates of the respective experimental groups. However, since the biological significance of these relatively small changes is not clear, we adopted a conservative approach and focused only on the proteins that were identified as important group discriminants by the random forest analysis and showed > 50% change in mean abundance among the treatment groups in subsequent analyses (Table 1).

Based on the overall relative abundance, 31 mitochondrial proteins changed in the same direction in hypoxia and reoxygenation compared with the normoxic controls, and 18 proteins showed the opposite patterns of change in hypoxia vs. reoxygenation (Table 1). The mitochondrial proteins with the strongest (> 4-fold) increase in abundance during hypoxic exposure included the assembly complex of the mitochondrial NADH dehydrogenase (Complex I), serine-threonine protein phosphatase PGAM5, and D-lactate dehydrogenase (Table 1).

Several other proteins (including ferrochelatase, ATP synthase subunit delta, mitochondrial import protein TIM44 and GTP: AMP phosphotransferase AK3) increased in abundance by > 2-fold in mitochondria of hypoxic oysters. Furthermore, cytochrome *c* oxidase subunit 5B and glycerol-3-phosphate dehydrogenase increased by 60–80% in mitochondria of hypoxia-exposed oysters. During reoxygenation, the assembly complex of the mitochondrial NADH dehydrogenase (Complex I), serine-threonine protein phosphatase PGAM5, and mitochondrial NADH dehydrogenase (Complex I) iron-sulfur protein 6 showed > 3-fold increase in abundance, while acetyl-CoA dehydrogenase 9, Ca^{2+}-binding mitochondrial carrier protein Aralar2, frataxin and carnitine O-palmitoyltransferase 2 increased by ≥2-fold in reoxygenation compared with the controls (Table 1). Furthermore, persulfide dehydrogenase ETHE1 increased by ~70% in mitochondria of oysters following 1 h of post-hypoxic reoxygenation.

The mitochondrial proteins downregulated during hypoxia included two mitochondrial carrier proteins (dicarboxylate carrier and mitochondrial carrier protein homolog 1), NAD^+-dependent isocitrate dehydrogenase subunit alpha, peroxiredoxin 6, mitochondrial import protein Tim9 and carnitine O-palmitoyltransferase 2 (Table 1). During reoxygenation, the downregulated proteins included a cytochrome *b*560 subunit, NAD + −dependent isocitrate dehydrogenase subunit alpha, calcium uptake protein 3, protein SCO1, kyneurine/alpha aminoadipate aminotransferase, proteins involved in fatty acid and amino acid oxidation (glutaryl-CoA dehydrogenase, short-chain specific acyl-CoA dehydrogenase, enoyl-CoA delta isomerase 2, and isovaleryl-CoA dehydrogenase), as well as ubiquinone biosynthesis protein COQ9, and GTP: AMP phosphotransferase AK3 (Table 1).

3.2. Phosphoproteins

Abundance of 36 mitochondrial phosphoproteins (out of the total of 379 identified mitochondrial phosphoproteins) changed in the gill mitochondria of oysters in response to hypoxia and reoxygenation (Table 2). Of these, 23 phosphoproteins showed a similar direction of change in hypoxia and reoxygenation, while 13 phosphoproteins showed opposite patterns of change in hypoxia vs. reoxygenation. The phosphorylated protein with the highest increase in abundance in hypoxia and reoxygenation (by ~6 and ~10-fold, respectively) was pyruvate dehydrogenase (PDH) E1 subunit alpha (with PDH E1 subunit beta also slightly increasing in abundance in reoxygenation by ~33%) (Table 2).

3.3. Mitochondrial function

Mitochondrial capacity of *C. gigas* was generally well preserved during the hypoxia-reoxygenation cycle. Exposure to hypoxia had no effect on the mitochondrial proton leak, ETS or OXPHOS capacity, and mitochondria maintained high RCR (> 4) (Fig. 1). Post-hypoxic reoxygenation (1 h) led to a decrease in the proton leak driven by NADH-dependent substrates (Fig. 1A) and a considerable (by 36%) decrease in the activity of the terminal ETS enzyme, COX (Fig. 1C). Despite this suppression of COX, the overall ETS activity did not change in the mitochondria after 1 h reoxygenation, indicating that COX is not a rate-limiting step under the conditions used in the present study. This is reflected in the relatively high (~2) ratio of COX to ETS activity, which is indicative of an apparent excess of the COX activity over the ETS flux (Fig. 1D). There was a slight decrease in OXPHOS activity in mitochondria of oysters exposed to post-hypoxic reoxygenation but this trend was not significant ($P > 0.05$) (Fig. 1A).

The mitochondrial membrane potential was lower in phosphorylating (~160 mV) compared with the resting (~190 mV) mitochondria and was not significantly affected by hypoxia-reoxygenation (Fig. 1B).

4. Discussion

Our present study on oyster mitochondria demonstrates considerable changes in the mitochondrial total and phosphoproteome during H/R stress, encompassing ~50 proteins involved in OXPHOS, substrate catabolism, mitochondrial quality control, protein homeostasis, and iron and redox homeostasis (Figs. 2, 3), yet very little change in mitochondrial function (Fig. 1). This indicates that proteome rearrangement might serve as a basis for the mitochondrial robustness in this hypoxia-tolerant intertidal bivalve, resulting in the adaptive shifts that protect against H/R-induced injury and dysfunction.

4.1. Oxidative phosphorylation

The mitochondrial OXPHOS pathways (including the ETS, substrate and adenylate transporters and ATP synthase) were strongly affected by H/R stress in oyster gills (Fig. 2). Among mitochondrial ETS complexes, Complexes I and IV were most responsive to H/R stress as shown by elevated abundance of a Complex IV (COX) subunit 5B as well as increased levels of the Complex I sulfur-iron protein and two proteins involved in Complex I assembly (the assembly factor 6 of NADH-ubiquinone dehydrogenase and acyl-CoA dehydrogenase family member 9) during H/R exposure. In mammals, mitochondrial Complexes I and IV are among the main targets of H/R-induced oxidative damage as well as important sites of the protective response against H/R injury [69–71]. Hypoxia results in a gradual transition of mammalian Complex I from active to de-active form (the so called A/D transition), most notably in metabolically active tissues such as brain and skeletal muscle [69,72]. Deactivation of Complex I in hypoxia is considered an adaptive mechanism that downregulates ETS activity during oxygen deficiency and prevents ROS burst upon reoxygenation [69,72]. Complex IV (COX) is also commonly deactivated during hypoxia and reoxygenation in mammals, mainly due to the reversible phosphorylation of COX and the loss of cytochrome *c* from the active site of the enzyme [71,73,74]. Notably, H/R stress led to a slight but statistically significant decrease in activities of Complex I and IV in oyster mitochondria (Fig. 1) despite increased abundance of several Complex I and IV subunits. The mechanisms of this mild suppression of Complex I and IV respiration in oyster mitochondria are not known. It is unlikely to be driven by the loss of cytochrome *c*, because experimental addition of cytochrome *c* to isolated mitochondria in our present study failed to significantly stimulate oxygen consumption, regardless of the experimental treatment group (data not shown).

Post-translational modifications such as protein phosphorylation may play a role in the observed changes in the ETS activity during H/R stress [69,75,76]. In our present study, the phosphorylation levels of several Complex I subunits (alpha subcomplex subunit 8, beta subcomplex subunit 5 and flavoprotein 1) considerably declined (by ~2–3 fold) during reoxygenation, in parallel with the decline in the resting mitochondrial respiration driven by Complex I activity. Unlike Complex I, no changes in the levels of phosphoproteins belonging to Complex IV were detected during H/R stress in oyster mitochondria. To date, the functional consequences of reversible phosphorylation of Complex I subunits are unknown, and future studies are required to show whether phosphorylation at these sites activates or suppresses Complex I activity in molluscan mitochondria. Furthermore, it would be important to investigate whether other types of post-translational modifications (such as acetylation, nitrosylation or succinylation) are involved in the functional modulation of ETS complexes during H/R stress. Nevertheless, our present data indicate that Complexes I and IV are strong candidates for such modulation, and suggest that upregulation of Complex I and IV proteins during reoxygenation (detected by the total mitoproteome assessment of oyster mitochondria) might be a compensatory mechanism to maintain ETS flux and counteract the H/R-induced loss of activity.

Notably, the abundance of one of Complex II subunits (SDH

Table 2
Differentially abundant mitochondrial phosphoproteins in the gills of control (C), hypoxia- (H) and reoxygenation-exposed (R) *C. gigas*.

Only the proteins identified by the random forest analysis as contributing to distinguishing between two and more of the experimental groups are shown. The last two columns of the table indicate relative change of the abundance of the respective phosphoprotein compared to its abundance in the control group (taken as 100%). ~ – change in abundance < 50%; single arrow – change between 1.5 and 2-fold; double arrows – change > 2-fold compared with the control. Changes larger than 50% compared with the control levels are highlighted in bold.

Accession #	MW (kDa)	Identified proteins	Separates groups	H	R	H, % control	R, % control
Upregulated in H and R							
XP_019929359	**24**	**acylpyruvase FAHD1, mitochondrial**	**C–R, H–R**	~	↑↑	**145**	**265**
XP_011433038	**20**	**frataxin, mitochondrial-like isoform X2**	**C–H**	↑	↑	**172**	**160**
XP_011433203	**32**	**pyruvate dehydrogenase E1 component subunit alpha, somatic form, mitochondrial isoform X1**	**C–R**	↑↑	↑↑	**631**	**981**
XP_011428081	**48**	**succinate-CoA ligase [GDP-forming] subunit beta, mitochondrial isoform X1**	**C–R, H–R**	~	↑↑	**109**	**241**
XP_011442314	39	pyruvate dehydrogenase E1 component subunit beta, mitochondrial	C–R	~	~	102	133
XP_011434348	46	2-amino-3-ketobutyrate coenzyme A ligase, mitochondrial	C–H	~	~	111	148
XP_011427320	47	ornithine aminotransferase, mitochondrial	C–R	~	~	109	109
Downregulated in H and R							
XP_011413278	**22**	**NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 8**	**C–H**	~	↓	**69**	**55**
XP_011452888	**27**	**NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 5, mitochondrial-like**	**C–R**	~	↓↓	**71**	**34**
XP_011424780	**51**	**NADH dehydrogenase [ubiquinone] flavoprotein 1, mitochondrial**	**C–R**	~	↓↓	**92**	**34**
XP_011421874	**19**	**methylmalonyl-CoA epimerase, mitochondrial-like**	**C–H**	↓	~	**65**	**77**
XP_011416756	**38**	**citrate lyase subunit beta-like protein, mitochondrial**	**C–R**	~	↓	**76**	**51**
XP_011428409	**49**	**short/branched chain specific acyl-CoA dehydrogenase, mitochondrial-like**	**C–R**	~	↓	**79**	**65**
XP_011427299	**47**	**isobutyryl-CoA dehydrogenase, mitochondrial**	**C–H**	↓↓	↓↓	**0**	**0**
XP_011438538	**52**	**metalloendopeptidase OMA1, mitochondrial-like**	**C–H**	↓↓	↓↓	**16**	**0**
XP_011412648	**25**	**mitochondrial import inner membrane translocase subunit Tim21**	**C–H, C–R**	~	↓	**88**	**53**
XP_011422285	**84**	**elongation factor G, mitochondrial-like**	**C–H**	↓↓	↓↓	**21**	**34**
XP_011445701	**20**	**28S ribosomal protein S24, mitochondrial-like**	**C–R**	~	↓↓	**86**	**48**
XP_011423098	**11**	**28S ribosomal protein S21, mitochondrial-like**	**C–H**	↓↓	↓↓	**20**	**0**
XP_011423012	**18**	**28S ribosomal protein S18a, mitochondrial-like**	**C–R**	↓	↓↓	**63**	**0**
XP_019927416	50	dihydrolipoyl dehydrogenase, mitochondrial-like	C–R	~	~	91	86
XP_011417326	25	probable 28S ribosomal protein S26, mitochondrial isoform X2	C–R	~	~	99	69
XP_011442596	17	28S ribosomal protein S14, mitochondrial	C–R	~	~	99	86
Upregulated in H, downregulated in R							
XP_011428996	**32**	**2-methoxy-6-polyprenyl-1,4-benzoquinol methylase, mitochondrial**	**C–H, C–R, H–R**	~	↓	**149**	**51**
XP_011441694	**70**	**probable arginine–tRNA ligase, mitochondrial isoform X1**	**C–R, H–R**	↑	↓↓	**187**	**0**
XP_011456793	**32**	**39S ribosomal protein L9, mitochondrial**	**C–R, H–R**	~	↓↓	**116**	**10**
XP_011428118	42	39S ribosomal protein L45, mitochondrial	C–R	~	~	140	82
XP_011418644	23	28S ribosomal protein S17, mitochondrial	C–R, H–R	~	~	107	71
XP_011444071	43	28S ribosomal protein S22, mitochondrial	C–R, H–R	~	~	126	88
XP_019928450	101	LETM1 and EF-hand domain-containing protein 1, mitochondrial	C–R	~	~	102	95
Downregulated in H, upregulated in R							
XP_011445975	32	succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial	C–R	~	~	97	119
XP_011439961	73	succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	C–R	~	~	99	129
XP_019924399	60	ATP synthase subunit alpha, mitochondrial-like	C–H	~	~	97	104
XP_011432206	39	phosphate carrier protein, mitochondrial-like isoform X1	C–H	~	~	85	105
NP_001292269	51	isocitrate dehydrogenase [NADP], mitochondrial-like	C–H, C–R	~	~	85	103
XP_011447164	83	trifunctional enzyme subunit alpha, mitochondrial-like	C–R	~	~	89	114

cytochrome *b*560) strongly decreased during reoxygenation in oyster mitochondria. This subunit anchors Complex II in the mitochondrial membrane and plays an important role in linking the mitochondrial ETS with the TCA cycle, transferring electrons from succinate to ubiquinone (coenzyme Q). Two key enzymes of the ubiquinone biosynthesis pathway (ubiquinone biosynthesis protein COQ9 and 2-methoxy-6-polyprenyl-1,4-benzoquinol methylase) were also strongly downregulated during reoxygenation in oyster mitochondria. These changes might assist in reducing Complex II activity and thus limiting the reverse flow of electrons from succinate to Complex I that results in elevated ROS production [77]. This potential protective mechanism might be especially important in bivalves that accumulate high levels of succinate during hypoxia-induced anaerobiosis [78–80], as it creates conditions conducive to the reverse electron flow.

Subunit delta of mitochondrial ATP synthase (Complex V) increased in abundance (by ~2-fold) in mitochondria of hypoxia-exposed oysters and rapidly returned to near-control levels during reoxygenation. This protein is a part of the catalytic core of the ATP synthase and plays a key role in Complex V assembly [81]. An increase of Complex V protein expression during hypoxia appears counterintuitive, because mitochondrial ATP synthase activity is typically suppressed during hypoxia in bivalves such as scallops and clams [11]. This decrease has been proposed to serve as an adaptive mechanism to prevent ATP wastage due to the reverse action of mitochondrial F_O, F_1-ATPase [82]. In contrast, in the white-leg shrimp *Litopenaeus vannamei* exposure to hypoxia results in elevated activity of mitochondrial ATP synthase [83]. Further investigations are needed to determine whether changes in the abundance of Complex V subunits in oyster mitochondria during hypoxia have functional consequences for the mitochondrial ATP synthase activity, and whether other proteins such as the ATPase Inhibitory Factor 1 [84] modulate ATP synthase and OXPHOS activity during H/R stress in hypoxia-tolerant bivalves such as oysters.

4.2. Intermediary metabolism and TCA cycle

Post-hypoxic reoxygenation resulted in a decreased abundance of enoyl-CoA-Δ isomerase (a key enzyme involved in β-oxidation of unsaturated fatty acids) and several enzymes from the acyl-CoA dehydrogenase (ACAD) family involved in fatty acid and amino acid catabolism (including glutaryl-CoA dehydrogenase, isovaleryl-CoA

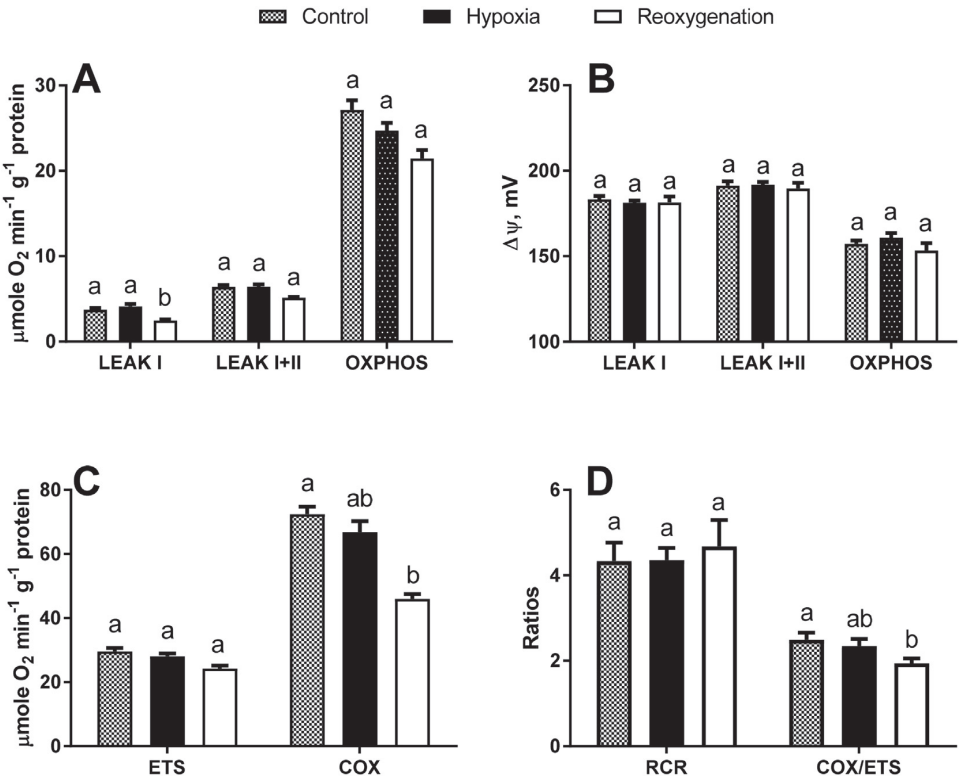

Fig. 1. Effects of hypoxia and reoxygenation on oxygen consumption rates and mitochondrial membrane potential of mitochondria from *C. gigas* gills.
A –rates of proton leak of mitochondria respiring on NADH-linked (LEAK I) and NADH- and FADH2-linked substrate (LEAK I + II), and oxidative phosphorylation rates (OXPHOS); B –mitochondrial membrane potential of resting mitochondria respiring on NADH-linked (LEAK I) and NADH- and FADH2-linked substrate (LEAK I + II), and actively phosphorylating mitochondria (OXPHOS); C –maximum rates of ETS and COX activities; D – respiratory control ratio and ratio of COX to ETS activity. Different letters indicate values that significantly differ from each other ($P < 0.05$). $N = 6$.

dehydrogenase, methylmalony-CoA epimerase, and short-chain specific acyl-CoA dehydrogenases) in oyster mitochondria (Fig. 2). Furthermore, hypoxic exposure suppressed the levels of very long-chain specific acyl-CoA dehydrogenase in oyster mitochondria. Mitochondrial acyl-CoA dehydrogenases play a key role in fatty acid and amino acid oxidation as they transfer electrons from corresponding CoA esters to ubiquinone on the mitochondrial Complex III with the help of the electron transferring protein (ETF) and ETF dehydrogenase [85]. Therefore, a decrease in the levels of enoyl-CoA-Δ isomerase and ACAD enzymes may indicate suppression of β-oxidation and amino acid oxidation. This suggests that oyster mitochondria more strongly rely on oxidation of carbohydrates than fatty acids during reoxygenation. A two-fold decrease in citrate lyase abundance during reoxygenation indicates that fatty acid biosynthesis might also be suppressed. Reoxygenation also resulted in higher phosphorylation levels (and thus presumably lower activity) of mitochondrial acylpyruvase FAHD1, an enzyme involved in catabolism of tyrosine and other aromatic amino acids [86,87]. Likewise, the levels of phosphorylated GDP-forming succinate-CoA ligase (SUCL), which catalyzes the reversible conversion of succinyl-CoA and GDP to succinate [88], increased during post-hypoxic recovery in oyster mitochondria. The functional consequences of β-subunit phosphorylation for SUCL activity are not known [89]; however, changes in the phosphorylation status of this enzyme are consistent with the modulation of CoA-dependent metabolic pathways by H/R stress in oyster mitochondria.

The H/R-induced changes in the oyster mitochondrial (phospho-) proteome indicate a regulated suppression of the substrate influx into the mitochondria during the H/R stress. Thus, hypoxia and reoxygenation resulted in a dramatic increase (by ~6 and ~10-fold, respectively) of the phosphorylation levels of an important gatekeeper enzyme, pyruvate dehydrogenase (PDH), in oyster mitochondria. PDH catalyzes conversion of pyruvate to acetyl coenzyme A (acetyl-CoA) and thereby connects glycolysis to the tricarboxylic acid (TCA) cycle and OXPHOS. Phosphorylation of E1 subunit of PDH deactivates the enzyme and thus interrupts the pyruvate supply into the TCA cycle [39,90]. Our present study shows that, similar to mammals [39,90–92] and yeast [90,93], PDH phosphorylation plays an important role in mitochondrial responses to H/R stress in the hypoxia-tolerant bivalve *C. gigas*. This supports the role of PDH phosphorylation as an ancient and highly conserved mechanism of metabolic regulation during oxygen deficiency. It remains open to speculation whether suppression of PDH activity during H/R stress is an adaptive response that limits influx of electrons into ETS and excessive ROS formation or a side effect of the hypoxia-induced shifts in the energy status and concentrations of metabolic intermediates [90,92,94]. Furthermore, mitochondrial dicarboxylate carrier (a transporter of malate and succinate) and carnitine O-palmitoyltransferase 2 (a fatty acid transporter) were downregulated during hypoxia in oyster mitochondria. Taken together, these changes could potentially limit the substrate uptake into the oyster mitochondria during oxygen limitation.

It is worth noting that reoxygenation resulted in an increased abundance of Ca^{2+}-binding mitochondrial carrier Aralar 2 in oyster mitochondria, which acts as a substrate transporter and Ca^{2+} signal transducer to regulate OXPHOS [95–97]. Aralar 2 is a Ca^{2+}-dependent transporter that provides mitochondria with glutamate and NADH in response to high ATP demand triggered by elevated cytosolic [Ca^{2+}]

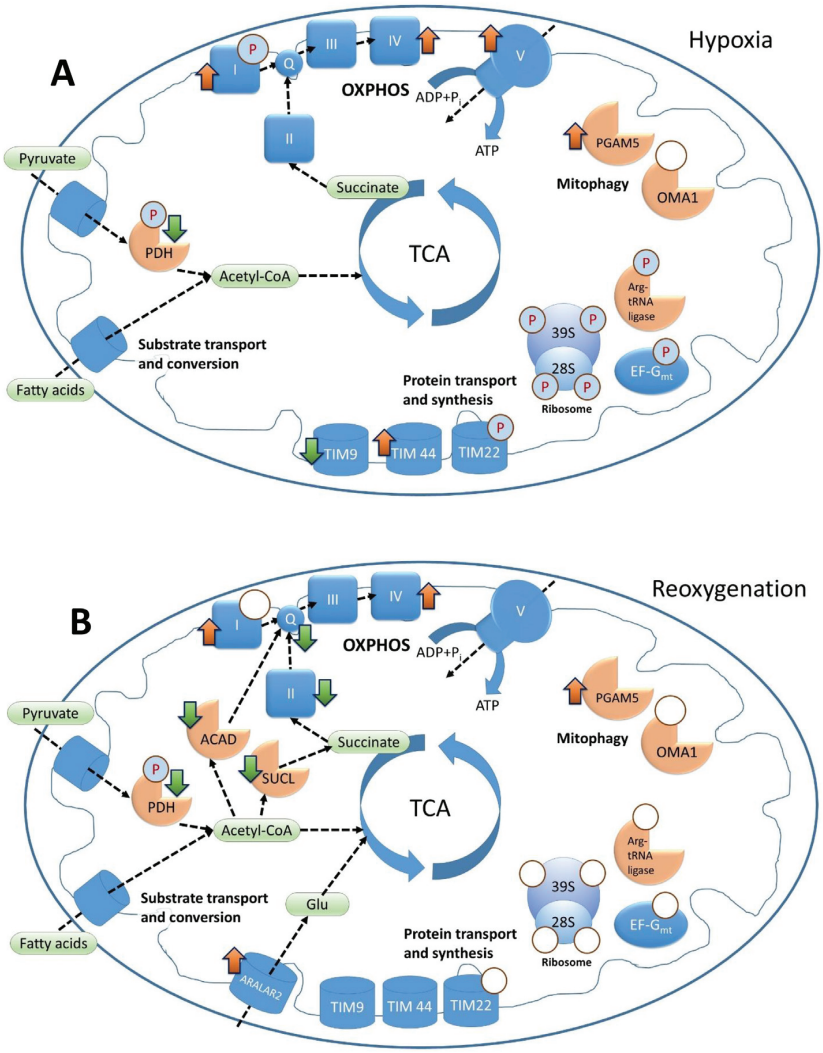

Fig. 2. Summary of the changes in the abundance and phosphorylation status of proteins involved in mitochondrial OXPHOS, protein transport and synthesis and quality control caused by hypoxia and reoxygenation in gill mitochondria of *C. gigas*.
A – hypoxia, B- reoxygenation. Squares indicate ETS complexes I, II, II and IV, cylinders indicate membrane transporters, cut-out circles indicate all other enzymes, and ovals indicate non-enzymatic proteins (such as transcription factors or ion-binding proteins). Filled circles with P indicate proteins whose phosphorylation levels increased, and empty circles indicate proteins whose phosphorylation levels decreased under the respective conditions. Short thick upward and downward arrows indicate proteins that increased or decreased in abundance compared with the control, respectively. Abbreviations: Q – ubiquinone; TCA – tricarboxylic acid cycle; PDH – pyruvate dehydrogenase; 39S and 28 S – 39S and 28S subunits of ribosome; EF-G_{mt} – mitochondrial elongation factor G; OMA1 - metalloendopeptidase OMA1; PGAM5 - Serine/threonine-protein phosphatase PGAM5; TIM9, TIM44 and TIM22 - mitochondrial import inner membrane translocase subunits 9, 44 and 22, respectively. Thin broken arrows show the flow of electrons and/or substrates.

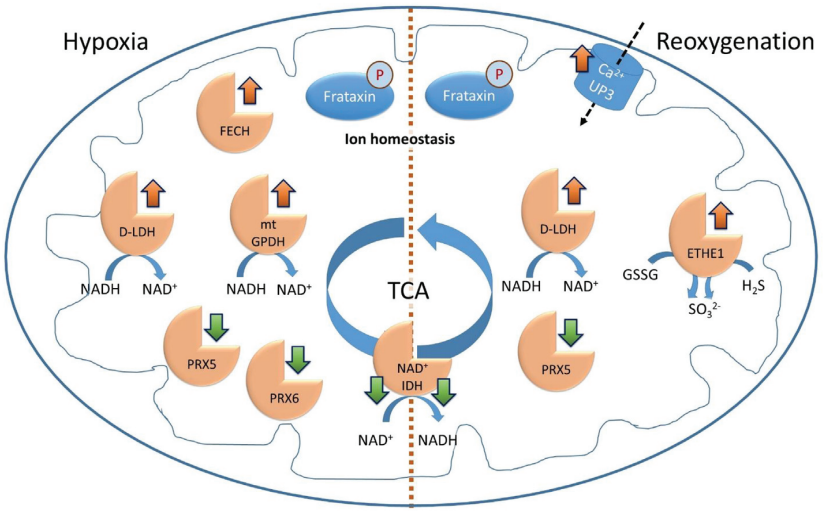

Fig. 3. Summary of the changes in the abundance and phosphorylation status of proteins involved in mitochondrial ion and redox homeostasis during hypoxia and reoxygenation in gill mitochondria of *C. gigas*.
Left-hand side – hypoxia, right-hand side – reoxygenation. Cylinders indicate membrane transporters, cut-out circles indicate enzymes, and ovals indicate non-enzymatic proteins (such as ion-binding proteins). Filled circles with P indicate proteins whose phosphorylation levels increased, and empty circles indicate proteins whose phosphorylation levels decreased under the respective conditions. Short thick upward and downward arrows indicate proteins that increased or decreased in abundance compared with the control, respectively. Abbreviations: FECH – ferrochelatase; Ca^{2+} UP3 – Ca^{2+} uptake protein 3; LDH – lactate dehydrogenase; IDH – isocitrate dehydrogenase; mtGPDH – mitochondrial Glycerol-3-phosphate dehydrogenase; PRX – peroxiredoxin; ETHE1 - persulfide dioxygenase ETHE1; GSSG – oxidized glutathione.

[95–97]. Elevated abundance of Aralar 2 during reoxygenation may indicate increased oxidation of glutamate (as compared with other substrates) during post-hypoxic recovery in oysters, or may be a compensatory response to restore mitochondrial redox homeostasis after prolonged hypoxia.

4.3. Mitochondrial quality control

Hypoxia and reoxygenation stress resulted in a significant upregulation (by ~4 and 3-fold, respectively) of the serine/threonine protein phosphatase PGAM5 in oyster mitochondria (Fig. 2). PGAM5 plays a central role in the quality control of mitochondria and regulation of mitophagy, as well as apoptosis and necroptosis [32,98–100]. In mammals, PGAM5 is engaged in the protective response against H/R-induced injury by stimulating mitophagy to selectively remove damaged mitochondria [98], preventing activation of apoptotic pathways [101], and acting as an essential partner of mitoprotective heme oxygenase 1 [102]. Increased abundance of PGAM5 in the oyster mitoproteome during H/R stress indicates upregulation of the mitochondrial quality control pathways. H/R stress also resulted in a dramatic decrease of phosphorylation levels of mitochondrial metalloendopeptidase OMA1 in oyster mitochondria. This protease plays a crucial role in mitochondrial proteostasis and is activated by mitochondrial stress such as a decline in ATP levels or mitochondrial depolarization [103,104]. The mechanisms of OMA1 regulation by stressors is not yet fully understood, but recent studies show that ATP deficiency activates OMA1, while exposure to high levels of ATP induces OMA1 degradation in mitochondria [105]. If reversible phosphorylation is involved in this ATP–dependent regulation of OMA1 activity [105], dephosphorylation of OMA1 would be expected to activate it, thereby fostering mitochondrial fragmentation and suppressing mitochondrial fusion in stress [103–105]. This putative mechanism of activation by dephosphorylation would explain a dramatic decrease in abundance of OMA1 phosphoprotein during H/R in oyster mitochondria; however, at present it remains speculative since the role of posttranslational modifications in OMA1 regulation is unknown and requires further investigations.

Notably, earlier studies in marine bivalves showed that hypoxia-tolerant species, such as oysters and hard shell clams, express high levels of mitochondrial proteases (especially ATP-dependent mitoproteases such as LON) that are involved in the protein quality control and selective removal of oxidatively damaged proteins from mitochondria [11,12,106]. Unlike hypoxia-tolerant clams that maintained high activity of mitochondrial proteases, hypoxia-sensitive bay scallops showed suppression of the mitochondrial protease activity during H/R stress that went hand-in-hand with the loss of the functional activity of mitochondria [11,12]. Interestingly, the maintenance of the high proteolytic activity during H/R stress in hypoxia-tolerant species of bivalves was limited to the mitochondria, since activity of cytosolic proteases (such as the 28S proteasome) was strongly suppressed (presumably as an energy-saving mechanism) in hypoxia [11,12]. Overall, these findings show that the mitochondrial quality control plays an essential role in H/R responses, and the ability to upregulate key enzymes at the quality control check points (such as PGAM5, OMA1 or LON protease) might be involved in adaptations to frequent H/R stress in hypoxia-tolerant intertidal bivalves.

4.4. Protein homeostasis

H/R stress had a major impact on the mitochondrial pathways involved in protein synthesis and homeostasis (Fig. 2). Exposure to prolonged hypoxia strongly suppressed the levels of subunit Tim9 of the mitochondrial inner membrane translocase (TIM) and increased the abundance of subunit Tim44 in oyster mitochondria, while reoxygenation led to a decrease in the levels of phosphorylated Tim22. Tim9 is an essential protein in the Tim9-Tim10 protein complex that aids in translocation, assembly and membrane insertion of transmembrane mitochondrial proteins [107,108] and protects other members of this complex from proteolytic degradation [109]. In contrast, Tim44 is involved as the terminal member of the translocase pathway that mediates mitochondrial import of protein precursors in an ATP- and HSP70-dependent manner [107,110]. The opposite direction of changes in Tim9 and Tim44 in response to hypoxia might thus indicate differential modulation of certain protein import pathways or general dysregulation of the protein import in oyster mitochondria. Regardless of the functional consequences of hypoxia-induced changes of Tim9 and Tim44 abundance, the mitochondrial levels of these proteins are quickly restored to baseline levels within 1 h of reoxygenation. Overall, our present data indicate that the mitochondrial TIM complexes are an important target for hypoxia-induced modulation in oysters and functional consequences of the observed changes in TIM proteins and/or phosphoproteins require further investigation.

Notably, reoxygenation led to a concerted decrease in the phosphorylation levels of multiple proteins involved in the mitochondrial protein synthesis, including proteins S17, S18a, S21, S22 and S24 of the 28S subunit and proteins L9 and L455 of the 39S subunit of mitochondrial ribosome. Typically, phosphorylation of the ribosomal proteins in the 28S subunit results in suppression of ribosomal translation [111,112]. In hypoxia-tolerant marine mollusks such as oysters, mussels and clams, hypoxia leads to a strong decrease of protein synthesis due to transcriptional and translational arrest [11,113–115]. Therefore, a decrease in the level of phosphorylation of ribosomal proteins during reoxygenation may assist in restoration of the proteosynthetic activity following the hypoxia-induced suppression of the protein synthesis. This correlates with a decrease in the abundance of the phosphorylated form of elongation factor G (by ~5-fold and 3-fold, respectively) during hypoxia and reoxygenation in oyster mitochondria. The mitochondrial elongation factor G (EF-G_{mt}) is a key factor in ribosomal recycling [116,117]. Dephosphorylation activates EF-G_{mt} and thereby increases ribosome turnover and facilitates protein synthesis [118]. Interestingly, phosphorylation levels of the mitochondrial arginine-tRNA ligase (involved in the activation of amino acid arginine for protein synthesis) increased during hypoxia and strongly declined during reoxygenation. This is consistent with the hypoxia-induced suppression of protein translation and its subsequent stimulation during reoxygenation. Taken together, these changes indicate that reversible phosphorylation of multiple proteins associated with ribosomal activity plays an important role in the post-hypoxic recovery and restoration of protein homeostasis of oyster mitochondria.

4.5. Ion homeostasis

H/R stress in oysters affected the phosphorylation levels of several key proteins involved in mitochondrial iron and calcium homeostasis including ferrochelatase, frataxin and calcium uptake protein (Fig. 3). Ferrochelatase is the terminal enzyme of heme biosynthesis which catalyzes the insertion of iron into protoporphyrin IX, yielding heme [119,120]. Ferrochelatase was strongly (~2-fold) upregulated in the mitochondria of hypoxia-exposed oysters and quickly returned to the control levels during reoxygenation. Ferrochelatase expression sensitively responds to cellular Fe^{3+} levels [121], and an increase in mitochondrial ferrochelatase in hypoxia-exposed oysters might reflect elevated Fe^{3+} loads due to the release of iron from the storage proteins and/or the active center of aconitase [106,122–124]. Similar to ferrochelatase, frataxin plays an important role in mitochondrial iron homeostasis and Fe—S cluster formation [125,126]. In mammals, frataxin protects hypoxia-sensitive tissues (such as heart) against ischemia-reperfusion injury by acting as an iron storage during hypoxia [125]. Post-hypoxic reoxygenation commonly induces a surge of reactive oxygen species in a variety of organisms, including mollusks [2,12,20,127,128]. Therefore, a strong increase in frataxin levels in oyster mitochondria after 1 h of post-hypoxic recovery may contribute

to anti-oxidant protection of oyster mitochondria by controlling the mitochondrial iron load. Notably, hypoxia and reoxygenation led to an increase of the relative levels of phosphorylated frataxin (by ~60–70%) in oyster mitochondria. Frataxin phosphorylation typically targets the protein for ubiquitination and subsequent degradation [129], and elevated levels of phosphorylated frataxin in oyster mitochondria may indicate increased turnover of this protein during the H/R-induced stress.

H/R stress also led to a decrease in abundance of calcium uptake protein 3 in oyster mitochondria. Ca^{2+} overload is a common mechanism responsible for H/R-induced mitochondrial injury in hypoxia-sensitive mammals [4,130], so that downregulation of mitochondrial Ca^{2+} uptake might protect oyster mitochondria from damage caused by excessive Ca^{2+} accumulation.

4.6. Redox balance

Maintenance of the mitochondrial redox balance and NADH/NAD^+ plays an important role in the tolerance to H/R stress by preventing electron slip from highly reduced electron carriers and resulting oxidative injury [1,3,69]. In oyster mitochondria, H/R stress resulted in a notable increase in the abundance of NADH-oxidizing enzymes (including mitochondrial D-lactate dehydrogenase and glycerol-3-phosphate dehydrogenase mtGPDH) and a decrease in the levels of NADH-generating isocitrate dehydrogenase (NAD^+-IDH subunit α) (Fig. 3).

Interestingly, the abundance of mitochondrial antioxidant enzymes peroxiredoxin 5 and 6 decreased during hypoxia in oysters, and in the case of peroxiredoxin 5, remained low during reoxygenation. This finding contradicts earlier reports of increase in the total antioxidant capacity in mitochondria of hypoxia-tolerant mollusks during hypoxia [1,12,131]. This increase has been proposed as an adaptive mechanism which prepares the cells to counteract ROS burst during reoxygenation [132]. A decrease in peroxiredoxins 5 and 6 during hypoxia in oysters might be due to the specific mechanisms of these enzymes that reduce hydrogen peroxide and organic peroxides to water and alcohols, so that suppression of peroxiredoxins might prevent accumulation of potentially toxic end products during hypoxia-induced valve closure in bivalves.

Reoxygenation also resulted in upregulation of persulfide dioxygenase ETHE1 in oyster mitochondria, an important oxygen-dependent enzyme that catabolizes hydrogen sulfide (H_2S) and organic persulfides derived from cysteine catabolism in the mitochondrial matrix [133–135]. H_2S production by anaerobic bacteria associated with the shells and tissue surfaces considerably shortens the survival time of bivalves in low oxygen conditions [136–138], and upregulation of H_2S-catabolizing enzymes might protect oyster mitochondria from H_2S toxicity during oxygen fluctuations. High activity of H_2S-catabolizing enzymes was earlier reported in mitochondria of other hypoxia-tolerant mollusks [139–142] and might represent a common adaptation to survive in periodically hypoxic environments such as estuaries or coastal dead zones [1].

4.7. Conclusions and outlook

Mitochondria of a hypoxia-tolerant intertidal bivalve *C. gigas* showed functional robustness (indicated by the ability to maintain normal ETS and OXPHOS capacity as well as the mitochondrial membrane potential) during H/R stress, similar to other hypoxia-tolerant bivalves and fish and unlike hypoxia-sensitive invertebrate and vertebrate species, where reoxygenation results in mitochondrial injury and functional collapse [review in [1]]. Functional stability of oyster mitochondria during H/R stress goes hand-in-hand with significant rearrangements of the mitochondrial proteome and phosphoproteome that start in hypoxia but become considerably more pronounced during reoxygenation, affecting multiple pathways (Figs. 2, 3). The observed patterns of protein up- and down-regulation during reoxygenation in oyster mitochondria indicate that post-hypoxic recovery is not a passive restoration of the normoxic status quo, but an active regulation responding to the unique physiological challenges caused by reinstated O_2 influx. These rearrangements involve upregulation of mitochondrial ETS proteins (most notably Complexes I and IV), suppression of pathways channeling electrons to ubiquinone, stimulation of mitochondrial quality control mechanisms and modulation of protein synthesis and transport. Overall, our study indicates that shifts in the mitochondrial proteome may play an important role in responses to intermittent hypoxia potentially underlying the functional mitochondrial reorganization elicited by H/R stress [11,12,20] and complementing adaptive shifts in anaerobic metabolism and metabolic rate depression [78,113,143].

Acknowledgements

This work was in part supported by the funding line *Strategic Networks of the Leibniz Association* within the scope of the Leibniz ScienceCampus Phosphorus Research Rostock (www.sciencecampus-rostock.de) and the German Research Foundation DFG award MA 6346/2-1. We thank Jana Matulla and Anne Brauer for technical assistance during protein sample preparation, Sebastian Grund for LTQ Orbitrap Classic measurements, and Elisa Kasbohm for introduction to the random forests analysis. We also thank two anonymous reviewers for their constructive comments on an earlier version of this manuscript.

References

[1] I. Sokolova, Mitochondrial adaptations to variable environments and their role in animals' stress tolerance, Integr. Comp. Biol. 58 (2018) 519–531 (icy017-icy017).
[2] M. Hermes-Lima, T. Zenteno-Savín, Animal response to drastic changes in oxygen availability and physiological oxidative stress, Comp. Biochem. Physiol. C 133 (2002) 537–556.
[3] S. Cadenas, ROS and redox signaling in myocardial ischemia-reperfusion injury and cardioprotection, Free Radic. Biol. Med. 117 (2018) 76–89.
[4] T. Kalogeris, C.P. Baines, M. Krenz, R.J. Korthuis, Cell Biology of Ischemia/Reperfusion Injury, Int. Rev. Cell Mol. Biol. 298 (2012) 229–317.
[5] Y. Wang, S.X. Zhang, D. Gozal, Reactive oxygen species and the brain in sleep apnea, Respir. Physiol. Neurobiol. 174 (2010) 307–316.
[6] A.A. Chiang, Obstructive sleep apnea and chronic intermittent hypoxia: a review, Chin. J. Phys. 49 (2006) 234–243.
[7] A. Bogdanova, B. Grenacher, M. Nikinmaa, M. Gassmann, Hypoxic responses of Na+/K+ ATPase in trout hepatocytes, J. Exp. Biol. 2005 (1793-1801) 208.
[8] P.W. Hochachka, L.T. Buck, C.J. Doll, S.C. Land, Unifying theory of hypoxia tolerance: molecular/metabolic defense and rescue mechanisms for surviving oxygen lack, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 9493–9498.
[9] L.T. Buck, M.E. Pamenter, Adaptive responses of vertebrate neurons to anoxia–Matching supply to demand, Respir. Physiol. Neurobiol. 154 (2006) 226.
[10] P.W. Hochachka, P.L. Lutz, Mechanism, origin, and evolution of anoxia tolerance in animals, Comp. Biochem. Physiol. B: Biochem. Mol. Biol. 130 (2001) 435–459.
[11] A.V. Ivanina, I. Nesmelova, L. Leamy, E.P. Sokolov, I.M. Sokolova, Intermittent hypoxia leads to functional reorganization of mitochondria and affects cellular bioenergetics in marine molluscs, J. Exp. Biol. 219 (2016) 1659–1674.
[12] A.V. Ivanina, I.M. Sokolova, Effects of intermittent hypoxia on oxidative stress and protein degradation in molluscan mitochondria, J. Exp. Biol. 219 (2016) 3794–3802.
[13] Z. Hilton, K.D. Clements, A.J.R. Hickey, Temperature sensitivity of cardiac mitochondria in intertidal and subtidal triplefin fishes, J. Comp. Physiol. B. 180 (2010) 979–990.
[14] R. Sussarellu, T. Dudognon, C. Fabioux, P. Soudant, et al., Rapid mitochondrial adjustments in response to short-term hypoxia and re-oxygenation in the Pacific oyster, Crassostrea gigas, J. Exp. Biol. 216 (2013) 1561–1569.
[15] Edward T. Chouchani, Victoria R. Pell, Andrew M. James, Lorraine M. Work, et al., A unifying mechanism for mitochondrial superoxide production during Ischemia-reperfusion injury, Cell Metab. 23 (2016) 254–263.
[16] L. Wang, S. Cui, Z. Liu, Y. Ping, et al., Inhibition of mitochondrial respiration under hypoxia and increased antioxidant activity after reoxygenation of Tribolium castaneum, PLoS One 13 (2018) e0199056.
[17] M.A. Geihs, M.A. Vargas, L.E.M. Nery, Damage caused during hypoxia and reoxygenation in the locomotor muscle of the crab Neohelice granulata (Decapoda: Varunidae), Comp. Biochem. Physiol. A Mol. Integr. Physiol. 172 (2014) 1–9.
[18] M.A. Geihs, F.E. Maciel, M.A. Vargas, B.P. Cruz, L.E.M. Nery, Effects of hypoxia and reoxygenation on the energetic metabolism of the crab Neohelice granulata (Decapoda, Varunidae), J. Exp. Mar. Biol. Ecol. 445 (2013) 69–78.
[19] E.J. Lesnefsky, Q. Chen, B. Tandler, C.L. Hoppel, Mitochondrial Dysfunction and Myocardial Ischemia-Reperfusion: Implications for Novel Therapies, Annu. Rev. Pharmacol. Toxicol. 57 (2017) 535–565.
[20] I.O. Kurochkin, A.V. Ivanina, S. Eilers, C.A. Downs, et al., Cadmium affects

metabolic responses to prolonged anoxia and reoxygenation in eastern oysters (Crassostrea virginica), Am. J. Phys. Regul. Integr. Comp. Phys. 297 (2009) R1262–R1272.
[21] L. Tomanek, Environmental proteomics: changes in the proteome of marine organisms in response to environmental stress, pollutants, infection, symbiosis, and development, Annu. Rev. Mar. Sci. 3 (2011) 373–399.
[22] P. Fields, C. Eurich, W. Gao, B. Cela, Changes on protein expression in the salt marsh mussel Geukensia demissa:evidence for a shift from anaerobic to aerobic metabolism, J. Exp. Biol. 217 (2014) 1601–1612.
[23] J. Mukherjee, K.K.W. Wong, K.H. Chandramouli, P.-Y. Qian, et al., Proteomic response of marine invertebrate larvae to ocean acidification and hypoxia during metamorphosis and calcification, J. Exp. Biol. 216 (2013) 4580–4589.
[24] H. Jiang, F. Li, Y. Xie, B. Huang, et al., Comparative proteomic profiles of the hepatopancreas in Fenneropenaeus chinensis response to hypoxic stress, Proteomics 9 (2009) 3353–3367.
[25] C.A.I. Bosworth, C.W. Chou, R.B. Cole, B.B. Rees, Protein expression patterns in zebrafish skeletal muscle: initial characterization and the effects of hypoxic exposure, Proteomics 5 (2005) 1362–1371.
[26] A. Groehler, S. Kren, Q. Li, M. Robledo-Villafane, et al., Oxidative cross-linking of proteins to DNA following ischemia-reperfusion injury, Free Radic. Biol. Med. 120 (2018) 89–101.
[27] T. Eismann, N. Huber, T. Shin, S. Kuboki, et al., Peroxiredoxin-6 protects against mitochondrial dysfunction and liver injury during ischemia-reperfusion in mice, Am. J. Physiol. Gastrointest. Liver Physiol. 296 (2009) G266–G274.
[28] T.A. Schiffer, M. Peleli, M.L. Sundqvist, B. Ekblom, et al., Control of human energy expenditure by cytochrome C oxidase subunit IV-2, Am. J. Phys. Cell Phys. 311 (2016) C452–C461.
[29] R. Fukuda, H. Zhang, J. Kim, L. Shimoda, C. Dang, HIF-1 regulates cytochrome oxidase subunits to optimize efficiency of respiration in hypoxic cells, Cell 129 (2007) 111–122.
[30] K.J. Cowan, K.B. Storey, Mitogen-activated protein kinases: new signaling pathways functioning in cellular responses to environmental stress, J. Exp. Biol. 206 (2003) 1107–1115.
[31] C. Koumenis, C. Naczki, M. Koritzinsky, S. Rastani, et al., Regulation of protein synthesis by hypoxia via activation of the endoplasmic reticulum kinase PERK and phosphorylation of the translation initiation factor eIF2{alpha}, Mol. Cell. Biol. 22 (2002) 7405–7416.
[32] G. Zhou, T. Golden, I.V. Aragon, R.E. Honkanen, Ser/Thr protein phosphatase 5 inactivates Hypoxia-induced activation of an apoptosis signal-regulating kinase 1/MKK-4/JNK signaling cascade, J. Biol. Chem. 279 (2004) 46595–46605.
[33] M.B. Smolinski, J.J.L. Mattice, K.B. Storey, Regulation of pyruvate kinase in skeletal muscle of the freeze tolerant wood frog, *Rana sylvatica*, Cryobiology 77 (2017) 25–33.
[34] S.R. Green, K.B. Storey, Regulation of crayfish, Orconectes virilis, tail muscle lactate dehydrogenase (LDH) in response to anoxic conditions is associated with alterations in phosphorylation patterns, Comp. Biochem. Physiol. B: Biochem. Mol. Biol. 202 (2016) 67–74.
[35] J.L. Lama, R.A.V. Bell, K.B. Storey, Hexokinase regulation in the hepatopancreas and foot muscle of the anoxia-tolerant marine mollusc, Littorina littorea, Comp. Biochem. Physiol. B: Biochem. Mol. Biol. 166 (2013) 109–116.
[36] S.P.J. Brooks, K.B. Storey, Glycolytic controls in estivation and anoxia: a comparison of metabolic arrest in land and marine molluscs, Comp. Biochem. Physiol. 118A (1997) 1103–1114.
[37] B. Michaelidis, K.B. Storey, Evidence for phosphorylation/dephosphorylation control of phosphofructokinase from organs of the anoxia-tolerant sea mussel Mytilus edulis, J. Exp. Zool. 257 (1991) 1–9.
[38] R.E. Whitwam, K.B. Storey, Pyruvate kinase from the land snail Otala lactea: regulation by reversible phosphorylation during estivation and anoxia, J. Exp. Biol. 154 (1990) 321–337.
[39] A.D. Zimmer, G. Walbrecq, I. Kozar, I. Behrmann, C. Haan, Phosphorylation of the PDH complex precedes HIF-1-mediated effects and PDK1 upregulation during the first hours of hypoxic treatment in HCC cells, Hypoxia 4 (2016) 135–145.
[40] C.J. Ramnanan, K.B. Storey, Suppression of Na^+/K^+-ATPase activity during estivation in the land snail Otala lactea, J. Exp. Biol. 209 (2006) 677–688.
[41] K.B. Storey, Regulation of hypometabolism: insights into epigenetic controls, J. Exp. Biol. 218 (2015) 150–159.
[42] N. Gedik, M. Krüger, M. Thielmann, E. Kottenberg, et al., Proteomics/phosphoproteomics of left ventricular biopsies from patients with surgical coronary revascularization and pigs with coronary occlusion/reperfusion: remote ischemic preconditioning, Sci. Rep. 7 (2017) 7629.
[43] M. Schönenberger, W. Kovacs, Hypoxia signaling pathways: modulators of oxygen-related organelles, Front. Cell Dev. Biol. 3 (2015).
[44] G.L.J. Galli, J.G. Richards, Mitochondria from anoxia-tolerant animals reveal common strategies to survive without oxygen, J. Comp. Physiol. B. 184 (2014) 285–302.
[45] K. Reise, C. Buschbaum, H. Büttger, K.M. Wegner, Invading oysters and native mussels: from hostile takeover to compatible bedfellows, Ecosphere 8 (2017) e01949-n/a.
[46] J.L. Gutiérrez, C.G. Jones, D.L. Strayer, O.O. Iribarne, Mollusks as ecosystem engineers: the role of shell production in aquatic habitats, Oikos 101 (2003) 79–90.
[47] G. Zhang, X. Fang, X. Guo, L. Li, et al., The oyster genome reveals stress adaptation and complexity of shell formation, Nature 490 (2012) 49–54.
[48] G. Le Moullac, I. Queau, P. Le Souchu, S. Pouvreau, et al., Metabolic adjustments in the oyster Crassostrea gigas according to oxygen level and temperature, Mar. Biol. Res. 3 (2007) 357–366.
[49] A.V. Ivanina, B. Froelich, T. Williams, E.P. Sokolov, et al., Interactive effects of cadmium and hypoxia on metabolic responses and bacterial loads of eastern oysters Crassostrea virginica Gmelin, Chemosphere 82 (2011) 377–389.
[50] V.S. Kennedy, R.I.E. Newell, A.F. Eble (Eds.), The Eastern Oyster Crassostrea virginica, 1996 A Maryland Sea Grant Book, College Park, Maryland.
[51] E.P. Sokolov, I.M. Sokolova, Compatible osmolytes modulate mitochondrial function in a marine osmoconformer Crassostrea gigas (Thunberg, 1793), Mitochondrion (2018), https://doi.org/10.1016/j.mito.2018.02.002.
[52] A.V. Ivanina, I.M. Sokolova, Effects of cadmium exposure on expression and activity of P-glycoprotein in eastern oysters, Crassostrea virginica Gmelin, Aquat. Toxicol. 88 (2008) 19–28.
[53] N.R. Sims, M.F. Anderson, Isolation of mitochondria from rat brain using Percoll density gradient centrifugation, Nat. Protoc. 3 (2008) 1228–1239.
[54] M. Morgenstern, S.B. Stiller, P. Lübbert, C.D. Peikert, et al., Definition of a high-confidence mitochondrial proteome at quantitative scale, Cell Rep. 19 (2017) 2836–2852.
[55] I.O. Kurochkin, M. Etzkorn, D. Buchwalter, L. Leamy, I.M. Sokolova, Top-down control analysis of the cadmium effects on molluscan mitochondria and the mechanisms of cadmium-induced mitochondrial dysfunction, Am. J. Phys. Regul. Integr. Comp. Phys. 300 (2011) R21–R31.
[56] E. Gnaiger, Mitochondrial pathways and respiratory control: An introduction to OXPHOS analysis, Mitochondrial Physiological Network (MiPNet) 17.18, OROBOROS MiPNet Publications, 2012Innsbruck. 2012, 1–64. Open access http://www.bioblast.at/index.php/Gnaiger_2012_MitoPathways.
[57] R.W. Estabrook, Mitochondrial respiratory control and the polarographic measurements in mitochondria, Methods Enzymol. 10 (1967) 41–47.
[58] Measurement of mitochondrial protonmotive force, in: M.D. Brand, G.C. Brown, C.E. Copper (Eds.), Bioenergetics. A Practical Approack, IRL Press, Oxford, New York, Tokyo, 1995, pp. 39–62.
[59] F. Haider, H. Falfushynska, A.V. Ivanina, I.M. Sokolova, Effects of pH and bicarbonate on mitochondrial functions of marine bivalves, Comp. Biochem. Physiol. A Mol. Integr. Physiol. 198 (2016) 41–50.
[60] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Chem. 72 (1976) 248–254.
[61] J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips, Nat. Protoc. 2007 (1896-1906) 2.
[62] P. Pjevac, D.V. Meier, S. Markert, C. Hentschker, et al., Metaproteogenomic profiling of microbial communities colonizing actively venting hydrothermal chimneys, Front. Microbiol. 9 (2018) 680.
[63] B. Zybailov, A.L. Mosley, M.E. Sardiu, M.K. Coleman, et al., Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae, J. Proteome Res. 5 (2006) 2339–2347.
[64] J.A. Vizcaino, A. Csordas, N. Del-Toro, J.A. Dianes, et al., Update of the PRIDE database and its related tools, Nucleic Acids Res. 2016 (44) (2016) D447–D456.
[65] A. Liaw, M. Wiener, Classification and regression by randomforest, R News 2 (2002) 18–22.
[66] R-Core-Team, https://www.r-project.org, (2017).
[67] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[68] R. Díaz-Uriarte, S.A. de Andrés, Gene selection and classification of microarray data using random forest, R News 2 (2002) 18–22.
[69] S. Dröse, A. Stepanova, A. Galkin, Ischemic a/D transition of mitochondrial complex I and its role in ROS generation, Biochim. Biophys. Acta Bioenerg. 1857 (2016) 946–957.
[70] E. Maklashina, Y. Sher, H.-Z. Zhou, M.O. Gray, et al., Effect of anoxia/reperfusion on the reversible active/de-active transition of NADH-ubiquinone oxidoreductase (complex I) in rat heart, Biochim. Biophys. Acta Bioenerg. 1556 (2002) 6–12.
[71] M. Hüttemann, S. Helling, T.H. Sanderson, C. Sinkler, et al., Regulation of mitochondrial respiration and apoptosis through cell signaling: cytochrome c oxidase and cytochrome c in ischemia/reperfusion injury and inflammation, Biochim. Biophys. Acta 2012 (1817) 598–609.
[72] M. Babot, A. Birch, P. Labarbuta, A. Galkin, Characterisation of the active/de-active transition of mitochondrial complex I, Biochim. Biophys. Acta Bioenerg. 1837 (2014) 1083–1092.
[73] S. Arnold, The power of life—Cytochrome c oxidase takes center stage in metabolic control, cell signalling and survival, Mitochondrion 12 (2012) 46–56.
[74] S. Srinivasan, J. Spear, K. Chandran, J. Joseph, et al., Oxidative stress induced mitochondrial protein kinase a mediates cytochrome C oxidase dysfunction, PLoS One 8 (2013) e77129.
[75] T.A. Prime, F.H. Blaikie, C. Evans, S.M. Nadtochiy, et al., A mitochondria-targeted S-nitrosothiol modulates respiration, nitrosates thiols, and protects against ischemia-reperfusion injury, Proc. Natl. Acad. Sci. 106 (2009) 10764–10769.
[76] D.S. Lark, L.R. Reese, T.E. Ryan, M.J. Torres, et al., Protein kinase a governs oxidative phosphorylation kinetics and oxidant emitting potential at complex I, Front. Physiol. 6 (2015).
[77] A.J. Lambert, M.D. Brand, Superoxide production by NADH:ubiquinone oxidoreductase (complex I) depends on the pH gradient across the mitochondrial inner membrane, Biochem. J. 382 (2004) 511–517.
[78] A.V. Ivanina, E.P. Sokolov, I.M. Sokolova, Effects of cadmium on anaerobic energy metabolism and mRNA expression during air exposure and recovery of an intertidal mollusk Crassostrea virginica, Aquat. Toxicol. 99 (2010) 330–342.
[79] I.M. Sokolova, C. Bock, H.O. Pörtner, Resistance to freshwater exposure in White Sea Littorina spp. I: Anaerobic metabolism and energetics, J. Comp. Physiol. B 170 (2000) 91–103.
[80] M.K. Grieshaber, I. Hardewig, U. Kreutzer, H.O. Pörtner, Physiological and metabolic responses to hypoxia in invertebrates, Rev. Physiol. Biochem. Pharmacol. 125 (1994) 43–147.
[81] M.-F. Giraud, J. Velours, The absence of the mitochondrial ATP synthase delta subunit promotes a slow growth phenotype of Rho− yeast cells by a lack of assembly of the catalytic sector F1, Eur. J. Biochem. 245 (2004) 813–818.
[82] J. St-Pierre, M.D. Brand, R.G. Boutilier, Mitochondria as ATP consumers: cellular treason in anoxia, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 8670–8674.
[83] O. Martinez-Cruz, Calderon de la Barca, A. M, S. Uribe-Carvajal, A. Muhlia-Almazan, The function of mitochondrial F_0F_1 ATP-synthase from the whiteleg

shrimp Litopenaeus vannamei muscle during hypoxia, Comp. Biochem. Physiol. B: Biochem. Mol. Biol. 162 (2012) 107–112.
[84] J. García-Bermúdez, J. Cuezva, M., the ATPase inhibitory factor 1 (IF1): a master regulator of energy metabolism and of cell survival, Biochim. Biophys. Acta Bioenerg. 1857 (2016) 1167–1182.
[85] Z. Swigoňová, A.-W. Mohsen, J. Vockley, Acyl-CoA Dehydrogenases: dynamic history of protein family evolution, J. Mol. Evol. 69 (2009) 176–193.
[86] H. Pircher, G.D. Straganz, D. Ehehalt, G. Morrow, et al., Identification of human Fumarylacetoacetate Hydrolase Domain-containing Protein 1 (FAHD1) as a Novel Mitochondrial Acylpyruvase, J. Biol. Chem. 286 (2011) 36500–36508.
[87] A.K.H. Weiss, J.R. Loeffler, K.R. Liedl, H. Gstach, P. Jansen-Durr, The fumarylacetoacetate hydrolase (FAH) superfamily of enzymes: multifunctional enzymes from microbes to mitochondria, Biochem. Soc. Trans. 46 (2018) 295–309.
[88] E. Ostergaard, Disorders caused by deficiency of succinate-CoA ligase, J. Inherit. Metab. Dis. 31 (2008) 226–229.
[89] M.E. Fraser, M.N.G. James, W.A. Bridger, W.T. Wolodko, Phosphorylated and dephosphorylated structures of pig heart, GTP-specific succinyl-CoA synthetase11Edited by D Ress, J. Mol. Biol. 299 (2000) 1325–1339.
[90] G.L. Semenza, Oxygen-dependent regulation of mitochondrial respiration by hypoxia-inducible factor 1, Biochem. J. 405 (2007) 1–9.
[91] B. O'Rourke, J.E.V. Eyk, D.B. Foster, Mitochondrial protein phosphorylation as a regulatory modality: implications for mitochondrial dysfunction in heart failure, Congest. Heart Failure 17 (2011) 269–282.
[92] K. Kobayashi, J.R. Neely, Effects of ischemia and reperfusion on pyruvate dehydrogenase activity in isolated rat hearts, J. Mol. Cell. Cardiol. 15 (1983) 359–367.
[93] U. Gey, C. Czupalla, B. Hoflack, G. Rödel, U. Krause-Buchholz, Yeast Pyruvate Dehydrogenase complex is regulated by a concerted activity of two kinases and two phosphatases, J. Biol. Chem. 283 (2008) 9759–9767.
[94] A. Thibodeau, X. Geng, L. Previch, Y. Ding, Pyruvate dehydrogenase complex in cerebral ischemia-reperfusion injury, Brain Circ. 2 (2016) 61–66.
[95] J. Satrústegui, L. Contreras, M. Ramos, P. Marmol, et al., Role of aralar, the mitochondrial transporter of aspartate-glutamate, in brain N-acetylaspartate formation and Ca2+ signaling in neuronal mitochondria, J. Neurosci. Res. 85 (2007) 3359–3366.
[96] F.N. Gellerich, Z. Gizatullina, T. Gainutdinov, K. Muth, et al., The control of brain mitochondrial energization by cytosolic calcium: the mitochondrial gas pedal, IUBMB Life 65 (2013) 180–190.
[97] Gellerich, Frank N., Gizatullina, Z., Trumbekaite, S., Korzeniewski, B., et al., Cytosolic Ca^2 regulates the energization of isolated brain mitochondria by formation of pyruvate through the malate–aspartate shuttle. Biochem. J. 2012, 443, 747–755.
[98] G. Chen, Z. Han, D. Feng, Y. Chen, et al., A regulatory signaling loop comprising the PGAM5 phosphatase and CK2 controls receptor-mediated mitophagy, Mol. Cell 54 (2014) 362–377.
[99] Z. Wang, H. Jiang, S. Chen, F. Du, X. Wang, The mitochondrial phosphatase PGAM5 functions at the convergence point of multiple Necrotic death pathways, Cell 148 (2012) 228–243.
[100] Z. Zhou, V. Han, J. Han, New components of the necroptotic pathway, Protein Cell 3 (2012) 811–817.
[101] C. Yang, X. Liu, F. Yang, W. Zhang, et al., Mitochondrial phosphatase PGAM5 regulates Keap1-mediated Bcl-xL degradation and controls cardiomyocyte apoptosis driven by myocardial ischemia/reperfusion injury, In Vitro Cell. Dev. Biol. Anim. 53 (2017) 248–257.
[102] J.-M. Hong, S.-M. Lee, Heme oxygenase-1 protects liver against ischemia/reperfusion injury via phosphoglycerate mutase family member 5-mediated mitochondrial quality control, Life Sci. 200 (2018) 94–104.
[103] K. Zhang, H. Li, Z. Song, Membrane depolarization activates the mitochondrial protease OMA1 by stimulating self-cleavage, EMBO Rep. 15 (2014) 576–585.
[104] M.J. Baker, P.A. Lampe, D. Stojanovski, A. Korwitz, et al., Stress-induced OMA1 activation and autocatalytic turnover regulate OPA1-dependent mitochondrial dynamics, EMBO J. 33 (2014) 578–593.
[105] T.K. Rainbolt, J. Lebeau, C. Puchades, R.L. Wiseman, Reciprocal degradation of YME1L and OMA1 adapts mitochondrial proteolytic activity during stress, Cell Rep. 14 (2016) 2041–2049.
[106] B. Sanni, K. Williams, E.P. Sokolov, I.M. Sokolova, Effects of acclimation temperature and cadmium exposure on mitochondrial aconitase and LON protease from a model marine ectotherm, Crassostrea virginica, Comp. Biochem. Physiol. C 147 (2008) 101–112.
[107] A. Chacinska, C.M. Koehler, D. Milenkovic, T. Lithgow, N. Pfanner, Importing mitochondrial proteins: machineries and mechanisms, Cell 138 (2009) 628–644.
[108] A. Vasiljev, U. Ahting, F.E. Nargang, N.E. Go, et al., Reconstituted TOM core complex and Tim9/Tim10 complex of mitochondria are sufficient for translocation of the ADP/ATP carrier across membranes, Mol. Biol. Cell 15 (2004) 1445–1458.
[109] Michael P. Spiller, L. Guo, Q. Wang, P. Tran, H. Lu, Mitochondrial Tim9 protects Tim10 from degradation by the protease Yme1, Biosci. Rep. 35 (2015) e00193.
[110] D.P. Hutu, B. Guiard, A. Chacinska, D. Becker, et al., Mitochondrial protein import motor: differential role of Tim44 in the recruitment of Pam17 and J-complex to the Presequence Translocase, Mol. Biol. Cell 19 (2008) 2642–2649.
[111] J.L. Miller, H. Cimen, H. Koc, E.C. Koc, Phosphorylated proteins of the mammalian mitochondrial ribosome: implications in protein synthesis, J. Proteome Res. 8 (2009) 4789–4798.
[112] E.C. Koc, H. Koc, Regulation of mammalian mitochondrial translation by post-translational modifications, Biochim. Biophys. Acta 2012 (1819) 1055–1066.
[113] I.M. Sokolova, A.A. Sukhotin, G. Lannig, Stress effects on metabolism and energy budgets in mollusks, in: D. Abele, T. Zenteno-Savin, J. Vazquez-Medina (Eds.), Oxidative Stress in Aquatic Ecosystems, Blackwell Wiley, Boston etc, 2011, pp. 263–280.
[114] K.B. Storey, J.M. Storey, Metabolic rate depression in animals: transcriptional and translational controls, Biol. Rev. 79 (2004) 207–233.
[115] M. Guppy, P. Withers, Metabolic depression in animals: physiological perspectives and biochemical generalizations, Biol. Rev. 74 (1999) 1–40.
[116] G.C. Atkinson, S.L. Baldauf, Evolution of Elongation factor G and the origins of mitochondrial and chloroplast forms, Mol. Biol. Evol. 28 (2011) 1281–1292.
[117] M. Tsuboi, H. Morita, Y. Nozaki, K. Akama, et al., EF-G2mt is an exclusive recycling factor in Mammalian mitochondrial protein synthesis, Mol. Cell 35 (2009) 502–510.
[118] A. Talavera, J. Hendrix, W. Versées, D. Jurènas, et al., Phosphorylation decelerates conformational dynamics in bacterial translation elongation factors, Sci. Adv. 4 (2018).
[119] G.C. Ferreira, Ferrochelatase, Int. J. Biochem. Cell Biol. 31 (1999) 995–1000.
[120] G.C. Ferreira, R. Franco, S.G. Lloyd, I. Moura, et al., Structure and function of ferrochelatase, J. Bioenerg. Biomembr. 27 (1995) 221–229.
[121] T. Shigeru, A. Yasushi, N. Yoshitsugu, Regulation of the expression of human ferrochelatase by intracellular iron levels, Eur. J. Biochem. 267 (2000) 4685–4692.
[122] A.S. Cherkasov, R.A. Overton Jr., E.P. Sokolov, I.M. Sokolova, Temperature-dependent effects of cadmium and purine nucleotides on mitochondrial aconitase from a marine ectotherm, Crassostrea virginica: a role of temperature in oxidative stress and allosteric enzyme regulation, J. Exp. Biol. 210 (2007) 46–55.
[123] P. Korge, P. Ping, J.N. Weiss, Reactive oxygen species production in energized cardiac mitochondria during hypoxia/reoxygenation: modulation by nitric oxide, Circ. Res. 103 (2008) 873–880.
[124] C.S. Powell, R.M. Jackson, Mitochondrial complex I, aconitase, and succinate dehydrogenase during hypoxia-reoxygenation: modulation of enzyme activities by MnSOD, Am. J. Phys. Lung Cell. Mol. Phys. 285 (2003) L189–L198.
[125] G. Nanayakkara, A. Alasmari, S. Mouli, H. Eldoumani, et al., Cardioprotective HIF-1α-frataxin signaling against ischemia-reperfusion injury, Am. J. Phys. Heart Circ. Phys. 309 (2015) H867–H879.
[126] F. Lupoli, T. Vannocci, G. Longo, N. Niccolai, A. Pastore, The role of oxidative stress in Friedreich's ataxia, FEBS Lett. 592 (2018) 718–727.
[127] J. Levraut, H. Iwase, Z.H. Shao, T.L. Vanden Hoek, P.T. Schumacker, Cell death during ischemia: relationship to mitochondrial depolarization and ROS generation, Am. J. Physiol. Heart Circ. Physiol. 284 (2003) H549–H558.
[128] G.A. Rivera-Ingraham, I. Rocchetta, S. Meyer, D. Abele, Oxygen radical formation in anoxic transgression and anoxia-reoxygenation: Foe or phantom? Experiments with a hypoxia tolerant bivalve, Mar. Environ. Res. 92 (2013) 110–119.
[129] F. Cherubini, D. Serio, I. Guccini, S. Fortuni, et al., Src inhibitors modulate frataxin protein levels, Hum. Mol. Genet. 24 (2015) 4296–4305.
[130] H.M. Piper, K. Meuter, C. Schäfer, Cellular mechanisms of ischemia-reperfusion injury, Ann. Thorac. Surg. 75 (2003) S644–S648.
[131] D.C. Moreira, L.P.R. Venancio, M.A.C.T. Sabino, M. Hermes-Lima, How widespread is preparation for oxidative stress in the animal kingdom? Comp. Biochem. Physiol. A Mol. Integr. Physiol. 200 (2016) 64–78.
[132] M. Hermes-Lima, J. Storey, M., Storey, K. B., Antioxidant defenses and metabolic depression. The hypothesis of preparation for oxidative stress in land snails, Comp. Biochem. Physiol. B: Biochem. Mol. Biol. 120 (1998) 437–448.
[133] J. Bełtowski, Hydrogen sulfide in pharmacology and medicine – an update, Pharmacol. Rep. 67 (2015) 647–658.
[134] L. Krüßel, J. Junemann, M. Wirtz, H. Birke, et al., The mitochondrial Sulfur Dioxygenase ETHYLMALONIC ENCEPHALOPATHY PROTEIN1 is required for Amino Acid Catabolism during Carbohydrate starvation and embryo development in Arabidopsis, Plant Physiol. 165 (2014) 92–104.
[135] V. Tiranti, C. Viscomi, T. Hildebrandt, I. Di Meo, et al., Loss of ETHE1, a mitochondrial dioxygenase, causes fatal sulfide toxicity in ethylmalonic encephalopathy, Nat. Med. 15 (2009) 200.
[136] A.D. Zwaan, J.M.F. Babarro, M. Monari, O. Cattani, Anoxic survival potential of bivalves: (arte)facts, Comp. Biochem. Physiol. A 131 (2002) 615–624.
[137] J.M.F. Babarro, A. de Zwaan, Influence of abiotic factors on bacterial proliferation and anoxic survival of the sea mussel Mytilus edulis L, J. Exp. Mar. Biol. Ecol. 273 (2002) 33–49.
[138] J.M.F. Babarro, A. de Zwaan, Factors involved in the (near) anoxic survival time of Cerastoderma edule: associated bacteria vs. endogenous fuel, Comp. Biochem. Physiol. C 128 (2001) 325–337.
[139] K.R. Olson, Mitochondrial adaptations to utilize hydrogen sulfide for energy and signaling, J. Comp. Physiol. B. 182 (2012) 881–897.
[140] D.W. Kraus, J.E. Doeller, Sulfide consumption by mussel gill mitochondria is not strictly tied to oxygen reduction: measurements using a novel polarographic sulfide sensor, J. Exp. Biol. 207 (2004) 3667–3679.
[141] J.E. Doeller, Cellular energetics of animals from high Sulfide environments, Am. Zool. 35 (1995) 154–162.
[142] M.A. Powell, G.N. Somero, Hydrogen Sulfide Oxidation is coupled to Oxidative Phosphorylation in Mitochondria of Solemya reidi, Science 233 (1986) 563–566.
[143] Peter W. Hochachka, Peter L. Lutz, Thomas J. Sick, Myron Rosenthal (Eds.), Surviving hypoxia: mechanisms of control and adaptation, CRC Press, 9780849342264, 1993, p. 592 CAT# 4226.

PUBLICATION IV

A new method to calculate stable carbon isotope ratios in communities from metaproteomics data is presented in the publication "Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities".

Supplementary data is provided in the CD.

Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities

Manuel Kleiner[a,b,1], Xiaoli Dong[a], Tjorven Hinzke[a,c,d], Juliane Wippler[e], Erin Thorson[a], Bernhard Mayer[a], and Marc Strous[a,1]

[a]Department of Geoscience, University of Calgary, Calgary, AB, Canada T2N 1N4; [b]Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695 ; [c]Department of Pharmaceutical Biotechnology, Institute of Pharmacy, University of Greifswald, 17489 Greifswald, Germany; [d]Institute of Marine Biotechnology, 17489 Greifswald, Germany; and [e]Symbiosis Department, Max Planck Institute for Marine Microbiology, 28359 Bremen, Germany

Edited by Margaret J. McFall-Ngai, University of Hawaii at Manoa, Honolulu, HI, and approved May 2, 2018 (received for review December 21, 2017)

Measurements of stable carbon isotope ratios ($\delta^{13}C$) are widely used in biology to address questions regarding food sources and metabolic pathways used by organisms. The analysis of these so-called stable isotope fingerprints (SIFs) for microbes involved in biogeochemical cycling and microbiota of plants and animals has led to major discoveries in environmental microbiology. Currently, obtaining SIFs for microbial communities is challenging as the available methods either only provide low taxonomic resolution, such as the use of lipid biomarkers, or are limited in throughput, such as nanoscale secondary ion MS imaging of single cells. Here we present "direct protein-SIF" and the Calis-p software package (https://sourceforge.net/projects/calis-p/), which enable high-throughput measurements of accurate $\delta^{13}C$ values for individual species within a microbial community. We benchmark the method using 20 pure culture microorganisms and show that the method reproducibly provides SIF values consistent with gold-standard bulk measurements performed with an isotope ratio mass spectrometer. Using mock community samples, we demonstrate that SIF values can also be obtained for individual species within a microbial community. Finally, a case study of an obligate bacteria–animal symbiosis shows that direct protein-SIF confirms previous physiological hypotheses and can provide unexpected insights into the symbionts' metabolism. This confirms the usefulness of this approach to accurately determine $\delta^{13}C$ values for different species in microbial community samples.

Protein-SIP | metaproteome | microbial ecology | microbiome | Q Exactive

Measurements of stable carbon isotope ratios ($^{13}C/^{12}C$, commonly called $\delta^{13}C$) are used in many different scientific fields, including atmospheric sciences, biology, paleoclimatology, oceanography, geology, environmental sciences, food and drug authentication, and forensic applications (1). In biology, stable isotope ratios (stable isotope fingerprints, SIFs) can be used to address at least two major questions. First, what is the food source of an organism? This question can be answered based on the principle that heterotrophic organisms usually have a SIF similar to their food source ("you are what you eat") (2). This has been used, for example, to assess the diet of animals (3) and to determine which microorganisms consume a specific carbon source (e.g., methane) in marine sediments (4). The second question, which can be addressed for those organisms that grow on C_1 carbon sources (bicarbonate, CO_2, or methane), is which metabolic pathway is used to assimilate the carbon source. This question can be answered based on the principle that most metabolic pathways/enzymes for C_1 assimilation discriminate against ^{13}C, which leads to characteristic carbon isotope fractionation effects. The extent of isotope fractionation of different C_1 assimilation pathways varies and thus the metabolic pathway used can be predicted based on the extent of the isotope fractionation (2, 5). This has, for example, been used in the past to distinguish plants with different carbon assimilation physiologies (6) and to predict differences in carbon fixation pathways used by symbionts of marine animals (7, 8).

Obtaining the SIFs of individual species in microbial communities is in theory a very promising tool to help unravel important abiotic and biotic interactions in global biogeochemical cycles as well as in microbiota of plants and animals. For example, if SIFs of individual species in the intestinal microbiota of humans were known, we could deduce which dietary components are used by different species in the intestine. However, there is currently no experimental approach to determine the specific SIFs of a large number of species in communities with reasonable effort and cost. The presently available approaches either have no or limited taxonomic resolution or are low-throughput. The most common approach for measuring $^{13}C/^{12}C$ ratios, isotope ratio mass spectrometry (IRMS), usually determines highly

Significance

To understand the roles that microorganisms play in diverse environments such as the open ocean or the human intestinal tract, we need an understanding of their metabolism and physiology. A variety of methods such as metagenomics and metaproteomics exist to assess the metabolism of environmental microorganisms based on gene content and gene expression. These methods often only provide indirect evidence for which substrates are used by a microorganism in a community. The direct protein stable isotope fingerprint (SIF) method that we developed allows linking microbial species in communities to the environmental carbon sources they consume by determining their stable carbon isotope signature. Direct protein-SIF also allows assessing which carbon fixation pathway is used by autotrophic microorganisms that directly assimilate CO_2.

Author contributions: M.K. and M.S. designed research; M.K., X.D., T.H., J.W., E.T., B.M., and M.S. performed research; M.K., X.D., and M.S. analyzed data; and M.K., B.M., and M.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The MS proteomics data and the protein sequence databases have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository for the pure culture data (dataset identifier PXD006762), the mock community data (dataset identifier PXD006118), and the *O. algarvensis* case study data (dataset identifier PXD007510).

[1]To whom correspondence may be addressed. Email: manuel_kleiner@ncsu.edu or mstrous@ucalgary.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1722325115/-/DCSupplemental.

Published online May 29, 2018.

accurate C isotope ratios for bulk organic samples that have been converted to the measurement gas CO_2 via thermal decomposition. In IRMS the measured C isotope ratios are reported as $\delta^{13}C$ values, which give the per mille (‰) deviation of the measured ratio from the internationally accepted standard V-PDB (Vienna Pee Dee Belemnite). If lipid biomarkers are separated followed by their C isotope analysis using IRMS, high-level taxonomic groups can be resolved (9). Recently separation of proteins has also been combined with C isotope analysis using IRMS (P-SIF) (10). The P-SIF approach theoretically allows assigning $\delta^{13}C$ values to 5–10 taxa per sample; however, the approach has only been used on two bacterial pure cultures so far as it is extremely low-throughput because the mass spectrometer run time required for peptide identification alone amounts to around 2 wk per sample. A final approach is the combination of fluorescence in situ hybridization and nanoscale secondary ion MS (nanoSIMS), which enables measurement of $\delta^{13}C$ values of individual cells (4). The nanoSIMS approach, however, is currently difficult to use to measure natural abundance stable isotope ratios because the fixation and labeling procedures required for taxonomic cell identification lead to a large addition of reagent-derived carbon, nitrogen, hydrogen, and oxygen, thus diluting the true sample SIF (11, 12). Additionally, nanoSIMS has a low throughput because specific fluorescently labeled probes have to be applied for each individual species or higher-level taxonomic group.

Here we present "direct protein-SIF" and the Calis-p (The CALgary approach to ISotopes in Proteomics) software package, a method that enables high-throughput measurement of $\delta^{13}C$ (SIF) values for individual species within microbiota and environmental microbial communities using metaproteomics. We use the word "direct" to highlight the fact that the SIF data are directly extracted from a standard metaproteomic dataset (i.e., the same mass spectrometry data are used for both peptide identification and SIF estimation). The direct protein-SIF workflow (Fig. 1 and *SI Appendix*, Fig. S1) consists of a standard proteomics sample preparation to produce peptide mixtures of the samples and a reference material, followed by acquisition of 1D or 2D liquid chromatography tandem MS (LC-MS/MS) data of the peptide mixture using a high-resolution Orbitrap mass spectrometer (for details see *Methods*). The MS/MS data are used as input for a standard proteomic database search to produce a table with scored peptide spectrum matches (PSMs). The PSM tables for the samples and the reference material plus the raw MS data (in mzML format) are used as input for the Calis-p software. In a first step the software finds the isotopic peaks for each PSM and sums their intensity across a retention time window. The isotope peak intensities are reported together with the sum formula of the identified peptides. Isotope peak intensity patterns with low intensities or low search engine scores are discarded. The remaining isotope patterns are used as input for the stable isotope fingerprinting. In this step, experimentally derived isotope peak distributions are fitted to theoretical isotope peak distributions computed for a range of $\delta^{13}C$ values with a fast Fourier transform method adapted from Rockwood et al. (13). Peptides with a poor goodness of fit are discarded. The remaining peptides are used to compute an average $\delta^{13}C$ value for each individual species and associated SEs in per mille. In a final step the species $\delta^{13}C$ values are corrected for instrument isotope fractionation by applying the offset determined using the reference material (see details in *Results*). Just like IRMS, our approach provides a single, robust, averaged $\delta^{13}C$ value, one for each species. It is well known that ^{13}C contents can vary between biomolecules and even between positions within a single biomolecule (5, 14–16). We show that given sufficient data the ^{13}C content of individual amino acids might be resolved from average per-peptide values by multivariate regression.

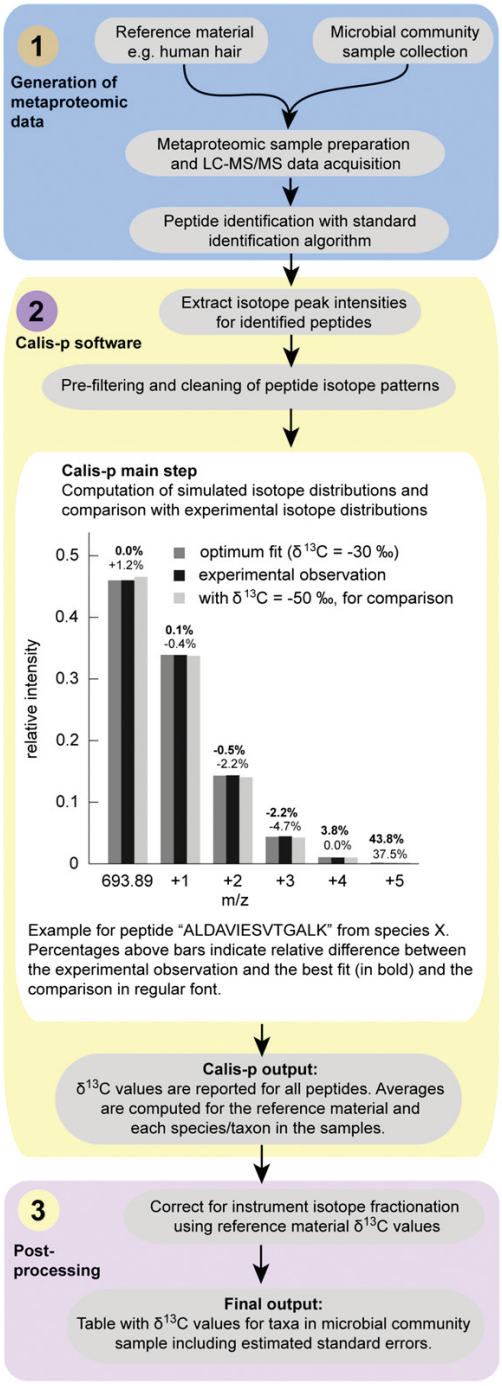

Fig. 1. Direct protein-SIF workflow. In the main step of the data analysis with the Calis-p software, the experimentally derived isotope distributions for peptides are compared with theoretical isotope peak distributions computed based on peptide molecular formulae with a fast Fourier transform method. The comparison with theoretical distributions is done for a specified range of $\delta^{13}C$ values in increments. Goodness of fit is calculated for all comparisons and the $\delta^{13}C$ value for the best fit is reported if a predetermined goodness of fit threshold is passed. A more detailed workflow can be found in *SI Appendix*, Fig. S1.

Results

Benchmarking with Pure Culture Data. For benchmarking, we measured the stable carbon isotope ratios of 20 pure culture species using both direct protein-SIF and continuous flow elemental analysis IRMS (CF-EA-IRMS). CF-EA-IRMS is the most commonly used method to determine highly accurate carbon isotopic compositions of organic bulk samples with measurement uncertainties of ±0.15‰ or less and can be considered the gold standard. The 20 pure cultures represented 18 bacterial, 1 archaeal, and 1 eukaryotic species (Datasets S1 and S2). For seven of the species we obtained technical replicate measurements to determine the precision of the direct protein-SIF method.

The protein-SIF $\delta^{13}C$ values for the technical replicate measurements of individual species were highly consistent with each other and generally deviated by less than ±1‰ from the mean (Dataset S1). The direct protein-SIF $\delta^{13}C$ values were linearly correlated to the CF-EA-IRMS values ($R^2 = 0.94$). The direct protein-SIF $\delta^{13}C$ values showed a systematic offset from the CF-EA-IRMS values of −15.4‰ (SD = 2.55) (Fig. 2*A*). This systematic offset is likely caused by isotope fractionation in the Orbitrap mass spectrometer, which has been recently described (17). The large range of $\delta^{13}C$ values measured with both direct protein-SIF and CF-EA-IRMS corresponds to the isotope ratios of the different carbon sources used to grow the pure cultures (Dataset S3) and the heterotrophic versus autotrophic lifestyle of the different microorganisms.

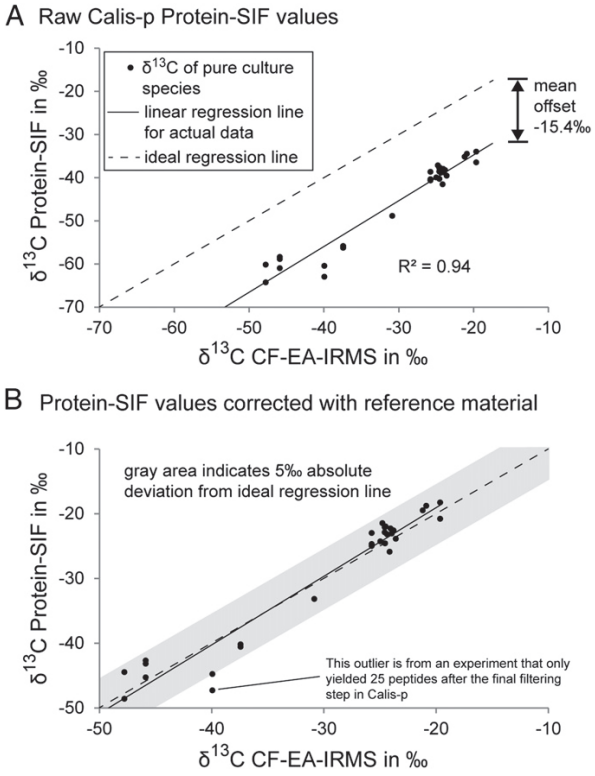

Fig. 2. Comparison of $\delta^{13}C$ measurements of pure cultures with protein-SIF and CF-EA-IRMS. Twenty pure cultures representing 18 bacterial, 1 archaeal, and 1 eukaryotic species were measured with both methods (detailed data in Datasets S1 and S2). For seven of the species technical replicate measurements were obtained. (*A*) The raw $\delta^{13}C$ values from the Calis-p software plotted against the IRMS-derived values. The average offset of protein-SIF values from the IRMS values is indicated. (*B*) Protein-SIF values after offset correction using the offset determined with reference material (human hair).

Correction of Instrument Isotope Fractionation with Reference Material. To correct for instrument isotope fractionation when using the direct protein-SIF method, we implemented the use of a reference material which is prepared, measured, and analyzed alongside the samples. With CF-EA-IRMS, reference materials with known isotopic compositions are also needed to correct for instrument isotope fractionation (18). We chose human hair as the reference material for the following reasons. (*i*) The bulk of its dry mass consists of protein, which allows directly correlating protein-SIF and IRMS $\delta^{13}C$ values. Interference from other biomolecules with potentially different isotope composition, such as lipids, is limited. (*ii*) It contains a large diversity of proteins providing hundreds to thousands of different peptides that can be measured for protein-SIF. (*iii*) The protein sequences required for peptide identification are known. (*iv*) It is easy to obtain in large batches (a few grams) to serve as a reference material for many years. Human hair with known $\delta^{13}C$ values measured by IRMS is also available from major chemical suppliers and the US Geological Survey (https://isotopes.usgs.gov/lab/referencematerials.html).

The offset between $\delta^{13}C$ values generated with direct protein-SIF and CF-EA-IRMS for the human hair reference material was −15.7‰ and thus almost identical to the average offset for the benchmarking pure culture $\delta^{13}C$ values (Fig. 2*A*). We corrected the pure culture protein-SIF $\delta^{13}C$ values with the reference material offset. After this, the absolute deviation of the protein-SIF $\delta^{13}C$ values from the CF-EA-IRMS $\delta^{13}C$ values was on average ±2.1‰ (SD = 1.5) (Fig. 2*B* and Dataset S2).

To explore whether the approach could also be used to estimate per-amino acid ^{13}C contents, we used multivariate regression to estimate the per-amino acid ^{13}C content for each of the 18 amino acids included in our analysis. *SI Appendix*, Fig. S2 shows consistent trends in ^{13}C content, with some amino acids strongly depleted in ^{13}C ($\delta^{13}C$ −78‰ for arginine and −64‰ for lysine) and others strongly enriched in ^{13}C ($\delta^{13}C$ +58‰ for asparagine and +50‰ for lysine) relative to an organism's average $\delta^{13}C$ value. *SI Appendix*, Fig. S2 also shows that these estimates are not very precise. Differences of 50‰ between replicates were not exceptional.

Detection Limit and Accuracy of Protein-SIF in Mock Communities. To determine the detection limit of our approach and to test how accurately we can measure SIFs in complex community samples, we used mock communities. These mock communities were created by mixing the 20 pure culture species used for benchmarking with 12 additional microbial strains and species so that the final community contained a total of 32 strains and species. We generated a total of 12 biological replicates of the community with differing species abundances and analyzed each replicate with two different LC-MS/MS methods (19). To evaluate the mock community direct protein-SIF data we only considered the 20 benchmarking species in the community for which IRMS and direct protein-SIF $\delta^{13}C$ values were known based on the results presented above.

The accuracy and detection limit of direct protein-SIF depends heavily on the available amount of data. We found that the number of peptides for each species that pass the final Calis-p filtering step has a large influence on the accuracy of the determined SIF values (Fig. 3). If fewer than 60 peptides were available for SIF value calculation, the protein-SIF $\delta^{13}C$ values differed from the IRMS-derived $\delta^{13}C$ values by more than 10‰ for more than half of the 20 species. If between 60 and 100 peptides were available, protein-SIF values differed less than 10‰ from IRMS values for six out of nine species, with only one large outlier. With more than 100 peptides available, over 95% of protein-SIF values differed less than 10‰ from $\delta^{13}C$ values determined by IRMS (mean deviation 4.7‰). As expected, the detection limit for direct protein-SIF depended on the amount of mass spectrometric data available, because to reach the necessary

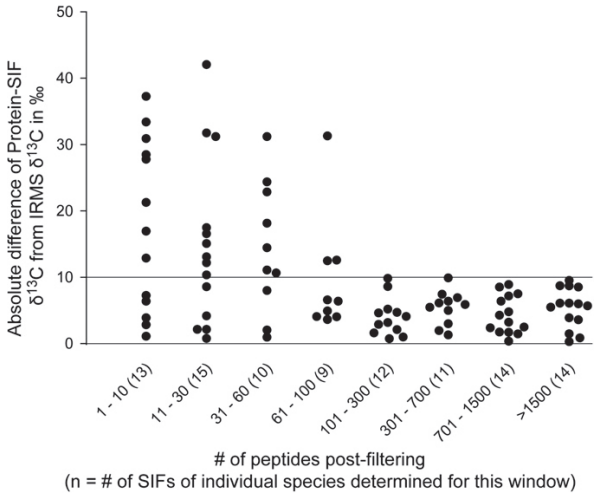

Fig. 3. Absolute difference between $\delta^{13}C$ values determined with direct protein-SIF of individual species in mock communities and IRMS of the corresponding pure cultures. Five mock community datasets with a total of 32 species and strains were analyzed. For 20 species, the $\delta^{13}C$ values were known from IRMS performed on pure cultures. For these species, the $\delta^{13}C$ values were determined with direct protein-SIF. Each dataset contained different amounts of data (Table 1). Different numbers of peptides were identified and passed the final Calis-p peptide filter for each species in each dataset. The absolute difference between $\delta^{13}C$ values obtained via protein-SIF and IRMS was calculated and sorted according to how many peptides were available for SIF calculation by Calis-p after filtering the peptides. The plot gives the absolute differences for different ranges of peptide numbers used for SIF calculation.

number of high-quality peptide isotopic patterns for low-abundant species more mass spectra are required (Table 1 and Dataset S4). For example, one 4-h LC-MS/MS run enabled determination of SIF values for 25% of the species in the mock community. These species were relatively abundant (proteinaceous biomass >5.7% of the total community). In contrast, combined data of 12 4-h runs enabled determination of SIF values for 75% of species, including species with lower (<1% of the total proteinaceous biomass) abundance (Table 1). We also found that the accuracy of SIF estimates of individual peptides strongly depended on the mass spectrometric peak intensities of the peptide (*SI Appendix*, Fig. S3). This could be explained by the presence of background noise, most likely originating from small interfering peaks that were not completely resolved from some of the analyzed peaks for a peptide, thus slightly changing individual isotopic peak intensities for peptides.

Case Study. To demonstrate the power and application of direct protein-SIF we applied it to a well-studied bacteria–animal symbiosis, the gutless marine oligochaete *Olavius algarvensis* (Fig. 4*A*). *O. algarvensis* lacks a digestive system. Instead, the worm relies on at least five bacterial symbionts under its cuticle for nutrition (Fig. 4 *B* and *C*). The metabolism and physiology of these symbionts and their interactions with the host have been extensively studied using metagenomics, metaproteomics, and physiological incubation experiments (20–22). Based on these previous studies, the carbon sources of the symbionts and host were thought to be as follows (Fig. 4*C*). Two sulfur oxidizers (γ1 and γ3) fix seawater-derived inorganic carbon using the Calvin–Benson–Bassham cycle with a Form IA RubisCO enzyme (20). Two sulfate reducers (δ1 and δ4) consume the host's organic waste products. For a spirochete, no data on metabolism and physiology were available. The host itself derives its carbon directly from the symbionts by digestion through endocytosis (23). Thus, the $\delta^{13}C$ values of the γ- and δ-symbionts as well as the host were expected to be similar, since all carbon is derived from the initial carbon fixation by the γ-symbionts. In a previous study the $\delta^{13}C$ value of complete worms was determined by IRMS to be −30.6‰ (21). Our expectation was therefore that all direct protein-SIF–derived $\delta^{13}C$ values would be somewhere in the range of −25 to −35‰. We assumed this range of possible $\delta^{13}C$ values because of the measurement uncertainty of direct protein-SIF, as well as the fact that different trophic levels in a heterotrophic food chain can vary in their $\delta^{13}C$ value by 0.5–2‰ from their food source (24, 25).

We put this model of carbon flow in *O. algarvensis* to the test by applying direct protein-SIF using a metaproteomic dataset obtained from multiple individual worms. To obtain sufficient peptides for measurement of $\delta^{13}C$ values of all of the symbionts, we combined 18 LC-MS/MS runs from a total of 14 individual worms. The $\delta^{13}C$ values obtained by direct protein-SIF of the host (−25.6‰), the two γ-symbionts (−32.9 and −31.8‰) and the δ1-symbiont (−34.8‰) were in the expected range (Fig. 4*D* and Dataset S5). However, the δ4-symbiont had a much higher $\delta^{13}C$ value (−17.9‰, SE ±1.8‰) than the δ1-symbiont. This was unexpected because the two δ-symbionts appear to be almost identical in terms of expressed carbon uptake and catabolic pathways. Both are characterized by many abundantly expressed high-affinity uptake transporters for sugars, amino acids, peptides, and organic acids, as well as pathways for the use of host-derived

Table 1. Detection limit of direct protein-SIF for species in mock communities depending on the amount of LC-MS/MS data available

	Total LC-MS/MS run time, h				
Experimental parameters and outcomes	92	52	31	8	4
Biological replicates	12	12	4	1	1
Gradient length, min	460	260	460	460	260
MS/MS spectra (in millions)	~2.04	~1.2	~0.68	~0.17	~0.1
Species (out of 20) with protein-SIF*	15	15	10	6	5
Lower species abundance limit for SIF determination†, %	0.82	0.82	0.92	5.65	5.79
Mean deviation of protein-SIF $\delta^{13}C$ from IRMS $\delta^{13}C$, ‰	3.6	5.4	4.7	4.4	5.8
Minimum deviation, ‰	0.2	0.6	0.9	1.5	2.8
Maximum deviation, ‰	9.4	9.7	8.8	8.4	9.8

Mock community samples with 32 species and strains were used (14). For 20 of these species the IRMS $\delta^{13}C$ values were known. The detailed data can be found in *SI Appendix*, Table S4.
*Number of species with a sufficient number of peptides for protein-SIF (>100) after final Calis-p filtering.
†Abundance (percentage of total community protein) of lowest abundance species for which determining protein-SIF value was possible (i.e., >100 peptides after final Calis-p filtering).

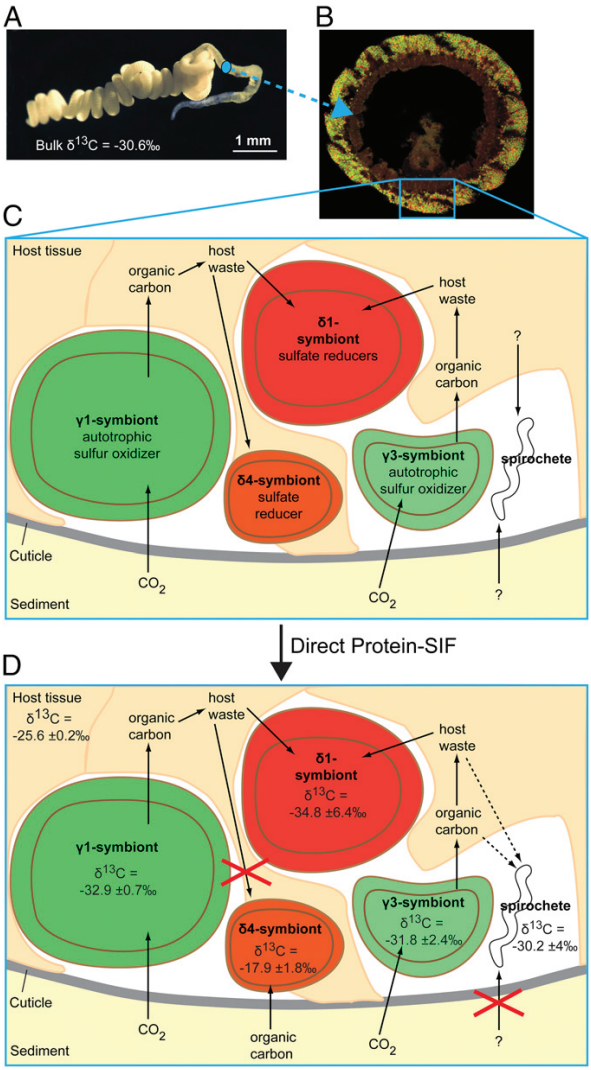

Fig. 4. Testing the model of physiological interactions in the *O. algarvensis* symbiosis using direct protein-SIF. (*A*) Live *O. algarvensis* specimen. The $\delta^{13}C$ value of bulk worms was determined by IRMS on six biological replicates in Kleiner et al. (21). Image courtesy of Christian Lott (photographer). (*B*) Cross-section through the worm. The bacterial symbionts right below the worm's cuticle are stained with specific fluorescence in situ hybridization probes (γ-symbionts in green, δ-symbionts in red). (*C*) Simplified model of carbon flow in the symbiosis based on previous metagenomic (22) and metaproteomic (26) studies. For the δ1- and δ4-symbionts the metaproteomic data suggested that these two symbionts are highly similar in terms of metabolism and physiology. (*D*) Adjusted model of carbon flow based on carbon sources predicted using direct protein-SIF–derived $\delta^{13}C$ values. Detailed protein-SIF data in Dataset S5. The ± value for each $\delta^{13}C$ value indicates the SE.

acetate and propionate (26). The direct protein-SIF data now point toward a functional difference between the two δ-symbionts. Apparently, the δ4-symbiont derives part or all of its carbon from external sources that have a different carbon isotopic composition. Currently we do not have an explanation of how the δ4-symbiont obtains external carbon and what the main physiological difference between the two δ-symbionts is. One potential mechanism for how the δ4-symbiont could obtain symbiosis-external carbon is that it lives in the sediment pore water and from there continuously infects *O. algarvensis*; however, continuous infection of the worm is unlikely as the worm does not have any orifices that would allow for such infection to occur and the δ4-symbiont genome does not encode for any known pathways that would allow it to penetrate the worm's cuticle. It is also possible that the two symbionts have differing affinities for specific substrates, which cannot be deduced from the transporter annotations. Future work using isotopically labeled substrates could help to address this hypothesis.

We were also able to measure a sufficient number of peptides to estimate a $\delta^{13}C$ value for the spirochetal symbiont, despite its low abundance. The spirochete $\delta^{13}C$ value of −30‰ was in the range of symbiosis-internal carbon, suggesting that this symbiont uses a symbiosis-internal carbon source. Currently, we are not aware of any symbiosis-external carbon sources that have a $\delta^{13}C$ value in the range of −30‰; however, we cannot exclude that the spirochete has access to a symbiosis-external source with such a negative $\delta^{13}C$ value. In summary, by applying direct protein-SIF to the *O. algarvensis* symbiosis we derived insights into carbon flow that were not suggested by any of the previous metaomics-based studies.

Discussion

The developed direct protein-SIF approach provides a means to directly and simultaneously access the SIFs (i.e., $\delta^{13}C$ values) of many individual species in microbial communities. As little as 4 h of LC-MS/MS time can be sufficient to estimate SIFs for the most abundant species in a sample. For this type of analysis only very small sample amounts are needed; recent advances in sample preparation allow for the production of metaproteomic data of sufficient quality from as little as 1 mg of wet weight cell mass. If the determination of SIFs for lower-abundant species is desired, more LC-MS/MS run time can provide the required metaproteomic depth. For longer or additional LC-MS/MS runs proportionally larger amounts of sample are needed (e.g., 2 mg for two 4-h runs). Additionally, enrichment of specific cell populations by filtration or centrifugation methods can be used to obtain better metaproteomic coverage (26, 27).

Differences Between IRMS and Direct Protein-SIF Measurements. We observed a range of differences between the $\delta^{13}C$ values measured by IRMS and direct protein-SIF when using both on single-species biomass (absolute deviation of the protein-SIF $\delta^{13}C$ values from the CF-EA-IRMS $\delta^{13}C$ values was on average ±2.1‰). Part of the observed variation between the two methods is likely due to a lower accuracy of direct protein-SIF using the highly complex metaproteomic mass spectrometric data. However, there are at least two other factors that might contribute to this variation in the observed $\delta^{13}C$ values.

First, direct protein-SIF measures the $\delta^{13}C$ value of proteins, while bulk IRMS measures all cell components such as protein, lipids, DNA, and metabolites, providing a weighted average. It has been shown that the $\delta^{13}C$ values of different cell components can differ. For example, lipids have a 1.6‰ lower $\delta^{13}C$ value than protein in *Escherichia coli* (28) and the lipids of Calvin–Benson–Basham cycle autotrophs have around 6‰ lower $\delta^{13}C$ values compared with the $\delta^{13}C$ values of the total biomass (5, 7). The possible difference between protein $\delta^{13}C$ values and bulk organic matter $\delta^{13}C$ values should thus be considered when interpreting direct protein-SIF results. In addition, different amino acids have different ^{13}C contents, and even carbon positions within amino acids differ in their $^{13}C/^{12}C$ ratios. These differences might also lead to a detectable bias if the amino acid composition is highly skewed or if only few peptides are available. Since protein usually makes up the majority of a cell in terms of mass, for example, 55% of *E. coli* dry weight (BioNumbers ID 104954) (29), protein $\delta^{13}C$ values will be a good approximation of bulk $\delta^{13}C$ values in most cases.

Second, the variation in the intensities of the peptide isotopic peaks used for direct protein-SIF is mostly due to variation in the ratio of ^{13}C to ^{12}C, because ^{13}C is, with a natural abundance of ~1.1%, the most abundant heavy stable isotope in the considered peptides. However, very large isotope fractionation of the three other elements (hydrogen, oxygen, and nitrogen) in the peptides that we consider (sulfur-containing peptides are excluded) could change the measured $\delta^{13}C$ values by several per mille. For example, the hydrogen isotope fractionation in photosynthate compared with the water used by the photosynthetic organism is $\varepsilon = -171‰$ (5). The $^{2}H/^{1}H$ ratio in the Vienna Standard Mean Ocean Water (VSMOW) reference is 0.00015576, which means that the $^{2}H/^{1}H$ ratio of photosynthate is 0.000129 if ocean water is the hydrogen source (i.e., a change in the fifth decimal of the fraction of heavy atoms). For comparison, carbon isotope fractionation influences the third and fourth decimal. As the $^{13}C/^{12}C$ ratio of V-PDB is 0.0111802, a $\delta^{13}C$ value of −47.2‰ corresponds to a $^{13}C/^{12}C$ ratio of 0.01065. Carbon isotope fractionation by only −3‰ would already yield a $^{13}C/^{12}C$ ratio of 0.01115, a shift in the fraction of heavy atoms similar to that produced by the fractionating hydrogen isotopes by −171‰. Or, to put it differently, a $\delta^{2}H$ value of −171‰ would change the estimated $\delta^{13}C$ value of −47.2‰ to −50.1‰ in direct protein-SIF if the reference material used had a $\delta^{2}H$ value similar to that of VSMOW (0‰). However, protein reference materials used for direct protein-SIF will have a much more negative $\delta^{2}H$ value compared with VSMOW [e.g., the $\delta^{2}H$ values of human hair range typically between −130 and −80‰ (30)], thus removing the most common hydrogen isotope fractionation effects when applying the reference material-based offset correction to the direct protein-SIF $\delta^{13}C$ values. Therefore, while nitrogen, hydrogen and oxygen isotope fractionation will usually not have a major effect on direct protein-SIF $\delta^{13}C$ values, it should always be considered during interpretation of direct protein-SIF results. Measurement of δ values for these three elements by IRMS bulk analyses of samples and reference material would provide insights into any major deviations that should be considered to achieve even better accuracy of the results.

Can This Approach Be Used for Protein-SIP? In recent years, metaproteomics has been successfully combined with stable isotope probing (protein-SIP) to follow the incorporation of isotopically labeled substrates by individual members in microbial communities (31–34). Several different algorithms have been used to compute isotope incorporation levels in protein-SIP and some of these algorithms employ approaches similar to our direct protein-SIF approach. All current protein-SIP approaches, however, can compute isotope incorporation levels only with low precision (i.e., several atom percent changes in isotope ratios are needed to obtain a clear readout). Naturally, the question arises if the direct protein-SIF approach could be also employed for protein-SIP. Answering this question will require further method development, but we predict that the direct protein-SIF approach should be applicable for protein-SIP approaches that use small percentages of labeled substrate or short labeling pulses. If we assume that the protein identification algorithm used correctly identifies peptides if the intensity of the monoisotopic peak is at least 33% of its original (unlabeled) intensity, the maximum amount of label assimilated that would still enable correct identification of >90% of all peptides would be ~1%. Since the Calis-p software resolves differences between $\delta^{13}C$ values of <10‰ units, the resulting resolution would be 0.01%. The Calis-p approach would thus not be a replacement for other protein-SIP approaches but would complement them by adding the capability to detect incorporation of very low amounts of heavy atoms into proteins.

Potential Limitations. There are two potential limitations of the direct protein-SIF approach when it comes to resolving $\delta^{13}C$ values for species-level microbial populations. First, strains within one species often share a high protein sequence identity, which makes it difficult to uniquely assign most peptides to individual strains within one species. This means that if we sample a microbial community that has multiple strains of the same species and these strains use different carbon substrates, the direct protein-SIF method will only report an "averaged" SIF for the species, which would hamper deductions about which carbon substrates were used. This strain resolution challenge could be addressed if a strain-resolved metagenomic database were available for the sample in question. With such a database, SIFs for individual strains could be resolved by filtering for strain-unique peptides before calculating the SIFs. Additionally, a careful evaluation of "expressed" metabolism and physiology of the species-level population using the metaproteomic data would provide insights into which and how many substrates might be used by a population and thus aid in the interpretation of the SIF data. Second, it is conceivable that individual cells of a population consisting of a single strain use different carbon substrates depending on their position in an environmental matrix or the same cell uses multiple substrates at once. This again would lead to an "averaged" SIF for the strain, confounding the signals from the carbon substrates used. In this second case, the use of the metaproteomic data to analyze the "expressed" metabolism would be essential for the SIF data interpretation.

Conclusions

The direct protein-SIF approach provides us with a key ability that no other method can provide at the moment. By measuring $\delta^{13}C$ values of individual species in microbial communities we can now make inferences about the food sources for these species as well as the metabolic pathways used for carbon assimilation. Ongoing technological development will further improve the accuracy, detection limits, and capabilities of the direct protein-SIF approach in the future in at least three ways. First, the fractionation of isotopic species in the mass spectrometer could be reduced with specialized methods and potentially with improvements of the instrumentation. Current approaches for reducing isotope fractionation in Orbitrap mass spectrometers are very promising even though they are not yet usable for metaproteomics (17), as these new approaches significantly reduce the number of MS/MS spectra acquired and thus reduce the number of identified peptides. Second, mass spectrometers with increased resolving power at high scan rates will make it possible to separate the isotopologues of peptides based on which element provides the heavy isotope (hydrogen, nitrogen, carbon, or oxygen). For example, the exchange of one ^{14}N atom with one ^{15}N atom in a peptide changes its mass by 0.99703489 Da, while the exchange of one ^{12}C with one ^{13}C changes the mass by 1.0033548378 Da. Once isotopic species can be resolved on a per-element basis, it should become feasible to not only determine carbon isotope ratios but also isotope ratios for other elements in peptides such as nitrogen, oxygen, and hydrogen. The necessary mathematical approach to calculate the required fine-structure peptide isotope patterns via a multidimensional Fourier transform has been recently demonstrated by Ipsen (35). Third, if sufficient numbers of peptides are measured for a species it is in theory possible to estimate amino acid-specific $\delta^{13}C$ values. $\delta^{13}C$ values of individual amino acids can provide additional information about carbon sources or biosynthetic pathways for a species (5, 14, 15).

Methods

Sample Preparation. The cultivation of the pure cultures and the creation of the mock communities are described in Kleiner et al. (19). For the case study, 14 *O. algarvensis* specimens were collected off the coast of Sant' Andrea Bay,

Elba, Italy (42°48′26″N, 010°08′28″E) in August 2015 from shallow-water (6- to 8-m water depth) sediments next to seagrass beds. Live worms were transported in native Elba sediment and seawater to the Max Planck Institute for Marine Microbiology in Bremen, Germany. The sand which was used for worm transport and storage was washed three times with clean freshwater followed by three washes with clean seawater to remove life and dead meiofauna and other potential sources of organic substrates. The worms were kept for 1 mo in the dark and at Elba marine sediment temperature. Worms were then carefully removed from the sediment and frozen at −80 °C until further processing.

The human hair used as a standard for correction of instrument isotope fractionation was obtained from M.K.

Peptide samples for proteomics were prepared and quantified as described by Kleiner et al. (19) following the filter-aided sample preparation protocol described by Wiśniewski et al. (36). The only modification for the *Olavius* samples compared with the pure cultures and the mock communities was that no bead beating step was used. The bead beating step was used for the human hair reference material.

One-Dimensional LC-MS/MS. Samples were analyzed by 1D LC-MS/MS as described in Kleiner et al. (19). One or two wash runs and one blank run were performed between samples to reduce carryover. For the 1D LC-MS/MS runs of pure culture samples, 2 μg of peptide were loaded onto a 5-mm, 300-μm i.d. C18 Acclaim PepMap 100 precolumn (Thermo Fisher Scientific) using an UltiMate 3000 RSLCnano Liquid Chromatograph (Thermo Fisher Scientific). After loading, the precolumn was switched in line with a 50-cm × 75-μm analytical EASY-Spray column packed with PepMap RSLC C18, 2 μm material (Thermo Fisher Scientific). For the 1D LC-MS/MS runs of the *Olavius* samples, 0.8–4 μg of peptide were loaded onto a 2-cm, 75-μm i.d. C18 Acclaim PepMap 100 precolumn (Thermo Fisher Scientific) using an EASY-nLC 1000 Liquid Chromatograph (Thermo Fisher Scientific) set up in two-column mode. The precolumn was also connected to a 50-cm × 75-μm analytical EASY-Spray column packed with PepMap RSLC C18, 2 μm material. In both cases the analytical column was connected via an Easy-Spray source to a Q Exactive Plus hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific). Peptides were separated on the analytical column using 260- or 460-min gradients and mass spectra were acquired in the Orbitrap as described by Petersen et al. (37). For the mock communities, the existing 1D LC-MS/MS data from Kleiner et al. (19) were used.

Input Data Generation, Algorithm, and Software Development for Protein-SIF.
Peptide identification. For peptide and protein identification of pure culture samples, protein sequence databases were created using the reference protein sequences for each species separately. The databases are available from the PRIDE repository (PXD006762). For the mock community samples, the database from the PRIDE repository (PXD006118) was used (19). For the Olavius samples, an existing *O. algarvensis* host and symbiont protein database from project PXD003626 (ftp://massive.ucsd.edu/MSV000079512/sequence/) was used. To this *Olavius* database we added additional symbiont protein sequences from recently sequenced metagenomes. The new *Olavius* database is available from the PRIDE repository (PXD007510). For the human hair standard, the human reference protein sequences from UniProt (UP000005640) were used. CD-HIT was used to remove redundant sequences from the multimember databases using an identity threshold of 95% (38). The cRAP protein sequence database (https://www.thegpm.org/crap/) containing protein sequences of common laboratory contaminants was appended to each database. One important consideration for creating protein sequence databases for use with the Calis-p software is that Calis-p will calculate taxon/population-specific $\delta^{13}C$ values based on accession number prefixes that indicate to which taxon a sequence belongs. The prefix should be separated from the accession number by an underscore (e.g., >TAX_00000). MS/MS spectra were searched against the databases using the Sequest HT node in Proteome Discoverer version 2.0.0.802 (Thermo Fisher Scientific) and peptide spectral matches were filtered using the Percolator node as described by Petersen et al. (37). The FidoCT node in Proteome Discoverer was used for protein inference and to restrict the protein-level false discovery rate to below 5% (FidoCT q-value < 0.05).
Input files. Examples for all input files are provided in PRIDE project PXD006762 alongside the raw data. The LC-MS/MS-produced raw files were converted into mzML format using MSConvertGUI via ProteoWizard (39) with the following options set: Output format: mzML, Binary encoding precision: 64-bit, Write index: checked, TPP compatibility: checked, Filter: Peak Picking, Algorithm: Vendor, MS Levels: 1 (The MS/MS scans are not needed for isotope pattern extraction). The peptide-spectrum match (PSM) files generated by Proteome Discoverer were exported in tab-delimited text format. The mzML files and the PSM files were used to extract isotopic patterns for all identified peptides.
Isotopic pattern extraction. The steps described here are carried out by the Calis-p software in a fully automated fashion upon provision of correctly formatted input files. The entries in the PSM files were excluded if they (*i*) had any identified posttranslational modifications, (*ii*) contained "M" or "C" in the peptide sequence, (*iii*) had a peptide confidence score not equal to "High," or (*iv*) had a PSM Ambiguity value equal to "Ambiguous." The remaining entries were subjected to the isotopic pattern extraction from the mzML mass spectrum input files. In a first step, the *m/z* value of the monoisotopic peak (A) of each peptide was searched for in the full scans in the mass spectrum file in a defined retention time window (peptide scan start time ±0.5). All peaks with the same *m/z* value in the retention time window were considered to be from the same peptide, because most peptides are analyzed multiple times in subsequent MS^1 scans in the defined RT window size. In a second step, each MS^1 scan of the same peptide was searched for the isotopic peaks of the peptide produced by replacement of single or multiple atoms in the peptide with heavier isotopes (i.e., A+1, A+2. . .). The isotopic peaks were defined as all of the peaks following the monoisotopic peak in the full scan mass spectrum with *m/z* values being n*(1/charge) distance removed from the monoisotopic *m/z* value (n is the peak count i.e., A+1, A+2. . .). Since the number of isotopic peaks found for a peptide in different MS^1 scans is variable due to decreasing signal intensity for the higher-number isotopic peaks, we chose to report the peak intensities only for the isotopic peaks detected in the majority of scans. To clarify, if we found the A, A+1, A+2 peaks for five MS^1 scans, A, A+1, A+2, A+3 peaks for three MS^1 scans, A, A+1, A+2, A+3, A+4 for seven MS^1 scans, and A, A+1, A+2, A+3, A+4, A+5 for three MS^1 scans, we only reported peak intensities for A to A+4. Scans that did not have all isotopic peaks to be reported were excluded (i.e., for the preceding example this would mean a scan that only had peaks A to A+3 would be excluded). To maximize carbon ratio calculation accuracy the intensities of each isotopic peak from all of the included scans of the same peptide were summed up. The filtered PSM entries with the identified isotopic patterns and aggregated intensities were written to an isotopicPattern text file in a tab delimited format. Additional information needed for SIF computation such as peptide sequence, peptide sum formula, charge, and peptide identification quality score (Xcorr) was written into the isotopicPattern file. For examples of isotopic pattern files see PRIDE project PXD006762.
SIF computation. Peptide isotope patterns were filtered to remove exact duplicates (peptides associated with identical spectra). Also, the following default parameters were used: peptides associated with more than three proteins or that were identified with low confidence (Xcorr < 2.0), or had fewer than three isotopic peaks in their spectra, or had a total intensity (sum of all peaks) of less than 0.5×10^8, were not considered further.

To calculate the $\delta^{13}C$ value for peptides, we compared the experimentally derived isotope peak intensities with simulated isotope peak intensities. For the peptides remaining after filtering, the absolute intensity of each peak (monoisotopic mass, A+1, A+2, etc.) was divided by the summed intensity of all peaks to obtain relative peak intensities. The $^{13}C/^{12}C$ ratio for each peptide was computed as follows. For $^{13}C/^{12}C$ ratios between 0 and 0.1 and fixed (natural) relative abundances of ^{15}N, ^{18}O, ^{17}O, and ^{2}H, the theoretical spectrum (relative intensity of each peak) was computed using a fast Fourier transform modified after Rockwood et al. (13) based on the peptide's molecular formula. In contrast to the approach used by Rockwood et al., we did not model the peaks as Gaussians because we used centroided data. Instead, the theoretical relative peak intensities were computed as described by Rockwood et al. (13), with the Gaussian peak shape function $s(\mu) = 1$:

$$\text{Peptide spectrum} = \bar{F}(m) = \mathbf{IFT}\left(\prod_E \left[\mathbf{FT}\left(\overline{RA}_E\right)\right]^{N_E} \right),$$

with **IFT** inverse Fourier transform, $\mathbf{\Pi}$ product (for each element C, N, O, and H), **FT** Fourier transform, RA_E relative abundances of the isotopes (e.g., ^{12}C, ^{13}C, and ^{14}C) for each element, and N_E number of atoms in the peptide for each element.

Because not all predicted peaks were detected experimentally, the predicted peak spectrum was truncated by discarding all peaks that were not detected experimentally and the relative peak intensity of the remaining peaks was renormalized, so that the predicted intensities of the remaining peaks was equal to one, as it was for the experimentally detected spectrum. Finally, the goodness of fit between the experimental and theoretical spectrum was calculated as the sum of squares of the difference between observed intensity and predicted intensity for each peak. The goodness of fit was optimized (lowest sum of squares) by decreasing the step size of the $^{13}C/^{12}C$ value from 0.01 to 0.0001 in four iterations. For each species

(population) in the sampled community (identified by an accession number prefix for the taxon), the average $^{13}C/^{12}C$ ratio was calculated as the intensity-weighted average of the $^{13}C/^{12}C$ ratio of each peptide attributed to that population. The intensity-weighted average was used because there was a clear correlation between spectral intensity and accuracy of the SIF estimate for individual peptides (*SI Appendix*, Fig. S3). These data were also used to estimate the SE and the $\delta^{13}C$ for each population. To eliminate results from peptides suffering from interference from other peptides (due to overlapping spectra), only results with a sum of squares lower than 0.00005 or those results with a goodness of fit in the upper 33% of the data for that population were used in the calculation of the average $\delta^{13}C$ values.

Correction of direct protein-SIF values with reference standard. To correct for isotope fractionation introduced during sample processing and analysis, mostly by isotope fractionation in the mass spectrometer (17) (see also *Results*), we used human hair as a standard. $\delta^{13}C$ values for the human hair standard were obtained both with direct protein-SIF and CF-EA-IRMS (discussed below) and the offset between the two measurement methods was calculated. The offset value was then used to correct the direct protein-SIF values obtained for the pure culture and for the community samples.

CF-EA-IRMS. Stable isotope ratios of carbon and nitrogen of bacterial strains, human hair, and culture media were determined (Datasets S2 and S3) by CF-EA-IRMS. Frozen cell pellets of bacterial strains were dried at 105 °C and stored at 4 °C. Dried cell pellets as well as carbon and nitrogen sources used for cultivation were weighed (~1 mg per sample) into tin capsules and stored in a desiccator. $\delta^{13}C$ and $\delta^{15}N$ values were measured using CF-EA-IRMS on a Delta V Plus (Thermo) mass spectrometer coupled to an ECS 4010 elemental analyzer (Costech). A Zero Blank autosampler (Costech) was used for dropping the samples onto a quartz tube combustion column heated to 1,000 °C. Simultaneously with sample dropping, an O_2 pulse was injected to flash-combust the samples. Ultra-high-purity helium was used as a carrier gas to transport the resulting gases through a water trap and a reduction furnace (600 °C) in which NOx species were reduced to N_2. The generated N_2 and CO_2 were separated on a GC column. A Finnigan ConFlo III (Thermo) interface was used to deliver the gases to the ion source of the mass spectrometer. Peak areas for the isotopic species of sample CO_2 and N_2 were compared with those of reference gas peaks to determine the $\delta^{13}C$ and $\delta^{15}N$ values of the samples. International standards (Dataset S3) were used at the beginning and end of the measurement sequence to normalize the measurements to the internationally accepted delta (δ) scale with respect to V-PDB for $\delta^{13}C$ and atmospheric N_2 for $\delta^{15}N$. Additionally, caffeine (C-0750; Sigma-Aldrich) was included at the beginning and gelatin (G-9382; Sigma-Aldrich) after each fifth sample as internal laboratory standards. The measurement uncertainty was equal or better than ±0.15‰ for $\delta^{13}C$ and $\delta^{15}N$ values determined by IRMS.

Data and Software Availability. The Calis-p software was implemented in Java and is freely available for download, use, and modification at https://sourceforge.net/projects/calis-p/. The MS proteomics data and the protein sequence databases have been deposited to the ProteomeXchange Consortium (40) via the PRIDE partner repository for the pure culture data with the dataset identifier PXD006762, dataset identifier PXD006118 for the mock community data published previously in Kleiner et al. (19), and the *O. algarvensis* case study data with dataset identifier PXD007510.

ACKNOWLEDGMENTS. We thank Jianwei Chen, Emmo Hamann, Marc Mussmann, Jessica Kozlowski, Sean Booth, Jessica Duong, Johanna Voordouw, Kenneth Sanderson, JoongJae Kim, Joenel Alcantara, Anupama P. Halmillawewa, Michael F. Hynes, and Heidi Gibson for donations of cultures for the mock communities; Christine E. Sharp for help with bacterial stock collection; Jackie Zorz, Maryam Ataeian, and Angela Kouris for help with proteomics sample preparation; Angela Kouris and Eric Miller for grammar and spell-checking of the manuscript; Christian Lott for providing the *O. algarvensis* image used in Fig. 4; Steve Taylor and his team at the University of Calgary Stable Isotope Laboratory for measurements using isotope ratio MS; and two anonymous reviewers for thoughtful comments and feedback. Sequencing of new *O. algarvensis* metagenomes was conducted by the US Department of Energy Joint Genome Institute and is supported by the Office of Science of the US Department of Energy under Contract DE-AC02-05CH11231. This study was supported by the Campus Alberta Innovation Chair Program (M.S. and X.D.), the Canadian Foundation for Innovation (M.S.), the North Carolina State Chancellor's Faculty Excellence Program Cluster on Microbiomes and Complex Microbial Communities (M.K.), a DAAD fellowship by the German Academic Exchange Service (to T.H.), and the Natural Sciences and Engineering Research Council (NSERC) of Canada through a Banting fellowship (to M.K.), an NSERC undergraduate student research award (to E.T.), and a NSERC Discovery grant (to M.S. and B.M.). We acknowledge the support of the International Microbiome Center as well as funding support from the Government of Canada through Genome Canada, the Government of Alberta through Genome Alberta, Genome Prairie, Research Manitoba, and Genome Quebec. This research was undertaken thanks in part to funding from the Canada First Research Excellence Fund.

1. Coplen TB, et al. (2006) New guidelines for $\delta^{13}C$ measurements. *Anal Chem* 78: 2439–2441.
2. Pearson A (2010) Pathways of carbon assimilation and their impact on organic matter values $\delta^{13}C$. *Handbook of Hydrocarbon and Lipid Microbiology*, ed Timmis KN (Springer, Berlin), pp 143–156.
3. DeNiro MJ, Epstein S (1978) Influence of diet on the distribution of carbon isotopes in animals. *Geochim Cosmochim Acta* 42:495–506.
4. Orphan VJ, House CH, Hinrichs K-U, McKeegan KD, DeLong EF (2001) Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. *Science* 293:484–487.
5. Hayes JM (2001) Fractionation of carbon and hydrogen isotopes in biosynthetic processes. *Rev Mineral Geochem* 43:225–277.
6. O'Leary MH, Madhavan S, Paneth P (1992) Physical and chemical basis of carbon isotope fractionation in plants. *Plant Cell Environ* 15:1099–1104.
7. Suzuki Y, et al. (2006) Host-symbiont relationships in hydrothermal vent gastropods of the genus *Alviniconcha* from the Southwest Pacific. *Appl Environ Microbiol* 72:1388–1393.
8. Beinart RA, et al. (2012) Evidence for the role of endosymbionts in regional-scale habitat partitioning by hydrothermal vent symbioses. *Proc Natl Acad Sci USA* 109:E3241–E3250.
9. Zhang CL (2002) Stable carbon isotopes of lipid biomarkers: Analysis of metabolites and metabolic fates of environmental microorganisms. *Curr Opin Biotechnol* 13:25–30.
10. Mohr W, Tang T, Sattin SR, Bovee RJ, Pearson A (2014) Protein stable isotope fingerprinting: Multidimensional protein chromatography coupled to stable isotope-ratio mass spectrometry. *Anal Chem* 86:8514–8520.
11. Woebken D, et al. (2015) Revisiting N_2 fixation in Guerrero Negro intertidal microbial mats with a functional single-cell approach. *ISME J* 9:485–496.
12. Musat N, et al. (2014) The effect of FISH and CARD-FISH on the isotopic composition of $^{(13)}C$- and $^{(15)}N$-labeled Pseudomonas putida cells measured by nanoSIMS. *Syst Appl Microbiol* 37:267–276.
13. Rockwood AL, Van Orden SL, Smith RD (1995) Rapid calculation of isotope distributions. *Anal Chem* 67:2699–2704.
14. Macko SA, Fogel ML, Hare PE, Hoering TC (1987) Isotopic fractionation of nitrogen and carbon in the synthesis of amino acids by microorganisms. *Chem Geol Isot Geosci Sect* 65:79–92.
15. McMahon KW, Fogel ML, Elsdon TS, Thorrold SR (2010) Carbon isotope fractionation of amino acids in fish muscle reflects biosynthesis and isotopic routing from dietary protein. *J Anim Ecol* 79:1132–1141.
16. Wolyniak CJ, Sacks GL, Pan BS, Brenna JT (2005) Carbon position-specific isotope analysis of alanine and phenylalanine analogues exhibiting nonideal pyrolytic fragmentation. *Anal Chem* 77:1746–1752.
17. Su X, Lu W, Rabinowitz JD (2017) Metabolite spectral accuracy on orbitraps. *Anal Chem* 89:5940–5948.
18. Bay LJ, Chan SH, Walczyk T (2015) Isotope ratio analysis of carbon and nitrogen by elemental analyser continuous flow isotope ratio mass spectrometry (EA-CF-IRMS) without the use of a reference gas. *J Anal At Spectrom* 30:310–314.
19. Kleiner M, et al. (2017) Assessing species biomass contributions in microbial communities via metaproteomics. *Nat Commun* 8:1558.
20. Kleiner M, Petersen JM, Dubilier N (2012) Convergent and divergent evolution of metabolism in sulfur-oxidizing symbionts and the role of horizontal gene transfer. *Curr Opin Microbiol* 15:621–631.
21. Kleiner M, et al. (2015) Use of carbon monoxide and hydrogen by a bacteria-animal symbiosis from seagrass sediments. *Environ Microbiol* 17:5023–5035.
22. Woyke T, et al. (2006) Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* 443:950–955.
23. Wippler 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.
24. McCutchan JH, Lewis WM, Kendall C, McGrath CC (2003) Variation in trophic shift for stable isotope ratios of carbon, nitrogen, and sulfur. *Oikos* 102:378–390.
25. Tiunov AV (2007) [Stable isotopes of carbon and nitrogen in soil ecological studies]. *Izv Akad Nauk Ser Biol* 34:475–489. Russian.
26. Kleiner M, et al. (2012) Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc Natl Acad Sci USA* 109:E1173–E1182.
27. Ponnudurai R, et al. (2017) Metabolic and physiological interdependencies in the *Bathymodiolus azoricus* symbiosis. *ISME J* 11:463–477.
28. Blair N, et al. (1985) Carbon isotopic fractionation in heterotrophic microbial metabolism. *Appl Environ Microbiol* 50:996–1001.
29. Milo R, Jorgensen P, Moran U, Weber G, Springer M (2010) BioNumbers–The database of key numbers in molecular and cell biology. *Nucleic Acids Res* 38:D750–D753.
30. Ehleringer JR, et al. (2008) Hydrogen and oxygen isotope ratios in human hair are related to geography. *Proc Natl Acad Sci USA* 105:2788–2793.
31. Jehmlich N, Vogt C, Lünsmann V, Richnow HH, von Bergen M (2016) Protein-SIP in environmental studies. *Curr Opin Biotechnol* 41:26–33.

32. Jehmlich N, Schmidt F, von Bergen M, Richnow H-H, Vogt C (2008) Protein-based stable isotope probing (Protein-SIP) reveals active species within anoxic mixed cultures. *ISME J* 2:1122–1133.
33. Pan C, et al. (2011) Quantitative tracking of isotope flows in proteomes of microbial communities. *Mol Cell Proteomics* 10:M110.006049.
34. Fischer CR, Bowen BP, Pan C, Northen TR, Banfield JF (2013) Stable-isotope probing reveals that hydrogen isotope fractionation in proteins and lipids in a microbial community are different and species-specific. *ACS Chem Biol* 8:1755–1763.
35. Ipsen A (2014) Efficient calculation of exact fine structure isotope patterns via the multidimensional Fourier transform. *Anal Chem* 86:5316–5322.
36. Wiśniewski JR, Zougman A, Nagaraj N, Mann M (2009) Universal sample preparation method for proteome analysis. *Nat Methods* 6:359–362.
37. Petersen JM, et al. (2016) Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation. *Nat Microbiol* 2:16195.
38. Li W, Godzik A (2006) Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659.
39. Chambers MC, et al. (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* 30:918–920.
40. Vizcaíno JA, et al. (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* 32:223–226.

PUBLICATION V

The article “Nitrogen fixation in a chemoautotrophic lucinid symbiosis” deals with the unexpected capability of the thiotrophic symbiont of the shallow-water bivalve *Codakia orbicularis* to fix atmospheric nitrogen.

Supplementary data is provided in the CD.

nature microbiology

ARTICLES

PUBLISHED: 24 OCTOBER 2016 | VOLUME: 2 | ARTICLE NUMBER: 16193

OPEN

Nitrogen fixation in a chemoautotrophic lucinid symbiosis

Sten König[1,2†], Olivier Gros[2†], Stefan E. Heiden[1], Tjorven Hinzke[1,3], Andrea Thürmer[4], Anja Poehlein[4], Susann Meyer[5], Magalie Vatin[2], Didier Mbéguié-A-Mbéguié[6], Jennifer Tocny[2], Ruby Ponnudurai[1], Rolf Daniel[4], Dörte Becher[5], Thomas Schweder[1,3] and Stephanie Markert[1,3*]

The shallow water bivalve *Codakia orbicularis* lives in symbiotic association with a sulfur-oxidizing bacterium in its gills. The endosymbiont fixes CO_2 and thus generates organic carbon compounds, which support the host's growth. To investigate the uncultured symbiont's metabolism and symbiont-host interactions in detail we conducted a proteogenomic analysis of purified bacteria. Unexpectedly, our results reveal a hitherto completely unrecognized feature of the *C. orbicularis* symbiont's physiology: the symbiont's genome encodes all proteins necessary for biological nitrogen fixation (diazotrophy). Expression of the respective genes under standard ambient conditions was confirmed by proteomics. Nitrogenase activity in the symbiont was also verified by enzyme activity assays. Phylogenetic analysis of the bacterial nitrogenase reductase NifH revealed the symbiont's close relationship to free-living nitrogen-fixing Proteobacteria from the seagrass sediment. The *C. orbicularis* symbiont, here tentatively named '*Candidatus* Thiodiazotropha endolucinida', may thus not only sustain the bivalve's carbon demands. *C. orbicularis* may also benefit from a steady supply of fixed nitrogen from its symbiont—a scenario that is unprecedented in comparable chemoautotrophic symbioses.

Mutualistic associations between marine invertebrates and sulfur-oxidizing (thiotrophic) bacteria are a well-documented and widespread phenomenon in a variety of sulfidic habitats ranging from hydrothermal vents to shallow-water coastal ecosystems[1–3]. Thioautotrophic symbionts generate energy through sulfide oxidation and provide their hosts with organic carbon. In the Lucinidae, a diverse family of marine bivalves, all members are obligatorily dependent on their bacterial gill endosymbionts after larval development and metamorphosis[4]. The shallow-water lucinid *Codakia orbicularis*, which lives in the sediment beneath the tropical seagrass *Thalassia testudinum* along the Caribbean and Western Atlantic coast[5], harbours a single species of endosymbionts in its gills[6]. The symbiont has been shown to be newly acquired by each clam generation[7,8] from a pool of free-living symbiosis-competent bacteria in the environment[9], rather than being inherited from clam parents. *C. orbicularis* appears not to release its endosymbionts, even under adverse conditions, but can digest them as a source of nutrition[10–12]. Moreover, bacterial cell division seems to be inhibited inside the host tissue. The majority of the symbiont population was shown to be polyploid (that is, containing multiple genome copies), while dividing symbiont cell stages are very rarely observed in host bacteriocytes[13]. The host undoubtedly benefits from the symbiont both by way of detoxification of its sulfidic environment and by supply of organic compounds through the bacterial Calvin–Benson cycle. It remains questionable, however, whether the symbiont gains any advantage from this association in evolutionary terms[11].

Biological nitrogen fixation (diazotrophy) is the conversion of molecular nitrogen (N_2) into ammonia[14]. It provides the basis for the biosynthesis of organic nitrogen compounds and thus of indispensable cellular components such as proteins and nucleic acids. Diazotrophy occurs exclusively in the prokaryotic kingdom[15]; nitrogen fixation in eukaryotes is therefore restricted to symbioses with bacteria[16]. Prokaryotic nitrogen fixers are associated with a diversity of eukaryotic hosts in a wide range of habitats, including the well-studied *Rhizobium* and *Bradyrhizobium* genera that form root-nodule symbioses with legumes[17] and the actinomycetal *Frankia* species associated with woody plants[18]. In marine environments, symbiotic nitrogen-fixers are known to be associated with a variety of invertebrates, such as wood-boring bivalves (shipworms)[19], corals, sponges and sea urchins[20].

Marine ecosystems are frequently nitrogen-limited and therefore considered to be favourable niches for diazotrophic organisms[21]. Coastal seagrass beds, such as those inhabited by *Codakia* species, are areas of particularly high microbial nitrogen fixation activity[14], as the plants' rhizosphere is densely populated by a complex diazotrophic bacterial community[22,23]. The majority of these nitrogen-fixers are anaerobic, heterotrophic sulfate reducers, which decompose decaying seagrass leaves, producing large amounts of toxic sulfide[24]. Thiotrophic members of the rhizosphere community oxidize and thus detoxify the sulfide[25]. Unlike most other animals, for which sulfide is toxic, lucinids thrive in this habitat. Moreover, through the sulfide-oxidation activity of their symbionts, the bivalves contribute to sulfide detoxification of the seagrass bed ecosystem and have therefore been suggested to be part of a beneficial association with the seagrass ('tripartite symbiosis')[26]. However, nitrogen fixation in their sulfur-oxidizing symbiont has to our knowledge never been reported for *C. orbicularis*, nor yet for any other thiotrophic marine invertebrate symbiosis, despite the broad scientific attention these systems have attracted[1–3].

[1]Department of Pharmaceutical Biotechnology, E.M.A. University of Greifswald, Institute of Pharmacy, Greifswald, Germany. [2]Département de Biologie, Institut de Biologie Paris-Seine, UMR 7138 – Evolution Paris-Seine, Equipe Biologie de la Mangrove, Université des Antilles, UFR des Sciences Exactes et Naturelles, BP 592, 97159 Pointe-à-Pitre cedex, Guadeloupe, France. [3]Institute of Marine Biotechnology, Greifswald, Germany. [4]Göttingen Genomics Laboratory, Göttingen, Germany. [5]E.M.A. University of Greifswald, Institute of Microbiology, Greifswald, Germany. [6]CIRAD – UMR QUALISUD, Station de Neufchateau, 97130 Capesterre Belle-Eau, Guadeloupe, France. †These authors contributed equally to this work. *e-mail: stephanie.markert@uni-greifswald.de

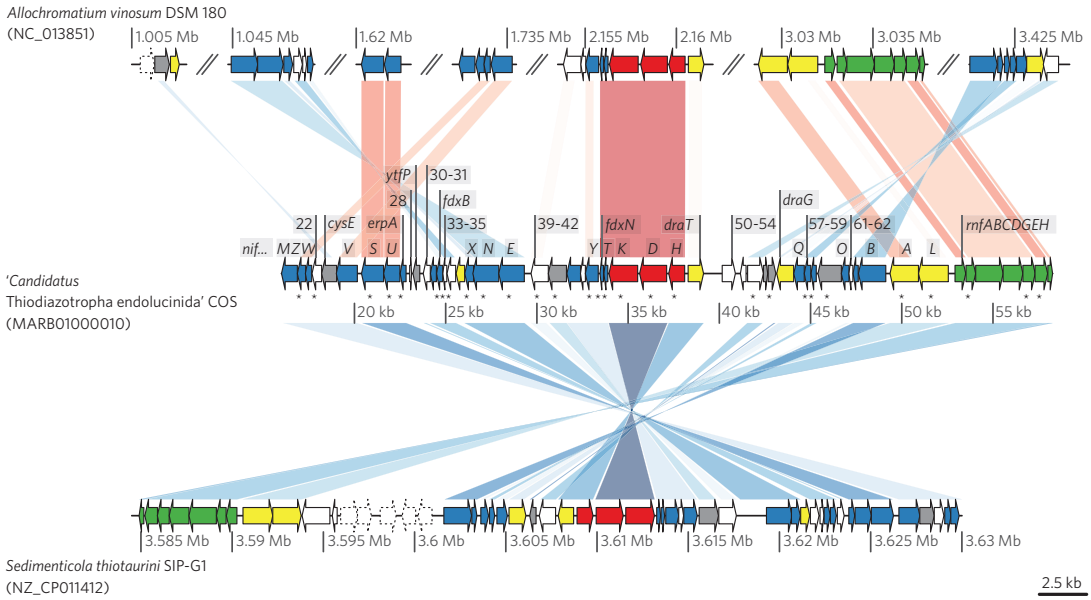

Figure 1 | Nitrogen fixation gene clusters in the *Codakia orbicularis* symbiont and related diazotrophic organisms. The '*Ca.* T. endolucinida' COS *nif* and *rnf* gene clusters (middle) were compared to the genomes of *Allochromatium vinosum* DSM 180 (top) and *Sedimenticola thiotaurini* SIP-G1 (bottom) using BI2seq (BLASTn, E-value 1e-5). Gene functions: red, nitrogenase and nitrogenase reductase; blue, electron transfer, cofactor biosynthesis and other nitrogen fixation-specific functions; yellow, transcriptional and post-translational regulation of nitrogen fixation; grey, function not necessarily related to nitrogen fixation; green, *rnf* genes. White arrows with solid outlines indicate unknown function, and white arrows with broken outlines indicate other functions (these genes of *A. vinosum* and *S. thiotaurini* do not have a homologue in the *C. orbicularis* symbiont's *nif* cluster). Sequence similarities are symbolized by red hues for direct comparisons and blue hues for reversed comparisons. Darker colours correspond to higher identities. For protein functions see Supplementary Table 3. *Identified as a protein in this study. Sequence data and the BLAST comparison files were drawn with the R package genoPlotR (ref. 57) and edited in Inkscape.

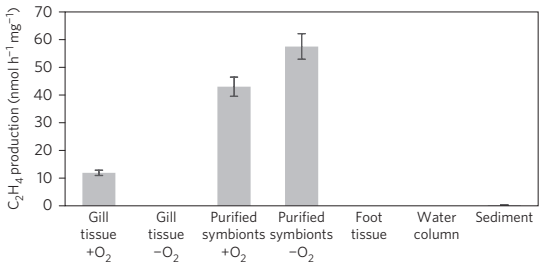

Figure 2 | Nitrogenase activity assay. The production of ethylene as a consequence of acetylene reduction by the nitrogenase enzyme complex was measured in gill tissue samples from freshly collected *C. orbicularis* specimens, in purified gill symbionts, in symbiont-free foot tissue, in the water column and in *T. testudinum* sediment. Gill tissue and isolated symbionts were assayed under aerobic ($+O_2$) conditions and under microaerobic ($-O_2$) conditions ('$+O_2$' and '$-O_2$' refer to initial oxygen concentrations at the start of the assay). All other sample types were incubated under aerobic conditions. Purified symbiont fractions contained approximately 5×10^8 to 8×10^8 bacteria per assay and gill tissue samples contained approximately 3×10^8 bacteria per assay (see Methods for details). The ethylene production rate is given in nmol h^{-1} mg^{-1} as average values ($n = 5$), and error bars show standard deviation. No ethylene production was detected in gill tissue samples under microaerobic conditions, in symbiont-free foot tissue and in the water column.

Here, we show, in a combined proteogenomic and experimental approach, that the thioautotrophic gill endosymbiont of *C. orbicularis*, '*Candidatus* Thiodiazotropha endolucinida', expresses nitrogen fixation proteins and exhibits nitrogenase activity under its natural ambient conditions in the host. This finding adds a novel feature to the picture of the *Codakia* symbiosis and enhances our understanding of its role in the seagrass ecosystem.

Results and discussion

Detection of nitrogen fixation in the *C. orbicularis* symbiont. To reconstruct the *C. orbicularis* symbiont's metabolism, we sequenced the genome of purified bacteria to provide a basis for subsequent global proteomic analyses. As expected, the symbiont's genome encoded all the enzymes necessary for sulfide oxidation via the Dsr-APS-Sat pathway and for carbon fixation through the Calvin–Benson cycle. The respective enzymes were expressed in high concentrations, as detected at the protein level (these results will be the subject of a separate publication). Interestingly, we also detected a complete gene cluster for the fixation of dinitrogen in the symbiont's genome (CODIS_20190–20650), including the highly conserved nitrogenase enzyme complex.

The *C. orbicularis* symbiont's nitrogen fixation (*nif*) cluster comprises 47 genes (Fig. 1), 26 of which were found to be expressed as proteins in this study (Supplementary Tables 3 and 4). The actual nitrogenase enzyme complex consists of two components: a dinitrogenase (component I), which is a molybdenum-iron protein and heterotetramer of two NifD and two NifK subunits, and a dinitrogenase reductase (component II), a homodimer of two NifH subunits[15]. All three proteins were found to be expressed in *C. orbicularis* symbiont samples isolated from freshly collected clams (see Supplementary Table 4 for protein concentrations). Nitrogen fixation-related electron transfer proteins, regulatory proteins, as well as proteins needed for nitrogenase cofactor biosynthesis are also encoded among the *nif* genes and many of them were

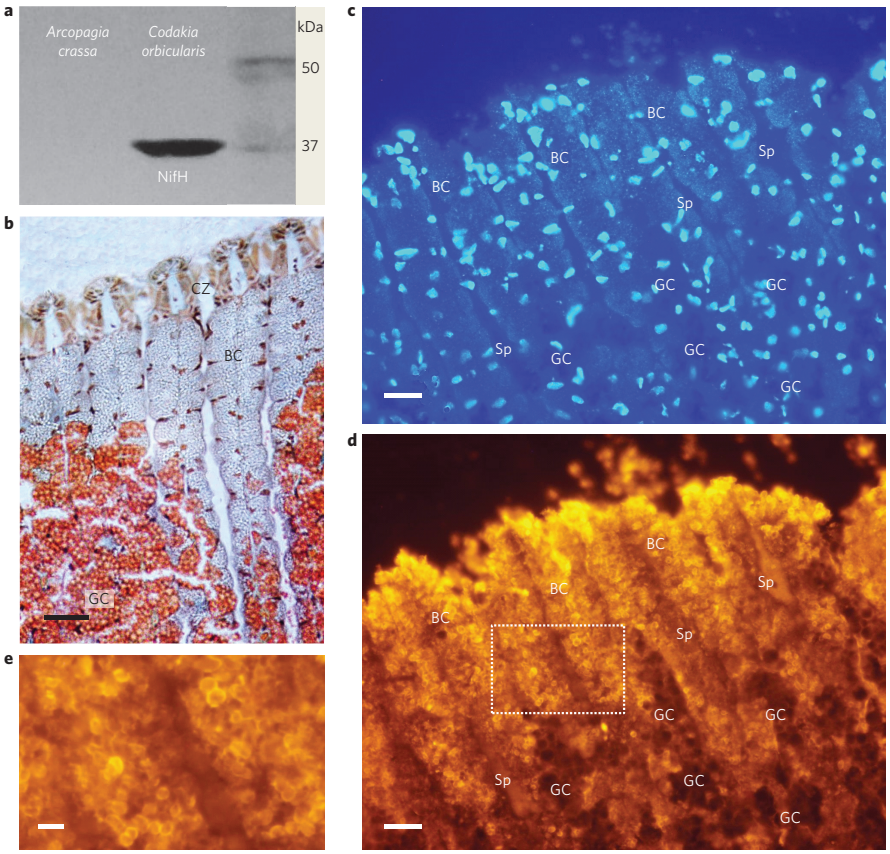

Figure 3 | Immunodetection of nitrogenase expression. a, Western blot analysis of NifH expression in gill extracts (50 mg of protein extract per lane) of the asymbiotic tellinid *Arcopagia crassa* (left, negative control) and of *C. orbicularis* (right). The NifH band migrates at around 34 kDa. The full western blot is displayed in Supplementary Fig. 6. **b**, Goldner staining of gill filaments from a freshly collected *C. orbicularis* bivalve. Each gill filament possesses a ciliated zone (CZ) devoid of bacteria. The lateral zone contains mostly bacteriocytes (BC) harbouring the bacterial symbionts and granule cells (GC, stained in orange), which are free of bacteria. Dark brown dots are *C. orbicularis* cell nuclei. Scale bar, 30 µm. **c**, DAPI staining of the lateral gill filament zone. Large bright blue dots indicate the location of nuclei in bacteriocytes and granule cells. Symbiont DNA is visible as small light blue dots all over the bacteriocyte area but not in the granule cell area. Sp: space between the gill filaments. Scale bar, 20 µm. **d**, Immunolocalization of the *C. orbicularis* symbiont's nitrogenase protein in the same gill section as in **c**, hybridized with the Alexa Fluor 546-labelled NifH antibody (yellow fluorescence). NifH fluorescence is visible in bacteriocytes but not in granule cells. Sp: space between gill filaments. Scale bar, 20 µm. **e**, Magnification of a detail from **d** (shown by a rectangle), highlighting the localization of NifH labelling in the bacteria. Scale bar, 2 µm.

detected as proteins under *in situ* conditions. In addition to these *nif*-encoded proteins, we also identified a number of genes and proteins that are apparently indirectly related to nitrogen fixation in the *C. orbicularis* symbiont (see Supplementary Results and Discussion for details). First, immediately adjacent to the *nif* genes, the *C. orbicularis* symbiont's genome encodes a set of Rnf electron transfer proteins (CODIS_20660–20720), which are putatively involved in electron transfer towards the nitrogenase complex. Second, the high abundance of glutamine synthetase (CODIS_03590) and the P-II signal transduction protein GlnK (CODIS_03150) suggests that nitrogen fixation-derived ammonium is assimilated through the GS/GOGAT pathway and that the symbionts analysed in this study may have experienced nitrogen limitation. The two-component nitrogen regulation (Ntr) system, which controls ammonium assimilation in Proteobacteria[27], appears also to be present and active in the *C. orbicularis* symbiont. Third, we also detected a rubrerythrin (Rbr, CODIS_04230) as one of the most abundant proteins in the '*Ca*. T. endolucinida' proteome. Rubrerythrins provide protection against hydrogen peroxide-mediated oxidative stress[28] and have been suggested to play a role in shielding the highly oxygen-sensitive nitrogenase protein from oxidative damage[29]. We observed that both the nitrogenase proteins and Rbr were detected with significantly lower protein concentrations during diazotrophy-inhibiting conditions (that is, energy limitation caused by sulfide starvation; Supplementary Fig. 5 and Supplementary Results and Discussion). These results may suggest a co-regulation of Rbr and nitrogenase expression and support the idea that Rbr might be involved in protecting the nitrogenase from oxidative stress.

The functionality of the nitrogenase enzyme complex was confirmed by an acetylene reduction assay[30]. We detected the nitrogenase-dependent conversion of acetylene to ethylene in *C. orbicularis* gill tissue samples and in purified symbionts obtained from freshly collected individuals (Fig. 2). Ethylene accumulation rates in gill samples ranged from 8.65 to 15.41 nmol h^{-1} mg^{-1} of tissue under aerobic conditions, while no ethylene production was detected in

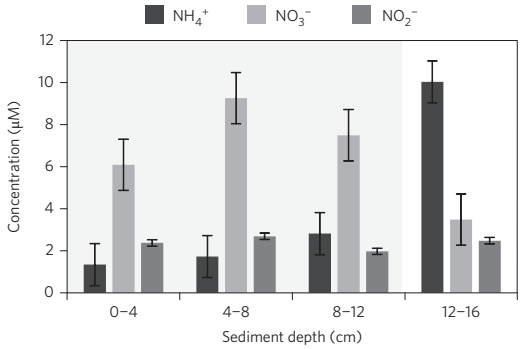

Figure 4 | *In situ* habitat concentrations of NH_4^+, NO_3^- and NO_2^-. Gas concentrations (in μM) in pore water samples from *T. testudinum* sediments are given as averages of six individual sediment cores ($n = 6$). Standard deviations are indicated by error bars. The light grey background indicates the sediment depth in which *C. orbicularis* burrows. Sediment samples were taken within a radius of 20 cm around *C. orbicularis* specimens.

gills incubated under low oxygen conditions. The latter might be due to the fact that host bacteriocytes probably die quickly in the absence of oxygen, which may also be detrimental for their intracellular symbionts. On the other hand, symbionts that were isolated from their bacteriocytes showed distinctly higher nitrogenase activities (Fig. 2) with an average ethylene production rate of 43 nmol h^{-1} mg^{-1} under aerobic conditions. Under microaerobic conditions, the nitrogenase activity of purified symbionts was even higher (57.5 nmol C_2H_4 h^{-1} mg^{-1}) than in the aerobic approach, suggesting that the thioautotrophic symbiont's nitrogenase enzyme complex may—as most other nitrogenases—be oxygen-sensitive and work more efficiently under low-oxygen conditions[16]. Low ethylene production (~0.24 nmol C_2H_4 h^{-1} mg^{-1}) was observed in crude sediment collected from *T. testudinum* seagrass beds, presumably indicating the presence of free-living nitrogen-fixing bacteria in the sediment. No ethylene production was detected after incubation of sea water with acetylene and after incubation of symbiont-free foot tissue with acetylene, respectively. This clearly indicates that nitrogenase activity is exclusively present in symbiont-containing tissue in *C. orbicularis*.

To support this observation, we performed a western blot analysis based on an antibody raised against NifH (~34 kDa). As displayed in Fig. 3a, hybridization of the NifH antibody was observed in extracts from freshly collected *C. orbicularis* gills, while no signal could be detected in the negative control, that is, gill tissue from the asymbiotic marine bivalve *Arcopagia crassa* (Tellinidae). As described by Frenkiel and Mouëza[31], gills of freshly collected individuals of *C. orbicularis* consist of an arrangement of filaments organized in three zones, the lateral of which contains the bacteriocytes (BCs), filled with intracellular bacterial symbionts, and granule-containing cells (GC) without endosymbionts (Fig. 3b). To allocate the expression of the nitrogenase protein NifH to the symbionts in the bacteriocytes, we conducted immunofluorescence microscopy on *C. orbicularis* gill histological sections using an anti-NifH antibody and 4′,6-diamidino-2-phenylindole (DAPI) counterstaining. The DAPI staining (Fig. 3c) allowed for the localization of host nuclei (large light blue dots) in bacteriocytes and granule cells all along the lateral gill filament zone but also of bacterial DNA (small light blue dots), only within the bacteriocytes. The NifH signals (yellow fluorescence, Fig. 3d) cover the bacteriocyte area but are absent from the granule cells and thus evidently emanate from the bacterial endosymbionts located in the cytoplasm of the bacteriocytes (Fig. 3e), even if not all the bacteria from a single bacteriocyte seem to be positive. In the asymbiotic tellinid *A. crassa*, which was used as a negative control, no NifH immunofluorescence was observed (Supplementary Fig. 2). Moreover, trophosome tissue sections of the deep-sea tube worm *Riftia pachyptila* (Vestimentifera, negative control 2), did not hybridize with the NifH antibody either. *R. pachyptila* harbours a sulfur-oxidizing symbiont in its trophosome, which is closely related to the *C. orbicularis* symbiont, but which is not capable of diazotrophy[32,33]. These results clearly reveal that *nifH* expression in *C. orbicularis* can be specifically attributed to the symbionts inside the host bacteriocytes.

Availability and use of other nitrogen sources in the seagrass sediment. Our sediment pore water analyses show that ambient inorganic nitrogen concentrations are comparatively low in the direct vicinity of *C. orbicularis* (that is, within a radius of 20 cm from the animals), suggesting that diazotrophy is of particular advantage for the bacterial symbiont in this habitat. In marine systems, the main sources of nitrogen available for bacterial growth (besides N_2) are NH_4^+ and free amino acids[34], as well as—to a lesser extent—NO_3^-, NO_2^- and urea[35]. In seagrass bed sediments in particular, NH_4^+ is considered to be the dominating nitrogen source, with concentrations of up to 175 μM in some areas[36]. We detected NH_4^+ levels from 1 to 10 μM in the *T. testudinum* sediment directly surrounding *C. orbicularis*, with the highest concentrations measured in the deeper layers of the sediment (Fig. 4). These values are low compared to previously reported NH_4^+ concentrations of up to 120 μM in *T. testudinum* sediments in subtropical Florida[37]. However, sediment pore water ammonium concentrations have been shown to vary substantially between sampling sites and times[36,38]. In tropical *T. testudinum* seagrass beds of the Caribbean, NH_4^+ levels were observed to be particularly low, that is, below 25 μM in the upper 10 cm of sediment[39]. In this presumably ammonium-limited environment, nitrogen fixation is probably the preferable strategy for bacterial nitrogen acquisition.

Interestingly, a number of genes required for uptake and assimilation of nitrate and/or nitrite were found to be expressed in the *C. orbicularis* symbiont (including a nitrate transporter and assimilatory nitrate and nitrite reductase subunits, Supplementary Table 4). Unlike ammonium, nitrate and nitrite are usually considered not to play a substantial role as nitrogen sources in seagrass sediments, with $NO_3^- + NO_2^-$ concentrations ranging from 0 to 10 μM (ref. 36). In accord with previous studies, we detected NO_3^- and NO_2^- concentrations of up to 9 μM and below 3 μM, respectively, in our pore water analysis in the immediate surroundings, that is, within a radius of 20 cm around *C. orbicularis* specimens (Fig. 4). The detection of assimilatory nitrate/nitrite reductase subunits indicates that, despite the relatively low ambient concentration, nitrate (and nitrite) may be relevant as nitrogen sources for the *C. orbicularis* symbiont after all, in addition to N_2 fixation. This seems to be surprising at first, as nitrogen fixation is often repressed in the presence of alternative nitrogen sources in free-living diazotrophs[40]. The simultaneous expression of both *nif* genes and nitrate assimilation genes might be due to less stringent regulation in *C. orbicularis* symbionts, as also suggested for other symbiotic nitrogen fixers, which provide for their hosts' nitrogen supply[41]. However, the most likely explanation would be that the environmental nitrate and nitrite levels are too low to trigger repression of the N_2 fixation genes.

Overall, in view of the specific nitrogen-limited ambient conditions in the seagrass bed habitat, nitrogen fixation in the *C. orbicularis* symbiont emerges as a most advantageous feature. Moreover, employing bacterial endosymbionts that can fix N_2 in this environment might also be highly profitable for *C. orbicularis*. Nitrogen incorporated by the symbiont through N_2 fixation most probably supplements the bivalve's nitrogen diet after symbiont

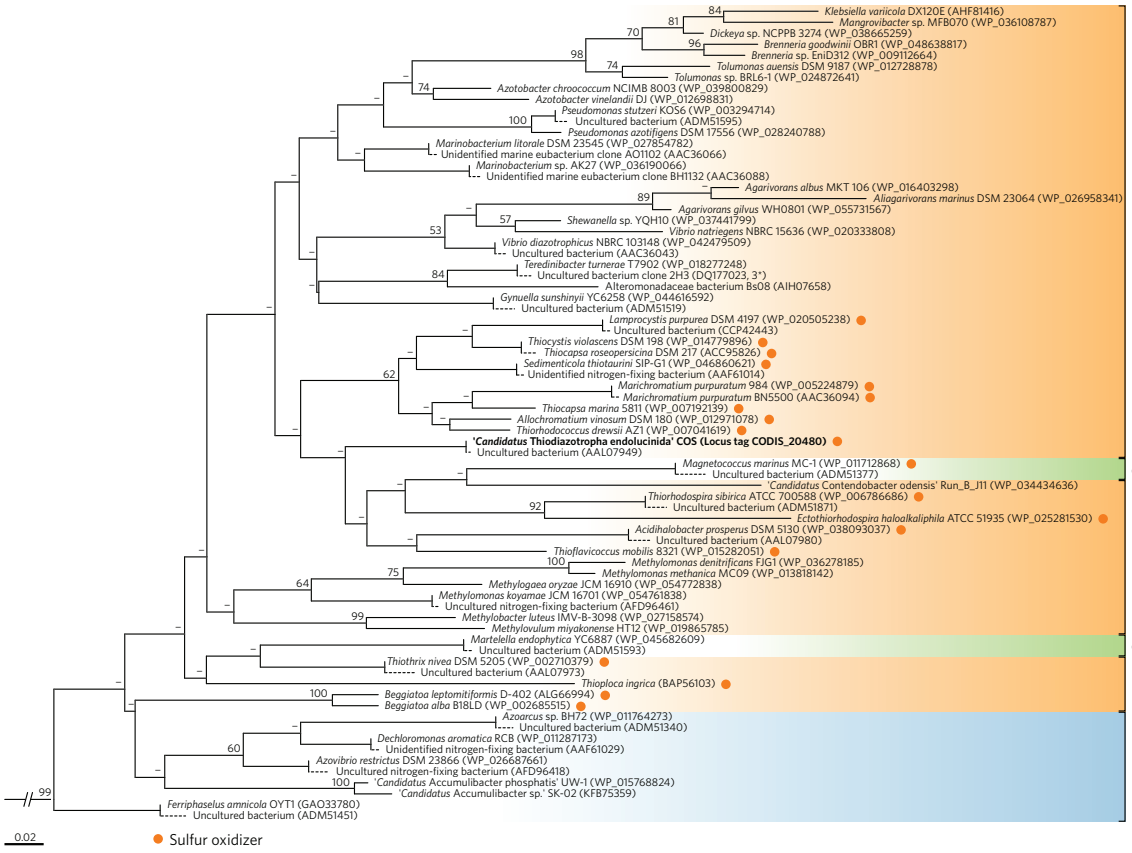

Figure 5 | NifH-based phylogenetic tree of the *Codakia orbicularis* symbiont and its closest relatives. The tree was inferred based on maximum likelihood. Numbers given on the branches are bootstrap proportions as a percentage of 1,000 replicates for values ≥50%. Taxonomic classes of the phylum Proteobacteria are given on the right. Although the original inference used full-length NifH amino acid sequences, partial NifH sequences from uncultured marine isolates were added by maximum parsimony (dotted branches). Note that some of the full-length sequences had up to 80 partial sequences affiliated to them. For clarity's sake, only one partial sequence per full-length sequence is displayed (that is, the one with highest sequence similarity). For a comprehensive list of all partial sequences that were affiliated to the *C. orbicularis* symbiont '*Candidatus* Thiodiazotropha endolucinida' COS see Supplementary Table 5. '*Ca.* T. endolucinida' COS (this study) is shown in bold. Orange dots denote known diazotrophic sulfur oxidizers. (Note that no information regarding potential thiotrophy is available for the uncultured seagrass bed isolates. Some of these clones might be sulfur oxidizers, although they are not marked with orange dots.) Accession numbers or locus tags in parentheses refer to the International Nucleotide Sequence Database Collaboration (INSDC). The sequence marked with an asterisk was only available as a nucleotide sequence and was translated for this analysis (frame 3). The NifH sequences of *Anabaena variabilis* ATCC 29413 (AAA93020) and *Frankia alni* ArI3 (AAA96262) were used as outgroup (shown in Supplementary Figure 4d).

digestion. Thus supplied, the host would circumvent competition with seagrass root tissues for other, limited nitrogen sources. (See Supplementary Fig. 3 for a model of interactions between bivalves, symbionts, seagrass and the free-living microbial community.)

Phylogeny of nitrogen fixation in the *C. orbicularis* symbiont. Our phylogenetic analyses of the nitrogenase reductase NifH and of the nitrogenase subunits NifD and NifK place the *C. orbicularis* symbiont in a group with free-living diazotrophic Gammaproteobacteria that are known to be sulfur oxidizers, such as *Sedimenticola thiotaurini*, *Thiorhodococcus drewsii* and *Allochromatium vinosum* (Fig. 5 and Supplementary Fig. 4b–d). Moreover, particularly high similarities were observed between the *C. orbicularis* symbiont's NifH and NifH sequences of a variety of uncultured diazotrophic isolates from seagrass beds and mangrove sediments, some of them classified as presumptive Beta- or Gammaproteobacteria[42] (see Supplementary Results and Discussion and Supplementary Table 5 for details). Although these uncultured isolates are not characterized with regard to their 16S rRNA genes (nor their genetic potential for thiotrophy), this pronounced NifH-based identity implies a high degree of physiological similarity to '*Ca.* T. endolucinida'. At the level of 16S rRNA sequence phylogeny, the *C. orbicularis* symbiont clusters most closely with other yet uncultured thiotrophic lucinid symbionts from seagrass bed habitats, many of which are known Gammaproteobacteria (Supplementary Fig. 4a). None of these related lucinid symbionts has been characterized with regard to potential diazotrophy, but considering their close phylogenetic relatedness and their very similar ecosystems it seems quite likely that these symbionts, too, are capable of nitrogen fixation.

These observations suggest the presence of a hitherto unrecognized group of sulfur-oxidizing diazotrophic Gammaproteobacteria in seagrass sediment ecosystems, whose members can be free-living, like the uncultured clones, or live in symbiosis with *C. orbicularis* and (possibly) other lucinids. The phylogenetic relatedness of

thiotrophic lucinid symbionts and seagrass sediment bacteria has been suggested previously[23,43]. The *C. orbicularis* symbiont exists as a free-living bacterium in the environment before colonization of its bivalve host[9] and therefore presumably belongs to this microbial seagrass community. Moreover, free-living sulfide oxidizers are known to be part of the complex bacterial population in the seagrass rhizosphere (Supplementary Fig. 3)[25], although their potential involvement in plant-associated nitrogen fixation has not been addressed so far. It will be highly interesting to investigate this hypothetical gammaproteobacterial community of thiotrophic diazotrophs in seagrass beds in further studies, to verify its existence, abundance and composition.

Conclusion

The results presented in this study provide evidence that the thiotrophic *C. orbicularis* gill symbiont '*Ca.* T. endolucinida' is a diazotroph. The bacterium is thus perfectly adapted to the prevalent low NH_4^+ and NO_3^- concentrations in its seagrass habitat. As *C. orbicularis* digests parts of its symbiont population for nutrition, it seems very likely that diazotrophy-derived nitrogen compounds are transferred from the symbiont to the host, supporting the bivalve's nitrogen diet. Future studies will be needed to confirm this proposition. Harbouring a 'multi-talented' symbiont, which not only detoxifies sulfide and sustains its host's carbon needs, but which presumably also provides nitrogen compounds, seems to be most advantageous for *C. orbicularis* and may enable the bivalve to prevail in its highly sulfidic and nitrogen-limited environment. The interesting question of how the symbiont might benefit from this association in evolutionary terms—considering the highly restrictive regime of its host—remains to be elucidated in future studies (Supplementary Discussion). As the specific ambient conditions in sulfidic seagrass beds are a common feature shared by many lucinid symbioses, it seems quite likely that future studies will reveal diazotrophy in other thiotrophic lucinid symbionts as well. Follow-up studies are required to elucidate the composition of a putative population of sulfur-oxidizing diazotrophic Gammaproteobacteria in the seagrass bed sediment and their specific role in the rhizosphere ecosystem.

Marine chemoautrophic symbioses have been studied intensively over the past decades, as has nitrogen fixation in marine and terrestrial habitats. However, the combination of both chemoautotrophy and nitrogen fixation in a microbial population or even in a single organism has rarely been considered. The recent discovery of diazotrophic *Thiothrix* species living as amphipod ectosymbionts in a sulfidic cave[44] and the coexistence of chemoautotrophy and diazotrophy in the microbial community of cold water corals[45], along with the results of this study, suggest that chemoautotrophic diazotrophs may be more widespread than previously anticipated. Other invertebrate models colonizing sulfidic environments may rely on chemoautotrophic symbionts that are also capable of diazotrophy.

Methods

Sampling. Adult individuals of *C. orbicularis* (Linné, 1758) were collected by hand at a depth of 5–10 cm in the sediment of *T. testudinum* seagrass beds in Guadeloupe (French West-Indies, Caribbean). For details of sampling sites, times and replicate numbers see Supplementary Table 1. Specimens designated for immediate dissection (for proteomics, genome sequencing, acetylene reduction assays, western blot and immunohistochemistry) were transported to the laboratory in seawater at ambient temperature within 1 h after collection. *C. orbicularis* individuals designated for incubation (starvation experiment, see below) were transported to the laboratory within 5–6 h. Sediment samples were collected off 'Îlet cochon', directly adjacent to *C. orbicularis* individuals (that is, not more than 20 cm away from the animals) using translucent polypropylene sediment corers (80 mm diameter). The cores (up to 16 cm in depth) were sliced at 4 cm intervals in the field, stored at 4 °C in an ice chest for transport to the laboratory and were subjected to pore water extraction within 1 h (see section 'Chemical analysis of pore water samples').

Incubation experiments with *C. orbicularis*. For incubation experiments, freshly collected *C. orbicularis* specimens were randomly separated into two batches (a control batch and a starvation batch). Specimens in the control batch were dissected immediately, while those in the starvation batch were kept in sterile (0.22 µm-filtered) sea water in 50 l plastic tanks at 26 °C for seven days. The water was oxygenated using an aquarium air pump. To simulate starvation conditions, no organic particles and no reduced sulfur compounds were added throughout the experiment. Bivalves were killed after one week of incubation, and symbionts were purified by density gradient centrifugation (see section 'Symbiont enrichment') and stored at −20 °C.

Chemical analysis of pore water samples. Nitrate, nitrite and ammonium concentrations in pore water from sediment cores were analysed as follows. Pore water was extracted from the sliced cores by filtration under vacuum using a 5 µm filter. Collected pore water samples were then passed through 0.45 µm membrane syringe filters, frozen immediately and stored frozen before colorimetric tests. Nitrate and nitrite concentrations were analysed according to the standard colorimetric Griess method[46] with the use of vanadium as a reduction agent[47], and ammonium was measured by the indophenol blue method[48].

Symbiont enrichment. Bacterial symbionts were purified from healthy *C. orbicularis* gills with a light beige colour using Percoll density gradient centrifugation as described previously[13,49]. Briefly, after dissection of the animals, gill tissue was homogenized in a Dounce homogenizer in cold sterile seawater (salinity 35 g l^{-1}) and the resulting homogenate was centrifuged at 30*g* for 1 min (4 °C) to remove crude tissue debris. The supernatant was centrifuged again (400*g*, 2 min, 4 °C) to pellet the bacterial cells, which were subsequently resuspended in sterile sea water and layered on top of a Percoll cushion (60% Percoll in imidazole-buffered saline, containing 490 mM NaCl, 30 mM $MgSO_4$, 11 mM $CaCl_2$, 3 mM KCl and 50 mM imidazole, pH 7.5). During centrifugation (4,000*g*, 10 min, 4 °C), the bacteria—because of their elemental sulfur inclusions—accumulate below the cushion, while host tissue fragments stay on top. The bacteria were collected, washed twice in sterile sea water, and stored at −20 °C.

Genome sequencing and genomic analysis. Symbionts for genome sequencing were isolated from one single *C. orbicularis* host individual immediately after collection of the animal (as described in section 'Sampling') in March 2012. Genomic DNA was isolated from the purified bacteria using the MasterPure DNA Purification Kit (Epicentre). The genome was sequenced using Illumina sequencing technology. A Nextera shot-gun library was generated for a 112 bp paired end (PE) sequencing run on a Genome Analyzer IIx. For an overview of all subsequent steps involved in assembly, binning and annotation, see Supplementary Fig. 1. Sequencing resulted in 2 × 6,108,111 raw reads, which were checked with FastQC version 0.11.3 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and subsequently trimmed with Trimmomatic version 0.35 (ref. 50) (settings: remove Nextera adapters and leading/trailing low quality bases, trim using MAXINFO:40:0.8 and keep sequences ≥36 nt). As preliminary assemblies had shown some eukaryotic contaminations in the sequenced data, trimmed reads (2 × 5,729,978 reads + 355,335 single reads) were assembled in a two-step process of binning and mapping to obtain a clean symbiont draft genome. First, Ray version 2.3.1 (ref. 51; k = 37) was used and the resulting 1,018 contigs were visually binned with VizBin (ref. 52; Bin 'COS': 83 contigs; Bin 'Other': 935 contigs). The trimmed reads used for assembly were then mapped against Bin COS with the help of BBSplit version 35.85. In a second step, SPAdes version 3.7.0 (ref. 53) used those mapped reads (2 × 5,131,806 reads + 336,019 single reads) to perform the final assembly (k = 21, 33, 55, 77, 'careful' option and minimum coverage of 30). This time, contigs were binned with MetaWatt version 3.5.2 (ref. 54), basically splitting contigs into bins with known and unknown taxon assignments. The final draft genome comprises 55 contigs (total length = 4,489,993 bp; largest contig = 297,909 bp; N50 = 164,003 bp), as evaluated by QUAST version 3.2 (ref. 55). The rapid annotation tool Prokka version 1.11 (ref. 56) was used for automatic open reading frame (ORF) calling and annotation.

Gene cluster comparison. All genes in the *nif* gene cluster of '*Ca*. T. endolucinida' COS (CODIS_20190–20650, Fig. 1) were compared to the genomes of *Allochromatium vinosum* DSM 180 (NC_013851) and *Sedimenticola thiotaurini* SIP-G1 (NZ_CP011412) using Bl2seq (BLASTn, E-value 1e-5). Sequence data and the BLAST comparison files were drawn with the R package genoPlotR version 0.8.4 (ref. 57) and edited in Inkscape version 0.91. Blast results were automatically edited, so that short hits contained in longer hits were removed. They were further curated manually, so that each of the '*Ca*. T. endolucinida' genes matched only one gene in its relatives.

Proteomic analyses. Mass spectrometry (MS) analyses of *C. orbicularis* symbiont proteins were performed in two parallel approaches: (1) in an identification-centred (that is, qualitative), gel-based approach and (2) in a quantitative gel-free approach (see section 'Gel-free protein quantification'). Measurements for the identification-centred approach were performed in three biological replicates (that is, for symbionts of three individual *C. orbicularis* hosts) for (1) soluble symbiont proteins from freshly collected *C. orbicularis*, (2) symbiont membrane proteins from freshly

collected hosts and (3) symbiont membrane proteins from *C. orbicularis* specimens collected after one week of starvation. For the quantitative approach, three measurements (three technical replicates) were acquired for each of the three biological replicates of (1) symbiont proteins from freshly sampled *C. orbicularis* specimens and (2) symbiont proteins from starved *C. orbicularis* after one week of starvation. For a detailed overview of replicate numbers in all MS measurements performed in this study, see Supplementary Table 2.

Protein extraction. Symbiont cell pellets for gel-based MS analysis (approach (1)) of soluble proteins and membrane proteins were resuspended in lysis buffer (10 mM Tris, 1 mM EDTA, Roche complete protease inhibitor cocktail) before cell disruption by sonication (3 × 20 s) under permanent cooling. Cell debris was removed by centrifugation (10 min, 4 °C, 12,000*g* for soluble protein extraction/8,000*g* for extraction of membrane proteins), leaving the protein raw extract in the supernatant. Raw extracts designated for the analysis of soluble proteins were subjected to acetone precipitation overnight (−20 °C), the resulting protein pellets were washed with ethanol, and soluble proteins were resuspended in 8 M urea/2 M thiourea solution. For membrane protein analysis, the protein raw extract was subjected to repeated ultracentrifugation and solubilization steps as described by Eymann *et al.*[58], before solubilization of the enriched membrane protein fraction in 50 mM TEAB (triethylammonium bicarbonate) buffer. For quantitative gel-free proteome analysis (approach (2)), symbiont cell pellets were resuspended in 50 mM TEAB buffer and cells were disrupted by sonication (3 × 20 s). In either case, protein concentrations were determined using a Bradford assay[59].

Gel-based protein identification. Aliquots of 20 µg soluble protein and 20 µg membrane protein fraction, respectively, were loaded onto 12% polyacrylamide mini gels and subjected to gel electrophoresis at 150 V for 1 h. After Coomassie staining, individual gel lanes were excised and divided into ten equally sized slices (subsamples) each. Gel pieces were destained (200 mM NH_4HCO_3, 30% acetonitrile) and dried before overnight digestion with trypsin solution (1 µg ml^{-1}, sequencing grade, Promega) at 37 °C. Peptides were eluted from the gel pieces in a sonication bath and purified using ZipTips (Millipore) according to the manufacturer's recommendations. As described previously[60], peptide mixes were subjected to separation using reversed-phase C18 column chromatography on a nanoACQUITY-UPLC system (Waters). Mass spectrometry (MS)- and tandem mass spectrometry (MS/MS) data were acquired using an online-coupled LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific). MS spectra were searched against a target–decoy protein sequence database including all sequences from the *C. orbicularis* symbiont database (see 'Genome sequencing and genomic analysis' section) and common laboratory contaminants using the SEQUEST Sorcerer platform (Sage-N). Scaffold (version 4.2, Proteome Software) was used to validate MS/MS-based peptide and protein identifications using the XCorr-based filter settings described previously[61]. Proteins were considered identified if they passed the filter criteria in at least one of the three biological replicates. Peptide false discovery rates (FDRs) were <2% and protein FDRs were 0.0% throughout all samples.

Gel-free protein quantification. As described by Muntel *et al.*[62], 250 µg protein (in TEAB buffer) was treated with RapiGest SF Surfactant solution (Waters), before reduction with TCEP (tris(2-carboxyethyl)phosphine, final concentration 5 mM), alkylation with iodacetamide (final concentration 10 mM) and digestion with activated trypsin (5 h, 37 °C). After removal of RapiGest by acidification and repeated centrifugation, peptides were purified using StageTips (Proxeon) as suggested by the manufacturer. Tryptically digested bovine haemoglobin (Waters, MassPREP Bovine Hemoglobin Standard) was added to the protein samples as an internal peptide standard at a final concentration of 50 fmol µl^{-1}. Samples were stored at −80 °C until MS analysis. Peptide mixtures were analysed using the nanoACQUITY-UPLC system (Waters), coupled to a Synapt G2 mass spectrometer (Waters), equipped with a NanoLockSpray source in positive mode. The analyser was set to resolution mode and operated with ion mobility separation as described by Huja and co-authors[63]. The ProteinLynx Global Server (PLGS) 2.5.3 was used for processing raw data and carrying out the database searches against the randomized *C. orbicularis* symbiont database with added common laboratory contaminants and the bovine haemoglobin sequence[63]. Proteins were considered identified if they were detected in one of the respective technical or biological replicates with a minimum of two peptides at a maximum FDR of 5% on the protein level. For quantification, only those proteins were included for which a minimum of two peptides were detected and which were identified in a minimum of two technical replicates in at least two biological replicates. For the proteins that met those criteria, average values of protein concentrations (in fmol ng^{-1}) were calculated in two subsequent steps, that is, first across all available technical replicates for each biological replicate and subsequently across all biological replicates using the average values from step 1 (ref. 64). Protein FDRs were below 1%. To identify significant differences in protein expression between fresh control samples and samples obtained after starvation, mean protein abundances calculated across all three technical replicates of each biological replicate were used for statistical testing in TM4 MeV (TM4 Software Suite's Multi-Experiment Viewer; http://www.tm4.org/mev.html). After *z*-score transformation, an unpaired permutational *t*-test with Welch approximation (using all permutations, *P* value of 0.01, adjusted Bonferroni correction) was applied.

Microscopy and immunohistochemistry. For an overview and histological information about the *C. orbicularis* gill ultrastructure, a gill (dissected from one *C. orbicularis* individual) was fixed in Bouin's fluid for 24 h at room temperature and then embedded in Paraplast. Tissue sections (7 µm thick) were stained by Goldner's trichrome[65] to allow for clear differentiation between granule cells and bacteriocytes.

For immunolocalization, gills of a freshly collected *C. orbicularis* specimen were dissected and fixed in 2% paraformaldehyde (in 0.22 µm-filtered seawater) and dehydrated in ethanol before being embedded in paraffin. Gill tissue of the asymbiotic bivalve *Arcopagia crassa* (familiy Tellinidae) was used as a negative control and treated accordingly. As an additional negative control, pieces of trophosome tissue of the vestimentiferan tube worm *Riftia pachyptila* were fixed in 4% paraformaldehyde immediately after collection (in November 2014, Supplementary Table 1) and stored in 70% ethanol before embedding in paraffin. Histological sections (7 µm thick) were deparaffinized before rehydration. After non-specific binding site saturation, sections were incubated with anti-NifH (nitrogenase iron protein) antibody from Agrisera, at a dilution of 1:500 in blocking solution containing 3% bovine serum albumin (BSA) in 1× PBS (phosphate-buffered saline) for 2 h at room temperature. Subsequently, sections were incubated in the secondary antibody, horseradish peroxidase-conjugated anti-hen IgY (Santa Cruz Biotechnology) at a 1:200 dilution in blocking solution for 1 h at room temperature. Slides were incubated in amplification buffer with Alexa Fluor 546, mounted with Vectashield Mounting Medium with 4′,6-diamidino-2-phenylindole (DAPI, VectorLabs) and observed under an epifluorescence microscope (Nikon Eclipse 80i).

Western blot analysis. Gill tissues from fresh *C. orbicularis* and *A. crassa* specimens (negative control) were crushed on ice before protein extraction in lysis buffer as described in the 'Proteomic analyses' section. The total protein raw extract was homogenized in Laemmli buffer (Sigma-Aldrich), and denatured at 95 °C for 10 min before storage at −20 °C. Proteins were separated on precast 4–12% NuPage polyacrylamide mini gels (Invitrogen) and blotted to PVDF membranes (Immobilon) for 1.5 h. The proteins were probed with a hen anti-NifH antibody (Agrisera) at a dilution of 1:10,000. Blots were incubated in the secondary antibody (horse radish peroxidase-conjugated anti-hen IgY) diluted to 1:50,000. HRP activity was detected using the ECL kit (GE Healthcare) according to the manufacturer's instructions and documented on Kodak X-OMAT autoradiography film (Fisher-Scientific). The specificity of the antibody raised against NifH was verified using *C. orbicularis* whole gill protein extracts.

Nitrogenase activity assay. Nitrogenase activity was tested in both dissected gills and in purified gill endosymbiont fractions obtained by the Percoll cushion method (see 'Symbiont enrichment' section). The isolated symbiont fractions contained ~5×10^8 bacteria (per assay), while the number of bacterial cells in the gill samples was approximately 3×10^8 per assay (assuming a bacterial density of 10^6 bacteria per mg of gill tissue; 242–352 mg of gill tissue was used for each assay). Tissue and cell samples, respectively, were incubated in 22 ml glass tubes containing 2 ml of 0.22 µm-filtered sea water in contact with a headspace (20 ml) of air supplemented with 20 µmoles of acetylene (aerobic conditions at the beginning of the incubation). A parallel incubation of gill tissue and purified symbionts was started under low-oxygen conditions, for which air in the head space was replaced by O_2-free argon and supplemented with the same quantity of acetylene. The samples were incubated at 24 °C and shaken continuously (60 r.p.m.) using a shaking water bath (SWB25 Thermo Scientific) throughout the experiment to equilibrate the gas and bacteria/tissue samples. After 24 h of incubation, ethylene production was measured as follows. A 500 µl gas sample was withdrawn from the headspace using a gas-tight hypodermic glass syringe (Interchim) and analysed using a CP-3800 gas chromatograph (Varian). The apparatus was equipped with a flame ionization detector (FID) for acetylene and ethylene detection, and thermal conductivity detectors (TCDs) for air nitrogen detection. Gas analysis was performed on a Rt-QS-Bond capillary PLOT fused-silica column (15 cm × 0.53 mm × 20 µm; Restek) under the following conditions: column input flow rate 9.1 ml min^{-1}, oven temperature 40 °C, FID temperature 150 °C, helium as gas carrier (30 ml min^{-1} flow) and make-up (20 ml min^{-1} flow), hydrogen at 4.5 bar and air at 300 ml min^{-1} flow. The column retention time for acetylene and ethylene was monitored with standards throughout the experiments. Sea water and symbiont-free *C. orbicularis* foot tissue were assayed as negative controls, as was a sediment sample from the *T. testudinum* bed (larger seagrass roots were removed before analysis). In addition, gill and foot tissue samples, purified symbionts, sea water and sediment samples were incubated without acetylene and assayed for ethylene production to exclude false positive results. No ethylene production was detected in any of these samples incubated without acetylene.

Phylogenetic analysis

NifH phylogeny. The NifH tree was created in two consecutive steps. First, full-length NifH amino acid sequences of the *C. orbicularis* symbiont and of related cultured diazotrophic organisms were aligned and the phylogeny was inferred based on maximum likelihood. Second, partial NifH sequences of environmental clones from various marine habitats were added using ARB's interactive parsimony (see later in this paragraph). The NifH protein of '*Ca.* T. endolucinida' was identified by

searching the Pfam[66] model Fer4_NifH (PF00142, available from http://pfam.xfam.org/family/PF00142/hmm) against all annotated proteins of the *C. orbicularis* symbiont's draft genome with hmmsearch (Expect value 1e-10) from HMMER version 3.1b2 (ref. 67). Its sequence was subsequently searched with HMMER's phmmer (Expect value 1e-10) against a local version of NCBI's nr database (version 2015-11-17), and the top 100 hits, that is, the 100 most closely related NifH sequences from other organisms, were retrieved. Additional (partial) NifH sequences of bacterial isolates from tropical seagrass beds[42], oligotrophic open ocean[68], salt marsh ecosystems[69,70], mangrove root sediments[71,72] and sulfidic cave water[44] (1,007 sequences in total) were included. Some of these NifH sequences[71] were only available as nucleotide sequences and were therefore translated using EMBOSS Transeq[73] and the correct reading frame was selected manually. Four additional NifH protein sequences from known thiotrophs and two sequences from the outgroup taxa *Anabaena variabilis* ATCC 29413 (NifH: AAA93020) and *Frankia alni* ArI3 (NifH: AAA96262) complemented the data set, giving a total of 1,114 sequences for the NifH-based phylogenetic analysis. Sequence entries from the PDB and PRF databases as well as sequences containing stop codons or ambiguous amino acids (X) were removed, leaving 1,073 sequences that were clustered at 100% identity with CD-HIT (refs 74,75) to remove redundancies. All 834 representative non-redundant sequences (93 full length, 741 partial length) were automatically aligned with MAFFT version 7.221 (2014/04/16)[76] using the L-INS-i algorithm and the 'leave gappy region' option. The alignment was imported into ARB version 5.5 (ref. 77) using a custom filter, and a protein database was created. A total of 55 full-length amino-acid sequences were selected and exported. ProtTest version 3 (refs 78,79) was employed to select the best-fit model of amino-acid replacement from 120 models (15 matrices: +G, +I or + G+I; +F) with a starting topology based on maximum likelihood. As the Bayesian information criterion (BIC; also known as the Schwarz criterion, or SC)[80], the corrected Akaike information criterion (AICc)[81,82] and the decision theory framework (DT)[83] favoured LG+I+G, the phylogeny was then inferred using RAxML version 8.2.4 (ref. 84) with the LG amino acid matrix[85], a gamma model of rate heterogeneity and an estimate of proportion of invariable sites. The best tree was chosen from 1,000 independent inferences that were executed on the original alignment (169 alignment patterns) using 1,000 distinct randomized maximum parsimony (MP) trees and was imported into ARB version 6.0.3. All 741 partial length NifH sequences were added by ARB's interactive parsimony. Sequences that could not be unambiguously inserted at a specific position in the tree were removed. To identify the partial length NifH sequences with the highest identity to the full-length sequences, a BLAST+ (ref. 86) protein database of all partial sequences was created and the full-length sequences that had partial sequences assigned by parsimony were searched against this database using BlastP version 2.2.31+. For each full-length sequence, only the best matching partial sequence (that is, the one with highest identity) was kept in the tree, and only if sequence identity between both was 95% or higher. The final tree (Fig. 5) thus comprises the *C. orbicularis* symbiont's NifH, 54 full-length NifH sequences of cultured organisms and 21 partial NifH sequences mostly from uncultured clones.

NifDK phylogeny. The NifD and NifK trees were constructed based on maximum likelihood as described above and include the respective amino acid sequences of the *C. orbicularis* symbiont and of the same cultured organisms that are displayed in the NifH tree, but with two exceptions: *Methylogaea oryzae* JCM 16910 is not part of the NifK tree and *Agarivorans gilvus* WH0801 could not be included in the NifD tree because the respective genes (locus tags JCM16910_RS02630 and AR383_RS01715) are disrupted. *Anabaena variabilis* ATCC 29413 (NifD: AAA93021; NifK: AAA93022) and *Frankia alni* ArI3 (NifD: AAA96263; NifK: AAA96264) were used as outgroup taxa. Note that the uncultured clones from seagrass and mangrove sediments (dotted branches in the NifH tree) are not included in the NifD and NifK trees, because no NifD or NifK sequence information was available for these isolates.

16S rRNA phylogeny. The 16S rRNA-coding region of the '*Ca*. T. endolucinida' genome was identified using barrnap version 0.6 (available from https://github.com/tseemann/barrnap) and automatically aligned against the SILVA SSU Ref NR 99 Release 123 database (available from http://www.arb-silva.de)[87] using the SILVA Incremental Aligner (SINA) version 1.2.11 (ref. 88). The SINA-aligned sequence was imported into the ARB software package version 5.5 (ref. 77). Besides the *C. orbicularis* symbiont's 16S ribosomal RNA nucleotide sequence and its closest homologues (by NCBI blastn), the data set contained additional 16S rRNA sequences of bivalve and gastropod symbionts, free-living sulfur oxidizers and some diazotrophic bacteria, whose NifH sequences were also included in the NifH tree. The alignment of these 58 sequences was manually refined taking into account the secondary structure information of the rRNA. Phylogenetic reconstruction was performed using a maximum likelihood method. The final tree was calculated with RAxML version 8.2.4 (GTRGAMMA model)[84] and was based on an alignment with 750 distinct alignment patterns. The best tree from 1,000 independent inferences using 1,000 distinct randomized maximum parsimony trees is presented. *Frankia alni* ACN14a (CT573213) was used as outgroup.

Accession codes and data availability. Sequence data that support the findings of this study, that is, the '*Ca*. T. endolucinida' Whole Genome Shotgun project, have been deposited at DDBJ/ENA/GenBank under project accession no. MARB00000000. The version described in this paper is version MARB01000000 (NCBI BioSample SAMN03435122, BioProject PRJNA284177). All nitrogen metabolism-related protein identifications obtained in this study are available in Supplementary Table 4. Additional data (for example, MS raw data) are available on request from the corresponding author (S.Ma.).

Received 14 March 2016; accepted 7 September 2016; published 24 October 2016

References

1. Cavanaugh, C. M., McKiness, Z. P., Newton, I. L. & Stewart, F. J. in *The Prokaryotes: Symbiotic Associations, Biotechnology, Applied Microbiology* Vol. 1 (ed. Dworkin, M. M.) 475–507 (Springer, 2006).
2. Fisher, C. R. Chemoautotrophic and methanotrophic symbioses in marine invertebrates. *Rev. Aquat. Sci.* **2,** 399–436 (1990).
3. Felbeck, H. & Distel, D. L. in *The Prokaryotes* Vol. 4 (eds Balows, A. *et al.*) 3891–3906 (Springer, 1992).
4. Taylor, J. D. & Glover, E. A. Lucinidae (Bivalvia)—the most diverse group of chemosymbiotic molluscs. *Zool. J. Linn. Soc.* **148,** 421–438 (2006).
5. Abbott, R. T. *American Seashells* 2nd edn (Van Nostrand Reinhold, 1974).
6. Berg, C. J. & Alatalo, P. Potential of chemosynthesis in molluscan mariculture. *Aquaculture* **39,** 165–179 (1984).
7. Gros, O., Darrasse, A., Durand, P., Frenkiel, L. & Moueza, M. Environmental transmission of a sulfur-oxidizing bacterial gill endosymbiont in the tropical lucinid bivalve *Codakia orbicularis*. *Appl. Environ. Microbiol.* **62,** 2324–2330 (1996).
8. Gros, O., Frenkiel, L. & Mouëza, M. Embryonic, larval, and post-larval development in the symbiotic clam *Codakia orbicularis* (Bivalvia: Lucinidae). *Invertebr. Biol.* **116,** 86–101 (1997).
9. Gros, O., Liberge, M., Heddi, A., Khatchadourian, C. & Felbeck, H. Detection of the free-living forms of sulfide-oxidizing gill endosymbionts in the lucinid habitat (*Thalassia testudinum* environment). *Appl. Environ. Microbiol.* **69,** 6264–6267 (2003).
10. Caro, A., Got, P., Bouvy, M., Troussellier, M. & Gros, O. Effects of long-term starvation on a host bivalve (*Codakia orbicularis*, Lucinidae) and its symbiont population. *Appl. Environ. Microbiol.* **75,** 3304–3313 (2009).
11. Brissac, T., Gros, O. & Merçot, H. Lack of endosymbiont release by two Lucinidae (Bivalvia) of the genus *Codakia*: consequences for symbiotic relationships. *FEMS Microbiol. Ecol.* **67,** 261–267 (2009).
12. König, S., Le Guyader, H. & Gros, O. 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 (2015).
13. Caro, A., Gros, O., Got, P., De Wit, R. & Troussellier, M. Characterization of the population of the sulfur-oxidizing symbiont of *Codakia orbicularis* (Bivalvia, Lucinidae) by single-cell analyses. *Appl. Environ. Microbiol.* **73,** 2101–2109 (2007).
14. Welsh, D. T. Nitrogen fixation in seagrass meadows: regulation, plant–bacteria interactions and significance to primary productivity. *Ecol. Lett.* **3,** 58–71 (2000).
15. Raymond, J., Siefert, J. L., Staples, C. R. & Blankenship, R. E. The natural history of nitrogen fixation. *Mol. Biol. Evol.* **21,** 541–554 (2004).
16. Kneip, C., Lockhart, P., Voß, C. & Maier, U.-G. Nitrogen fixation in eukaryotes—new models for symbiosis. *BMC Evol. Biol.* **7,** 55 (2007).
17. Pawlowski, K., Ribeiro, A. & Bisseling, T. in *Biotechnol. Ann. Rev.* Vol. 2 (ed. Raafat El-Gewely, M.) 151–184 (Elsevier, 1996).
18. Santi, C., Bogusz, D. & Franche, C. Biological nitrogen fixation in non-legume plants. *Ann. Bot. (Lond.)* **111,** 743–767 (2013).
19. Distel, D. L., Beaudoin, D. J. & Morrill, W. Coexistence of multiple proteobacterial endosymbionts in the gills of the wood-boring Bivalve *Lyrodus pedicellatus* (Bivalvia: Teredinidae). *Appl. Environ. Microbiol.* **68,** 6292–6299 (2002).
20. Fiore, C. L., Jarett, J. K., Olson, N. D. & Lesser, M. P. Nitrogen fixation and nitrogen transformations in marine symbioses. *Trends Microbiol.* **18,** 455–463 (2010).
21. Riemann, L., Farnelid, H. & Steward, G. F. Nitrogenase genes in non-cyanobacterial plankton: prevalence, diversity and regulation in marine waters. *Aquat. Microb. Ecol.* **61,** 235–247 (2010).
22. Ingemann Jensen, S., Kühl, M. & Priemé, A. Different bacterial communities associated with the roots and bulk sediment of the seagrass *Zostera marina*. *FEMS Microbiol. Ecol.* **62,** 108–117 (2007).
23. Crump, B. C. & Koch, E. W. Attached bacterial populations shared by four species of aquatic angiosperms. *Appl. Environ. Microbiol.* **74,** 5948–5957 (2008).
24. Jørgensen, B. B. Mineralization of organic matter in the sea bed—the role of sulphate reduction. *Nature* **296,** 643–645 (1982).
25. Küsel, K., Trinkwalter, T., Drake, H. L. & Devereux, R. Comparative evaluation of anaerobic bacterial communities associated with roots of submerged macrophytes growing in marine or brackish water sediments. *J. Exp. Mar. Biol. Ecol.* **337,** 49–58 (2006).
26. van der Heide, T. *et al.* A three-stage symbiosis forms the foundation of seagrass ecosystems. *Science* **336,** 1432–1434 (2012).
27. Merrick, M. J. & Edwards, R. A. Nitrogen control in bacteria. *Microbiol. Rev.* **59,** 604–622 (1995).
28. Kurtz, D. M. Jr. Avoiding high-valent iron intermediates: superoxide reductase and rubrerythrin. *J. Inorg. Biochem.* **100,** 679–693 (2006).

29. Zhao, W., Ye, Z. & Zhao, J. RbrA, a cyanobacterial rubrerythrin, functions as a FNR-dependent peroxidase in heterocysts in protection of nitrogenase from damage by hydrogen peroxide in *Anabaena* sp. PCC 7120. *Mol. Microbiol.* **66,** 1219–1230 (2007).
30. Myrold, D. *Quantification of Nitrogen Transformations* 3rd edn (ASM, 2007).
31. Frenkiel, L. & Mouëza, M. Gill ultrastructure and symbiotic bacteria in *Codakia orbicularis* (Bivalvia, Lucinidae). *Zoomorphology* **115,** 51–61 (1995).
32. Robidart, J. C. *et al.* Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed through metagenomics. *Environ. Microbiol.* **10,** 727–737 (2008).
33. Gardebrecht, A. *et al.* Physiological homogeneity among the endosymbionts of *Riftia pachyptila* and *Tevnia jerichonana* revealed by proteogenomics. *ISME J.* **6,** 766–776 (2012).
34. Jørgensen, N. O. G., Kroer, N., Coffin, R. B., Yang, X.-H. & Lee, C. Dissolved free amino acids, combined amino acids, and DNA as sources of carbon and nitrogen to marine bacteria. *Mar. Ecol. Prog. Ser.* **98,** 135–148 (1993).
35. Mulholland, M. R. & Lomas, M. W. in *Nitrogen in the Marine Environment* 2nd edn (eds Capone, D. G., Bronk, D. A., Mulholland, M. R. & Carpenter, E. J.) Ch. 7, 303–384 (Academic, 2008).
36. Touchette, B. W. & Burkholder, J. M. Review of nitrogen and phosphorus metabolism in seagrasses. *J. Exp. Mar. Biol. Ecol.* **250,** 133–167 (2000).
37. Fourqurean, J. W., Zieman, J. C. & Powell, G. V. N. Relationships between porewater nutrients and seagrasses in a subtropical carbonate environment. *Mar. Biol.* **114,** 57–65 (1992).
38. Lee, K.-S. & Dunton, K. H. Inorganic nitrogen acquisition in the seagrass *Thalassia testudinum*: development of a whole-plant nitrogen budget. *Limnol. Oceanogr.* **44,** 1204–1215 (1999).
39. Williams, S. L. Experimental studies of Caribbean seagrass bed development. *Ecol. Monogr.* **60,** 449–469 (1990).
40. Masepohl, B. & Forchhammer, K. in *Biology of the Nitrogen Cycle* 131–145 (Elsevier, 2007).
41. Dixon, R. & Kahn, D. Genetic regulation of biological nitrogen fixation. *Nat. Rev. Microbiol.* **2,** 621–631 (2004).
42. Bagwell, C. E. *et al.* Molecular diversity of diazotrophs in oligotrophic tropical seagrass bed communities. *FEMS Microbiol. Ecol.* **39,** 113–119 (2002).
43. Green-García, A. M. & Engel, A. S. Bacterial diversity of siliciclastic sediments in a *Thalassia testudinum* meadow and the implications for *Lucinisca nassula* chemosymbiosis. *Estuar. Coast. Shelf Sci.* **112,** 153–161 (2012).
44. Desai, M. S., Assig, K. & Dattagupta, S. Nitrogen fixation in distinct microbial niches within a chemoautotrophy-driven cave ecosystem. *ISME J.* **7,** 2411–2423 (2013).
45. Middelburg, J. J. *et al.* Discovery of symbiotic nitrogen fixation and chemoautotrophy in cold-water corals. *Sci. Rep.* **5,** 17962 (2015).
46. Green, L. C. *et al.* Analysis of nitrate, nitrite, and [^{15}N]nitrate in biological fluids. *Anal. Biochem.* **126,** 131–138 (1982).
47. Schnetger, B. & Lehners, C. Determination of nitrate plus nitrite in small volume marine water samples using vanadium(III) chloride as a reduction agent. *Mar. Chem.* **160,** 91–98 (2014).
48. Solórzano, L. Determination of ammonia in natural waters by the phenolhypochlorite method. *Limnol. Oceanogr.* **14,** 799–801 (1969).
49. Elisabeth, N. H. *et al.* Comparative modifications in bacterial gill–endosymbiotic populations of the two bivalves *Codakia orbiculata* and *Lucina pensylvanica* during bacterial loss and reacquisition. *FEMS Microbiol. Ecol.* **89,** 646–658 (2014).
50. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).
51. Boisvert, S., Laviolette, F. & Corbeil, J. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J. Comput. Biol.* **17,** 1519–1533 (2010).
52. Laczny, C. C. *et al.* VizBin—an application for reference-independent visualization and human-augmented binning of metagenomic data. *Microbiome* **3,** 1 (2015).
53. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19,** 455–477 (2012).
54. Strous, M., Kraft, B., Bisdorf, R. & Tegetmeyer, H. E. The binning of metagenomic contigs for microbial physiology of mixed cultures. *Front. Microbiol.* **3,** 410 (2012).
55. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29,** 1072–1075 (2013).
56. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30,** 2068–2069 (2014).
57. Guy, L., Roat Kultima, J. & Andersson, S. G. E. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26,** 2334–2335 (2010).
58. Eymann, C. *et al.* A comprehensive proteome map of growing *Bacillus subtilis* cells. *Proteomics* **4,** 2849–2876 (2004).
59. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal. Biochem.* **72,** 248–254 (1976).
60. Xing, P. *et al.* Niches of two polysaccharide-degrading *Polaribacter* isolates from the North Sea during a spring diatom bloom. *ISME J.* **9,** 1410–1422 (2015).
61. Heinz, E. *et al.* The genome of the obligate intracellular parasite *Trachipleistophora hominis*: new insights into microsporidian genome dynamics and reductive evolution. *PLoS Pathog.* **8,** e1002979 (2012).
62. Muntel, J., Hecker, M. & Becher, D. An exclusion list based label-free proteome quantification approach using an LTQ Orbitrap. *Rapid Commun. Mass Spectrom.* **26,** 701–709 (2012).
63. Huja, S. *et al.* Fur is the master regulator of the extraintestinal pathogenic *Escherichia coli* response to serum. *mBio* **5,** e01460 (2014).
64. Silva, J. C., Gorenstein, M. V., Li, G.-Z., Vissers, J. P. C. & Geromanos, S. J. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol. Cell. Proteomics* **5,** 144–156 (2006).
65. Luna, L. G. *Histopathologic Methods and Color Atlas of Special Stains and Tissue Artifacts* 1st edn (American Histolabs, 1992).
66. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44,** D279–D285 (2016).
67. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comp. Biol.* **7,** e1002195 (2011).
68. Zehr, J. P., Mellon, M. T. & Zani, S. New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (*nifH*) genes. *Appl. Environ. Microbiol.* **64,** 3444–3450 (1998).
69. Lovell, C. R., Piceno, Y. M., Quattro, J. M. & Bagwell, C. E. Molecular analysis of diazotroph diversity in the rhizosphere of the smooth cordgrass, *Spartina alterniflora*. *Appl. Environ. Microbiol.* **66,** 3814–3822 (2000).
70. Lovell, C. R. & Davis, D. A. Specificity of salt marsh diazotrophs for vegetation zones and plant hosts: results from a North American marsh. *Front. Microbiol.* **3,** 84 (2012).
71. Flores-Mireles, A. L., Winans, S. C. & Holguin, G. Molecular characterization of diazotrophic and denitrifying bacteria associated with mangrove roots. *Appl. Environ. Microbiol.* **73,** 7308–7321 (2007).
72. Dias, A. C. F. *et al.* Abundance and genetic diversity of *nifH* gene sequences in anthropogenically affected Brazilian mangrove sediments. *Appl. Environ. Microbiol.* **78,** 7960–7967 (2012).
73. Rice, P., Longden, I. & Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **16,** 276–277 (2000).
74. Li, W. & Godzik, A. CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22,** 1658–1659 (2006).
75. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28,** 3150–3152 (2012).
76. Katoh, K., Misawa, K., Kuma, K.-i. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30,** 3059–3066 (2002).
77. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32,** 1363–1371 (2004).
78. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52,** 696–704 (2003).
79. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27,** 1164–1165 (2011).
80. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **6,** 461–464 (1978).
81. Sugiura, N. Further analysts of the data by Akaike's information criterion and the finite corrections. *Commun. Stat. A* **7,** 13–26 (1978).
82. Hurvich, C. M. & Tsai, C.-L. Regression and time series model selection in small samples. *Biometrika* **76,** 297–307 (1989).
83. Minin, V., Abdo, Z., Joyce, P. & Sullivan, J. Performance-based selection of likelihood models for phylogeny estimation. *Syst. Biol.* **52,** 674–683 (2003).
84. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–1313 (2014).
85. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25,** 1307–1320 (2008).
86. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10,** 421 (2009).
87. Yilmaz, P. *et al.* The SILVA and 'All-species Living Tree Project (LTP)' taxonomic frameworks. *Nucleic Acids Res.* **42,** D643–D648 (2013).
88. Pruesse, E., Peplies, J. & Glöckner, F. O. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28,** 1823–1829 (2012).

Acknowledgements

The authors acknowledge the technical assistance of J. Matulla and S. Grund and thank K. Riedel, L. Westhoff and M. Bengtson for discussions. This work was supported by the EU in the framework of the Marie Curie ITN 'Symbiomics' project (no. 264774) and by the German Science Foundation (DFG project no. MA6346/2-1). O.G.'s group was partially supported by the French National Program EC2CO Biohefect/Ecodyn/Dril/MicrobiEn 'Fixation de l'azote chez les bivalves marins de la famille des Lucinidae: impact sur le fonctionnement de ces organismes et des écosystèmes tropicaux côtiers associés'.

Author contributions

S.K. and O.G. planned and performed the experiments with *C. orbicularis*. Sediment chemistry analyses were conducted by J.T., western blots and immunohistochemistry by M.V., and acetylene reduction measurements by O.G. and D.M.-M. Proteome analyses were planned by S.K., T.S., S.Ma. and D.B., and conducted by S.K., S.Me. and D.B. A.T., A.P. and R.D. performed the genome sequencing and submitted the sequences to NCBI. S.K. and S.E.H. assembled and annotated the genome. S.E.H. performed the phylogenetic analyses and gene cluster comparisons. O.G., T.S. and S.Ma. coordinated and supervised the laboratory work. S.K. and S.Ma. analysed the proteome data. S.Ma. wrote the

manuscript. T.H. was involved in manuscript writing, figure generation and, together with R.P., in MS data evaluation.

Additional information

Supplementary information is available for this paper.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to S.Ma.

How to cite this article: König, S. *et al.* Nitrogen fixation in a chemoautotrophic lucinid symbiosis. *Nat. Microbiol.* **2,** 16193 (2016).

Competing interests

The authors declare no competing financial interests.

PUBLICATION VI

The book chapter “Centrifugation-based enrichment of bacterial cell populations for metaproteomic studies on bacteria-invertebrate symbioses” gives a comprehensive overview and in-depth method description of centrifugation-based enrichment of morphologically different cell populations.

Chapter 22

Centrifugation-Based Enrichment of Bacterial Cell Populations for Metaproteomic Studies on Bacteria–Invertebrate Symbioses

Tjorven Hinzke, Manuel Kleiner, and Stephanie Markert

Abstract

Owing to high sample complexity, metaproteomic investigations on bacteria–animal symbioses with two or more uncultured partners can be challenging. A selective isolation or enrichment of distinct (sub-)populations within those consortia can solve this problem. Subsequent discrete proteomic analyses benefit from increased sample purity and higher proteome coverage for each of the individual organisms. Here, we describe centrifugation-based methods that allow for a separation of the host and its bacterial symbiont population(s), or even for an enrichment of distinct symbiotic cell cycle stages in the deep-sea mussels *Bathymodiolus azoricus* and *B. thermophilus*, the gutless oligochaete *Olavius algarvensis* and the deep-sea tube worm *Riftia pachyptila*, respectively.

Key words Bacteria–invertebrate symbiosis, Metaproteome, Differential pelleting, Rate-zonal density gradient centrifugation, Isopycnic density gradient centrifugation, Percoll®, HistoDenz™

1 Introduction

The study of symbioses and especially of symbiotic bacteria, most of which have not been cultivated to date, often requires a reduction of sample complexity in order to maximize gain of information: The number of expressed proteins from all partners is comparatively high and host proteins are usually massively overrepresented. Consequently, a separation of the partners is necessary to be able to detect specific molecules or functions and to attribute them to one of the organisms involved. As symbiont-bearing animals often harbor more than one bacterial species or several morphotypes of one species (e.g., [1, 2]), it can be advantageous to not only separate host from symbionts but also different symbiont populations from each other for further studies, such as proteomic analyses.

Several culture-independent methods can in principle be exploited to separate or enrich certain types of bacterial cells from

Dörte Becher (ed.), *Microbial Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1841,
https://doi.org/10.1007/978-1-4939-8695-8_22,

their surroundings, such as centrifugation, filtration, dilution, micromanipulation, use of optical tweezers and flow cytometry [3].

Centrifugation as one of these possibilities is comparatively straightforward and easily applicable even in the field, as it does not require high-end equipment. Moreover, it is fast and allows for relatively high quantities of sample to be treated at one time. A large range of centrifugation methods is available to separate cells and even subcellular compartments (for a comprehensive overview, see for example [4–10]). The most basic and fastest centrifugation procedure is differential pelleting (Fig. 1a). Here, particles are separated predominantly based on size and, to some extent, on density. Larger/denser particles sediment faster than smaller/less dense ones, which means that the latter will mostly remain in the supernatant if *g*-force and centrifugation duration are chosen

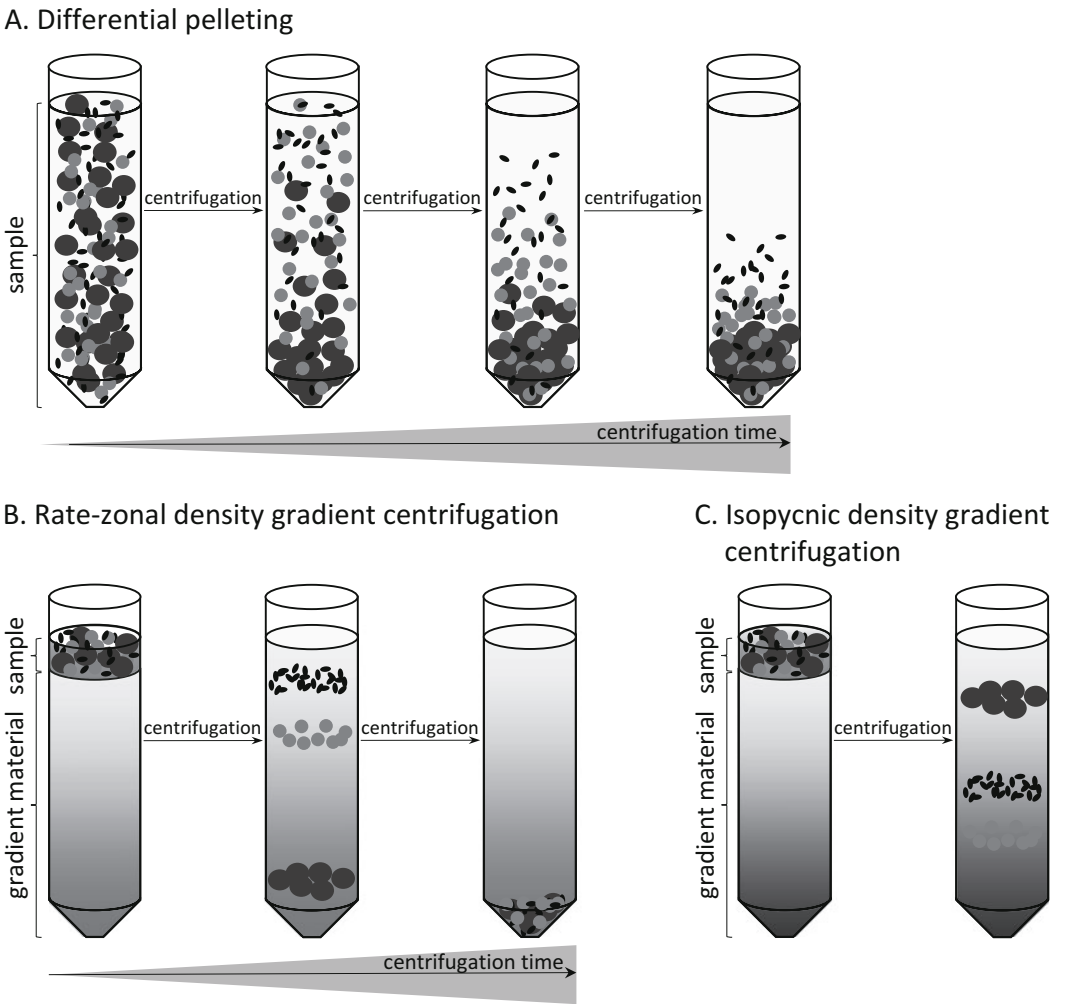

Fig. 1 Overview of centrifugation-based cell separation methods. (**a**) Differential pelleting, (**b**) Rate-zonal density gradient centrifugation, (**c**) Isopycnic density gradient centrifugation

appropriately (although some of them will sediment as well, contaminating the pellet). Thus, differential pelleting leads to a relatively crude separation of cells. To increase separation resolution, which is especially desirable for more complex separation tasks, density gradients can be employed. Two general types of density gradient-based methods can be distinguished and used according to the specific experimental question [4–12]: On the one hand, a rate-zonal centrifugation allows for separation of particles of similar density based on their size and shape, which determine their respective sedimentation speeds while moving through a shallow density gradient (Fig. 1b). During relatively short centrifugation times at low speed, smaller (i.e., slower-moving) particles accumulate in the upper part of the gradient tube, and larger (i.e., faster-moving) particles in the lower part of the tube. As for differential pelleting, all particles will pellet eventually, depending on centrifugation time and *g*-force. Particles with different densities, on the other hand, can be separated by isopycnic density gradient centrifugation. During comparatively longer centrifugation times at high speed, the particles collect in the gradient zone that matches their own density (Fig. 1c).

All centrifugation methods mentioned above have been successfully used for enrichment and separation of symbionts from marine bacteria–animal symbioses and subsequent proteomic analyses. In order to perform these experiments, whole animals (in the case of *Olavius algarvensis*) or their symbiont-containing tissues (i.e., the gills of *Bathymodiolus azoricus* and *B. thermophilus* and the trophosome of *Riftia pachyptila*) were homogenized and the homogenized tissue was subjected to one of the centrifugation methods described below (*see* overview in Table 1).

Table 1
Summary of presented centrifugation methods and previous applications

Method	Application and symbiotic system	For method description *see* Subheading
Differential pelleting	Separation of host from symbiont(s) in *Bathymodiolus azoricus* [16] and *B. thermophilus* (Ponnudurai et al., in preparation)	3.2
Isopycnic density gradient centrifugation	Separation of host and symbiont in *Riftia pachyptila* (Percoll® cushion) [14, 15]	3.3
	Enrichment of multiple symbionts in *Olavius algarvensis* [1]	3.5
Rate-zonal density gradient centrifugation	Enrichment of *Riftia pachyptila* symbiont morphotypes (Hinzke et al., in preparation)	3.4

Based on a protocol developed by Distel and Felbeck [13], a separation of host tissue and the total monospecific bacterial symbiont population of the deep sea tube worm *R. pachyptila* was obtained via isopycnic density gradient centrifugation using the density gradient medium Percoll® [14, 15].

Beyond that, an enrichment of individual life-stage dependent morphotypes of the *R. pachyptila* symbiont was recently achieved by rate-zonal density gradient centrifugation with HistoDenz™ as a density gradient medium (Fig. 2).

In recent studies with the deep-sea mussels *Bathymodiolus azoricus* and *B. thermophilus*, host and symbionts were successfully separated by differential pelleting ([16], Ponnudurai et al., in preparation). Furthermore, an effective enrichment of different symbiont species residing in the marine oligochaete *Olavius algarvensis* was achieved by means of isopycnic density gradient centrifugation [1].

If necessary symbiont enrichment resulting from the respective centrifugation approaches was analyzed and verified via fluorescence in situ hybridization (FISH), allowing for the selection of samples for subsequent proteomic analyses [1, 16, Hinzke et al., in preparation, Ponnudurai et al., in preparation].

Here, we present a compendium of these centrifugation methods to be used for bacterial symbiont enrichment and separation (*see* Table 1).

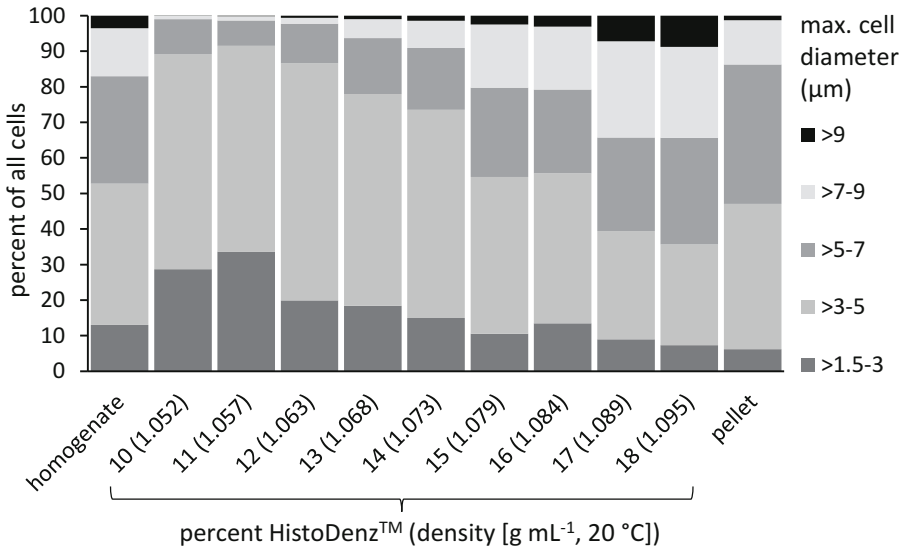

Fig. 2 Distribution of *Riftia pachyptila* symbiont cell sizes in trophosome homogenate (leftmost column) and in HistoDenz fractions of increasing concentrations after rate-zonal density gradient centrifugation (Hinzke et al., in preparation). Smaller symbiont cells (with a diameter of up to 5 m) are enriched in the less dense fractions (left columns), while larger cells of up to >9 m accumulate in the fractions with higher densities (right columns)

2 Materials

All solutions are prepared using deionized water and analytical grade reagents. Reagents are prepared at room temperature and stored at 4 °C.

2.1 Homogenization of Tissue Samples

1. 2.5× Imidazole-buffered saline (IBS): 1.225 M NaCl, 0.075 M $MgSO_4$, 0.0275 M $CaCl_2$, 0.0075 M KCl, 0.125 M imidazole, ph 7.1 [13]. Add about 300 mL water in a beaker. Weigh 28.64 g NaCl, 6.85 g $MgSO_4 \cdot 6\ H_2O$, 1.62 g $CaCl_2 \cdot 2\ H_2O$, 0.224 g KCl, 3.4 g imidazole and add reagents subsequently to the beaker. Allow each reagent to dissolve before adding the next. Adjust pH with HCl to 7.1. Fill up to 400 mL with water (*see* **Note 1**).
2. IBS: 0.49 M NaCl, 0.03 M $MgSO_4$, 0.011 M $CaCl_2$, 0.003 M KCl, 0.05 M imidazole, ph 7.1 [13]. Mix 400 mL of 2.5× IBS with 600 mL water.
3. Homogenizer: Use a 15 mL Dounce glass tissue grinder with a small and a large pestle, or a 15 mL Duall glass tissue grind tube and pestle (*see* **Note 2**).
4. Optional: rough nylon mesh (*see* **Note 3**).
5. Material for FISH analyses (*see* Subheading 2.6).

2.2 Differential Pelleting

1. IBS: *see* **items 1** and **2** in Subheading 2.1.
2. Refrigerated centrifuge with swing-out rotor for 15 mL tubes and fixed-angle rotor for 15 mL tubes (optional: fixed-angle rotor for 2 mL tubes).
3. Material for FISH analyses (*see* Subheading 2.6).

2.3 Isopycnic Gradient Centrifugation with PERCOLL® for Simple Separation of Symbionts from Host Tissue (Percoll® Cushion)

1. IBS: *see* **items 1** and **2** in Subheading 2.1.
2. 60% Percoll®: Mix 60 mL Percoll® with 40 mL of 2.5× IBS. Prepare 10 mL aliquots in 50 mL tubes. Store at 4 °C.
3. Tris–EDTA buffer (TE buffer): 10 mM Tris–HCl, 1 mM EDTA; pH 7.5. Prepare a 1 M stock of Tris–HCl by adding 60.6 g Tris to 400 mL water, adjusting pH with HCl to 7.5 and filling up to 500 mL with water. Additionally, prepare a 0.5 M EDTA stock by adding 18.6 g EDTA-$Na_2 \cdot 2H_2O$ to 80 mL water, adjusting pH to 8.0 with NaOH and filling up to 100 mL with water (*see* **Note 4**). Make TE buffer from stock solutions by adding 10 mL 1 M Tris–HCl (pH 7.5) and 2 mL 0.5 M EDTA (pH 8.0) to 950 mL water and filling up to 1 L with water.
4. 10 mL syringe with syringe needles (approx. 0.9 mm diameter, 18–20 G)

5. Refrigerated centrifuge with fixed-angle rotor for 50 mL tubes and fixed-angle rotor for 2 mL tubes (optional: fixed-angle rotor for 15 mL tubes, depending on sample volume).

2.4 HistoDenz™ Gradients for Rate-Zonal Density Gradient Centrifugation

1. IBS: *see* **items 1** and **2** in Subheading 2.1.
2. 60% (w/v) HistoDenz™ stock solution: Weigh 12 g HistoDenz™ (synonyms: Nycodenz®, iohexol) and transfer it into a 50 mL tube. Fill up to a total volume of 20 mL with IBS. Mix thoroughly by vortexing and allow for the HistoDenz™ to dissolve. Add more IBS if necessary to keep the total volume at 20 mL. The stock solution can be kept at 4 °C for several days.
3. HistoDenz™ gradients for rate zonal centrifugation (8–18%, Δ 1%): Use the 60% HistoDenz™ stock solution and IBS to prepare 7 mL of each HistoDenz™ dilution in 15 mL tubes (sufficient for six gradients; *see* Table 2 for pipetting scheme). Vortex each dilution thoroughly. HistoDenz™ dilutions can be stored at 4 °C for several days. Assemble the gradient from bottom (18%) to top (8%), starting with the highest concentration. Use 1 mL of each dilution step. Freeze gradient after assembly (*see* **Note 5**).
4. TE buffer: *see* **item 3** in Subheading 2.3.

Table 2
Pipetting scheme for dilutions to prepare 8–18% HistoDenz™ gradients for rate-zonal density gradient centrifugation

	For six gradients	
Percent HistoDenz™ in IBS (w/v)	60% HistoDenz™ stock solution (mL)	IBS (mL)
18	2.10	4.90
17	1.98	5.02
16	1.87	5.13
15	1.75	5.25
14	1.63	5.37
13	1.52	5.48
12	1.40	5.60
11	1.28	5.72
10	1.17	5.83
9	1.05	5.95
8	0.93	6.07

5. Refrigerated centrifuge with swing-out rotor for 15 mL tubes and fixed-angle rotor for 2 mL tubes.
6. Material for FISH analyses (*see* Subheading 2.6).

2.5 HistoDenz™ Gradients for Isopycnic Density Gradient Centrifugation

1. IBS: *see* **items 1** and **2** in Subheading 2.1.
2. 60% (w/v) HistoDenz™ stock solution: *see* **item 2** in Subheading 2.4.
3. 70% (w/v) HistoDenz™ stock solution: Weigh 14 g HistoDenz™ and transfer it into a 50 mL tube. Fill up to a total volume of 20 mL with IBS. Mix thoroughly by vortexing and allow for the HistoDenz™ to dissolve. Add more IBS if necessary to keep the total volume at 20 mL. The stock solution can be kept at 4 °C for several days.
4. HistoDenz™ gradients for isopycnic centrifugation (10–70%): Use the 60% HistoDenz™ stock solution and IBS to prepare 3.5 mL of each of the 60–20% HistoDenz™ dilutions and 2 mL each of the 15% and 10% HistoDenz™ dilutions in 15 mL tubes (sufficient for five gradients; *see* Table 3 for pipetting scheme). Vortex thoroughly. HistoDenz™ dilutions can be stored at 4 °C for several days. Assemble the gradient by adding 1.5 mL of 70% HistoDenz™ first, then add 0.6 mL of

Table 3
Pipetting scheme for dilutions to prepare 10–70% HistoDenz™ gradients for isopycnic density gradient centrifugation

	For five gradients		
Percent HistoDenz™ in IBS (w/v)	70% HistoDenz™ stock solution (mL)	60% HistoDenz™ stock solution (mL)	IBS (mL)
70	8.0	0	0
60	0	3.50	0
50	0	2.92	0.58
40	0	2.33	1.17
30	0	1.75	1.75
28	0	1.63	1.87
26	0	1.52	1.98
24	0	1.40	2.10
22	0	1.28	2.22
20	0	1.17	2.33
15	0	0.50	1.50
10	0	0.33	1.67

each of the subsequent 60–20% dilution steps (*see* **Note 6**). Use 0.3 mL of the 15% and the 10% dilution step (*see* **Note 7**).

5. TE buffer: *see* **item 3** in Subheading 2.3.
6. Refrigerated centrifuge with swing-out rotor for 15 mL tubes and fixed-angle rotor for 2 mL tubes.
7. Material for FISH analyses (*see* Subheading 2.6).

2.6 Preparation of Subsamples for FISH Analyses (See Note 8)

1. IBS, containing 1% paraformaldehyde (PFA): Prepare 2.5× IBS (*see* **item 1** in Subheading 2.1). Add 240 mL of 2.5× IBS, 75 mL of 8% PFA and fill up with water to 600 mL. Prepare 1 mL aliquots in tubes. The solution can be stored at 4 °C for several weeks.
2. IBS, containing 2% PFA: Prepare 2.5× IBS (*see* **item 1** in Subheading 2.1). Add 4 mL of 2.5× IBS, 2.5 mL of 8% PFA and fill up with water to 10 mL. Prepare 0.5 mL aliquots in tubes. The solution can be stored at 4 °C for several weeks.
3. Optional: Isopore™ polycarbonate, track-etched hydrophilic screen filter membranes (25 mm) with 0.2 μm pore size and filter apparatus.

3 Methods

All procedures are carried out at 4 °C or on ice.

3.1 Homogenization of Tissue Samples

1. Fill a Dounce homogenizer or a Duall homogenizer with 6 mL of prechilled (4 °C) IBS (or PBS, *see* **Note 1**). Add approximately 3 g of tissue. Alternatively, immerse sample in IBS so that it is just completely covered (*see* **Note 9**).
2. In the case of a Dounce homogenizer, use the smaller pestle first to grind the material with 5–6 strokes. Afterward, repeat with the second, larger pestle. For a Duall homogenizer, homogenize for approximately 1 min or about 20 strokes (*see* **Note 10**).
3. If necessary, filter the homogenate through a rough nylon mesh (*see* **Note 11**) into a 15 or 50 mL tube (depending on sample volume).
4. Take a 15 μL subsample of (filtered) homogenate for FISH analysis (*see* Subheading 3.6), to evaluate the enrichment procedure.

3.2 Differential Pelleting

1. Precool a refrigerated centrifuge with a swing-out rotor for 15 mL tubes to 4 °C. Additionally, precool a fixed-angle rotor for 15 mL tubes. (Optional: precool a fixed-angle rotor for 2 mL tubes).
2. Fill up tissue homogenate with IBS to a total volume of 15 mL.

3. Centrifuge tissue homogenate in swing-out rotor for 5 min at 500 × *g* and 4 °C (*see* **Notes 12** and **13**).
4. Transfer supernatant into a new 15 mL tube.
5. Take a FISH subsample (*see* Subheading 3.6) from the remaining pellet by picking approximately 5 μL out of the pellet with a pipette tip. Freeze pellet at −80 °C. (Optional: resuspend pellet in 0.5–1 mL IBS, transfer to a 2 mL tube and centrifuge at 14,000 × *g* and 4 °C for 5 min, discard supernatant before freezing.)
6. Repeat **steps 3–5** with the supernatant from **step 3** (*see* **Note 13**).
7. Use the chilled fixed-angle rotor for 15 mL tubes to centrifuge the supernatant for 10 min at 15,000 × *g* and 4 °C (*see* **Note 14**).
8. Transfer the supernatant to a new 15 mL tube. Take a 0.5 mL subsample of the supernatant for FISH analysis (*see* Subheading 3.6). Freeze supernatant at −80 °C.
9. Pick about 0.5 μL of the remaining pellet with a pipette tip as FISH subsample (*see* Subheading 3.6) without removing too much of the pellet.
10. Freeze pellet at −80 °C (*see* **Note 15**).

3.3 Isopycnic Gradient Centrifugation with Percoll® for Simple Separation of Symbionts from Host Tissue (Percoll® Cushion)

1. Precool a refrigerated centrifuge with a fixed-angle rotor for 50 mL tubes to 4 °C. Additionally, precool a fixed-angle rotor for 2 mL tubes. (Optional: also precool a fixed-angle rotor for 15 mL tubes.)
2. Centrifuge a 50 mL tube with 10 mL of chilled (4 °C) 60% Percoll® for 30 min at 5,400 × *g* (or higher) and 4 °C to allow the Percoll to self-assemble into a gradient (*see* **Note 16**).
3. Carefully layer about 5 mL homogenate on top of the Percoll® gradient.
4. Centrifuge for up to 30 min at 4 °C and 5,400 × *g* (*see* **Note 17**).
5. Remove the pelleted symbiont fraction carefully from the bottom of the tube using a 10 mL syringe with needle. Take care not to disturb the host layer on top.
6. Place the symbiont fraction in a 15 mL tube or 50 mL tube (depending on volume). To remove residual Percoll®, dilute the sample 1:2 with IBS and mix. Centrifuge for 10 min at 4 °C and 5,400 × *g*. Discard supernatant containing host tissue debris.
7. Resuspend the pellet in 5 mL TE buffer. Aliquot symbiont suspension in 1 mL portions in 2 mL tubes. Pellet cells for

5 min at 4 °C and 14,000 × *g*. Discard supernatant. (Optional: repeat washing step with 1 mL TE per tube.)

8. Freeze pellets at −80 °C.

3.4 Rate-Zonal Gradient Centrifugation with HistoDenz™ for Separation of Multiple Symbionts or Cell Types

1. Let HistoDenz™ gradients for rate-zonal centrifugation thaw at room temperature, then place them at 4 °C or on ice.
2. Precool a refrigerated centrifuge with a swing-out rotor for 15 mL tubes to 4 °C. Also precool a fixed-angle rotor for 2 mL tubes.
3. Layer 0.5 mL of homogenate on top of the gradient (*see* **Note 18**).
4. Centrifuge the gradients using the swing-out rotor. 1,000 × *g* and 5 min at 4 °C can already be sufficient (*see* **Note 19**).
5. Disassemble the gradient in 0.3–1 mL fractions by holding it upright and placing the pipette tip in the middle of the meniscus. Pipette slowly while following the meniscus downward (*see* **Note 20**). Resuspend the pellet—should there be any—in 0.5 mL IBS or in remaining gradient fluid, respectively. Transfer each zone (fraction) into an individual 2 mL tube. Place the tubes on ice.
6. Take a 20 µL subsample of each fraction for FISH analysis (*see* Subheading 3.6).
7. Add 1 mL TE buffer to each fraction to remove remaining HistoDenz™. Vortex thoroughly. Centrifuge for 5 min at 14,000 × *g* and 4 °C. Discard supernatant. (Optional: repeat washing step.) Freeze cell pellets at −80 °C.

3.5 Isopycnic Centrifugation with HistoDenz™ for Separation of Multiple Symbionts or Cell Types

1. Let HistoDenz™ gradients for isopycnic density gradient centrifugation thaw at room temperature (if frozen before use), then place them at 4 °C or on ice.
2. Precool a refrigerated centrifuge with a swing-out rotor for 15 mL tubes to 4 °C. Additionally, precool a fixed-angle rotor for 2 mL tubes.
3. Layer 1 to 2.5 mL of homogenate on top of the gradient (*see* **Note 21**).
4. Centrifuge the gradients for at least 30 min at maximum speed (5,400 × *g* or higher) and 4 °C (*see* **Note 22**).
5. Disassemble the gradient as described in Subheading 3.4, **steps 5–7**.

3.6 FISH Subsampling and Further Analyses

1. Transfer subsamples for FISH immediately into tubes with 1 mL IBS, containing 1% PFA (0.5 mL of IBS containing 2% PFA when taking 0.5 mL subsamples).
2. Mix thoroughly, fix at 4 °C overnight.

3. After fixation, dilute sample with 5 mL IBS and filter onto a 0.22 μm polycarbonate filter, wash filter once by adding additional 5 mL IBS, dry and label filter and store at −20 °C. Alternatively, freeze sample in fixative at −80 °C and process as soon as possible (*see* **Note 23**).
4. Select cell pellets for the specific research question and analysis method (e.g., proteomics) according to the results of the FISH analysis.

4 Notes

1. Alternatively, it is also possible to use phosphate-buffered saline (PBS) instead of IBS for all methods described here, depending on the organism's salinity preferences (for *Riftia*, IBS or another buffer with similar osmolarity is required). PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4. Fill approximately 800 mL water in a beaker. Weigh 8.0 g NaCl, 0.2 g KCl, 1.42 g Na_2HPO_4, 0.24 g KH_2PO_4. Allow each salt to dissolve before adding the next. Adjust pH with HCl to 7.4. Fill up to a final volume of 1 L.
2. The ball-shaped ends of the pestles used with Dounce homogenizers have a clearance that is designed to leave the symbiont cells intact, but destroy the host cells. Duall homogenizers with a ground inner surface are better suited for tougher material (such as *Bathymodiolus* gill tissue), whereas Dounce homogenizers are a good choice for softer tissue, e.g., the *Riftia* trophosome. *See* [10, 17] for further information regarding tissue homogenization.
3. Use for example nylon screening material with approximately 2 mm pore width. This is only necessary if the homogenate still contains chunks of host material, which cannot be crushed by homogenization (e.g., blood vessels), and is not recommended for low homogenate quantities, as some loss of material is inevitable during filtering.
4. The EDTA will not completely dissolve as long as pH is below approximately 8. Furthermore, solving will need time and proceeds better with vigorous stirring and moderate heat.
5. Hold the tube at an angle of 30–45° and slowly pipette each dilution by holding the pipette tip against the tube wall and as near to the meniscus of the previous layer as possible without touching it. It is strongly recommended to practice the gradient assembly and centrifugation process without adding a sample at first. Staining the density steps with different food colors helps to judge the quality of the gradient produced—the zones should mix as little as possible. Also try out different piston-

operated pipettes and serological pipettes with rubber two- or three-valve pipette fillers (*see* also Chap. 4 in [6]). Avoid rotating (vertically and, of course, horizontally) the gradient, as zones can get mixed otherwise. Tilt the gradient very slowly so as not to disturb the zonation due to the friction of the zones at the tube wall. To avoid mixing, it might help to "fix" each zone by freezing the tube standing upright at −80 °C or in dry ice before adding the next dilution step. The gradients can be prepared in advance and stored frozen at −80 °C for at least several months. Freezing and thawing of the gradient will cause the discontinuous gradient to become continuous (*see* Chap. 4 in [6], [12]). The ions in the gradient can change their distribution during freezing and thawing. Thus, the resulting gradient will not be isotonic [5] and will not necessarily be isosmotic (*see* Chap. 4 in [6]), but this did not seem to impair separation of the *Riftia* symbiont morphotypes (Hinzke et al., in preparation). A continuous gradient is necessary for rate-zonal centrifugation to allow for smooth particle movement (*see* Chap. 1 in [6]), while in a discontinuous gradient particles can get stuck at the interfaces between the density steps [5, 7, 12]. Nevertheless, as the shape of the gradient depends both on the tube volume and on the freeze and thaw rates, these factors should be as similar as possible throughout all experiments. Moreover, it is not recommended to repeatedly freeze and thaw the gradient, as this might alter its shape and reduce reproducibility (*see* [5], Chap. 4 in [6]). If the cells (or some of them) have the same density as one of the density steps in the gradient, they will not move further and will thus be separated by density, not by size. Therefore, the maximum density of the gradient has to be lower than the density of the cells (*see* Chap. 1 in [6], [7]). The density steps (HistoDenz™ dilutions) described here have been successfully used for enrichment of individual *Riftia* symbiont morphotypes (Hinzke et al., in preparation). However, other organisms may require other densities. Individual adaptations of the protocol are therefore indispensable for new applications.

6. Assemble the gradient as described in **Note 5**. Note that, as the originally discontinuous gradient becomes continuous during freezing and thawing and also during diffusion when it stands for longer time (*see* [5], Chap. 4 in [6]), it might be better to assemble the gradient immediately before use. Discontinuous gradients are well suited for preparative purposes, as they allow for a good separation of a restricted number of particle types with varying, known densities if the gradient densities are adapted accordingly. However, discontinuous gradients give less precise results, if a large number of different particle types are present. If a higher resolution for sample separation is

needed, a continuous gradient may work better (*see* Chap. 1 in [6]). An additional potential limitation of discontinuous gradients is that a barrier can form if many particles collect at a density border, causing denser, slower sedimenting material to get stuck [5, 7, 12].

7. It is also possible to use larger volumes of each dilution step and thus generate longer gradients. Collecting precise fractions might be easier, if density steps are larger. However, longer gradients come with higher cost for HistoDenz™ and the centrifugation time increases as it takes longer for the cells to reach the density that corresponds to their own buoyant density.
8. It is advisable to take FISH samples of the obtained centrifugation fractions in order to verify whether and to which degree the respective enrichment procedure was successful. Depending on the research question, it might be desirable to also take other subsamples, such as for DNA or protein analyses, in addition to or instead of FISH subsamples.
9. As a rough estimate of 3 g tissue: The total volume of IBS in the homogenizer should increase by 3 mL when adding the tissue sample. The amount of buffer should be adjusted for the respective type of tissue to ensure that the homogenate is not too viscous but contains enough material for subsequent analyses.
10. Use rotating movements while moving the pestle. The first few strokes with each pestle may be very difficult, depending on the texture of the tissue used. Also, it might be necessary to use more strokes for sufficient homogenization of the respective sample. But do not "overhomogenize" so as not to damage the bacterial cells.
11. Fold the nylon mesh into a cone and place it tip down in the opening of the falcon tube. It might be necessary to use more than one piece of mesh for filtering the whole material, depending on the viscosity of the homogenate.
12. These conditions were used for enrichment of symbionts from *B. thermophilus* (Ponnudurai et al., in preparation). In the case of *B. azoricus*, host and symbiont fractions were separated by centrifugation for 10 min at 4 °C and 700 × *g* [16]. The exact conditions have to be adapted for the sample used. It is always important to centrifuge relatively slowly in this step so as not to remove too many symbiont cells from the supernatant. For the *B. azoricus* symbiosis, which contains two bacterial symbionts, the larger and denser methanotrophic symbiont is enriched in the first pellet (R. Ponnudurai, personal communication). Differential pelleting can also be used to separate symbionts

containing storage granules (e.g., sulfur or polyhydroxyalkanoates) from those without granules.

13. This step serves to remove (remaining) large host tissue fragments and host nuclei [16].
14. This centrifugation step serves to pellet small tissue fragments, symbionts and mitochondria, whereas the final supernatant is enriched in cytosolic host proteins [16].
15. Alternatively, it is possible to resuspend the pellet in a low buffer volume (to obtain a concentrated suspension; approximately 0.5–1.5 mL) and use the material for subsequent density gradient centrifugation to achieve further enrichment. However, a rate-zonal centrifugation that was performed for *B. azoricus* after differential pelleting did not increase symbiont purity significantly (R. Ponnudurai, personal communication).
16. Shorter centrifugation times are possible, if the used centrifuge allows for higher speed—the higher the *g*-force, the lower the centrifugation time needed.
17. **Note 16** applies here, too—but a maximum of 25,000 × *g* is recommended for gradients loaded with sample material so as not to damage the symbionts. If the cell pellet is too soft, increase centrifugation time and/or speed. Note that this simplified version of an isopycnic density gradient centrifugation only works for symbionts with dense storage inclusions, such as sulfur granules. Less dense cells may not be sufficiently enriched by this method.
18. Do not use too much homogenate for one gradient as overloading the gradient will lead to poor separation—the starting position of the cells should be as uniform as possible, i.e., the total volume of homogenate added should be small for good results in rate-zonal centrifugation. With regard to pipetting the homogenate on top of the gradient, the same remarks as for assembling the gradient do apply here (*see* **Note 5**). Upon loading, the homogenate may sink down approximately 0.5 cm below the surface of the gradient, depending on concentration of the homogenate and density of the uppermost density gradient step. Diluting the homogenate might help circumvent this, though this sinking seems not to significantly hamper cell separation (Hinzke et al., in preparation).
19. This step might take some adjustments in order to adapt centrifugation conditions for the type of sample used. It is important that, after centrifugation, the cell material is distributed over the whole gradient and thus maximally separated. If centrifugation time and/or speed is too low, the material is still mostly in the upper part of the gradient, while a too long centrifugation time and/or too high speed will cause the cell material to pellet at the bottom of the tube.

The appearance of the gradient gives some hints whether the centrifugation conditions where appropriate: a slight turbidity also in the lower part of the gradient and a nonvisible or (very) small pellet at the bottom indicate an equal distribution of cells and thus a good separation (but there might be no turbidity at all). Decrease the centrifuge's acceleration and deceleration to approximately 2/3 of the highest possible value in order to reduce perturbations of the gradient.

20. The fraction size depends on the desired sample resolution. A fraction size of 0.5 mL was found to allow for good resolution while keeping expenditure of time and material acceptable (Hinzke et al., in preparation). First remove a volume equivalent to the amount of sample that was layered on top of the gradient. Then proceed with disassembling the gradient. The same pipette tip can be used for the whole gradient as the HistoDenz™ density increases downward. Alternatively, piercing a hole in the tube wall at the bottom of the tube and letting the gradient flow out stepwise (for example through a syringe needle) into subsequent tubes might work, too, but is not recommended for higher HistoDenz™ densities of 50% and more, as the high viscosity will hamper the flow. Practice gradient disassembly before working with samples, e.g., with colored density steps (*see* also **Note 5**).
21. Unlike in rate-zonal centrifugations (*see* **Note 18**), in isopycnic centrifugation the amount of homogenate is mainly limited by the remaining free space in the tube, although loading too much or too dense homogenate might impair separation here, too. The homogenate may sink down a bit under the surface.
22. Decrease the centrifuge's acceleration and deceleration to approximately 2/3 of the highest possible value in order to reduce perturbations of the gradient. Once the cells have reached the gradient density that matches their own buoyant density they will not move further downward. Therefore, it has to be made sure that centrifugation time is long and speed is high enough for the cells to reach the respective zones. Longer centrifugation times will then not alter the particles' positions any further. If a higher speed is used, centrifugation time can be reduced.
23. It is recommended to perform FISH analyses as soon as possible after taking the samples to avoid overfixation. Nevertheless, we obtained good results with samples stored in fixative at −80 °C for 6 months and longer.

Acknowledgments

We thank Horst Felbeck for valuable help with sample preparation, Ruby Ponnudurai for data collection and analysis, and Thomas Schweder for helpful input. This work was supported by the Deutsche Forschungsgemeinschaft (DFG grant MA 6346/2-1).

References

1. Kleiner M, Wentrup C, Lott C et al (2012) Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. Proc Natl Acad Sci U S A 109:E1173–E1182
2. Bright M, Sorgo A (2003) Ultrastructural reinvestigation of the trophosome in adults of *Riftia pachyptila* (Annelida, Siboglinidae). Invertebr Biol 122:345–366
3. Hugenholtz P (2002) Exploring prokaryotic diversity in the genomic era. Genome Biol 3: reviews0003.1–reviews0003.8
4. Hinton R, Dobrota M (1976) Density gradient centrifugation. Elsevier/North-Holland Biomedical Press, Amsterdam
5. Sharpe PT (1988) Chapter 3: Centrifugation. In: Methods of cell separation. Elsevier Science Publishers B.V. (Biomedical Division), Amsterdam
6. Graham J (2001) Biological centrifugation. BIOS Scientific Publishers Limited, Oxford
7. Hinton RH, Mullock BM (1997) Isolation of subcellular fractions. In: Graham JM, Rickwood D (eds) Subcellular fractionation: a practical approach. The practical approach series. Oxford University Press, Oxford, pp 31–69
8. Spragg SP, Steensgaard J (1992) Theoretical aspects of practical centrifugation. In: Rickwood D (ed) Preparative centrifugation: a practical approach. The practical approach series. Oxford University Press, Oxford, pp 1–42
9. Dobrota M, Hinton R (1992) Conditions for density gradient separations. In: Rickwood D (ed) Preparative centrifugation: a practical approach. The practical approach series. Oxford University Press, Oxford, pp 77–142
10. Evans WH (1992) Isolation and characterization of membranes and cell organelles. In: Rickwood D (ed) Preparative centrifugation: a practical approach. The practical approach series. Oxford University Press, Oxford, pp 233–270
11. Brakke MK (1951) Density gradient centrifugation: a new separation technique. J Am Chem Soc 73:1847–1848
12. Pertoft H (2000) Fractionation of cells and subcellular particles with Percoll. J Biochem Biophys Methods 44:1–30
13. Distel DL, Felbeck H (1988) Pathways of inorganic carbon fixation in the endosymbiont-bearing lucinid clam *Lucinoma aequizonata*. Part 1. Purification and characterization of the endosymbiotic bacteria. J Exp Zool 247:1–10
14. Markert S, Arndt C, Felbeck H et al (2007) Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila*. Science 315:247–250. https://doi.org/10.1126/science.1132913
15. Markert S, Gardebrecht A, Felbeck H et al (2011) Status quo in physiological proteomics of the uncultured *Riftia pachyptila* endosymbiont. Proteomics 11:3106–3117
16. Ponnudurai R, Kleiner M, Sayavedra L et al (2017) Metabolic and physiological interdependencies in the *Bathymodiolus azoricus* symbiosis. ISME J 11:463–477
17. Graham JM (1997) Homogenization of tissues and cells. In: Graham JM, Rickwood D (eds) Subcellular fractionation: a practical approach. The practical approach series. Oxford University Press, Oxford, pp 1–29

PUBLICATION VII

The manuscript “Bacterial symbiont subpopulations have different roles in a deep-sea symbiosis” describes the physiological roles of symbionts of different cell sizes in the *Riftia pachyptila* symbiosis as revealed by combining gradient centrifugation with metaproteomics analysis.

Supplementary data is provided in the CD.

Submitted

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

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

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

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

Abstract

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

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

Introduction

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

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

These presumptive subpopulations are likely congruent with symbionts of different cell sizes: Individual Endoriftia cells exhibit pronounced morphological diversity, ranging from small rods to small and large cocci in ultimate proximity to each other within the same host specimen (Hand, 1987, Bright *et al.*, 2000, Bright and Sorgo, 2003). In individual trophosome lobules, which measure approximately 200-500 µm in diameter, the smallest, rod-shaped symbiont cells are located close to the central blood vessel, while towards the lobule periphery, symbionts gradually increase in size and become coccoid, before they are degraded in the outermost lobule zone. Only small Endoriftia cells and the host bacteriocytes in which they reside appear to undergo cell division, indicating that small and large symbionts belong to a common cell cycle (Bright *et al.*, 2000, Bright and Sorgo, 2003). Previous microscopy-based studies indicated that small and large *Riftia* symbionts differ not only with regard to their frequency of cell division, but also with regard to carbon incorporation rates, amount of stored glycogen, and area of sulfur storage vesicles (Bright *et al.*, 2000, Sorgo *et al.*, 2002, Bright and

Sorgo, 2003, Pflugfelder *et al.*, 2005). This suggests that individual cell sizes may indeed have dissimilar metabolic properties.

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

Material and Methods

Sample collection and enrichment of symbiont subpopulations

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

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

To enrich symbiont cells of varying sizes (i.e., morphologically distinct symbiont subpopulations), *Riftia* specimens were dissected onboard the research vessel immediately after recovery of the worms and approximately 3 ml trophosome tissue were homogenized in a Dounce glass homogenizer in 6 ml imidazole-buffered saline (IBS, 0.49 M NaCl, 0.03 M MgSO4, 0.011 M CaCl2, 0.003 M KCl, 0.05 M imidazole). As described in Hinzke *et al.* (2018), the homogenate was subjected to rate-zonal density gradient centrifugation, which allows to separate particles based on their size (Graham, 2001). In brief, an 8-18% Histodenz™ density gradient was created using a dilution series of Histodenz™ in IBS (1% steps, 1 ml per step), which was stacked in a 15 ml tube so that Histodenz™ concentration was highest at the bottom and lowest at the top. 0.5 ml tissue homogenate was layered on top of this gradient and the gradient was centrifuged (1,000 x g, 5 min, 4 °C) in a swing-out rotor. Smaller symbiont cells were thus enriched in less dense gradient fractions (lower Histodenz™ concentrations) in the upper part of the gradient, while larger cells migrated to lower gradient fractions. After centrifugation, gradients were disassembled by carefully fractionizing the entire gradient volume into 0.5 ml subsamples, giving a total of 24 fractions. Enrichment of distinct symbiont subpopulations in these subsamples was confirmed using catalyzed reporter deposition-fluorescence in situ hybridization (CARD-FISH, see below). For this purpose, 20 µl of each gradient fraction subsample and 15 µl of homogenate was fixed in 1% PFA in IBS, and symbiont cells were subsequently filtered onto GTTP polycarbonate filters (pore size 0.2 µm, Millipore) as described previously (Ponnudurai *et al.*, 2017).

CARD-FISH

Enrichment of symbiont cell sizes in gradient fractions was analyzed employing fluorescence microscopy with samples labelled by CARD-FISH. CARD-FISH labeling was performed as previously described (Ponnudurai *et al.*, 2017), using the probe Rif445 (Nussbaumer *et al.*, 2006) and Alexa Fluor® 594-labeled tyramide. For counterstaining, 0.1% (w/v) 4,6-diamidino-2-phenylindole (DAPI) was added to the embedding medium (4:1 Citifluor AF1 (Citifluor) and Vectashield (Vector Laboratories)). CARD-FISH filters were analyzed using an Axio Imager.M2 fluorescence microscope (Carl Zeiss Microscopy GmbH). For semi-automated cell counting and to measure the longest cell dimension, we used a custom Fiji (Schindelin *et al.*, 2012) macro with the Fiji plugins Enhanced Local Contrast (CLAHE; Saalfeld, 2010) and Bi-

exponential edge preserving smoother (BEEPS; Thévenaz *et al.*, 2012). After image processing, we excluded objects with a size of less than 2 µm (as these were mainly artifacts) and set the maximum object size to 20 µm. To assign cell sizes to size classes (i.e., cell size ranges) we used a quartile split: We calculated quartiles of cell sizes in non-enriched homogenate samples (i.e., 25% of all cells in homogenate samples were assigned to each class). This resulted in the four calculative size classes very small (≥2 µm – <3.912 µm), small (≥3.912 µm – <5.314 µm), medium (≥5.314 µm – <6.83275 µm) and large (≥6.83275 µm – 20 µm). The majority of cells in all size classes were coccoid. Rod-shaped cells were almost exclusively present in the smallest size class. Individual gradient fractions (subsamples) were screened for their respective share of cells in each size class and the subsample with the highest percentage of cells in the respective quartile was chosen for metaproteomic analysis. For example, if of all 24 subsamples of a sample, fraction 5 had the highest percentage of very small cells, i.e. most of the cells in fraction 5 were between 2 µm and 3.912 µm in diameter (as measured by our Fiji macro), this fraction was chosen as representative of very small symbiont cells in the respective biological replicate (worm). The fraction containing the highest percentage of very small cells will be referred to as "fraction XS" in the following. The fractions containing the highest percentages of small, medium and large symbiont cells will be referred to as "S", "M", and "L", respectively. If the same subsample had the highest percentage of cells in two size classes, this subsample was chosen as representative for one of these size classes, and for the other size class, the subsample with the second highest percentage of cells in that class was used as representative. Cell size distributions in the four size class representatives are summarized in Figure 1.

Transmission electron microscopy (TEM)

Trophosome samples used for TEM in this study (see Supplementary Table S1 for details) were prepared and analyzed as described previously (Hinzke *et al.*, 2019). Tissue sections were recorded on sheet films (Kodak electron image film SO-163, Plano GmbH, Wetzlar) as described by Petersen *et al.* (2020). To create a composite high-resolution TEM image of a trophosome lobule (Figure 5A), we merged 50 individual micrographs of one section using Serif Affinity Photo (https://affinity.serif.com/en-us/photo/). All 50 partially overlapping images were loaded and the fully automated "Panorama Stitching" technique was applied, resulting in a panorama image still showing some vignette marks caused by inhomogeneous

exposure at the former edges of individual images. The global smooth frequencies reflecting these exposure errors were removed using the frequency separation filter with a large radius. The gradation curve was manually corrected. For acquisition of the images in Figure 5B, a wide-angle dual speed CCD camera Sharpeye (Tröndle, Moorenweis, Germany) was used, operated by the ImageSP software. All micrographs were edited using Adobe Photoshop CS6.

HCR-FISH and confocal laser scanning microscopy (CLSM)

A gradient fraction enriched in large symbiont cells (see Supplementary Table S1 for details) that was fixed for CARD-FISH and immobilized on GTTP polycarbonate filters as described above was used for hybridization chain reaction FISH (HCR-FISH) according to Choi *et al.* (2014). We used a HCR-FISH v2.0 Custom Kit (Molecular Technologies) according to the manufacturer's instructions. Probes targeted the *Riftia* symbiont's 16S rRNA (fluorescence marker: Alexa Fluor® 488), and the mRNAs of ATP-citrate lyase subunit AclB (Alexa Fluor® 647) and RubisCO (Alexa Fluor® 594; see Supplementary Table S2 for the probe sequences). In brief, filter sections were washed twice with 50% hybridization buffer (50% formamide, 5x sodium chloride sodium citrate (SCSC, 0.75 M NaCl, 75 mM $Na_3C_6H_5O_7$), 9 mM citric acid, pH 6.0, 0.1% Tween 20, 50 µg/ml heparin, 1x Denhardt's solution, 10% dextran sulfate) in 2x mPBS (89.8 mM Na_2HPO_4, 10.2 mM NaH_2PO_4, 0.9 M NaCl) at 45°C for 30 min for pre-hybridization, and incubated overnight (16 h, 45°C) with probe solution (1 pmol of each probe in 500 µl hybridization buffer). Excess probes were removed with several washing steps in 75-25% probe wash buffer (50% formamide, 5x SCSC, 9 mM citric acid, pH 6.0, 0.1% Tween 20, 50 µg/ml heparin in 5x SCSC) for 15 min at 45°C, 300 rpm, and subsequently in 5x SCSC for 30 min at 45°C and 300 rpm. Samples were pre-amplified with DNA amplification buffer (5x SCSC, 0.1% Tween 20, 10% dextran sulfate). Hairpins were activated by snap-cooling and added to the samples. After overnight incubation (16 h, room temperature) with the hairpin solution, samples were washed with 5x SCSC, containing 0.05% Tween 20 (room temperature, 300 rpm, four times 5 min, two times 30 min) and embedded in Mowiol 4-88 (Carl Roth GmbH) embedding medium prepared according to the manufacturer's instructions. Confocal microscopy was performed on a Zeiss LSM510 meta equipped with a 100x/1.3 oil immersion objective. Probes were excited with laser lines 633 (ATP-citrate lyase), 561 (RubisCO) and 488 (16S rRNA) and signals were detected with filters suitable for dye maximal emissions at 670

nm, 595 nm and 527.5 nm, respectively. Signal intensities and cell sizes (from 8 frames showing a total of 33 cells) were quantified using the Fiji software package (Schindelin *et al.*, 2012). Individual cells were defined as regions of interest (ROI), in which signal intensity per pixel was recorded. Mean pixel intensity of ROI was calculated and background was corrected. Global background values were calculated for every channel based on up to six ROIs randomly placed in each image frame. The following cell size parameters were calculated: (i) Feret's Diameter (the longest distance between any two points along the boundary of the ROI) and (ii) the area of the ROI.

Flow cytometry

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

procedure, FACS-sorted cell populations as well as unsorted subsamples of homogenate and gradient fractions were examined using an Axio Imager.M2 fluorescence microscope (Carl Zeiss Microscopy GmbH).

Peptide sample preparation

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

LC-MS/MS analysis

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

Protein identification and function prediction

Proteins were identified by searching the MS/MS spectra against the *Riftia* host and symbiont database (Hinzke *et al.*, 2019), which was constructed from the host transcriptome and three

symbiont genome assemblies, i.e., NCBI project PRJNA60889 (endosymbiont of *Riftia pachyptila* (vent Ph05)), NCBI project PRJNA60887 (endosymbiont of *Tevnia jerichonana* (vent Tica)), and JGI IMG Gold Project Gp0016331 (endosymbiont of *Riftia pachyptila* (vent Mk28)). The cRAP database containing common laboratory contaminants (The Global Proteome Machine Organization) was added to complete the database. Database search was conducted using Proteome Discoverer v. 2.0.0.802 with the Sequest HT node as described in Kleiner *et al.* (2018) with a false discovery rate of 5% (FidoCT Protein Validator node, q-value <0.05).
To systematically screen the *Riftia* symbiont metagenome for dissimilatory sulfur metabolism-related proteins, candidates identified in different studies were searched against the *Ca.* E. persephone metaproteome database using bioedit (Hall, 1999; Supplementary Table S9). Host proteins were additionally annotated using the same tools as in Hinzke *et al.* (2019). Symbiont hydrogenase sequences were classified using HydDB (Søndergaard *et al.*, 2016).

Statistical evaluation of metaproteomics data and abundance quantification

Filtering and normalization

For samples from sulfur-rich specimens, four replicates for each of the four size classes were used (resulting in 16 samples); for analysis of symbionts from sulfur-depleted specimens, three replicates were available per size class (giving a total of 12 samples). For comparisons of protein abundance (i) across different samples, e.g., to determine a protein's abundance trend across gradient fractions XS to L, edgeR-RLE-normalized spectral counts were calculated (see below), while (ii) %orgNSAF values were used for abundance comparisons of different proteins within one sample, e.g., to determine the "most abundant" proteins in a sample.
(i) To allow for comparisons of protein abundance across different samples, spectral count data were first filtered so that they included only proteins that had at least five spectral counts in at least four out of 16 (S-rich specimens) or three out of 12 (S-depleted specimens) samples. The filtered dataset was then normalized using Relative Log Expression (RLE) normalization with the package edgeR v.3.24.3 (Robinson *et al.*, 2010) in R v. 3.5.1 (R Core Team, 2018; Supplementary Table S3a). The filtering and normalization step was included to avoid biasing the analysis towards symbiont proteins that were only detected in the high-density fractions M and L (enriched in larger symbiont cells), but which were absent in fractions of lower density (XS and S, containing primarily smaller cells). Fractions S and particularly XS

contained relatively more host proteins, leading to a lower total number of detectable symbiont proteins. (Note that these values were not normalized to protein size, so that a protein's relative abundance changes can be followed across different samples, but abundances cannot be compared between proteins). We tested for significant differences in symbiont protein abundance between individual gradient factions (representing enrichments of different cell size classes) using two methods, i.e., hierarchical and profile clustering and random forests (see below).

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

STEM analysis

For protein expression profile clustering, we employed the Short Time Series Expression Miner (STEM; Ernst and Bar-Joseph, 2006) v. 1.3.11., which fits gene expression profiles in ordered short series datasets (like the cell cycle stages of *Ca.* E. persephone), to model profiles representing different expression patterns. Filtered and RLE-normalized data were log-normalized, repeat data were defined to be from different time points and data were clustered using the STEM method with default options. For STEM filtering, the minimum correlation between repeats and the minimum absolute expression change were set to 0.5. All permutations were used. For correction, the false discovery rate (FDR) was set to 0.05. Profiles were clustered with a minimum correlation percentile of 0.5. Other parameters were left at default values. Proteins which were assigned to model profiles, i.e., all proteins which were not removed by filtering and showed a consistent trend in all replicates, were used for further analysis. This means that differences in protein abundance patterns were considered significant if proteins were detected with a consistent abundance trend across all replicates (increase, decrease or alternating increase and decrease of abundance from fraction XS to L).

Random forests

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

Significant differences in protein abundance

Proteins that showed significant abundance differences as determined by STEM analysis or by random forest analysis (see above) or by both methods were included in a common list. Please note that this approach of determining significant protein abundance differences was not based on individual p-values. For proteins with significant abundance differences, we clustered the z-scored mean abundances using hierarchical clustering (Pearson correlation, complete linkage) in R to visualize their abundance trends (Supplementary Figure S1). For this purpose, we employed the R base package stats (R Core Team, 2018) as well as the packages cluster (Maechler *et al.*, 2018) and ComplexHeatmap (Gu *et al.*, 2016). For comparison of S-rich and S-depleted symbionts of the same size class, we used the R package edgeR v. 3.24.3 (Robinson *et al.*, 2010), which uses a Bayes-moderated Poisson model for count data analysis,

with an overdispersion-adapted analogon to Fisher's exact test for detecting differentially expressed genes (Robinson *et al.*, 2010).

Host proteins

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

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (ProteomeXchange – ProteomeCentral) via the PRIDE partner repository (Vizcaíno *et al.*, 2016) with the dataset identifier PXD016986 (access for reviewers: https://www.ebi.ac.uk/pride/, username reviewer69719@ebi.ac.uk, password LWF5rqEW).

Results

Enrichment of individual symbiont cell sizes by gradient centrifugation

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

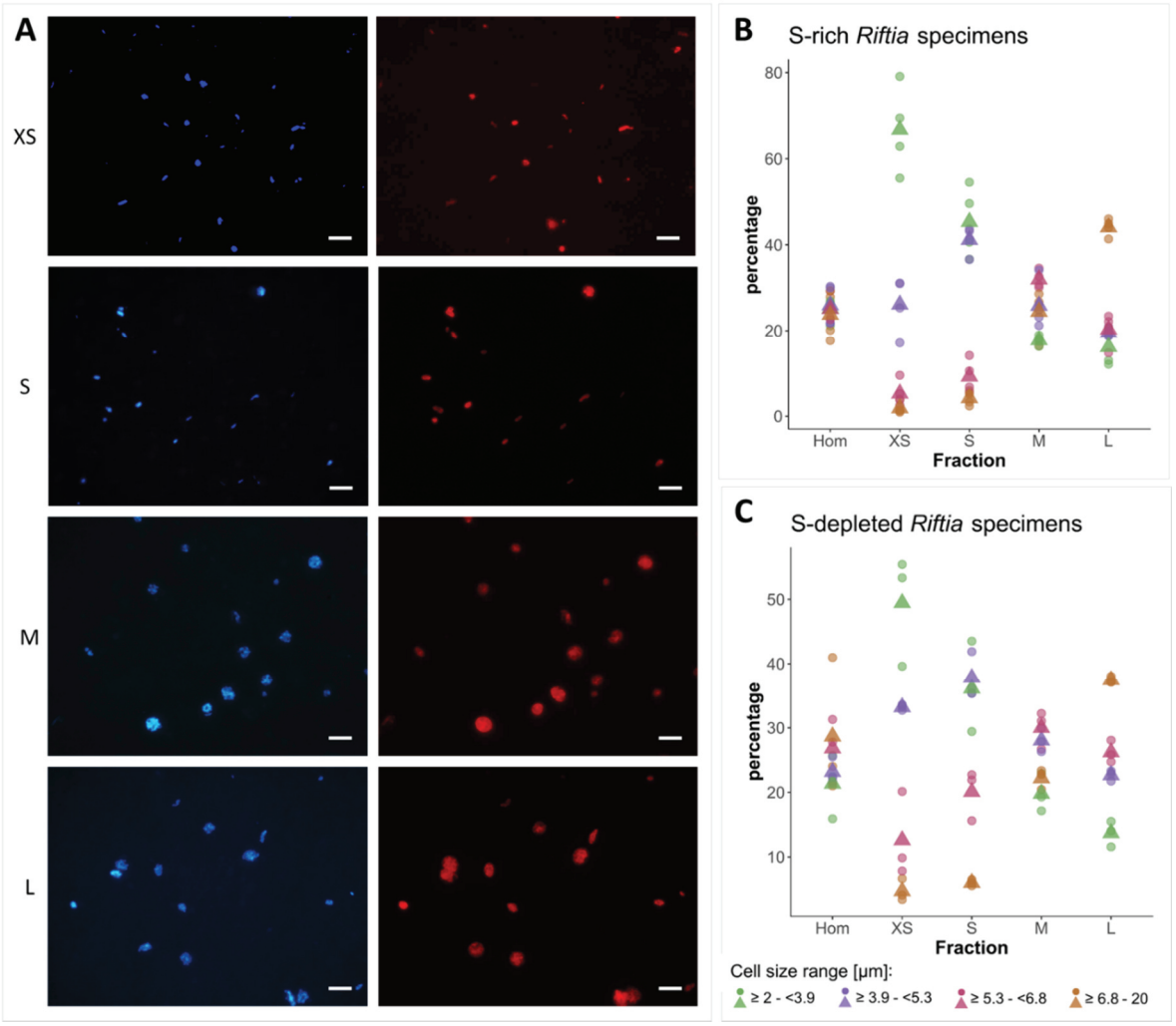

Figure 1: A) Catalyzed reporter deposition-fluorescence in situ hybridization (CARD-FISH) images of Riftia symbiont cells after density gradient centrifugation of trophosome homogenate. After the enrichment procedure, small bacterial cells had accumulated in the upper, less dense gradient fractions (top),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

Symbiont DNA quantification

Flow cytometry and fluorescence-activated cell sorting (FACS) indicated that DNA content in large *Riftia* symbionts is up to 10-fold higher compared to small symbionts. To identify

distinguishable bacterial cell populations, we examined Syto9-stained cells in *Riftia* trophosome homogenate and in gradient fractions from the upper and lower parts of the gradient (enriched in smaller and larger symbionts, respectively) with regard to their light scattering properties. Forward scatter (FSC) and side scatter (SSC) usually correlate with cell size and cell granularity, respectively (Bouvier *et al.*, 2001, Tracy *et al.*, 2010). Amongst a number of particle groups with different properties (see also Supplementary Results and Discussion C), we found two populations, 1 and 2, which were abundantly detected in non-enriched trophosome homogenate, but showed very dissimilar frequencies in fractions enriched in larger or smaller symbionts (Figure 2, Supplementary Figure S2): While population 1, which exhibited relatively lower FSC and SSC signals (indicative of smaller cell size and lower cell complexity), was highly abundant in fractions enriched in smaller symbionts, this population was notably less prominent in fractions enriched in larger symbionts. Simultaneously, population 2, which gave higher FSC and SSC signals (indicative of larger cell size and higher complexity), was highly abundant in fractions enriched in larger symbionts but nearly absent in gradient fractions enriched in smaller symbionts. This suggests that populations 1 and 2 consist of smaller and larger symbionts, respectively. This assumption was verified by FACS-separation of both populations from trophosome homogenate, and examination of the sorted cell suspensions by fluorescence microscopy along with unsorted enriched gradient fractions and homogenate samples for reference (Supplementary Figure S2). For quantification of DNA in smaller and larger symbionts, we compared median fluorescence intensities (MFI) per particle between population 1 and 2 in non-enriched homogenate and in enriched gradient fractions. In all sample types, MFI per particle was notably lower in population 1 (between 186 and 1,994 relative fluorescence units, rfu) than in population 2 (2,712 – 10,723 rfu). On average, MFI was 9.7-fold higher in population 2 than in population 1 (Supplementary Table S8).

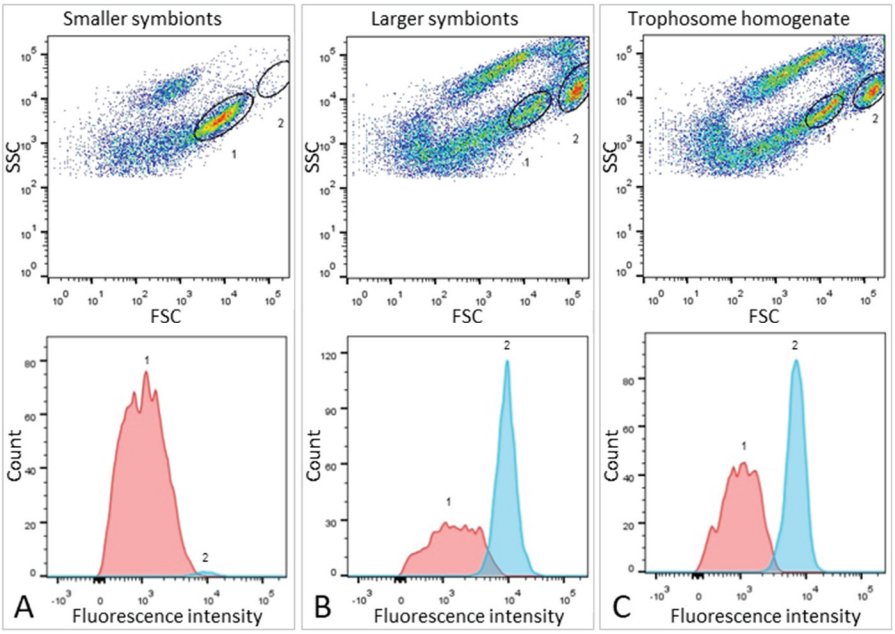

Figure 2: Flow cytometry for DNA quantification of *Riftia* symbionts. **A)** Dot plot of forward scatter (FSC) and side scatter (SSC), and histogram with fluorescence signal counts and fluorescence intensity per particle of a gradient fraction enriched in smaller symbionts. **B)** Gradient fraction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

Protein identifications and relative protein abundance

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

were consistent across all replicates in the context of our statistical approach). In Figure 3 and Supplementary Table S3, proteins that showed such significant changes in relative abundance are marked with asterisks.

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

Symbiont protein functions

Cell cycle, DNA topology, replication and repair

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

Chaperones and stress proteins

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

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

Transport

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

Central metabolism

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

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

Nitrogen metabolism: Relative abundance of all three respiratory membrane-bound nitrate reductase subunits, NarGHI, decreased significantly from fraction XS to L, as

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

Other categories

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

Symbiosis-specific host proteins

Our density gradient fractionation procedure allowed not only for the identification of symbiont proteins with differential abundance across different Endoriftia size ranges, but also enabled us to single out host proteins that are potentially involved in direct interactions with the symbionts. As host proteins that are attached to the symbionts are pulled down with the symbiont cells during gradient centrifugation, these proteins should be significantly more abundant in symbiont-enriched fractions compared to the non-enriched trophosome homogenate (Figure 4, Supplementary Table S5). Besides many ribosomal and mitochondrial host proteins, which were also enriched, putatively symbiont-associated host proteins included the host's peptidoglycan-recognition protein SC1a/b, beta carbonic anhydrase 1, digestive proteins involved in protein- and carbohydrate degradation, e.g., acid phosphatase, digestive proteases and glycan degradation enzymes, as well as hypoxia up-regulated proteins, a thiosulfate sulfurtransferase, and transmembrane protein 214-B.

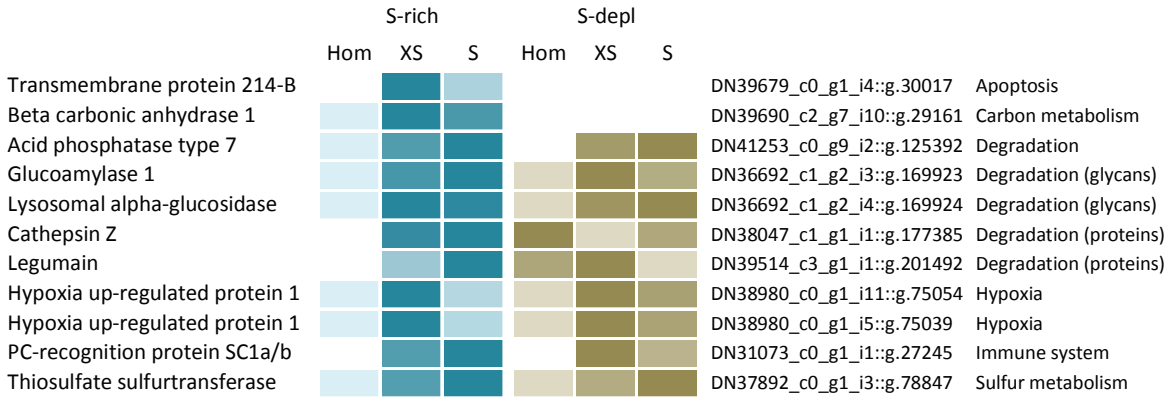

Figure 4: Selected Riftia host proteins with significantly higher relative abundance in the symbiont-enriched fractions XS and S compared to the non-enriched trophosome tissue homogenate (Hom) in sulfur-rich (S-rich) and sulfur-deleted (S-depl) Riftia specimens. Relative abundance trends are indicated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

Discussion

Symbiont growth and differentiation

Cell division plays a more prominent role in small symbionts

As indicated by the significant decrease in abundance of the cell division key protein FtsZ from fraction XS to fraction L, small Endoriftia are more engaged in cell division than larger symbionts. In accordance with the microscopy-based hypothesis of Bright and Sorgo (2003), the smallest symbionts, which are *in situ* localized in the trophosome lobule center (Figure 5), thus apparently function as stem cells of the symbiont population. During cell division, FtsZ forms the Z ring, to which the other division-related proteins are successively recruited (reviewed in Weiss, 2004). Cell size and cell division therefore likely depend on the amount of FtsZ available (Chien *et al.*, 2012). This correlation is, for example, also reflected by a decrease of FtsZ concentration during differentiation of vegetative cells into non-dividing larger heterocysts in the cyanobacterium *Anabaena* (Klint *et al.*, 2007). Interestingly, while FtsZ abundance decreased across fractions, many other proteins which interact with FtsZ during

cell division were detected with increasing abundance from fraction XS to L. This indicates that these proteins are also involved in processes other than cell division, e.g., in determining cell shape and stabilization. ZapD, for example, is involved in FtsZ filament organization, and its overexpression leads to cell filamentation (Durand-Heredia *et al.*, 2012). DamX overexpression, too, was observed to induce filamentation in *E. coli* (Lyngstadaas *et al.*, 1995), while overexpression of the cell shape determination protein CcmA in *E. coli* and *P. mirabilis* lead to enlarged, ellipsoidal cells (Hay *et al.*, 1999), and FtsEX is required for cell elongation rather than cell division in *B. subtilis* (Domínguez-Cuevas *et al.*, 2013). The actin homolog MreB is pivotal for rod-shape formation in bacteria and for cell stiffness in *E. coli*, could negatively regulate cell division, and participates in chromosome segregation (Wachi and Matsuhashi, 1989, Kruse *et al.*, 2006, Wang *et al.*, 2010, reviewed in Reimold *et al.*, 2013). In large Endoriftia, these proteins might therefore be involved in stabilizing growing symbiont cells. SlmA, which was only detected in fractions M and L in our study, was shown to disassemble FtsZ polymers, thus acting as a cell division inhibitor (Cho *et al.*, 2011), which supports the idea of relatively less cell division in large *Riftia* symbionts. Although Endoriftia‘s major cell division protein FtsZ was notably (1.75x) less abundant in fraction L (compared to fraction XS), it was not completely absent. This may indicate that cell division is reduced with increasing cell size, but not abandoned altogether, or it may point to additional FtsZ functions, besides cell division (as also suggested for *Anabeana* (Klint *et al.*, 2007) and *E. coli* (Thanedar and Margolin, 2004)).

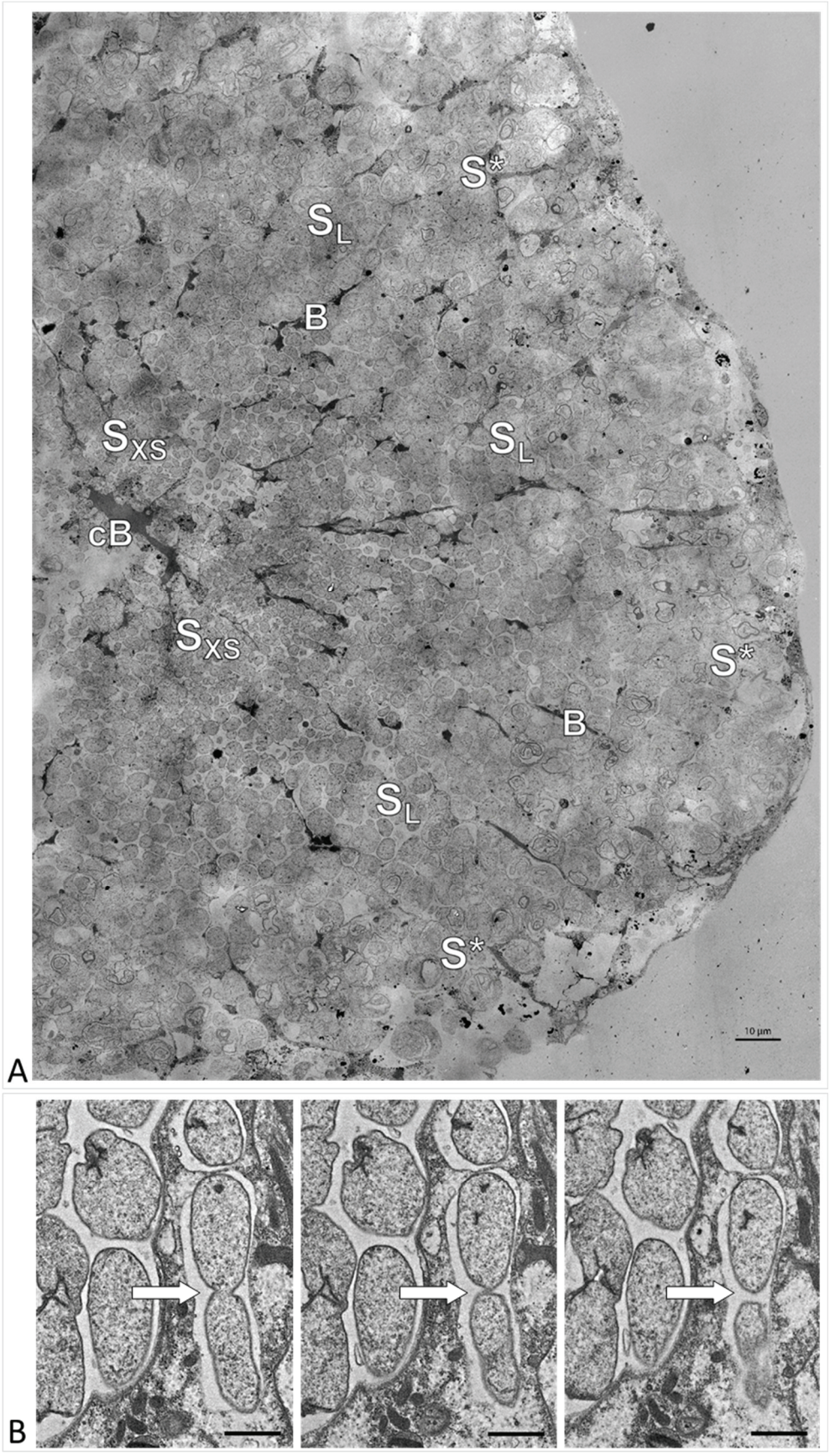
S*
S_L
B
S_XS
S_L
cB
S_XS
S*
B
S_L
S*
10 µm
A
B

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

Large symbionts have more genome copies and less compact chromosomes

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

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

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

Since DNA condensation may function as a DNA protection mechanism (Ohniwa *et al.*, 2006, Mukherjee *et al.*, 2008, Yoshikawa *et al.*, 2008, Takata *et al.*, 2013), less condensed DNA might be more prone to various kinds of damage and require the enhanced expression of DNA repair mechanisms. This would explain the observed higher abundance of several DNA repair proteins in fraction L, which was enriched in larger, older symbiont cells with (presumably) larger quantities of less condensed DNA, compared to the smaller symbiont cells. RadA, RdgC,

RecCN, UvrAB, and Mfd, which are known to be involved in DNA recombination and repair in many bacteria (Kowalczykowski, 2000, Beam *et al.*, 2002, Tessmer *et al.*, 2005, Drees *et al.*, 2006, Truglio *et al.*, 2006, Deaconescu *et al.*, 2007), may compensate for this elevated vulnerability. In eukaryotes, chromatin decondensation was shown to facilitate access of the DNA damage response to double strand breaks, thus allowing for more efficient repair (Murga *et al.*, 2007).

Small symbionts may be exposed to elevated stress levels

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

Possibly, such host-induced stress may also involve the production of ROS in symbiont-containing bacteriocytes, similar to animal and plant hosts, which generate ROS to defend themselves against pathogenic bacteria (Heath, 2000, Lynch and Kuramitsu, 2000, D'Haeze and Holsters, 2004). Small symbionts, which are relatively loosely packed in their host cell vesicles (Figure 5A) and have a comparatively high surface-to-volume ratio, might be particularly exposed to this presumptive ROS stress, while larger symbionts, which are more tightly packed, may face lower ROS levels. This would explain the observed higher abundance of the ROS scavengers rubrerythrin (Rbr2), superoxide dismutase (SodB), and alkylhydroperoxide reductase (AhpC) in small symbionts. In line with this assumption, a superoxide dismutase and also the chaperones ClpB, HtpG, and DnaK were suggested to be involved in ROS protection in *Serratia symbiotica* (Renoz *et al.*, 2017), and ClpB protease expression has been

shown to increase during oxidative stress in the intracellular pathogen *Francisella tularensis* (Twine *et al.*, 2006).

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

Host-microbe interactions may be particularly important in small Endoriftia

Abundant Endoriftia membrane proteins might play a key role in host interaction in small symbionts. Particularly, the high and differential abundance of porin Sym EGV52132.1, the most abundant symbiont protein in all fractions, which was nearly 3-times more abundant in fraction XS (11.7 %orgNSAF) than in fraction L (4.0 %orgNSAF), suggests that this protein may be of varying relative importance throughout the symbiont's differentiation process. Porins are water-filled channels in the outer membrane, through which small hydrophilic molecules can diffuse (Fernández and Hancock, 2012). In the oyster pathogen *Vibrio splendidus*, the porin OmpU serves as adhesin or invasin and is involved in recognition by the host cell (Duperthuy *et al.*, 2011), while in *Neisseria gonorrhoeae*, a porin inhibits phagocytosis by human immune cells (Mosleh *et al.*, 1998, Lorenzen *et al.*, 2000). Interestingly, the phagocytosis-inhibiting action of *N. gonorrhoeae* porin apparently involves interference with the host's oxidative burst, i.e., the porin allows the pathogen to evade killing by host-produced ROS (Lorenzen *et al.*, 2000). Although the exact function of Endoriftia porin has not been elucidated yet, we suggest that it may have a similar function in resistance against host stress or ROS. This would be in line with elevated levels of ROS scavengers in small *Riftia* symbionts (see above). Porins are furthermore not only known to be involved in recognition by the host (e.g., in the squid symbiont *Vibrio fischeri* (Nyholm *et al.*, 2009)), but were also shown to be involved in survival in and communication with the host in other intracellular and pathogenic bacteria, rendering *Vibrio cholerae* and *Xenorhabdus nematophila* more resistant against antimicrobial compounds (Mathur and Waldor, 2004, van der Hoeven and Forst, 2009). As *Riftia* trophosome tissue has antimicrobial effects (Klose *et al.*, 2016), and considering that *Riftia* might employ histone-derived antimicrobial peptides to modulate the symbiont's cell division (Hinzke *et al.*, 2019),

Endoriftia porin may enable the symbionts to reject antimicrobial compounds produced by the host. This would be of particular importance for small symbionts, as it would ensure survival of the symbiont stem cell subpopulation and sustain their division capability.

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

Microbe-host interactions with particular relevance in smaller Endoriftia may furthermore also be mediated by chaperones and stress proteins, which were most abundant in fraction XS (see above). Chaperones have been shown to play a role in host interaction and intracellular survival in several pathogenic and symbiotic bacteria. For example, DnaK appears to be essential for growth of *Brucella suis* in phagocytes (Köhler *et al.*, 1996), while HtpG seems to be involved in virulence and intracellular survival of *Leptospira* (King *et al.*, 2014), *Salmonella* (Verbrugghe *et al.*, 2015) and *Edwardsiella tarda* (Dang *et al.*, 2011). Mutations in the post-transcriptional regulator *hfq* often lead to reduced fitness and virulence in bacterial pathogens (reviewed in Chao and Vogel, 2010). Moreover, ClpB in *Listeria* is apparently specifically involved in virulence (Chastanet *et al.*, 2004), as are ClpX and ClpP in *Staphylococcus aureus* (Frees *et al.*, 2003). In the insect symbiont *Wolbachia*, HU beta was suggested to directly interact with the host (Beckmann *et al.*, 2013). Additional symbiont proteins that may protect small *Endoriftia* from host interference, and particularly so in S-depleted *Riftia* specimens, included an ankyrin protein and an FK506-binding protein (see Supplementary Results and Discussion B).

Interaction-specific host proteins

We detected a number of 'symbiosis-specific' *Riftia* proteins, which were co-enriched with symbiont cells in fractions XS and/or S and may thus facilitate direct host-microbe interactions

or enable the host to provide optimal conditions for the symbiont. PGRPs, for example, are involved in innate immunity (Kang *et al.*, 1998) and have previously been shown or suggested to participate in symbiotic interactions (Troll *et al.*, 2009, Wang *et al.*, 2009, Royet *et al.*, 2011, Wippler *et al.*, 2016). Since oxygen concentrations in the trophosome might be comparatively low (benefitting the microaerophilic symbionts; Hinzke *et al.*, 2019), the hypoxia up-regulated *Riftia* proteins we detected may present a protective adaptation of the host to these hypoxic conditions. In support of this idea, Hyou1 was shown to have a protective function during hypoxia in human cells (Ozawa *et al.*, 1999). Moreover, enrichment of beta carbonic anhydrase 1, which interconverts bicarbonate and CO_2, suggests that this host protein serves to optimally provide the symbionts with CO_2 for fixation. The host transmembrane protein 214-B (TMP214-B), which was exclusively detected in symbiont-enriched fractions (but not in trophosome homogenate) may be involved in cell death of symbiont-containing bacteriocytes by an apoptosis-related mechanism. This would be in line with our previous suggestion that apoptosis-related proteins may play a role in symbiont and bacteriocyte cell death (Hinzke *et al.*, 2019), and is further supported by the fact that TMP214-B was shown to be involved in apoptosis caused by endoplasmic reticulum stress (Li *et al.*, 2013). The detection of degradation proteins such as cathepsin Z, legumain, glucoamylase 1 and lysosomal alpha-glucosidase in fractions XS and S furthermore implies that the host digests not only large symbiont cells in the degradative trophosome lobule zone (see Figure 5A), but that small symbionts might also be exposed to host digestion.

Metabolic diversity among symbiont size classes

Large symbionts focus on carbon fixation and biosynthesis

Highest individual abundances of various carbon fixation and biosynthesis-related enzymes as well as highest overall abundances of all biosynthetic categories (including carbon-, amino acid-, lipid-, nitrogen- and cofactor metabolism; Supplementary Table S7) in fraction L suggests that large Endoriftia cells are relatively more engaged in the production of organic material than smaller symbiont cells. In support of this idea, we observed notably higher RubisCO mRNA signal intensity in large symbiont cells than in smaller *Riftia* symbionts in our HCR-FISH analysis (Supplementary Results and Discussion D, Supplementary Figure S3). This concurs with an autoradiographic study of Bright *et al.* (2000), who observed highest ^{14}C

carbon incorporation in the *Riftia* trophosome lobule periphery and lowest short-term incorporation in the lobule center. As previously suggested (Hand, 1987), these and other observed differences might be due to a biochemical gradient, which could be caused by the direction of blood flow (from the lobule periphery to the lobule center; Felbeck and Turner, 1995). This presumptive concentration gradient may lead to differential availability of inorganic carbon (and other substrates, see below), which in turn likely results in differential regulation of bacterial gene expression, such as highest abundance of CO_2 incorporation enzymes in large symbionts. Large *Riftia* symbionts thus presumably not only benefit from higher CO_2 levels, but also have more biosynthetic capacities at their disposal than small symbionts: Small Endoriftia need to maintain cell division and, consequently, invest a considerable part of their resources in growth-related processes and the expression of putative host interaction-related proteins that ensure survival of the stem cell population (see above). In contrast, large symbionts apparently divide less frequently and may be less endangered of host interference (before they reach the degenerative lobule zone) and can thus allocate more energy to production of organic material. This would eventually benefit the host, which digests the larger symbionts at the trophosome lobule periphery. Our observation that most cofactor- and vitamin metabolism-related proteins were more abundant in fractions M and/or L than in fractions XS or S supports the idea of relatively more biosynthesis in large Endoriftia. Higher abundances of glycogen-producing enzymes in fraction L furthermore suggest that large symbionts invest relatively more of their biosynthetic capacities in storage of fixed carbon in the form of glycogen than smaller symbionts. This is in accordance with a previous study (Sorgo *et al.*, 2002), which noted a glycogen gradient in the symbiont cells, with increasing glycogen density from the lobule center towards the periphery, i.e., towards larger symbiont cells.

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

Smaller symbionts produce relatively more sulfur globules for sulfur storage than larger symbiont cells, as indicated by relatively higher abundance of sulfur globule proteins in fraction XS (Supplementary Figure S4). This is in agreement with observations of Hand (1987), who noted more sulfur deposits in central (small) than in peripheral (large) *Riftia* symbionts. Although this finding was not supported by a subsequent study (Pflugfelder *et al.*, 2005), our

results do point to different amounts of S storage in different Endoriftia subpopulations. As shown for the free-living thiotrophic model bacterium *Allochromatium vinosum*, activation of stored sulfur involves trafficking proteins such as TusA, which is involved in sulfur transfer to DsrEFH and DsrC (Stockdreher *et al.*, 2014). In our study, the highly abundant TusA, several DsrC copies as well as DsrEFH were all detected with highest abundances in fraction XS, thus supporting the idea of relatively more re-mobilization of sulfur and subsequent utilization of reduced sulfur compounds in small Endoriftia. As the highly abundant adenylylsulfate reductase AprAB, the ATP-sulfurylase SopT and sulfide dehydrogenase subunit FccB were also detected with higher abundances in fractions XS or S than in M or L, one might conclude that sulfide oxidation itself also plays a more prominent role in smaller symbionts than in larger symbionts. However, as we detected the dissimilatory sulfite reductase DsrAB, the third key enzyme of cytoplasmic sulfide oxidation, with a rather ambiguous abundance pattern (Supplementary Table S4e), this idea remains speculative and requires further analysis.

In large symbionts, thiosulfate oxidation plays a more prominent role

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

and were proposed to be involved in thiosulfate oxidation (Hensen *et al.*, 2006, Welte *et al.*, 2009).

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

Hydrogen oxidation is more relevant in large symbionts

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

Endoriftia and why it might be relatively more relevant in larger symbionts remains to be discussed.

Denitrification in Riftia *symbionts appears to be modular*

Our results suggest that small *Riftia* symbionts rely relatively more on the NarGHI-mediated first step of respiratory nitrate reduction to nitrite, while all subsequent steps of nitrite reduction via NO and N_2O to N_2 seem to be more prominent in larger symbionts. Since expression of nitrate reduction genes is usually inhibited by oxygen (Payne, 1973), high NarGHI abundance in fraction XS suggests that O_2 levels might be particularly low in the trophosome lobule center.

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

Regulation of gene expression may be less stringent in large symbionts

Relative abundance of the RNA polymerase sigma factor RpoD decreased from fraction XS to L (Figure 3, Supplementary Table S4h) in S-rich and S-depleted samples, pointing to relatively more growth-related activities in small Endoriftia (see also Supplementary Results and

Discussion B). RpoD is the primary sigma factor for vegetative growth (σ70), which regulates transcription of most genes involved in exponential growth in many bacteria (Helmann and Chamberlin, 1988, Fujita *et al.*, 1994, Ishihama, 2000). This would be in agreement with the idea of small *Riftia* symbionts being mainly occupied with cell division and proliferation in a quasi-exponential growth phase, while large symbionts function as biosynthetic 'factories', focusing on carbon fixation and biomass production. Interestingly, RpoS, the master transcriptional regulator of stationary phase gene expression and antagonist of RpoD, was not detected in any of our samples (although it is encoded in the symbiont genome). RpoS abundance increases upon stress and limitation during transition to the stationary phase in free-living model bacteria (Hengge-Aronis, 1993, Fujita *et al.*, 1994, Ishihama, 2000). Its absence in the *Riftia* symbiont's proteome suggests that, unlike free-living bacteria, the symbiont does not experience a stationary phase-like growth arrest even in later developmental stages, probably because it is ideally supplied with all necessary substrates by the host. This 'lack' of stress or limitation possibly results in less stringent regulation of symbiont gene expression, which could explain the metabolic diversity we observed particularly in large symbionts, such as multiple ways of energy generation (thiosulfate- and hydrogen oxidation in addition to sulfide oxidation) and two CO_2 fixation pathways. Under these premises, the previously observed simultaneous expression of seemingly redundant metabolic pathways in *Riftia* symbionts (Markert *et al.*, 2011) very likely reflects this presumptive "de-regulation" of gene expression in large parts of the symbiont population, which allows Endoriftia to fully exploit its versatile metabolic repertoire to the advantage of the symbiosis.

Conclusion

Our results show that Endoriftia cells of different differentiation stages likely employ distinct metabolic profiles, thus confirming our initial hypothesis. Whereas small Endoriftia ensure survival of the symbiont population, large Endoriftia are primarily engaged in biomass production. The driving force behind this differentiation remains to be elucidated. For *Rhizobium*, a steep O_2 concentration gradient inside legume nodules was proposed to be involved in signaling for symbiont differential gene expression (Soupene *et al.*, 1995). Similarly, some of the differences we observed in small and large Endoriftia might also be

connected to the availability of electron donors or acceptors, and hence differentiation of Endoriftia cells might depend on substrate availability. Symbiont differentiation in *Riftia* might furthermore be induced by specific host effectors, e.g., histone-derived antimicrobial peptides, which were recently proposed to play a role in symbiont cell cycle regulation (Hinzke *et al.*, 2019), or other compounds that allow *Riftia* to modulate the symbiont's expression of certain metabolic pathways. Besides such direct interference, *Riftia* likely also exerts indirect influence on symbiont gene expression by providing copious amounts of all necessary substrates to the bacterial partner. We speculate that this constantly high nutrient availability inside the host causes Endoriftia's biosynthetic pathways to be regulated less stringently (compared to what we would expect in free-living bacteria). This would explain the previously observed metabolic versatility of symbionts in the same host: Large Endoriftia can afford to employ much of their metabolic repertoire at the same time. Such an 'advantageous deregulation', i.e., unhindered expression of multiple – even redundant – metabolic pathways, likely enables high symbiont productivity during symbiosis. We suggest that a symbiont species' differentiation into functionally dissimilar subpopulations in the same host, as shown here for the *Riftia* symbiont, might be an important key to success of microbe-host associations in various environments.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Levine, C., H. Hiasa and K. J. Marians (1998). DNA gyrase and topoisomerase IV: Biochemical activities, physiological roles during chromosome replication, and drug sensitivities. *Biochimica et Biophysica Acta - Gene Structure and Expression* **1400**: 29-43. DOI: 10.1016/S0167-4781(98)00126-2.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Petersen, J. M., F. U. Zielinski, T. Pape, R. Seifert, C. Moraru, R. Amann, S. Hourdez, P. R. Girguis, S. D. Wankel, V. Barbe, E. Pelletier, D. Fink, C. Borowski, W. Bach and N. Dubilier (2011). Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* **476**: 176-180. DOI: 10.1038/nature10325.

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

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

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

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

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

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

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

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

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

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

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

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

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

Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak and A. Cardona (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**(7): 676-682. DOI: 10.1038/nmeth.2019.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Acknowledgements

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

Author contributions

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

Conflicts of interest

The authors declare no conflicts of interest.

6 Bibliography

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, R106. doi:10.1186/gb-2010-11-10-r106.

Ankney, J. A., Muneer, A., and Chen, X. (2018). Relative and absolute quantitation in mass spectrometry–based proteomics. *Annu. Rev. Anal. Chem.* 11, 49–77. doi:10.1146/annurev-anchem-061516-045357.

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., Hentschel, U., Hirt, H., Hülter, N., Lachnit, T., Picazo, D., Pita, L., Pogoreutz, C., Rädecker, N., Saad, M. M., Schmitz, R. A., Schulenburg, H., Voolstra, C. R., Weiland-Bräuer, N., Ziegler, M. and Bosch, T. C. G. (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.

Berg, C. J., and Alatalo, P. (1984). Potential of chemosynthesis in molluscan mariculture. *Aquaculture* 39, 165–179. doi:10.1016/0044-8486(84)90264-3.

Blein-Nicolas, M., and Zivy, M. (2016). Thousand and one ways to quantify and compare protein abundances in label-free bottom-up proteomics. *Biochim. Biophys. Acta* 1864, 883–895. doi:10.1016/j.bbapap.2016.02.019.

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.

Brissac, T., Gros, O., and Merçot, H. (2009). Lack of endosymbiont release by two Lucinidae (Bivalvia) of the genus *Codakia*: consequences for symbiotic relationships. *FEMS Microbiol. Ecol.* 67, 261–267. doi:10.1111/j.1574-6941.2008.00626.x.

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.

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

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.

Cho, I., and Blaser, M. J. (2012). The human microbiome: at the interface of health and disease. *Nat. Rev. Genet.* 13, 260–270. doi:10.1038/nrg3182.

D'Haeze, W., and Holsters, M. (2004). Surface polysaccharides enable bacteria to evade plant immunity. *Trends Microbiol.* 12, 555–561. doi:10.1016/j.tim.2004.10.009.

Degenhardt, F., Seifert, S., and Szymczak, S. (2017). Evaluation of variable selection methods for random forests and omics data sets. *Brief. Bioinform.* 20, 492–503. doi:10.1093/bib/bbx124.

DeNiro, M. J., and Epstein, S. (1978). Influence of diet on the distribution of carbon isotopes in animals. *Geochim. Cosmochim. Acta* 42, 495–506. doi:10.1016/0016-7037(78)90199-0.

Díaz-Uriarte, R., and de Andrés, S. A. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics* 7, 3. doi:10.1186/1471-2105-7-3.

Distel, D. L., Lane, D. J., Olsen, G. J., Giovannoni, S. J., Pace, B., Pace, N. R., Stahl, D. A., and Felbeck, H. (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.

Douglas, A. E. (2014a). Symbiosis as a general principle in eukaryotic evolution. *Cold Spring Harb. Perspect. Biol.* 6, a016113. doi:10.1101/cshperspect.a016113.

Douglas, A. E. (2014b). The molecular basis of bacterial-insect symbiosis. *J. Mol. Biol.* 426, 3830–3837. doi:10.1016/j.jmb.2014.04.005.

Duperthuy, M., Schmitt, P., Garzón, E., Caro, A., Rosa, R. D., Le Roux, F., Lautrédou-Audouy, N., Got, P., Romestand, B., de Lorgeril, J., Kieffer-Jaquinod, S., Bachère, E., and Destoumieux-Garzón, D. (2011). Use of OmpU porins for attachment and invasion of *Crassostrea gigas* immune cells by the oyster pathogen *Vibrio splendidus*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2993–2998. doi:10.1073/pnas.1015326108.

Durand, P., Gros, O., Frenkiel, L., and Prieur, D. (1996). Phylogenetic characterization of sulfur-oxidizing bacterial endosymbionts in three tropical Lucinidae by 16S rDNA sequence analysis. *Mol. Mar. Biol. Biotechnol.* 5, 37–42.

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

Ernst, J., Nau, G. J., and Bar-Joseph, Z. (2005). Clustering short time series gene expression data. *Bioinformatics* 21, i159–i168. doi:10.1093/bioinformatics/bti1022.

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., and Turner, P. J. (1995). CO_2 transport in catheterized hydrothermal vent tubeworms, *Riftia pachyptila* (Vestimentifera). *J. Exp. Zool.* 272, 95–102. doi:10.1002/jez.1402720203.

Fernández, L., and Hancock, R. E. W. (2012). Adaptive and mutational resistance: role of porins and efflux pumps in drug resistance. *Clin. Microbiol. Rev.* 25, 661–681. doi:10.1128/CMR.00043-12.

Flores, J. F., Fisher, C. R., Carney, S. L., Green, B. N., Freytag, J. K., Schaeffer, S. W., and Royer Jr., W. E. (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., Wollherr, A., Kabisch, J., Le Bris, N., Lehmann, R., Daniel, R., Liesegang, H., Hecker, M., and Schweder, T. (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.

Girguis, P. R., Lee, R. W., Desaulniers, N., Childress, J. J., Pospesel, M., Felbeck, H., and Zal, F. (2000). Fate of nitrate acquired by the tubeworm *Riftia pachyptila*. *Appl. Environ. Microbiol.* 66, 2783–2790. doi:10.1128/AEM.66.7.2783-2790.2000.

Gros, O., Darrasse, A., Durand, P., Frenkiel, L., and Moueza, M. (1996). Environmental transmission of a sulfur-oxidizing bacterial gill endosymbiont in the tropical lucinid bivalve *Codakia orbicularis*. *Appl. Environ. Microbiol.* 62, 2324–2330.

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.

Harmer, T. L., Rotjan, R. D., Nussbaumer, A. D., Bright, M., Ng, A. W., DeChaine, E. G., and Cavanaugh, C. M. (2008). Free-living tube worm endosymbionts found at deep-sea vents. *Appl. Environ. Microbiol.* 74, 3895–3898. doi:10.1128/AEM.02470-07.

Heath, M. C. (2000). Nonhost resistance and nonspecific plant defenses. *Curr. Opin. Plant Biol.* 3, 315–319. doi:10.1016/S1369-5266(00)00087-X.

Hinzke, T., Kleiner, M., Breusing, C., Felbeck, H., Häsler, R., Sievert, S. M., Schlüter, R., Rosenstiel, P., Reusch, T. B. H., Schweder, T., and Markert, S. (2019a). Host-microbe interactions in the chemosynthetic *Riftia pachyptila* symbiosis. *mBio* 10, e02243-19. doi:10.1128/mBio.02243-19.

Hinzke, T., Kleiner, M., and Markert, S. (2018). "Centrifugation-based enrichment of bacterial cell populations for metaproteomic studies on bacteria-invertebrate symbioses", in *Microbial Proteomics. Methods in Molecular Biology*, ed. D. Becher (Humana Press, New York, NY), 319–334. doi:10.1007/978-1-4939-8695-8_22.

Hinzke, T., Kleiner, M., Meister, M., Schlüter, R., Hentschker, C., Pané-Farré, J., Hildebrandt, P., Felbeck, H., Sievert, S. M., Bonn, F., Völker, U., Becher, D., Schweder, T., and Markert, S. Bacterial symbiont subpopulations have different roles in a deep-sea symbiosis. *Submitted*, preprint available at bioRxiv. doi:10.1101/2020.04.08.032177.

Hinzke, T., Kouris, A., Hughes, R.-A., Strous, M., and Kleiner, M. (2019b). 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.

Hinzke, T., and Markert, S. (2017). Efficient protein extraction for proteomics and metaproteomics (also suitable for low biomass samples). *protocols.io*. doi:dx.doi.org/10.17504/protocols.io.kg6ctze.

Hollants, J., Leliaert, F., De Clerck, O., and Willems, A. (2013). What we can learn from sushi: a review on seaweed-bacterial associations. *FEMS Microbiol. Ecol.* 83, 1–16. doi:10.1111/j.1574-6941.2012.01446.x.

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.

Jackson, J. B. C. (1973). The ecology of molluscs of *Thalassia* communities, Jamaica, West Indies . I. Distribution, environmental physiology, and ecology of common shallow-water species. *Bull. Mar. Sci.* 23, 313–350.

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. (2019). Metaproteomics: Much more than measuring gene expression in microbial communities. *mSystems* 4, e00115-19. doi:10.1128/msystems.00115-19.

Kleiner, M., Dong, X., Hinzke, T., Wippler, J., Thorson, E., Mayer, B., and Strous, M. (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., and Strous, M. (2017). Assessing species biomass contributions in microbial communities via metaproteomics. *Nat. Commun.* 8, 1558. doi:10.1038/s41467-017-01544-x.

Klose, J., Polz, M. F., Wagner, M., Schimak, M. P., Gollner, S., and Bright, M. (2015). Endosymbionts escape dead hydrothermal vent tubeworms to enrich the free-living population. *Proc. Natl. Acad. Sci. U.S.A.* 112, 11300–11305. doi:10.1073/pnas.1501160112.

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

König, S., Gros, O., Heiden, S. E., Hinzke, T., Thürmer, A., Poehlein, A., Meyer, S., Vatin, M., Mbéguié-A-Mbéguié, D., Tocny, J., Ponnudurai, R., Daniel, R., Becher, D., Schweder, T., and Markert, S. (2016). Nitrogen fixation in a chemoautotrophic lucinid symbiosis. *Nat. Microbiol.* 2, 16193. doi:10.1038/nmicrobiol.2016.193.

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.

Lindemann, C., Thomanek, N., Hundt, F., Lerari, T., Meyer, H. E., Wolters, D., and Marcus, K. (2017). Strategies in relative and absolute quantitative mass spectrometry based proteomics. *Biol. Chem.* 398, 687–699. doi:10.1515/hsz-2017-0104.

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

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

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

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

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

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

McFall-Ngai, M. (2014). Divining the essence of symbiosis: insights from the squid-vibrio model. *PLoS Biol.* 12, e1001783. doi:10.1371/journal.pbio.1001783.

McFall-Ngai, M., Hadfield, M. G., Bosch, T. C. G., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S. F., Hentschel, U., King, N., Kjelleberg, S., Knoll, A. H., Kremer, N., Mazmanian, S. K., Metcalf, J. L., Nealson, K., Pierce, N. E., Rawls, J. F., Reid, A., Ruby, E. G., Rumpho, M., Sanders, J. G., Tautz, D. and Wernegreen, J. J. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad. Sci. U.S.A.* 110, 3229–3236. doi:10.1073/pnas.1218525110.

Mosleh, I. M., Huber, L. A., Steinlein, P., Pasquali, C., Günther, D., and Meyer, T. F. (1998). *Neisseria gonorrhoeae* porin modulates phagosome maturation. *J. Biol. Chem.* 273, 35332–35338. doi:10.1074/jbc.273.52.35332.

Muth, T., Benndorf, D., Reichl, U., Rapp, E., and Martens, L. (2013). Searching for a needle in a stack of needles: challenges in metaproteomics data analysis. *Mol. BioSyst.* 9, 578–585. doi:10.1039/c2mb25415h.

Nahnsen, S., Bielow, C., Reinert, K., and Kohlbacher, O. (2013). Tools for label-free peptide quantification. *Mol. Cell. Proteomics* 12, 549–556. doi:10.1074/mcp.r112.025163.

Nesvizhskii, A. I., and Aebersold, R. (2005). Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteomics* 4, 1419–1440. doi:10.1074/mcp.R500012-MCP200.

Ning, K., Fermin, D., and Nesvizhskii, A. I. (2012). Comparative analysis of different label-free mass spectrometry based brotein abundance estimates and their correlation with RNA-Seq gene expression data. *J. Proteome Res.* 11, 1–4. doi:10.1021/pr201052x.

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

Pearson, A. (2010). "Pathways of carbon assimilation and their impact on organic matter values $\delta^{13}C$", in *Handbook of Hydrocarbon and Lipid Microbiology*, ed. K. N. Timmis (Berlin, Heidelberg: Springer), 143–156. doi:10.1007/978-3-540-77587-4_9.

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

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

Pjevac, P., Meier, D. V., Markert, S., Hentschker, C., Schweder, T., Becher, D., Gruber-Vodicka, H. R., Richter, M., Bach, W., Amann, R., and Meyerdierks, A. (2018). Metaproteogenomic profiling of microbial communities colonizing actively venting hydrothermal chimneys. *Front. Microbiol.* 9, 680. doi:10.3389/fmicb.2018.00680.

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

Ponnudurai, R., Kleiner, M., Sayavedra, L., Petersen, J. M., Moche, M., Otto, A., Becher, D., Takeuchi, T., Satoh, N., Dubilier, N., Schweder, T., and Markert, S. (2017). Metabolic and physiological interdependencies in the *Bathymodiolus azoricus* symbiosis. *ISME J.* 11, 463–477. doi:10.1038/ismej.2016.124.

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

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

Sagan, L. (1967). On the origin of mitosing cells. *J. Theor. Biol.* 14, 225–274. doi:10.1016/0022-5193(67)90079-3.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., and Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. doi:10.1038/nmeth.2019.

Schneider, T., and Riedel, K. (2010). Environmental proteomics: studying structure and function of microbial communities. *Proteomics* 10, 785–798. doi:10.1002/pmic.200900450.

Schneider, T., Schmid, E., de Castro Jr., J. V., Cardinale, M., Eberl, L., Grube, M., Berg, G., and Riedel, K. (2011). Structure and function of the symbiosis partners of the lung lichen (*Lobaria pulmonaria* L. Hoffm.) analyzed by metaproteomics. *Proteomics* 11, 2752–2756. doi:10.1002/pmic.201000679.

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

Sokolov, E. P., Markert, S., Hinzke, T., Hirschfeld, C., Becher, D., Ponsuksili, S., and Sokolova, I. M. (2019). Effects of hypoxia-reoxygenation stress on mitochondrial proteome and bioenergetics of the hypoxia-tolerant marine bivalve *Crassostrea gigas*. *J. Proteomics* 194, 99–111. doi:10.1016/j.jprot.2018.12.009.

Sorgo, A., Gaill, F., Lechaire, J.-P., Arndt, C., and Bright, M. (2002). Glycogen storage in the *Riftia pachyptila* trophosome: contribution of host and symbionts. *Mar. Ecol. Prog. Ser.* 231, 115–120. doi:10.3354/meps231115.

Sowell, S. M., Abraham, P. E., Shah, M., VerBerkmoes, N. C., Smith, D. P., Barofsky, D. F., and Giovannoni, S. J. (2011). Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. *ISME J.* 5, 856–865. doi:10.1038/ismej.2010.168.

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

Timmins-Schiffman, E., May, D. H., Mikan, M., Riffle, M., Frazar, C., Harvey, H. R., Noble, W. S., and Nunn, B. L. (2017). Critical decisions in metaproteomics: achieving high confidence protein annotations in a sea of unknowns. *ISME J.* 11, 309–314. doi:10.1038/ismej.2016.132.

VerBerkmoes, N. C., Denef, V. J., Hettich, R. L., and Banfield, J. F. (2009). Functional analysis of natural microbial consortia using community proteomics. *Nat. Rev. Microbiol.* 7, 196–205. doi:10.1038/nrmicro2080.

Verheggen, K., Ræder, H., Berven, F. S., Martens, L., Barsnes, H., and Vaudel, M. (2017). Anatomy and evolution of database search engines-a central component of mass spectrometry based proteomic workflows. *Mass Spectrom. Rev.*, 1–15. doi:10.1002/mas.21543.

Wang, D.-Z., Xie, Z.-X., and Zhang, S.-F. (2014). Marine metaproteomics: current status and future directions. *J. Proteomics* 97, 27–35. doi:10.1016/j.jprot.2013.08.024.

Wilmes, P., and Bond, P. L. (2004). The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environ. Microbiol.* 6, 911–920. doi:10.1111/j.1462-2920.2004.00687.x.

Wilmes, P., Heintz-Buschart, A., and Bond, P. L. (2015). A decade of metaproteomics: where we stand and what the future holds. *Proteomics* 15, 3409–3417. doi:10.1002/pmic.201500183.

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

Woyke, T., Teeling, H., Ivanova, N. N., Huntemann, M., Richter, M., Gloeckner, F. O., Boffelli, D., Anderson, I. J., Barry, K. W., Shapiro, H. J., Szeto, E., Kyrpides, N. C., Mussmann, M., Amann, R., Bergin, C., Ruehland, C., Rubin, E. M., and Dubilier, N. (2006). Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* 443, 950–955. doi:10.1038/nature05192.

Xiong, W., Abraham, P. E., Li, Z., Pan, C., and Hettich, R. L. (2015a). Microbial metaproteomics for characterizing the range of metabolic functions and activities of human gut microbiota. *Proteomics* 15, 3424–3438. doi:10.1002/pmic.201400571.

Xiong, W., Giannone, R. J., Morowitz, M. J., Banfield, J. F., and Hettich, R. L. (2015b). Development of an enhanced metaproteomic approach for deepening the microbiome characterization of the human infant gut. *J. Proteome Res.* 14, 133–141. doi:10.1021/pr500936p.

Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C., and Yates III, J. R. (2013). Protein analysis by shotgun/bottom-up proteomics. *Chem. Rev.* 113, 2343–2394. doi:10.1021/cr3003533.

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.

Appendix C: List of Publications and Scientific Contributions

Peer-reviewed manuscripts

Assié, A., Leisch, N., Meier, D. V., Gruber-Vodicka, H., Tegetmeyer, H. E., Meyerdierks, A., Kleiner, M., **Hinzke, T.**, Joye, S., Saxton, M., Dubilier, N., Petersen, J. M. Horizontal acquisition of a patchwork Calvin cycle by symbiotic and free-living Campylobacterota (formerly Epsilonproteobacteria). ISME J 14, 104-122 (2020). doi: 0.1038/s41396-019-0508-7.

Ponnudurai, R., Heiden, S. E., Sayavedra, L., **Hinzke, T.**, Kleiner, M., Hentschker, C., Felbeck, H., Sievert, S. M., Schlüter, R., Becher, D., Schweder, T., Markert, S. Comparative proteomics of related symbiont mussel species reveals high variability of host-symbiont interactions. ISME J 14, 649-656 (2020). doi: 0.1038/s41396-019-0517-6.

Hinzke, T., Kleiner, M., Breusing, C., Felbeck, H., Häsler, R., Sievert, S.M., Schlüter, R., Rosenstiel, P., Reusch, T. B. H., Schweder, T., Markert, S. Host-microbe interactions in the chemosynthetic Riftia pachyptila symbiosis. mBio 10, e02243-19 (2019). doi: 10.1128/mBio.02243-19

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

Sokolov, E. P., Markert, S., **Hinzke, T.**, Hirschfeld, C., Becher, D., Ponsuksili, S., Sokolova, I. M. Effects of hypoxia-reoxygenation stress on mitochondrial proteome and bioenergetics of the hypoxia-tolerant marine bivalve *Crassostrea gigas*. J Proteomics 194, 99-111 (2019). doi: 10.1016/j.jprot.2018.12.009

Gruber-Vodicka, H. R., Leisch, N., Kleiner, M., **Hinzke, T.**, Liebeke, M., McFall-Ngai, M., Hadfield, M. G., Dubilier, N. Two intracellular and cell type-specific bacterial symbionts in the placozoan *Trichoplax* H2. Nat. Microbiol. 4, 1465-1474 (2019). doi: 10.1038/s41564-019-0475-9

Kleiner, M., Dong, X., **Hinzke, T.**, Wippler, J., Thorson, E., Mayer, B., Strous, M. Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. PNAS 115, E5576-E5584 (2018). doi: 10.1073/pnas.1722325115

Unfried, F., Becker, S., Robb, C. S., Hehemann, J.-H., Markert, S., Heiden, S.E., **Hinzke, T.**, Becher, D., Reintjes, G., Krüger, K., Avcı, B., Kappelmann, L., Hahnke, R.L., Fischer, T., Harder, J., Teeling, H., Fuchs, B., Barbeyron, T., Amann, R.I., Schweder, T. Adaptive mechanisms that provide competitive advantages to marine bacteroidetes during microalgal blooms. ISME J 12, 2894-2906 (2018). doi: 10.1038/s41396-018-0243-5.

König, S., Gros, O., Heiden, S. E., **Hinzke, T.**, Thürmer, A., Poehlein, A., Meyer, S., Vatin, M., Mbéguié-A-Mbéguié, D., Tocny, J., Ponnudurai, R., Daniel, R., Becher, D., Schweder, T., Markert, S. Nitrogen fixation in a chemoautotrophic lucinid symbiosis. Nat. Microbiol. 2, 16193 (2016). doi: 10.1038/nmicrobiol.2016.193

Karl, I., Becker, M., **Hinzke, T.**, Mielke, M., Schiffler, M., Fischer, K. Interactive effects of acclimation temperature and short-term stress exposure on resistance traits in the butterfly *Bicyclus anynana*. Physiol. Entomol. 39, 222-228 (2014). doi: 10.1111/phen.12065

Book chapter

Hinzke, T., Kleiner, M., Markert, S. Centrifugation-based enrichment of bacterial cell populations for metaproteomic studies on bacteria-invertebrate symbioses. In Becher, D. (ed), Methods in Molecular Biology: Microbial Proteomics, Humana Press, New York, 2018. doi: 10.1007/978-1-4939-8695-8_22

Submitted manuscript

Hinzke, T., Kleiner, M., Meister, M., Schlüter, R., Hentschker, C., Pané-Farré, J., Hildebrandt, P., Felbeck, H., Sievert, S. M., Bonn, F., Völker, U., Becher, D., Schweder, T., Markert, S. Bacterial symbiont subpopulations have different roles in a deep-sea symbiosis. Submitted. (Preprint available at bioRxiv, doi: 10.1101/2020.04.08.032177)

Protocol

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

Conferences and Scientific Presentations

11.2018	**Invited Talk:** Symposium "Let's talk about symbiosis" (LTAS), Vienna, Austria
08.2018	**Talk**: Symposium "Microbial Interactions in Marine Systems" (MIMAS II), Greifswald, Germany
04.2018	**Talk**: "International Symposium on Microbial Sulfur Metabolism" (**ISMSM 5**), Vienna, Austria
11.2017	**Invited Talk**: Symbiosis Group (Prof. Nicole Dubilier), MPI Bremen
08.2017	**Poster**: Symposium "Chemosynthesis-Based Ecosystems" (CBE6), Woods Hole, USA
07.2015	**Poster**: "International Symbiosis Society Congress" (ISS 8), Lisbon, Portugal

Appendix D: Note of Thanks

This thesis was made possible by the support and assistance of many people, to whom I wish to express my deepest gratitude.

I would like to thank Prof. Dr. Thomas Schweder for the opportunity to work with the fascinating *Riftia* symbiosis and for giving me the freedom to pursue my own ideas and strategies to tackle the arising research questions, also by enabling me to attend international workshops and conferences.

I am indebted to Dr. Stephanie Markert, who was always supportive during the work on my thesis, teached me wetlab and writing skills, was an example for me for organization and planning, and made it possible for me to participate in not one, but two amazing international research cruises.

I am honored to have seen the *Riftia* symbiosis in its natural habitat, the deep sea. With this thesis I strove to follow in the footsteps of those who made and are making the fascination of *Riftia*, and of symbioses in general, available to all.

My colleagues deserve my thanks for all the time spend together in the pursuit of knowledge. Without the assistance of Jana Matulla, I would probably still be totally lost in the lab(s). Or in the storage room. Frank Unfried and Dr. Ruby Ponnudurai have eased my start in the lab and in the office, and brought color and spirit to some otherwise probably rather quiet office hours. Dr. Norma Welsch and Sandy Gerschler showed that even Endoriftia proteins yield to heterologous gene expression. Thanks to Dr. Marie Zühlke for many inspiring discussions, especially late on Fridays. Herr Bartosik (aka Daniel), I will miss your dust-dry humor! Stefan Heiden, thank you for helping me with my PC issues. Many thanks also to Beate Zillmer and Vera Leesch, who helped me with regard to shipments, formulas, travel applications, contracts ...

I would like to thank Prof. Dr. Dörte Becher, Dr. Florian Bonn, and Dr. Christian Hentschker for the MS analyses of many, many *Riftia* samples.

Prof. Dr. Manuel Kleiner supported my thesis from the start with ideas, insights, and interest, and, together with Prof. Dr. Marc Strous, enabled me an unforgettable and fruitful research stay in Canada. Jackie Zorz, Maryam Ataeian, and Dr. Angela Kouris had no small part in making this stay unforgettable – together with all of EBG. Thank you! And the mass spec is still (well, again) running!

Dr. Stefan Sievert, Prof. Dr. Horst Felbeck, the crews of RV Atlantis, DSV Alvin, and ROV Jason, and the science parties of several research cruises to the East Pacific Rise have not only enabled me to obtain my *Riftia* samples, but also, during the two cruises which I attended, given me a lifetime's worth of enthusiasm, stories, and making-do-(on-a-ship-)know-how. My special thanks to Horst for picking a table cloth which not only hid blood stains (of *Riftia*, of course), but also made finding tube lids an adventure!

With regard to the *Riftia* host database, I owe my thanks to Dr. Corinna Breusing, Prof. Dr. Thorsten Reusch, Prof. Dr. Robert Häsler, Prof. Dr. Philip Rosenstiel, and the IKMB team. And thank you, Corinna, for patiently and with incredible speed answering each and every of our questions!

Insights into the actual tubeworm symbiosis were made possible with the help of Dr. Rabea Schlüter, Dr. Jan Pané-Farré, Annette Meuche, Alexander Graf, and Mareike Meister. Thank you for making my objects of study visible!

Thanks to Prof. Dr. Uwe Völker and Dr. Petra Hildebrandt, we were able to check our polyploidy hypothesis with flow cytometry analyses.

I am grateful to Prof. Dr. Inna Sokolova for giving me the opportunity to participate in yet another intriguing project, which also brought my attention to random forests, and thanks to Elisa Kasbohm for introducing this technique to me!

Many thanks to the KarriereWegeMentoring program, especially to Annette Ehmler, Angela Hoppe, and my peer group, as well as to my mentor Prof. Dr. Nicole Dubilier and her group for enthusiastic and inspiring discussions. The meetings and visits in Bremen were a great source of optimism and motivation.

Above all, my gratitude goes to my family. Especially to my daughter – many a part of this thesis was written together with you. And to Elischa, for your relentless support, encouragement, and patience. Thank you for being here.

Thank you!

Funding Acknowledgements

This thesis was supported by the Deutsche Forschungsgemeinschaft DFG (grant MA 6346/2-1 to Stephanie Markert), the Institute of Marine Biotechnologie e.V. IMaB (stipends to Tjorven Hinzke) and the Deutscher Akademischer Austauschdienst DAAD (travel grant to Tjorven Hinzke). For funding of the individual studies, please refer to the respective publications.