Evolution and physiology of the
Paracatenula symbiosis

Dissertation

Zur Erlangung des Grades eines

Doktors der Naturwissenschaften

- Dr. rer. nat. -

dem Fachbereich Biologie/Chemie der

Universität Bremen

vorgelegt von

Oliver Jäckle

Bremen

November 2018

Die Untersuchungen zur vorliegenden Doktorarbeit wurden in der Abteilung Symbiose am Max-Planck-Institut für Marine Mikrobiologie unter der Leitung von Prof. Dr. Nicole Dubilier und Dr. Harald Gruber-Vodicka als direktem Betreuer durchgeführt.

Gutachter:

Prof. Dr. Nicole Dubilier

Prof. Dr. Andreas Schramm

Tag des Promotionskolloquiums:

10.12.2018

„The good thing about science is that it's true whether or not you believe in it."

– Neil deGrasse Tyson –

Contents

Summary

My thesis focuses on the symbiosis between the marine flatworm host *Paracatenula*, which lacks mouth and digestive system, and the alphaproteobacterial symbiont *Candidatus* Riegeria. Both host and symbiont are unique partners across all known symbioses – no other flatworm besides *Paracatenula* hosts chemosynthetic symbionts, and *Ca.* Riegeria is the only known chemosynthetic symbiont within the *Alphaproteobacteria*. Contained in specialized cells within the trophosome of *Paracatenula*, the *Ca.* Riegeria symbionts can comprise up to a half of the total host volume, representing the highest ratio of symbiont-to-host biomass among animal-bacteria symbioses. *Ca*. Riegeria have been transmitted vertically between host generations for at least 500 million years, leading to the presumption that their genomes contain signatures of reductive genome evolution. Yet only little is known about the genetic repertoire of *Ca*. Riegeria, their ecophysiology and how reductive genome evolution progressed over their evolutionary history.

I have characterized key aspects of the *Paracatenula* symbiosis emphasizing the physiology of *Ca.* Riegeria and how it compares to "typical" gammaproteobacterial chemoautotrophs, and describe the consequences of reductive genome evolution on retained functions in the symbionts. To perform an integrated multi-disciplinary study of the *Paracatenula-Ca.* Riegeria physiology (Chapter 2), I used a highly abundant *Paracatenula* species from Mediterranean sediments (*Paracatenula* sp. standrea). I succeeded to cultivate stable populations of this species under laboratory conditions. The genome of its symbiont, *Ca.* R. standrea, is drastically reduced when compared to its closest alphaproteobacterial relatives and yet, they are fully autonomous for energy generation and biomass formation. *Ca.* R. standrea has a unique physiology which unites versatility and energy efficiency and therefore differs from all other chemosynthetic symbionts with similarly reduced genomes. Genomic predictions were supported by correlative imaging, expression data, metabolomics and physiological experiments, indicating that despite its compact genome, the symbiont can provide flexible carbon and energy storage support for its host. Analyses of symbiont and host gene expression along with ultrastructural analysis of the host suggest that the flatworm obtains its nutrition through digestion of either vesicles derived from the symbionts or entire symbiont cells.

For the first time, I describe the processes of genome reduction in geographically and phylogenetically diverse chemosynthetic symbiont clades (Chapter 3). I deciphered the gene fragmentations and genome rearrangements of three *Ca.* Riegeria clades which have been separated hundreds of millions of years ago. My results provide evidence for divergent trajectories of genome evolution in these clades. I conclude that genome reduction patterns in *Ca.* Riegeria unlikely reflect subsequent stages of the reduction process like it has been proposed previously for other symbionts.

Characteristic for many flatworms including *Paracatenula* is their enormous regenerative capacity which is associated with asexual reproduction. I investigated the role of the symbionts in tissue regeneration of the host (Chapter 4). Surprisingly, I observed massive changes of gene expression in symbionts during head regeneration of host animals.
Overall, I demonstrate that *Paracatenula* and *Ca.* Riegeria share an intimate biological connection that has been shaped over their long evolutionary history.

Zusammenfassung

Die vorliegende Doktorarbeit beschreibt die Symbiose zwischen dem mund- und darmlosen marinen Plattwurm *Paracatenula* und seinem alphaproteobakteriellen Symbionten *Candidatus* Riegeria (*Ca*. Riegeria). Sowohl der Wirt als auch die Symbionten sind unter den Symbiosen einzigartig. *Paracatenula* ist der einzige bekannte Plattwurm der in Symbiose mit chemosynthetischen Symbionten lebt und diese sind wiederum die einzigen chemosynthetischen Bakterien aus der Gruppe der Alphaproteobakterien. Die Symbionten, die in spezialisierten Zellen innerhalb des Trophosoms leben, können bis zur Hälfte der Biomasse des Wirts ausmachen – mehr als in jeder anderen bekannten Symbiose zwischen einem Tier und Bakterien. Da die Symbionten seit mindestens 500 Millionen Jahre von einer Wirtsgeneration zur nächsten übertragen wurden, ist zu erwarten, dass die Symbiontengenome auf für die Symbiose essentiellen Funktionen hin angepasst wurden. Sowohl über die Ökophysiologie und das genetische Repertoire von *Ca.* Riegeria als auch über die Prozesse, die zur Genomreduktion über einen solch langen Zeitraum führen, ist bisher wenig bekannt.

Ich habe Schlüsselaspekte der *Paracatenula* Symbiose untersucht. Dabei lag das Hauptgewicht meiner Arbeit auf Untersuchungen zur genomabgeleiteten Physiologie von *Ca.* Riegeria, vor allem im Vergleich zu „typischen" gammaproteobakteriellen Symbionten und auf einer Analyse der Genomreduktionsvoränge in verschiedenen Kladen. Es ist mir gelungen, Plattwürmer einer abundanten *Paracatenula* Art (*Paracatenula* sp. standrea) aus mediterranen Sedimenten über Jahre hinweg im Labor in Kultur zu halten und mit multidisziplinären Ansätzen zu untersuchen. Trotz des im Vergleich zu freilebenden Nahverwandten stark reduzierten Genoms konnte ich zeigen, dass *Ca.* R. standrea sowohl bezüglich der Energiegewinnung als auch Biomasseproduktion autonom ist (Kapitel 2). Einzigartig im Vergleich zu anderen chemoautotrophen Symbionten mit ähnlich stark reduziertem Genom ist *Ca.* R. standrea in Bezug auf metabolische Flexibilität und Energieffizienz. Die aus den Genomdaten abgeleiteten Schlussfolgerungen wurden durch Ergebnisse korrelativer Bildgebung, Expressionsdaten, Metabolomik und durch physiologische Experimente unterstützt. Ein wichtiger Befund ist, dass die Symbionten trotz eines kompakten Genoms Kohlenstoff und Energiespeicherstoffe dem Wirt zur Verfügung stellen können. Symbionten- und Wirtsexpressionsdaten kombiniert mit Ultrastukturanalysen zeigten außerdem, dass die Symbionten ihren Wirt entweder über Membranvesikel mit Nahrung versorgen oder dass sich der Wirt durch Verdau der gesamten Symbiontenzellen mit Nahrung bzw. Energie versorgt.

Genomreduktionen wurden in geographisch und phylogenetisch unterschiedlichen chemosynthetischen Symbiontenkladen erstmalig am Beispiel von drei *Ca.* Riegeria Kladen untersucht, die schon seit einigen hundert Millionen Jahren getrennt sind (Kapitel 3). Die beschriebenen Genfragmentierungen bzw. Genumlagerungen deuten auf unterschiedliche Ablaufprozesse seit der Trennung der Kladen hin. Ich postuliere daher, dass die Genomreduktion in *Ca.* Riegeria Symbionten nicht nach einem systematischen Einzelschrittschema verläuft, wie dies für andere Symbionten beschrieben ist, sondern unabhängig und Kladen-spezifisch.

Ein charakteristisches Merkmal vieler Plattwürmer, einschließlich *Paracatenula*, ist ihre enorme Fähigkeit zur Regeneration, die häufig mit asexueller Fortpflanzung einhergeht. Ich habe die Notwendigkeit und Rolle der *Paracatenula* Symbionten bei der Regeneration ihres Wirten untersucht (Kapitel 4). Überraschenderweise wurden während der Regeneration des Wirts deutliche Genexpressionsänderungen bei den Symbionten festgestellt, die funktionelle Rückschlüsse erlauben.

Insgesamt weisen meine Ergebnisse darauf hin, dass *Paracatenula* und *Ca.* Riegeria eine intime biologische Gemeinschaft darstellen, die durch einen langen gemeinsam Evolutionsprozess geformt und optimiert wurde.

List of abbreviations

3-HPB	3-hydroxypropionate bi-cycle
AAI	Amino acid identity
ATP	Adenosine triphosphate
CARD-FISH	Catalyzed reporter deposition-fluorescence *in situ* hybridization
CBB cycle	Calvin-Benson-Bassham cycle
COG	Clusters of Orthologous Genes
DIC	Differential interference contrast
dsrAB	Dissimilatory sulfite reductase genes A and B
GC-MS	Gas chromatography-mass spectrometry
gGC	Genomic GC content
HGT	Horizontal gene transfer
KEGG	Kyoto Encyclopedia of Genes and Genomes
LCA	Last common ancestor
Mb	Megabase pairs
mRNA	Messenger ribosomal ribonucleic acid
NMDS	Non-metric multidimensional scaling
NSAF	Normalized spectral abundance factor
OMV	Outer membrane vesicle
PHA	Polyhydroxyalkanoates
PHB	Polyhydroxybutyrate
PP$_i$	Pyrophosphate
rDSR	Reverse-acting dissimilatory sulfite reductase
rRNA	Ribosomal ribonucleic acid
rTCA cycle	Reverse TCA cycle
RuBisCO	Ribulose-1,5-bisphosphate carboxylase/oxygenase
SNP	Single-nucleotide polymorphism
SOX	Sulfur oxidation pathway
TCA cycle	Tricarboxylic acid (TCA) cycle
TEM	Transmission electron microscopy
TMM	Trimmed mean of M-values
TPM	Transcripts per million
tRNA	Transfer RNA

Chapter I: Introduction

An introduction into symbiosis

Interactions between organisms can be short when they involve competition for food or predation. In other cases, two or more different species do not only coexist but rather live in close association and interdependently for longer periods – a lifestyle referred to as "symbiosis" (de Bary, 1879). The first perception of symbiosis was more than 100 years ago in lichens, which are composed of two independent organisms, i.e. a fungus and algae, that live together.

The word symbiosis derives from the Greek *syn* "together" and *bios* "life". It was coined in 1877 by Albert Bernhard Frank with the word *symbiotismus* (Frank, 1877) and is used when two species establish a long-term biocoenosis. A more elaborated concept by de Bary included any associations between different species with persistent physical contact and can broadly be divided into groups that have positive (= mutualism), neutral (= commensalism) or negative (= parasitism) impacts on the partners involved (de Bary, 1879). Although de Bary explicitly included all three kinds of interaction, the term "symbiosis" was often misinterpreted by biologists as a synonym for mutualistic and cooperative interactions (Martin and Schwab, 2012). Another source of inconsistency is the definition of symbiosis as an association for a "longer period". At the present time, it is widely accepted that symbiotic organisms live together for at least a substantial portion of their lives but not necessarily their complete lifespan (Douglas, 2010). In my thesis I will use the term "symbiosis" when I refer to mutualistic interactions between a larger organism such as the animal host, and a smaller organism such as a bacterial partner to which I refer as symbionts.

A symbiotic lifestyle represents a successful biological strategy that has evolved among all kinds of living organisms and is ubiquitous in terrestrial, freshwater and marine communities dominating the biota of many habitats (Moran, 2006; Douglas, 2010). Symbiosis often has played a key role in the evolution of complex and multicellular life. For example, mitochondria and chloroplasts evolved from free-living bacteria through symbiosis within a eukaryotic host cell (Margulis, 1970; Margulis and Fester, 1991). Such endosymbiosis events paved the way for the evolution of a broad eukaryotic diversity in animals, plants as well as fungi.

Most commonly studied interactions include symbioses between bacterial symbionts and their hosts, but numerous symbioses also exist between higher organisms (Figure 1A–I). In the

famous symbiosis between sea anemones and clown fish, for example, the anemone host serve the fish as shelter where the fish is protected by the stingy nematocysts in the tentacles of the host (Figure 1A). In return, the clown fish allow their host to flourish. Another mutualistic symbiosis involves the leafcutter ants. They inoculate pieces of leaves to serve as the nutritional substrate for their fungal cultivates. In return for a mulchy compost of fresh leaves, the fungus breaks down the plant material to produce nutritious, edible structures which the ants can digest. In return, the ants ensure that the fungus stays disease-free (Figure 1C). Moreover, nitrogen fixation *via* bacterial symbionts allows legumes to inhabit soils with limited nitrogen availability, and at the same time establishes a distinctive bacterial community with essential effects on plant growth and the plant´s ecological performance (Figure 1E) (Zgadzaj *et al.*, 2016).

The association with symbiotic bacteria allows the hosts to thrive and adapt to new niches that would not be possible aposymbiotically (Little, 2010). One reason for these adaptations is the enhancement of the host´s ability to obtain nutrients from bacterial symbionts (Moran, 2006).
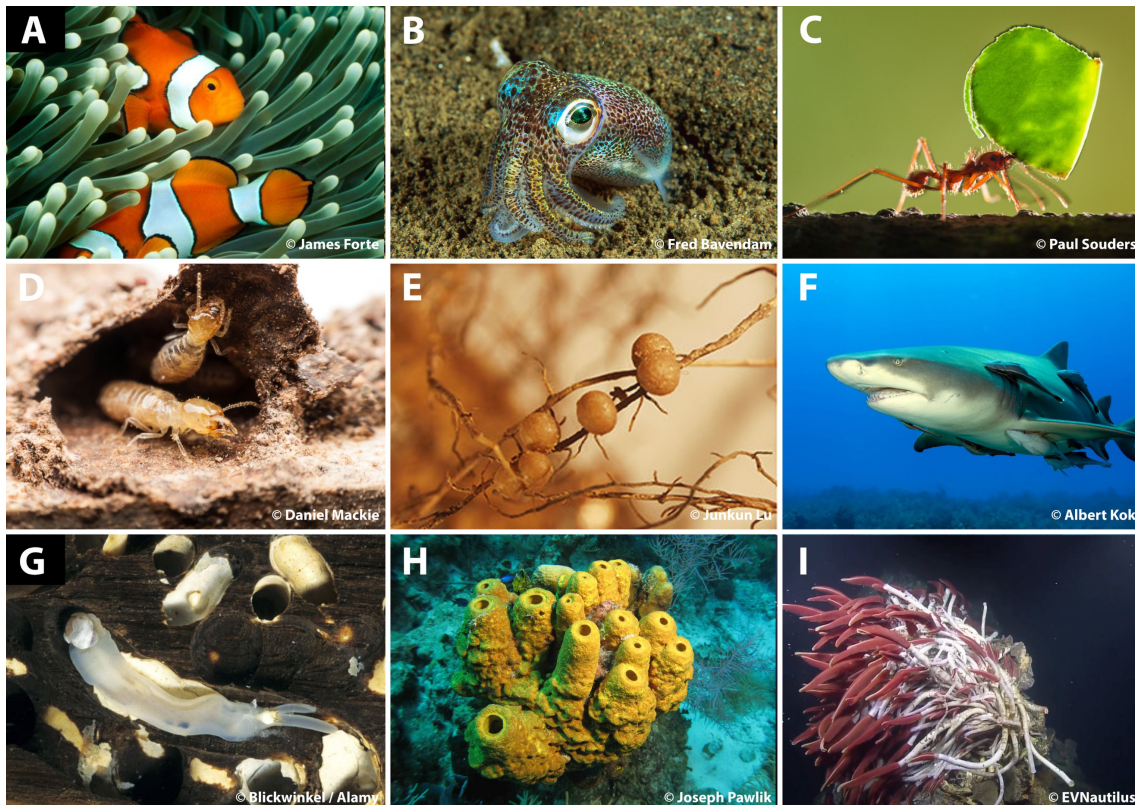
Figure 1. An overview of the diversity of symbiotic interactions. ***A***, Clown fish living in protective anemones. ***B***, Squid with bacterial symbionts located in the light organ that allow camouflage. ***C***, Leafcutter ant that provides fungi they cultivate with leaves. ***D***, ***E***, Termites and root nodules that both harbor nitrogen-fixing symbionts. ***F***, Hitch-riding remora fishes on a shark. ***G***, Ship-boring bivalves that harbor cellulose-digesting symbionts. ***H***, Sponge harbor bacteria with various functions ranging from antibiotic resistance to nutrition. ***I***, Giant tubeworms harbor chemosynthetic symbionts. Source of the photographs is indicated on the images.

The definition of chemosynthesis

Shortly after the term symbiosis was coined, a process called chemosynthesis was discovered by the biologist Sergei Winogradsky (Winogradsky, 1887). He was puzzled by the presence of sulfur in cells of the bacterium *Beggiatoa* and he asked what it might be used for (Winogradsky, 1887; Dworkin, 2012). His discoveries were pioneering as he suggested that: "[…] *the sulfur in these organisms is the sole respiratory source, and in that sense plays the same role as that of carbohydrate in other organisms* […]" (Winogradsky, 1887). This was surprising at that time since it was believed that all life on Earth was dependent on sunlight as the sole source of energy and on photosynthesis driving the primary production of organic carbon. By now, however, sulfur is also well known to function in various biogeochemical reactions that affect both the carbon and oxygen cycles (Hurtgen, 2012). In fact, free-living bacterial communities are capable of using organic compounds to reduce the oxidized form of sulfur – which is primarily sulfate – to sulfide which serves as terminal electron acceptor in the absence of

oxygen. In the oxidative portion of the sulfur cycle, sulfur-oxidizing bacteria such as *Beggiatoa* are capable of synthesizing organic compounds from the oxidation of reduced sulfur species, closing the sulfur cycle in another way of primary production (Van Dower, 2000). This microbial process is referred to as chemosynthesis, since chemical energy instead of sunlight drives the synthesis of organic carbon through the fixation of inorganic carbon (Figure 2). The biogeochemical significance of chemosynthesis, and thus giving the answer to Winogradsky´s question, was finally understood with the discovery of chemosynthetic-driven primary production at deep-sea hydrothermal vent systems (Jannasch, 1985; Van Dower, 2000).
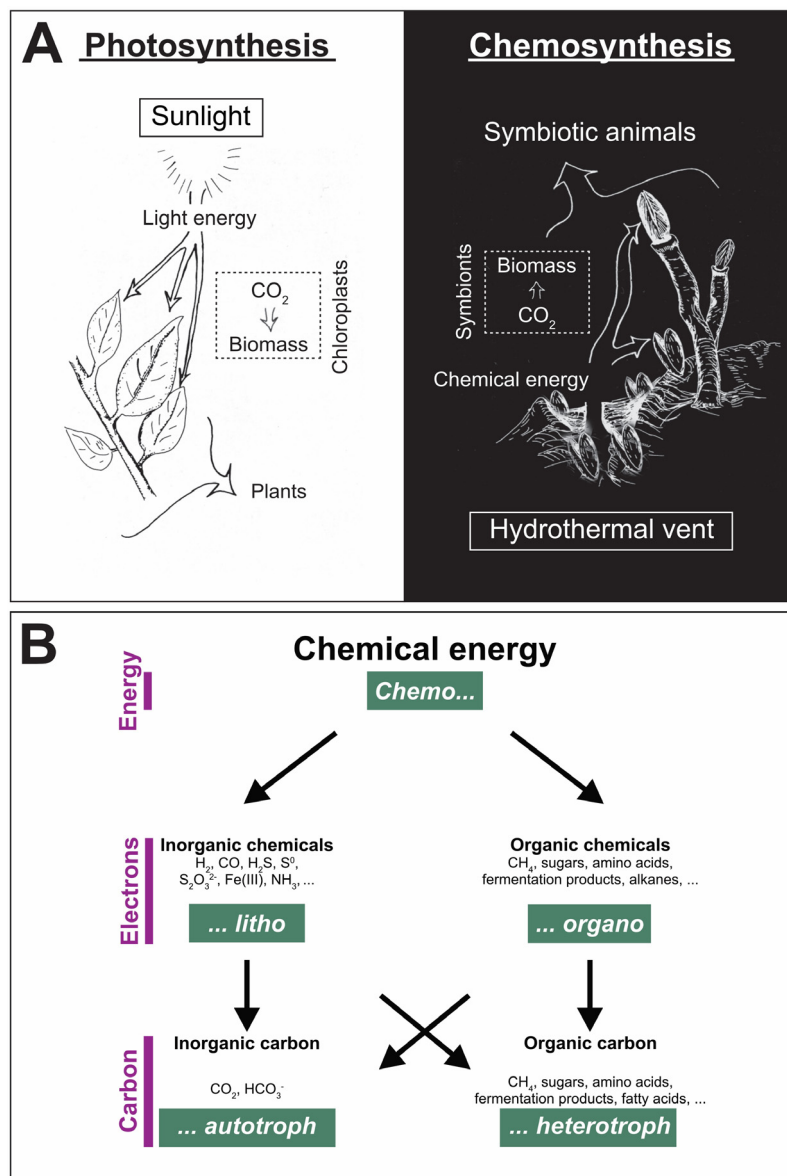

Figure 2. Strategies of energy conservation. A, Photosynthesis *vs.* chemosynthesis. Modified after Kreutzmann, 2017 and Somero, 1984. **B**, Metabolic options and substrates for the conservation of energy in chemosynthetic organisms.

Marine chemosynthetic symbioses and the benefits to both partners

Chemosynthetic symbioses were discovered in the late 1970s in the deep sea at hydrothermal vents (Cavanaugh *et al.*, 1981; Felbeck, 1981). These findings changed the previously common view that the deep sea is a desert-like habitat with limited food resources. The animals thriving at these vents were found to be associated with abundant bacterial symbionts. One of the dominating animals, the vestimentiferan tubeworm *Riftia pachyptila*, reduced mouth and digestive organs, and instead harbors bacterial symbionts in its trophosome (Cavanaugh *et al.*, 1981). Consistent with this observation, the vestimentiferans tubeworms were first suggested to live on "molecular food" as they were shown to host endosymbiotic bacteria with chemosynthetic activities (Cavanaugh *et al.*, 1981; Felbeck, 1981; Jones, 1981). It is reasonable to assume that these endosymbiotic bacteria feed their host *via* chemosynthesis, even though the exact mechanisms of the nutrient transfer are still unknown (Bright *et al.*, 2000).

Chemosynthetic symbioses were also discovered in environments other than the deep-sea vents and in diverse marine invertebrates (Cavanaugh *et al.*, 1981; Cavanaugh, 1983). Convergent evolution has led to the establishment of chemosynthetic symbioses in several animal phyla and ciliates (Dubilier *et al.*, 2008). The evolution of symbioses between bacteria and their various hosts appears to have often resulted in the complete loss of mouth and digestive system – similar to the vetimentiferan tubeworms from the deep sea. Such organs reductions cause symbiont dependencies, resulting in new organisms that evolved through symbiosis (Kiers and West, 2015).

The benefits of a symbiosis between invertebrates and chemosynthetic bacteria can be diverse for the two partners. For example, the symbionts allow their hosts to thrive in "hostile" environments such as sulfide-rich habitats or even deep-sea gas and oil seeps (Cavanaugh, 1983; Rubin-Blum *et al.*, 2017). In such environments, host-internal symbionts help to overcome carbon limitation if they are capable of using inorganic carbon that is not accessible for the host, and incorporate it into biomass. This biochemical conversion provides an essential supply of carbon to deep-sea animals as only 0.1 to 1% of the carbon-containing material from the surface reaches down into the deep sea (Bopp *et al.*, 2015).

Chemosynthetic symbioses are often found in sulfide-rich environments. At certain concentrations, sulfide can be deadly because it can bind to proteins of the respiratory chain and thereby inhibits its normal function (Jiang *et al.*, 2016). Chemosynthetic symbionts are hypothesized to detoxify sulfide concentration in the host, but it is also possible that the hosts are simply insensitive to sulfide (Powell, 1989; Hentschel *et al.*, 1999; Jan *et al.*, 2014). Besides carbon, animals additionally require other essential resources such as nitrogen, which can also be provided by their symbionts (Petersen *et al.*, 2016). A typical way of acquiring nitrogen from the environment is through the uptake of ammonium or nitrate by the symbionts (Lee *et al.*, 1992; Lee and Childress, 1994). Certain symbionts even allow the recycling of host waste products, which is advantageous in symbiotic systems which exist in nutrient-poor environments (Kleiner, Wentrup, *et al.*, 2012).

The benefits for chemosynthetic bacterial symbionts are just as versatile as they are for the hosts. For such bacteria, the respective host provides a sheltered environment with reduced competition, and at the same time serves to protect them from predators (Ott *et al.*, 2005). Free-living bacteria often encounter limited energy resources in their natural environment. For example, sediment bacteria have to cope with distinct spatial distributions of reduced and oxidized nutrients. Bacteria that live in symbiosis with a mobile meiofaunal organism, however, can overcome such limitations by using their host as a "shuttle service" that moves along the redoxcline longitudinally (Giere *et al.*, 1991; Ott *et al.*, 1991; Jäckle, 2018). In the case of sessile organisms, like the vestimentiferan tubeworms, specific hemoglobin molecules allow the storage of both oxygen and sulfide in numerous tentacles to provide their bacterial symbionts with conditions needed for their growth (Jones, 1981; Arp and Childress, 1983; Zal *et al.*, 1998). Collectively, these few arguments indicate that hosts provide a stable environment in which the symbionts – and in return also the host – can thrive and that the symbiotic partnership also increases the fitness of both symbiotic partners. In addition, these arguments suggest that symbiosis allows organisms to survive and to populate habitats which are otherwise hostile to both partners.

Environments of chemosynthetic symbioses

Convergent evolution of chemosynthetic symbioses in different lineages of marine invertebrates with various bacterial taxa argues for a successful and omnipresent life strategy

(Dubilier *et al.*, 2008). Since the discovery that the deep-sea vent fauna is often driven by chemosynthesis, it became clear that chemosynthetic symbioses are common to different types of ecosystems (Cavanaugh, 1983; Dubilier *et al.*, 2008). These ecosystems include whale and wood falls, cold seeps, mud volcanoes, continental margins and omnipresent shallow-water coastal sediments (reviewed by Dubilier *et al*. 2008). At a first glance, these ecosystems appear to be highly distinct from one another in terms of their abiotic factors. For example, the fauna of hydrothermal vents is exposed to high pressure which is a very different feature when compared to shallow-water sediments. In spite of the different abiotic factors, chemosynthetic ecosystems share key features that include a continuous supply of reduced energy sources serving as electron donors for chemosynthetic bacteria and the presence of oxygen that is normally essential for the animal host (Dubilier *et al.*, 2008).

Hydrothermal vents represent areas in the sea floor where water that is heated by volcanic activity reaches the surface at tectonic plate boundaries. Vents are distributed along the mid-ocean ridges and they can be found at sites where two tectonic plates diverge, an event that causes new oceanic crust to be formed. Abiotic processes cause the enrichment of different electron donors – the fuel for chemosynthesis – in seawater by interactions with basalt rocks. As the heated water travels through fissures, it creates a water flow resulting in the formation of hydrothermal vents (Jannasch, 1985; Van Dower, 2000). During the travel through basalt rocks, the chemical composition of the seawater is altered and, in turn, it causes changes in chemical properties as well as resulting in temperatures of $\geq$ 460 °C (Edmond *et al.*, 1982; Perner *et al.*, 2014). Vent fluids are typically enriched in inorganic substrates such hydrogen, hydrogen sulfide and methane which all represent potential energy sources for a broad vent community of living matter. At the border to the overlaying seawater, iron sulfides precipitate, resulting in the formation of so-called chimneys which are typical for distinct formations called black smokers. In addition to the already mentioned giant vestimentiferan tubeworms, communities at hydrothermal vents are dominated by vesicomyid clams, bathymodiolin mussels and shrimps which all live in symbiosis with chemosynthetic bacteria (Dubilier *et al.*, 2008).

At sea areas more closely to the water surface, whale and wood falls represent dynamic ecosystems and the chemosynthetic faunal community can adapt quickly. These ecosystems can be colonized by both free-living bacteria and chemosynthetic symbioses as both

14

environments are obviously limited by the locally and temporally restricted input of massive organic material (Bienhold *et al.*, 2013). The colonizers of these environments include besides free-living filamentous sulfur bacteria also a variety of marine invertebrates such as vestimentiferan tubeworms, vesicomyid clams, mytilids, limpets, snails, polynoid polychaetes and ciliates which all also rely on sulfur-oxidizing symbionts (Feldman *et al.*, 1998; Laurent *et al.*, 2009). Reducing conditions by elevated sulfide concentrations in the sediments are caused by the degradation of the organic materials by the free-living microbial communities which are established adjacent to the organic falls (Laurent *et al.*, 2009; Treude *et al.*, 2009).

No matter whether one considers the deep or the shallow areas of the sea, it is the chemosynthetic fauna that dominates marine sediments worldwide. In the beginning of the 1970s and before the discovery of vent chemosynthetic symbioses, the "sulfide system" of marine shallow-water sediments was discovered as a sheltered environment for marine invertebrates (Fenchel and Riedl, 1970; Sterrer and Rieger, 1974). At the time, however, it has already been known by bacteriologists that these sediments consist of an oxidized layer and an anaerobic and reduced black zone. This knowledge was almost completely ignored by marine zoologists. Clearly, shallow-water ecosystems are dependent on organic carbon that originates from photosynthetic-driven primary production, an important aspect which stands in contrast to the situation at hydrothermal vents (Fenchel and Riedl, 1970; Dubilier *et al.*, 2008). In sediments, reduced sulfur mostly originates from sulfate-reducing bacteria which are responsible for up to 29% of demineralized organic matter and are of key importance for the sulfur cycling (Bowles *et al.*, 2014; Wasmund *et al.*, 2017). Several lineages of marine invertebrates have been identified in alike enriched sediments including platyhelminths, nematodes, oligochaetes and also marine ciliates, and one can assume that the diversity is much larger than currently known and as previously anticipated (Fenchel and Riedl, 1970; Dubilier *et al.*, 2008). The early observations on nutritional speciation to bacterial food have led the basis for studying chemosynthesis, as already in the 70´s meiofaunal organisms were suggested to feed on sulfur bacteria or were discovered to lack digestive tracks, but contain "little granular bodies" (Fenchel and Riedl, 1970; Sterrer and Rieger, 1974).

Inorganic energy sources utilized by chemoautotrophic symbionts

In chemosynthetic symbioses, the symbionts produce biomass by the oxidation of reduced inorganic compounds. Although the first and the most described chemosynthetic symbioses to date rely on the oxidation of reduced sulfur compounds, a variety of additional substrates for energy conservation are known (Cavanaugh *et al.*, 1981; Felbeck, 1981). A few years after the discovery of thioautotrophy in symbionts of vestimentiferan tubeworms, the first methane-driven symbionts were discovered in deep-sea mussels of the genus *Bathymodiolus* (Childress *et al.*, 1986; Cavanaugh *et al.*, 1987). Methane, which can serve as energy- and carbon source, is assumed to be a widely used chemosynthetic substrate (Pimenov *et al*. 2000; Borowski *et al*. 2002; Schmaljohann & Flugel 1987; Rubin-Blum *et al*. submitted manuscript). Besides sulfur and methane, hydrogen also represents an electron donor with comparably high energy yield for biomass production by chemosynthetic symbionts (Amend and Shock, 2001). This energy source was shown to be utilized by the symbionts of for example *Bathymodiolus* that inhabit vent sites with high hydrogen concentrations in the environment (Petersen *et al.*, 2011). Hydrogen is not only restricted to deep-sea vents but also occurs together with carbon monoxide in seagrass sediments of marine shallow waters, where both substrates can also be utilized by the thiotrophic symbionts of gutless oligochaetes (Kleiner, Wentrup, *et al.*, 2012; Kleiner *et al.*, 2015).

Biodiversity of chemosynthetic symbioses

The diversity of chemosynthesis is not only represented by the ecosystems in which various host groups thrive. It is also reflected by several symbiont lineages, the location of the symbionts either inside or outside the host organism, in the way how the symbionts are acquired or passed on to new generations and finally also reflected in the electron donors they can utilize. Not surprisingly, chemosynthetic animals often undergo drastic developmental, physiological and behavioral adaptations, some of these features will be described more detailed in this thesis. I will first introduce some selected symbiotic systems which are of primary interest for my study (Figure 3). This introduction will be followed by a detailed description of the marine flatworm genus *Paracatenula* which was used for the study presented.

The vestimentiferan tubeworm *Riftia pachyptila* – often referred to as the giant tubeworm – represents the largest known chemosynthetic invertebrate. *Riftia* tubeworms grow fast compared to other marine invertebrates and reach tube lengths of up to 1.5 m (Jones, 1981; Lutz *et al.*, 1994; Gaill *et al.*, 1997). *Riftia* has four body regions including a tentacular plume, the vestimentum, trunk and an opisthosoma (Jones, 1981). *Riftia* lacks both mouth and gut. Thus, for nutrition, they rely on symbionts which are located in the so-called trophosome (Cavanaugh *et al.*, 1981; Felbeck, 1981; Cavanaugh, 1983). The trophosome is connected to the site of gas exchange in the obturacular plume and its vessels are filled with blood (Cavanaugh *et al.*, 1981; Jones, 1981). The blood has a high viscosity and contains hemoglobin that simultaneously binds sulfide and oxygen. Both substrates are provided to the symbionts to be used for chemosynthesis (Jones, 1981; Arp and Childress, 1983; Arp *et al.*, 1987).

Clams of the genus *Calyptogena* were characterized as abundant fauna of hydrothermal vents and seeps (Boss and Turner, 1980). Representatives of the species *Calyptogena magnifica* reach sizes of ~26 cm in length, possess a mouth and host symbionts in the gill tissues (Boss and Turner, 1980; Cavanaugh, 1983). Although *Calyptogena* has highly reduced labial palps and retained a short gut, its digestive system was suggested to be at least partially functional since the animals´ stomach has been reported to be filled with particles (Boss and Turner, 1980; Le Pennec *et al.*, 1995).

Clams of the genus *Solemya* are also symbiotic animals. However, they are found in shallow-water ecosystems (Figure 3). For more than a hundred years, these organisms were of interest to researchers because of their vestigial or even absent digestive systems, and their nutritional mode remained a mystery until the discovery of chemosynthetic symbioses (Pelseneer, 1891; Reid and Bernard, 1980). As observed with other bivalves, *Solemya* clams host symbionts in their fleshy gills and were proposed to completely rely on these symbionts for nutrition (Cavanaugh, 1983; Stewart and Cavanaugh, 2006). In fact, the solemyid class of clams has developed a unique behavioral strategy to access sulfide and oxygen simultaneously by digging in a Y-shaped burrow spanning the redoxcline. This way, the clam can position itself at the junction of the Y, where it can either pump oxygenated water from the surface layers or it can access sulfide that is diffusing upward from the deeper layers of the sediment. Most likely, this

strategy is an adaptation to ensure that the symbionts are provided with their redox requirements (reviewed in Stewart & Cavanaugh 2006).

In addition to the above-mentioned clams of the genus *Calyptogena,* mussels of the genus *Bathymodiolus* also belong to the common deep-sea fauna and are highly abundant at hydrothermal vents and seeps (Kenk and Wilson, 1985; Lorion *et al.*, 2013). The anatomy of these mussels largely resembles the body plan of other mytilid relatives, but they show unique characteristics such as the large gills filled with bacterial symbionts and a reduced digestive tract (Le Pennec and Hily, 1984; Fiala-Médioni *et al.*, 1986).

Furthermore, gutless oligochaetes are part of the typical marine meiofauna of tropical and subtropical shallow-water sediments (Figure 3) (Bright and Giere, 2005; Dubilier *et al.*, 2006; Giere, 2009). Although the first species has already been described in the 1970s, the absence of both gut and mouth along with the presence of thiotrophic bacteria was only noted after the gutless vestimentiferan deep-sea tubeworms were identified (Giere, 1985). Gutless oligochaetes also lack excretory organs such as nephridia (Giere and Erséus, 2002). Below the cuticle of the gutless oligochaetes, as many as six different bacterial symbionts were found and assigned to essential nutritional function and waste recycling (Blazejak *et al.*, 2005, 2006; Ruehland *et al.*, 2008, Kleiner, Wentrup, *et al.*, 2012).
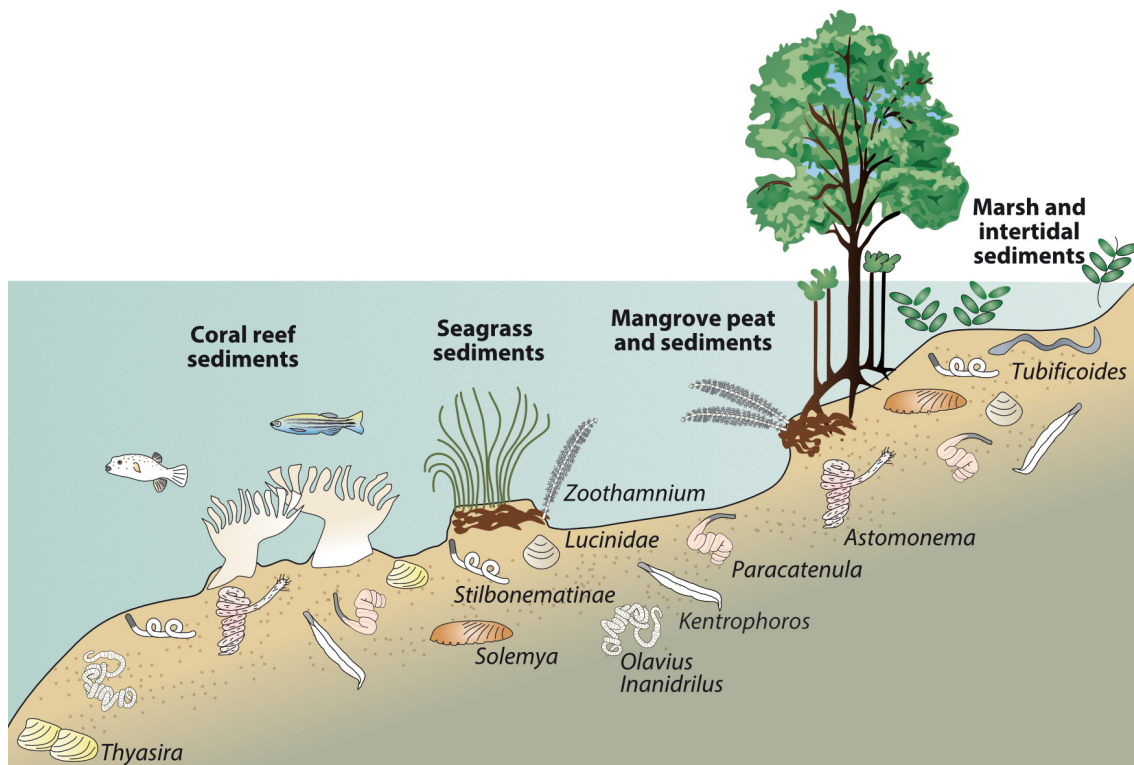
Figure 3. Diversity of shallow-water meiofauna. Modified after Dubilier *et al.*, 2008 and Gruber-Vodicka (2011). Selected logos were kindly provided by Brandon Seah. Animals and plants are not drawn to scale.

External and internal association of chemosynthetic symbionts

Chemosynthetic symbionts can be associated with their host in several ways. Their association can be either inside or outside the host as endo- or ectosymbionts, respectively. In endosymbiotic associations, the symbionts either live extracellularly – outside of cells – or intracellularly within the host organism. An example for endosymbionts are the symbionts of the deep-sea tubeworm *Riftia* which are located intracellularly within the trophosome tissue in specialized host cells called bacteriocytes (Cavanaugh *et al.*, 1981; Jones, 1981; Hand, 1987). In vestimentiferans including *Riftia*, the trophosome tissue has derived from mesodermal tissue and it had been proposed that it evolved convergently from different tissues among the siboglinid tubeworms and flatworms (Bright and Sorgo, 2003; Nussbaumer *et al.*, 2006; Dirks *et al.*, 2011). In other cases like the gutless oligochaetes, endosymbionts can be hosted extracellularly and are restricted to the epidermal-cuticle interface (Giere, 1981; Giere and Erséus, 2002).

The location of symbionts within or on a host does not allow to predict their long-term stability and specificity. In ectosymbioses such as nematodes, shrimps and ciliates, the symbionts are

attached to the outside of their hosts (Dubilier *et al.*, 2008). Even though one could assume that the attachment of ectosymbionts to hosts is unspecific, the opposite, i.e. specificity, was found for the nematode *Leptonemella* spp. where the attached ectosymbionts were found to be highly specific (Zimmermann *et al.*, 2016). Likewise, the mode of transmission cannot be predicted on the basis of the external or internal location of the symbionts. For example, both the endosymbionts of the vestimentiferan tubeworm *Riftia* and the bathymodiolin mussel *Bathymodiolus* are localized inside their hosts but taken up from surrounding seawater without evidence for simultaneous diversifications (Won *et al.*, 2003; Nussbaumer *et al.*, 2006; Fontanez and Cavanaugh, 2014; Zimmermann *et al.*, 2016). Another mode is exemplified by some nematode ectosymbionts which can be transmitted vertically from one host generation to another (Zimmermann *et al.*, 2016).

Diversity of chemosynthetic symbionts and their transmission

The known diversity of chemosynthetic symbionts includes primarily *Proteobacteria* species (Dubilier *et al.*, 2008). Typically, the symbionts belong to *Gamma-* and *Epsilonproteobacteria* which are either vertically transmitted or acquired horizontally through the environment (Dubilier *et al.*, 2008; Assié *et al.*, 2016; Seah *et al.*, 2017). Most sulfur-oxidizing bacteria, both free-living and symbiotic, belong to *Gammaproteobacteria* (Dubilier *et al.*, 2008). At least nine phylogenetically distinct gammaproteobacterial symbiont clades are identified, most of which were intermixed phylogenetically with free-living bacteria (Dubilier *et al.*, 2008).

Although the various hosts are dependent on their symbionts for nutrition, many of these acquire their symbionts horizontally from the pool of free-living representatives (Nussbaumer *et al.*, 2006; Wentrup *et al.*, 2014; Russell *et al.*, 2017). The horizontal acquisition of symbionts indicates a facultative symbiotic lifestyle in which at least one part of their lifestyle is symbiont-free, a phenomenon not seen with vertically transmitted symbionts (Bright and Bulgheresi, 2010). For example, the gammaproteobacterial thiotrophic symbionts of *Bathymodiolus* mussels are closely related and group with the cluster of free-living SUP05 (Ponnudurai *et al.*, 2017). It is known that these symbionts phylogenetically intermix with free-living organisms such as *Candidatus* Thioglobus, and it is postulated that these symbionts have established their symbiotic associations with multiple host species of deep-sea mytilid mussels independently (Petersen *et al.*, 2012). The exact mechanisms of horizontal symbiont uptake

in *Bathymodiolus* mussels is yet not known (Wentrup *et al*. 2014, M. Franke personal communication). In the case of *Riftia* tubeworms the gammaproteobacterial symbionts are also horizontally acquired and, notably, they possess a symbiont-free larval stage (Di Meo *et al.*, 2000; Nussbaumer *et al.*, 2006). To date, there is only one exception known where alphaproteobacterial chemosynthetic symbionts are associated with a marine flatworm, *Paracatenula* (Gruber-Vodicka *et al.*, 2011) which will be described separately below since this symbiosis is the subject of my studies described here.

Thiotrophic symbionts have diverse physiological profiles

The symbionts of chemosynthetic invertebrates often vary in genome sizes and in their physiological capabilities (Kleiner, Petersen, *et al.*, 2012) (Figure 4). Their genome sizes can range from being highly reduced with sizes no more than 1.02 to 4.88 Mb and being metabolically more flexible (Kuwahara *et al.*, 2007; Newton *et al.*, 2007; Petersen *et al.*, 2016). These differences suggest diverse metabolic as well as ecological strategies. The most reduced genomes of thiotrophic symbionts can be found in the symbionts of vesicomyid clams with sizes of 1.02–1.16 Mb (Kuwahara *et al.*, 2007; Newton *et al.*, 2007). The genome of the symbiont *Ca.* R. magnifica was characterized as encoding all metabolic pathways that are typical of free-living chemoautotrophs (Newton *et al.*, 2007). These pathways include the PP_i-dependent Calvin-Benson-Bassham (CBB) cycle for carbon fixation, the reverse-acting dissimilatory sulfite reductase (rDSR) for sulfur oxidation as well as synthesis pathways for nitrogen assimilation, amino acid production, cofactor and vitamin synthesis (Newton *et al.*, 2007; Kleiner, Wentrup, *et al.*, 2012) (Figure 4A). The genomes of the *Bathymodiolus* symbionts are only slightly larger with sizes between 1.7–2.3 Mb and have extended metabolic capabilities (Sayavedra *et al.*, 2015). Besides encoding for components of the sulfur oxidation pathway (SOX) for sulfide and thiosulfate oxidation, these symbionts have the ability to acquire energy from the oxidation of hydrogen when compared to most other known symbionts (Petersen *et al.*, 2011; Sayavedra *et al.*, 2015). Additionally, the *Bathymodiolus* symbionts express a wide repertoire of toxin-related genes that might function in interactions with their host and serve as a protection against parasites (Sayavedra *et al.*, 2015). In contrast to the clam symbionts, the *Bathymodiolus* symbionts encode substrate transporters involved in the uptake of organic compounds (Newton *et al.*, 2007; Sayavedra *et al.*, 2015). However, key enzymes of the tricarboxylic acid (TCA) cycle appear to be missing in the coding capacity of

these two symbionts, leaving open whether the symbionts can actually utilize sugars or are obligate autotrophs (Wood *et al.*, 2004; Kuwahara *et al.*, 2007; Newton *et al.*, 2007; Sayavedra *et al.*, 2015).

A larger metabolic flexibility is given in symbionts of the clam *Solemya*, the gutless oligochaete *Olavius* and the tubeworm *Riftia* with genome sizes of ~2.5 up to 4.5 Mb (Markert *et al.*, 2011; Gardebrecht *et al.*, 2012; Dmytrenko *et al.*, 2014, J. Wippler, personal communication). The intracellular symbionts of *Solemya* were proposed to have a broad thiotrophic and sugar metabolism based on a complete TCA cycle, besides a large number of transporters for exchanges with their environment (Dmytrenko *et al.*, 2014). The Gamma-1 (now referred to as *Thiosymbion*) have the additional ability to recycle waste products of the host by the assimilation of e.g. acetate, propionate and malate through a 3-hydroxypropionate bi-cycle (Kleiner, Wentrup, *et al.*, 2012). In addition, this pathway is proposed to be linked to the synthesis of storage compounds such as polyhydroxyalkanoates (PHA) (Kleiner, Wentrup, *et al.*, 2012). The Gamma-3 symbiont of *Olavius*, however, lacks these abilities but encodes a full TCA cycle and was suggested to use carbon monoxide as energy source. The tubeworm symbionts of the genera *Riftia* and *Tevnia* are unique since both encode an additional autotrophic pathway, the reverse TCA (rTCA) cycle in addition to the CBB cycle which is thought to bring increased metabolic flexibility (Markert *et al.*, 2007, 2011; Gardebrecht *et al.*, 2012; Li *et al.*, 2018). Until recently, chemosynthetic symbionts were suggested as being nutritional symbionts, although the analyses were primarily focused on the ability of the symbionts to make inorganic carbon sources available. The chemosynthetic symbionts of the clam *Loripes* and the nematode *Laxus* extend the typical functions of thiotrophic symbionts by nitrogen fixation (Petersen *et al.*, 2016). Their symbionts possess the so far largest known genomes with sizes between 4.33–4.88 Mb (Petersen *et al.*, 2016). Their characterization indicates that they otherwise resemble functionally the genomes of other chemosynthetic symbionts (Petersen *et al.*, 2016). Since symbioses involving these symbionts are often found in nutrient-poor environments, nitrogen-fixation is thought to provide an additional source of substrates to the host and to the ecosystem they live in (Petersen *et al.*, 2016).

Taken together, the variable metabolic makeup of the different symbionts genomes suggests a high degree of metabolic flexibilities. However, up to now, it has not been possible to reveal the essential functions of thiotrophic symbionts on which their host animals depend on.
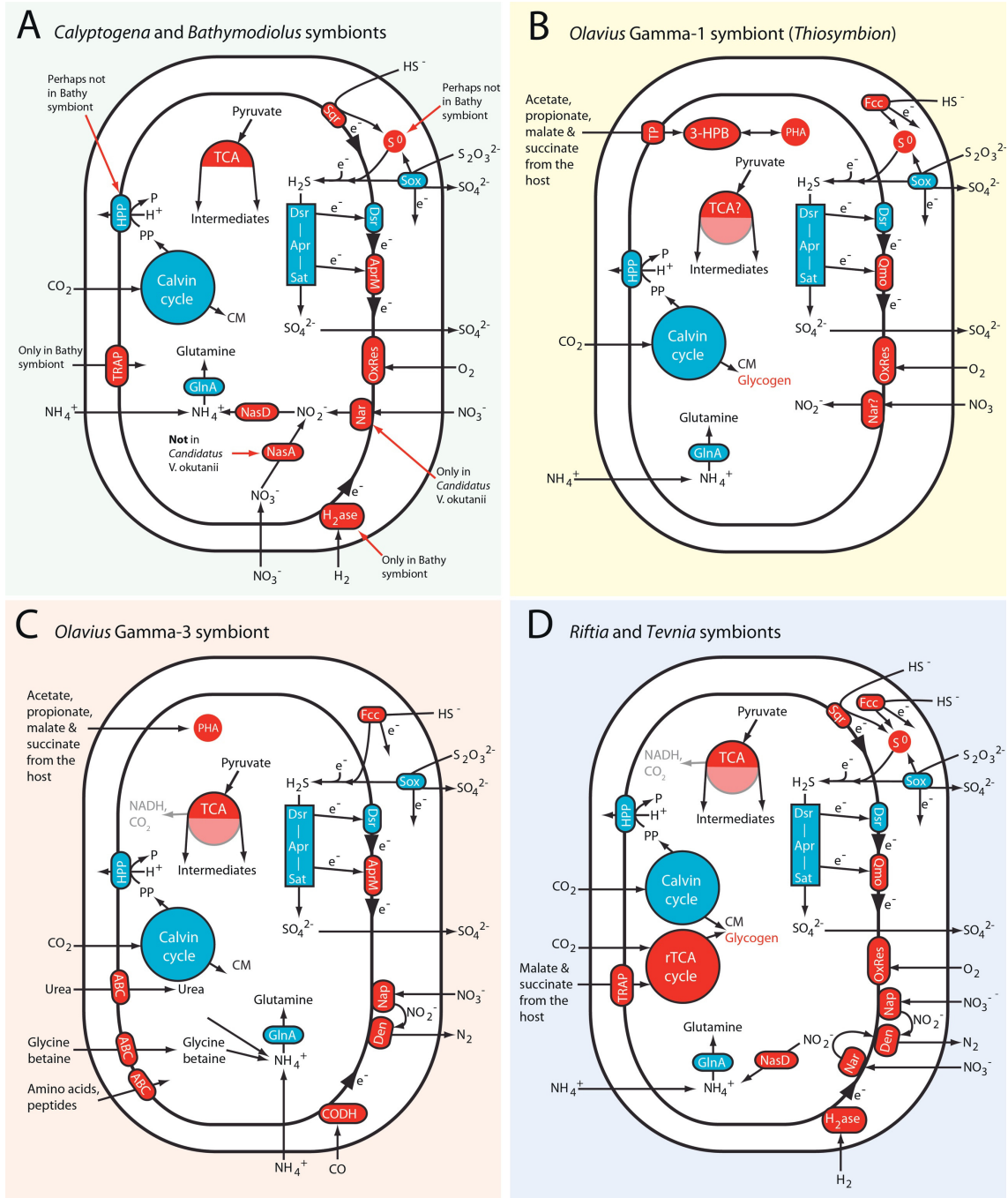
Figure 4. Metabolic pathways of selected gammaproteobacterial thiotrophic symbionts. *A*, The reduced symbionts of *Calyptogena* and *Bathymodiolus*. ***B*,** *Thiosymbion* of the gutless oligochaete *Olavius algarvensis*. ***C*,** Gamma-3 symbiont that often co-occurs in *O. algarvensis*. ***D*,** Symbionts of *Riftia* and *Tevnia*. The figure was adapted from Kleiner, Petersen, *et al.*, 2012, and modified according to Sayavedra *et al.*, 2015.

Symbiont transmission and impacts on genome sizes: processes of reductive genome evolution

The mode of symbiont transmission can influence the genome structure, the genetic repertoire and consequently genome size during the evolution of a long-term symbiotic partnership (Bright and Bulgheresi, 2010; McCutcheon and Moran, 2012). The genomes of continuously horizontally transmitted bacteria from the environment are typically large and genetically diverse. Organisms are continuous targets of genome mutations such as base substitutions, deletions and other genetic changes. However, it appears that such mutations do usually not accumulate in horizontally transmitted symbionts. Although numerous mutations might appear in the genomes of such symbionts, the majority of them will be lost within a few generations as new symbionts will be horizontally acquired by the hosts and thus, only a limited fraction of mutations can contribute to the host-symbiont evolution (Ohta, 1992). This can be explained by large effective population sizes where it takes a very long time for a mutation to be fixed (Ohta, 1992). An increased chance for the accumulation and fixation of mutations in the active population is therefore more likely when a small and rather homogeneous population of symbionts is present in the host and this population is then transmitted to the next host generation as in the case of vertically transmitted symbionts (Ohta, 1973; Funk *et al.*, 2001; Bright and Bulgheresi, 2010). The reduction in symbiont numbers transferred to the next generation represents a genetic bottleneck that is known to have critical consequences for the genetic and ecological evolution of symbiotic bacteria, which results in higher substitution rates (Mira and Moran, 2002).

Purifying selection against naturally occurring mutations has only minor effects on vertically transmitted symbionts but usually results in genome deletions (McCutcheon and Moran, 2012). In such organisms an irreversible ratchet mechanism – usually referred to as Muller´s ratchet – prevents selection of organisms that lack recombination (Muller, 1964). In the processes of Muller´s ratchet, the genomes constantly accumulate normally deleterious mutations – a trigger for the processes of reductive genome evolution if the affected individuals and their offspring survive (Moran, 1996; McCutcheon and Moran, 2012).

Genome reduction is universally seen with most bacteria and archaea, and can be found in free-living organisms, thermophilic archaea, insect symbionts as well as in chemosynthetic symbioses (Moran, 2002; Giovannoni *et al.*, 2005; Kuwahara *et al.*, 2007; Newton *et al.*, 2007;

Wolf and Koonin, 2013; Nicks and Rahn-Lee, 2017; Tian *et al.*, 2017). Intracellular bacterial symbionts appear to represent prime examples for these processes as they reproduce asexually, have small effective population sizes and are exposed to narrow bottlenecks for their survival (McCutcheon and Moran, 2012). Most models for reductive genome evolution are derived from studies on insect symbioses and usually illustrate a linear progression from a free-living state to streamlined obligate symbiotic organisms (Figure 5A). In insects, the co-evolution and vertical transmission of bacterial symbionts for millions of years repeatedly resulted in highly specialized interactions going along with a downstream into tiny genomes (McCutcheon *et al.*, 2009a; McCutcheon and Moran, 2012). In these cases, the endosymbionts are inherited from the parental hosts to their offspring and they provide various functions to their host that include the provision of essential amino acids, vitamins and cofactors, i.e. nutrients that are missing in the animal-host diet (Shigenobu *et al.*, 2000; Wu *et al.*, 2006; McCutcheon and Moran, 2007; Moya *et al.*, 2008). The switch from a free-living state to an intracellular lifestyle implies the loss of unnecessary genes over time, when genetic material of the symbiont is lost without detrimental effects on the organism (Moya *et al.*, 2008). The mentioned bottlenecks favor a random genetic drift that often affects non-essential genes and the loss of DNA repair and recombination genes, whereas genes that are critical for survival of the symbionts in intracellular environments must be retained in the genome (Gonza *et al.*, 2003; Moya *et al.*, 2008; McCutcheon and Moran, 2012; Shimamura *et al.*, 2017). Most drastic losses are usually found in metabolic genes, but only those which are not essential for the survival and fitness benefits of both the host and the symbiont in a given environment (Moya *et al.*, 2008).

The switch of bacteria from a free-living state to an obligate symbiont involves dynamic processes involving gene content reductions (Moya *et al.*, 2008). During the first stages of this process, mobile elements within the genome proliferate. Such elements are typically abundant in the genome of bacterial symbionts that were acquired recently by their hosts (Moran and Plague, 2004; Moya *et al.*, 2008; Plague *et al.*, 2008). Mobile elements integrate into genomes and have been implicated as a source for genome rearrangements and gene inactivation (Moya *et al.*, 2008). Side effects of mobile elements proliferation include scattered homology along the chromosome causing increased spontaneous rates of homologous recombination as well as activation of gene expression for various types of neighboring genes (Moran and Plague,

2004). Clear-cut signs for bacterial genomes in transition to an obligate symbiont stage are the accumulations of pseudogenes, a phenomenon that is well-investigated in the genomes of the insect symbionts *Serratia* (Figure 5B) (Moran and Plague, 2004; McCutcheon and Moran, 2012; Manzano-Marín and Latorre, 2016). Later stages of genome reduction include gradual gene losses along the genome, starting with single-nucleotide erosions. The process continues with rapid reduction in gene length until they are finally completely eroded (Gómez-Valero *et al.*, 2007; Moya *et al.*, 2008; McCutcheon and Moran, 2012; Manzano-Marín and Latorre, 2016). Accumulations of pseudogenes and the reduction of metabolic redundancies result in a continuous loss of genes, along with the decrease in mobile elements and the fixation of rearrangements, which result in a stable genome structure and streamlined genomes (Andersson, 2006; Moya *et al.*, 2008; McCutcheon and Moran, 2012).

Another side effect of reductive genome evolution are changes in the base composition (McCutcheon and Moran, 2012). Small genomes usually have shifted their genomic GC (gGC) contents from G+C to A+T, resulting in a gGC of ~13.5% in the case of spittlebug symbionts, which is low compared to its free-living relative *E. coli* which has a gGC of ~50% (Sueoko, 1962; Moya *et al.*, 2008; Hershberg and Petrov, 2010; Hildebrand *et al.*, 2010; McCutcheon and Moran, 2010; Raghavan *et al.*, 2012; Venton, 2012). The reasons for and consequences of these shifts are still debated, especially because there are exceptions such as small symbiont genomes that do not follow the trend of an increased A+T content (McCutcheon *et al.*, 2009b; McCutcheon and Moran, 2012; Venton, 2012). In more general terms, bacteria favor increased G+C contents despite a mutational bias towards A+T (Raghavan *et al.*, 2012). Multiple forces are thought to be the drivers of GC shift. For example, the loss of DNA repair and recombination genes during the genome reduction lead to changes in the G+C *versus* the A+T content because mutations are no longer repaired (Lind and Andersson, 2008; Moran *et al.*, 2008; McCutcheon and Moran, 2012). This allows mutations towards increasing the A+T content in tiny genomes, and shifts of the average gGC towards higher AT content values (McCutcheon and Moran, 2012).

Although it is fascinating to study genome reductions in insect symbioses, such processes also occur in symbionts of chemosynthetic animals which are less intensively studied. More than a decade ago, the chemosynthetic symbionts of vesicomyid clams were shown to have reduced genomes (Kuwahara *et al.*, 2007; Newton *et al.*, 2007). Just recently, the chemosynthetic

symbionts of a deep-sea sponge were shown to have reduced genomes as well (Tian *et al.*, 2017). The co-adaption of the vesicomyid clam symbionts and their host resulted in reduced symbiont genomes where genes coding for essential metabolic functions for both partners are retained. However, genes encoding non-essential functions such as diverse sets of substrate transporters, cell division genes (e.g. *ftsZ*) and DNA repair genes, as well as mobile elements were lost from their genomes (Kuwahara *et al.*, 2007, 2008; Newton *et al.*, 2007). Although these initial studies showed clear indications for reductive genome evolution, most of the follow-up studies focused mostly on the analysis of homologous recombination events and the gradual loss of DNA repair genes (Kuwahara *et al.*, 2011; Shimamura *et al.*, 2017; Tian *et al.*, 2017). General predictions, and how these processes compare to those of insect symbionts, are not yet addressed.
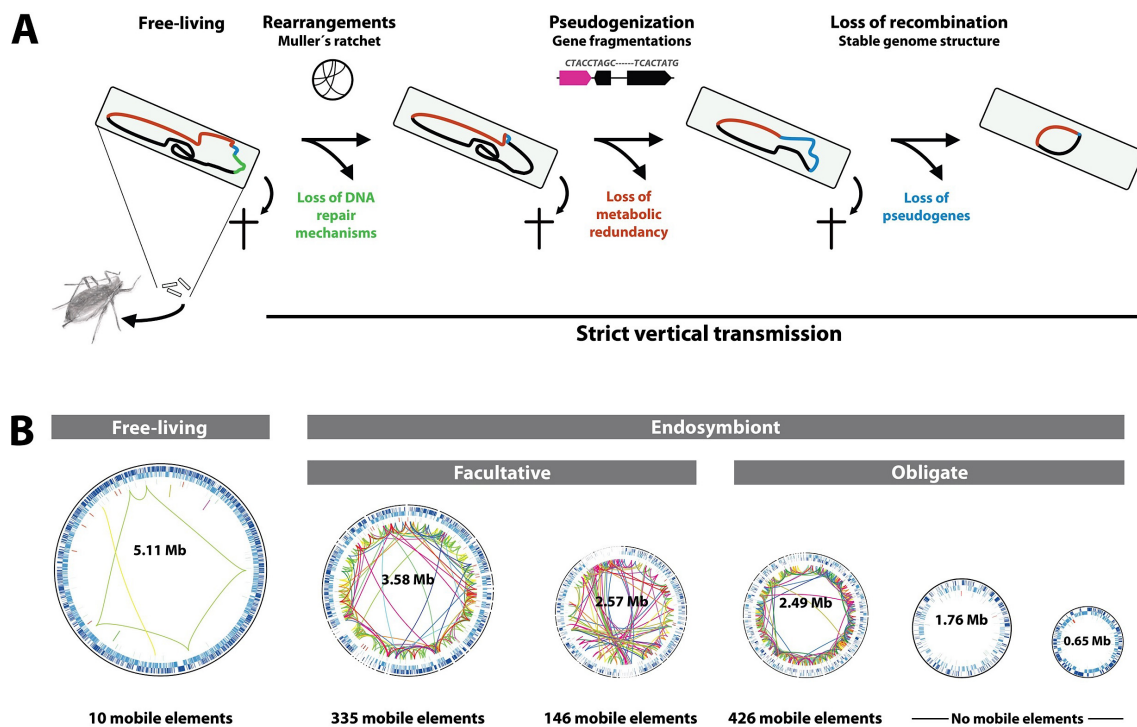

Figure 5. Principles of reductive genome evolution deduced from insect symbioses. *A*, General principles of ongoing processes, redrawn after Nicks & Rahn-Lee 2017 and Toft & Andersson 2010. ***B*,** "Snapshots" of genome reduction in the insect symbionts *Serratia symbiotica*. Colored lines indicate types of mobile elements scattered throughout the genomes. Modified after Manzano-Marín & Latorre 2016.

The *Paracatenula* symbiosis

Marine flatworms of the genus *Paracatenula* are of particular interest and, for multiple reasons, ideal subjects to study the host-symbiont relationship in detail. In the 1970s, these flatworms were identified as typical meiofaunal representatives of sheltered sulfidic sediments of warm

temperate to tropical regions (Sterrer and Rieger, 1974). They belong to the order *Catenulida* of the phylum *Platyhelminthes* (Larsson and Jondelius, 2008). The name *Paracatenula* originates from the latin *catenula* "small chains" since they are known to reproduce asexually by paratomy (Borkott, 1970; Larsson and Jondelius, 2008; Dirks, Gruber-Vodicka, Leisch, *et al.*, 2012). *Paracatenula* were morphologically classified as members of the family *Retronectidae* comprised of two genera, *Retronectes* and *Paracatenula*, which both can be found in sulfidic habitats (Sterrer and Rieger, 1974). While *Rectronectes* flatworms still have mouth, pharynx, gut lumen and protonephridia, *Paracatenula* lack all these organs. Although the phylogenetic placement of *Paracatenula* within the *Catenulida* is robust and was confirmed by molecular studies, there is no molecular data available for *Retronectes* (Figure 6) (Larsson and Jondelius, 2008; Littlewood, 2008; Gruber-Vodicka *et al.*, 2011; Ngamniyom and Panyarachun, 2016).
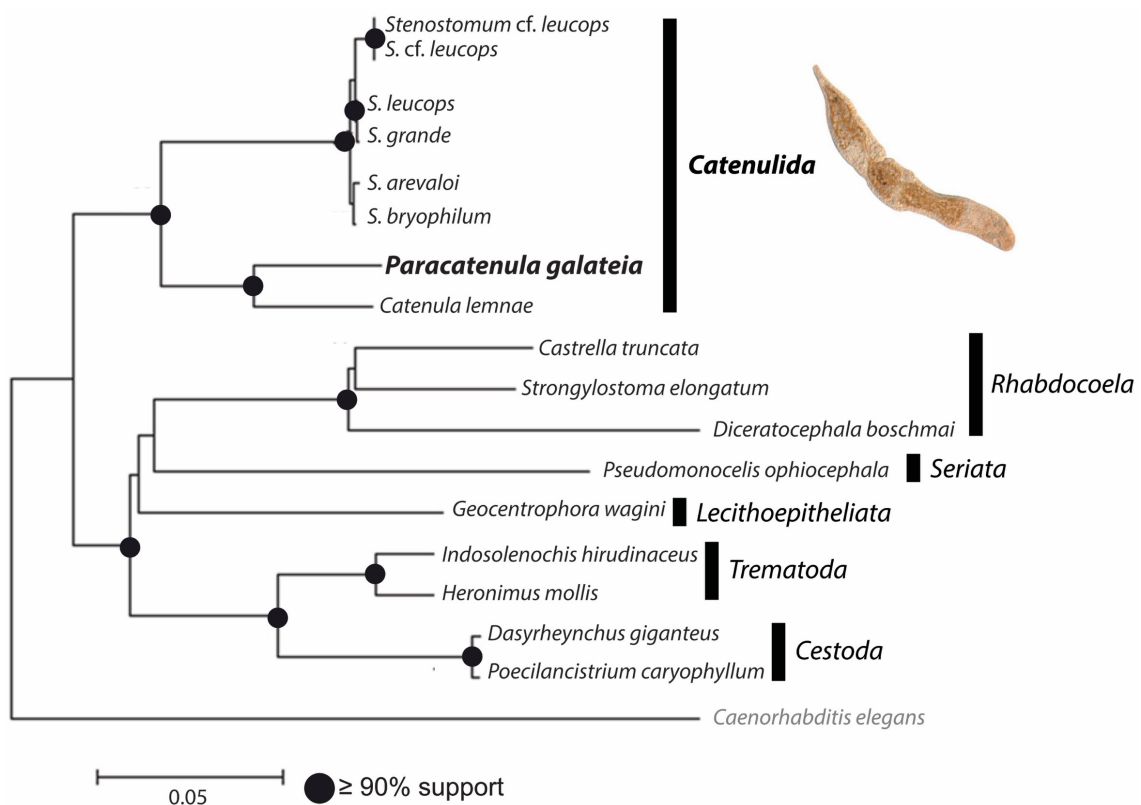

Figure 6. A phylogenetic tree based on 18S rRNA sequences covering a broad diversity of *Platyhelminthes* indicate the clustering of *Paracatenula* within the order *Catenulida*. The tree was modified from Ngamniyom & Panyarachun 2016. The micrographs show a representative of the genus *Stenostomum*, adapted from Egger *et al*. 2017. *Caenorhabditis elegans* was used as outgroup. The scale bar represents estimated sequence divergence in %, the black dots indicate bootstrap proportions ≥ 90%.

To date, five *Paracatenula* species have been described, and more species were identified morphologically and molecularly from sampling spots distributed among the world´s sediments

(Sterrer and Rieger, 1974; Dirks *et al.*, 2011; Gruber-Vodicka *et al.*, 2011). The extraordinary morphological structures of the mouthless *Paracatenula* flatworms were in an early need of explanations as it was unclear whether food uptake might occur *via* the epidermis similar to what is described for parasites. This feeding scenario, however, turned out to be rather unlikely (Sterrer and Rieger, 1974). Instead it was also hypothesized that being mouth- and gutless could represent a temporary stage in their life cycle (Sterrer and Rieger, 1974). However, already early drawings of live *Paracatenula* specimens documented their lifestyle by showing that the flatworms were filled with up 10 μm long rod-like inclusions in their gut rudiment which were later identified as microorganisms (Figure 7) (Sterrer and Rieger, 1974; Ott *et al.*, 1982). Although the functional role of the microorganisms was unclear at the time, a mutualistic symbiosis was proposed (Ott *et al.*, 1982).
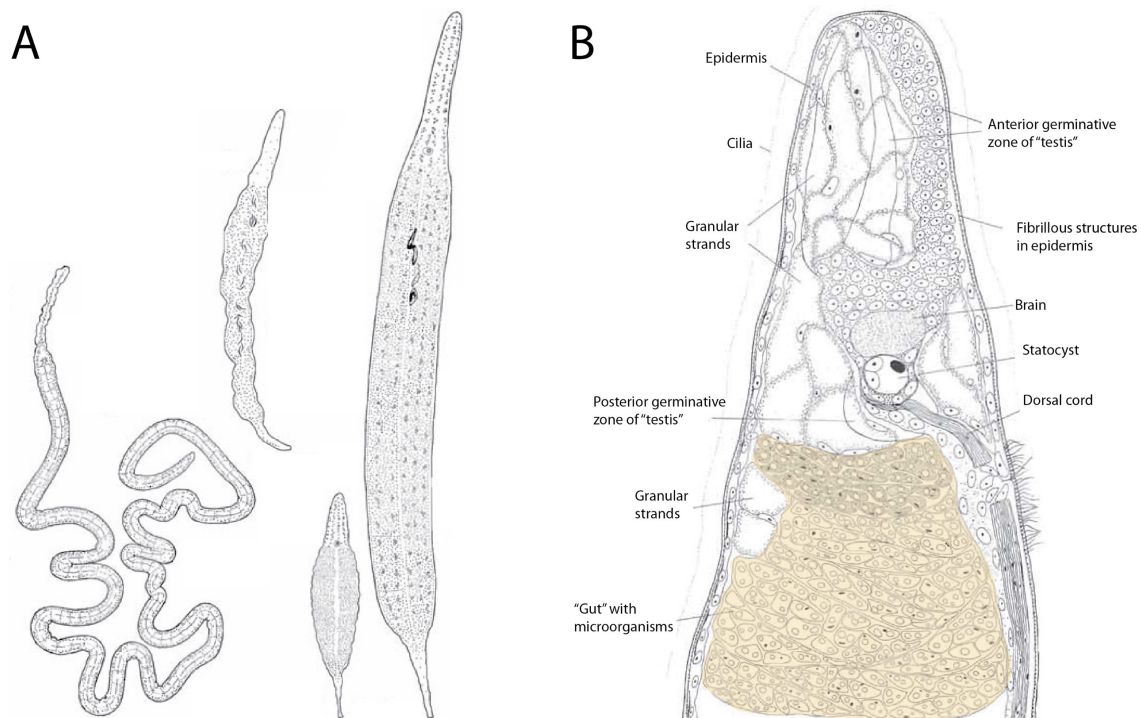

Figure 7. Morphological drawings of *Paracatenula* specimens in 1974. *A*, General morphology drawn from live *Paracatenula* spp. individuals. ***B*,** Orientation of anterior body region, the rostrum and part of the trophosome that is highlighted in orange. The figures were modified after Sterrer & Rieger 1974.

Almost 30 years later, studies on the *Paracatenula* using a variety of different methods such as ultrastructural visualizations as well as molecular and biochemical techniques were performed on both partners of the symbiosis. Furthermore, after including two additional species descriptions, the diverse *Paracatenula* hosts were shown to form a well-separated monophyletic clade within the *Catenulida,* indicating the presence of at least 16 species

(Sterrer and Rieger, 1974; Larsson and Jondelius, 2008; Dirks *et al.*, 2011; Gruber-Vodicka *et al.*, 2011; Leisch *et al.*, 2011). Morphological studies have revealed that *Paracatenula* are composed of a symbiont-free anterior body part, called rostrum, and a posterior part filled with symbiont-containing bacteriocytes which form the trophosome in analogy to the vestimentiferans (Figure 7, 8, Dirks *et al*. 2011).

Gene sequences of the reversely operating dissimilatory sulfite reductase (*dsrAB*) from *Paracatenula* samples were identified to cluster with *Alphaproteobacteria* (Loy *et al.*, 2009). At the time, contaminations could not be excluded. However, few years after this discovery the sequences could be assigned to originate from the genome of the symbionts (Gruber-Vodicka, 2011; Gruber-Vodicka *et al.*, 2011). The symbionts of *Paracatenula* are the bacteria *Candidatus* Riegeria that form a monophyletic group within the order *Rhodospirillales*, which is a basally branching order within the class *Alphaproteobacteria* (Williams *et al.*, 2007; Gruber-Vodicka *et al.*, 2011). The phylogeny of *Ca.* Riegeria could not be precisely resolved based on 16S rRNA marker genes (Figure 8D) but the chemoautotrophic lifestyle of *Ca.* Riegeria could be determined on the basis of identified key genes for the oxidation of sulfur (*dsrAB*, *aprA*) and the fixation of carbon (*cbbM*), combined with the storage of elemental sulfur (Gruber-Vodicka *et al.*, 2011). Until now, *Ca.* Riegeria are the only known alphaproteobacterial chemosynthetic symbionts.

Evidence for strict vertical transmission of the symbionts originates from congruent phylogenies of symbionts and hosts over a long evolutionary time (Figure 8E) (Gruber-Vodicka *et al.*, 2011). *Paracatenula* is thought to completely rely on their endosymbionts for nutrition, which make up half of the total volume (Ott *et al.*, 1982; Dirks *et al.*, 2011; Gruber-Vodicka *et al.*, 2011). The symbiosis is ancient and likely represents the oldest known association between animals and chemoautotrophic bacteria with an estimated age of more than 500 million years (Gruber-Vodicka *et al.*, 2011). A decreased average GC content of the coding sequences of three symbiont genes compared to other representatives of the family *Rhodospirillales* suggested population bottlenecks and a high genetic drift pointing to ongoing genome reduction processes. Apart from this, nothing was known about the genomes of *Ca*. Riegeria, neither with respect to their genetic repertoires, their functionality in the *Paracatenula* symbiosis, their ecophysiology nor their evolution over millions of years.
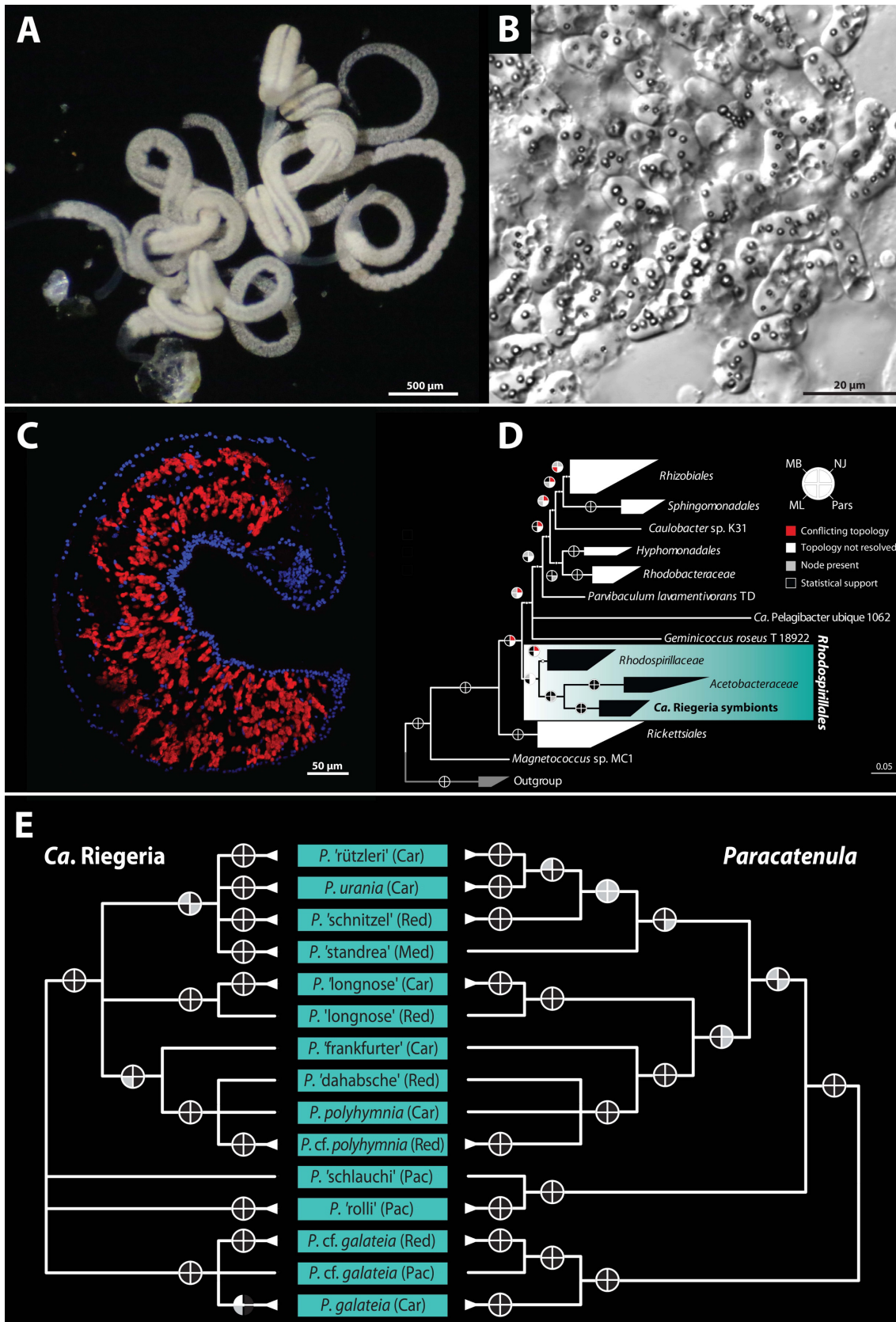
Figure 8. *Paracatenula* and its alphaproteobacterial symbionts. *A*, Habitus of *Paracatenula* individuals kept in laboratory conditions for 1066 days. ***B***, Microscopy image of *Ca.* Riegeria symbionts indicating multiple intracellular inclusions. ***C***, Visualization of *Paracatenula* intracellular symbionts on sections (red). Host nuclei were shown in blue. ***D***, Phylogenetic placement of *Ca.* Riegeria symbionts within

the *Alphaproteobacteria*. Scale bar indicates 5% sequence divergence. ***E***, Coevolution of *Ca.* Riegeria symbionts and their respective *Paracatenula* hosts. The trees are based on 16S rRNA for symbiont, and 18S and 28S rRNA for host phylogeny. Image shown in panel ***C*** was published in Kiers & West 2015. Phylogenetic trees shown in panel ***D*** and ***E*** were adapted from Gruber-Vodicka *et al*. 2011.

Regeneration events in *Paracatenula*

Paracatenula shares a fascinating feature with other *Platyhelminthes* that reproduce asexually. After cutting off their rostrum, means decapitation, the flatworms can regenerate their rostrum (Dirks, Gruber-Vodicka, Leisch, *et al.*, 2012; Wanninger, 2015). For this process, which takes about two weeks in *Paracatenula galateia*, the so-called neoblasts were shown to be essential (Dirks, Gruber-Vodicka, Leisch, *et al.*, 2012). Neoblasts have stem cell-like properties and they are distributed throughout the flatworm´s trophosome region (Wanninger, 2015). Interestingly, these cells are the source for both rostrum-regenerating cells and the symbiont-containing bacteriocytes (Dirks, Gruber-Vodicka, Leisch, *et al.*, 2012).

The processes of tissue and rostrum regeneration include three steps: (i) closing the wound, (ii) proliferation of cells and the formation of a blastema, and (iii) differentiation of cells leading to morphogenesis (Figure 9) (Dirks, Gruber-Vodicka, Leisch, *et al.*, 2012; Wanninger, 2015). In studies where tissue regeneration was induced, the locomotion of rostrum-decapitated *Paracatenula galateia* is first affected, but can be restored within five days after regeneration (Dirks, Gruber-Vodicka, Egger, *et al.*, 2012). While longitudinal nerves end blindly in the wound area after decapitation, first indications of the reorganization can be observed after seven days (Dirks, Gruber-Vodicka, Egger, *et al.*, 2012). The time for the complete rostrum regeneration in *Paracatenula galateia*, which takes two weeks, is long compared to other flatworms such as a *Catenula* species (~60 h), *Microstomum lineare* (~45 h) or *Paracatenula* cf. *polyhymnia* (48–72 h) (Moraczewski, 1977; Palmberg, 1991; Dirks, Gruber-Vodicka, Leisch, *et al.*, 2012). This difference can be explained on the basis of the different morphologies, *Paracatenula galateia* has to generate a larger body fragment than the other flatworms which are comparably smaller with diameters less than 100 µm (Dirks, Gruber-Vodicka, Egger, *et al.*, 2012).

While a gradient of morphogens along the body in tissue development is known to be important for re-establishing the body axes in tissue-regenerating flatworms, it is not well understood how the symbiont abundance and function influences their regeneration abilities (Adell *et al.*, 2010; Dirks, Gruber-Vodicka, Egger, *et al.*, 2012).
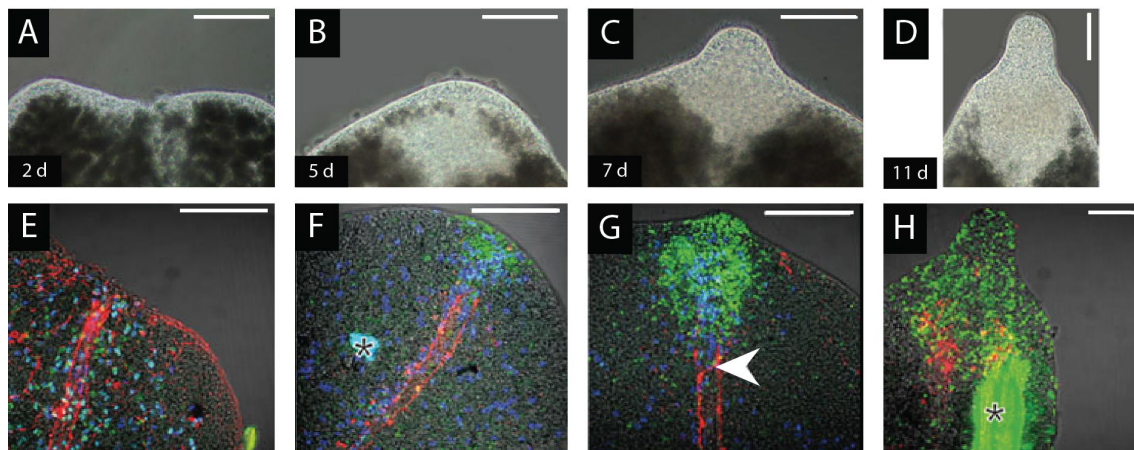
Figure 9. Rostrum regeneration in the species *Paracatenula galateia*. Images ***A–D*** show interference contrast micrographs of different rostrum stages between 2 and 11 days after sectioning. Images ***E–H*** show overlays of fluorescence micrographs of serotonergic nerves (red, labeled by serotonin staining) and neoblasts (green, labeled by EdU) labeled specimens. The images were modified from Dirks, Gruber-Vodicka, Egger, *et al*. 2012.

Mouth and gut reduction – what do chemosynthetic animals feed on?

The relationship of chemosynthetic symbionts with their animal hosts are usually considered to be nutrition-based. This implies that the primary role of the symbionts is to serve as a source of nutrition for their host (Dubilier *et al.*, 2008). Chemosynthetic symbionts of gammaproteobacterial origin were indeed shown to perform this core function in terms of thioautotrophy, i.e. to provide organics to the host *via* the fixation of inorganic carbon into biomass by using reduced sulfur compounds as energy source (Kleiner, Petersen, *et al.*, 2012). In *Gammaproteobacteria*, the same mode of nutrition has evolved convergently multiple times where key genes for thioautotrophy were taken up by horizontal gene transfer. In this case, the convergent evolution of host and symbionts is supported by incongruent phylogenies of symbionts and hosts (Kleiner, Petersen, *et al.*, 2012). Despite the partial understanding of these thioautotrophic features, there are still open questions concerning which metabolites are made accessible for the host, to what extent the symbionts contribute to the hosts proliferation and how similar global functions are among different host and symbiont groups.

One way to explain the nutrient transfer from symbiont to host is *via* transport through membranes, since some symbionts are known to harbor a broad variety of exporters for organic substrates whereas others lack most of them (Kuwahara *et al.*, 2007; Newton *et al.*, 2007; Kleiner, Wentrup, *et al.*, 2012; Dmytrenko *et al.*, 2014). Most recent studies focused on carbon fixation and the transfer of the products to the host. These studies indicate that in

addition to carbon, inorganic sources of nitrogen from the environment can be made accessible by the symbionts to be used by the hosts (Bright *et al.*, 2000; König *et al.*, 2016; Petersen *et al.*, 2016).

The question of the metabolic contribution of the symbionts to the host metabolism is even more complex in cases where hosts have partially functioning digestive systems or have the ability to filter-feed. For example, in *Bathymodiolus* mussels which still contain an attenuated gut, filter-feeding can occur to a certain degree (Page *et al.*, 1990; Pile and Young, 1999). This way, supplemental nitrogen can be provided while phosphorous and other minerals are thought to be additionally supplemented by the symbionts (Page *et al.*, 1990; Pile and Young, 1999). Similarly to *Bathymodiolus*, the vesicomyid clams kept their ability to filter-feed, supported by stomachs that were filled with various particles, although the gut itself is reduced (Le Pennec *et al.*, 1995). In both cases the enzymatic digestion of symbionts is an additional way to make substrates available to the host (Fiala-Médioni *et al.*, 1986; Newton *et al.*, 2007; Kádár *et al.*, 2008).

Although there is a first and partial understanding what symbionts have to provide in terms of nutrition for the host, it is still not understood what essential nutritional functions the symbionts have to provide to qualify as a long-term partner. Also, close to nothing is known to what extent the nutrient exchange has to occur.

Aims of this thesis

My thesis focuses on several unexplored aspects of the symbiosis between the marine flatworm *Paracatenula* and its symbiotic bacteria *Ca.* Riegeria. My goal was to reduce the gap left in our knowledge of this symbiotic association. In my study I aimed towards a better understanding of convergent evolutionary patterns in terms of metabolism and the function of an unusual alphaproteobacterial thioautotrophic symbiont. In addition, I present data concerning the patterns of reductive genome evolution and the gene reduction processes. Part of the results were obtained in collaboration to apply a large set of techniques such as correlative imaging approaches and physiological experiments, and to generate expression data to address the host and its symbionts functions that were unexplored in the *Paracatenula-Ca.* Riegeria symbiosis. Parts of my studies also involved fieldwork in the Mediterranean and Belize, which included the morphological classification of meiofauna organisms and the performance of

physiological experiments. All laboratory experiments and most of my studies were performed on *Paracatenula* specimens collected from Elba, Italy, for which I optimized the cultivation conditions to keep them alive for by now more than three years.

In the following three chapters, I present data concerning (i) the physiological characterization of the nutritional symbiosis between the mouthless *Paracatenula* flatworm and its symbiont *Ca.* Riegeria, (ii) the processes of reductive genome evolution in the *Ca.* Riegeria symbionts and (iii) the potential role of *Ca.* R. standrea in the rostrum regeneration processes of the host.

In the first part of my thesis, I focus on the physiology of the *Paracatenula* symbionts. The *Paracatenula* symbiosis is suggested to be nutritional, but besides a phylogenetic placement of the host and its symbionts, the observation of strict coevolution and thioautotrophic potential, exact functions remained unclear. Together with my collaborators, I have analyzed the metabolism and ecophysiology of the symbionts of one *Paracatenula* species that occurs in the marine sediments of Elba, Italy. A variety of culture-dependent and culture-independent approaches were optimized to the experimental needs and applied, mostly on single specimen. The methods included the cultivation of *Paracatenula* specimens, isotope-labelling combined with bulk measurements and microautoradiography, transmission electron microscopy (TEM), fluorescence *in situ* hybridization (FISH) and confocal laser-scanning microscopy, Raman spectroscopy, and current techniques of genomics, transcriptomics (for symbiont and host expression), metabolomics and proteomics.

Since *Ca.* Riegeria symbionts are unusual in their phylogenetic placement within the *Alphaproteobacteria*, we characterized and compared the core functions of chemosynthetic symbionts across divergent host and symbiont taxa. The data indicated that *Ca*. Riegeria standrea expresses a versatile physiology with a broad ability for the accumulation and mobilization of intracellular inclusions, despite having a highly reduced genome compared to its free-living relatives. The mode of nutrition seems to be through outer membrane vesicles or digestion of symbiont cells. The combination of its high abundance within the *Paracatenula* host along with the large carbon and energy stocks, indicates a so far undescribed role of symbionts not only as biosynthetic factories but also as primary energy storage for its host. The manuscript of this study is submitted to PNAS.

The second part of my thesis concerns the processes of reductive genome evolution in the *Ca.* Riegeria symbionts. The symbionts have highly reduced genome sizes when compared to their next relatives which have a 3–4 times larger genomes. Although chemosynthetic symbionts were shown to have reduced genomes, the processes underlying reductive genome evolution have not yet received much attention. Additionally, nothing was known on how these reduction processes compare to those in insect symbioses or in clades of closely related organisms.

To study footsteps of reductive genome evolution, the metagenomes of 35 *Paracatenula* organisms, including those that had been collected during my fieldtrips, were established and analyzed. Together with collaborators, I used bioinformatic tools to study gene conservations as well as gene losses over a large symbiont clade. I could show that in addition to the shared core metabolism, the symbionts revealed differential signs of gene losses. Gene losses affected parts of the DNA repair and cell division machinery and were lineage-specific, appeared stochastic and ranged from base pair erosions to full gene deletions. Genome syntenies between organisms were lineage-specific and highly variable, with one *Ca*. Riegeria clade displaying several 100-fold more rearrangements than other clades of similar phylogenetic distance. Furthermore, contrasting and atypical trajectories among the symbiont clade such as differential gene-length fragmentations and ancient genome rearrangements were identified and compared to the linear reductive trajectories known from insect symbioses.

A manuscript containing the results of this study is currently in the pre-submission state. Both co-authors have already reviewed the manuscript.

In the third part of my thesis, I address the potential role of *Ca.* R. standrea in the rostrum regeneration processes of their host. While the regeneration process is well studied in flatworms in general, literally nothing is known about the needs and contributions of the bacterial symbionts to this process. To at least partially fill this gap of knowledge, a transcriptome of decapitated *Paracatenula* was generated together with collaborators, after the flatworms have regenerated their rostrum. The results of this study were surprising and showed that more than 10% of the symbiont genes are differentially regulated. The majority of these genes were related to basic cellular processes such as translation, transcription and carbon and energy metabolism, implying that the metabolism of the symbionts is affected by induced tissue regeneration.

The research of this study represents a draft version of a manuscript.

Bibliography Introduction

Adell, T., Cebrià, F. and Saló, E. (2010) 'Gradients in planarian regeneration and homeostasis', *Cold spring harbor perspectives in biology*, 2:a000505.

Amend, J. P. and Shock, E. L. (2001) 'Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria', *FEMS Microbiology Reviews*, **25**, pp. 175–243.

Andersson, S. G. E. (2006) 'The bacterial world gets smaller', *Science*, **314**, pp. 259–260.

Arp, A. J. and Childress, J. J. (1983) 'Sulfide binding by the blood of the hydrothermal vent tube worm *Riftia pachyptila*', *Science*, **219**, pp. 295–297.

Arp, A. J., Childress, J. J. and Vetter, R. D. (1987) 'The sulphide-binding protein in the blood of the vestimentiferan tube-worm, *Riftia pachyptila*, is the extracellular haemoglobin', *Journal of Experimental Biology*, **128**, pp. 139–158.

Assié, A., Borowski, C., van der Heijden, K., Raggi, L., Geier, B., Leisch, N., Schimak, M. P., Dubilier, N. and Petersen, J. M. (2016) 'A specific and widespread association between deep-sea *Bathymodiolus* mussels and a novel family of *Epsilonproteobacteria*', *Environmental Microbiology Reports*, **8**, pp. 805–813.

de Bary, A. (1879) 'Vortrag, gehalten auf der Versammlung deutscher Naturforscher und Aerzte zu Cassel. Strassburg, Verlag von Karl J. Trübner'.

Bienhold, C., Pop Ristova, P., Wenzhöfer, F., Dittmar, T. and Boetius, A. (2013) 'How deep-sea wood falls sustain chemosynthetic life', *PLoS ONE*, **8**, pp. 1–17.

Blazejak, A., Erséus, C., Amann, R. and Dubilier, N. (2005) 'Coexistence of bacterial sulfide oxidizers, sulfate reducers, and spirochetes in a gutless worm (Oligochaeta) from the Peru margin', *Applied and environmental microbiology*, **71**, pp. 1553–1561.

Blazejak, A., Kuever, J., Erséus, C., Amann, R. and Dubilier, N. (2006) 'Phylogeny of 16S rRNA, ribulose 1,5-bisphosphate carboxylase/oxygenase, and adenosine 5'-phosphosulfate reductase genes from gamma- and alphaproteobacterial symbionts in gutless marine worms (Oligochaeta) from Bermuda and the Bahamas', *Applied and environmental microbiology*, **72**, pp. 5527–5536.

Bopp, L., Bowler, C., Guidi, L., Karsenti, E. and de Vargas, C. (2015) 'The ocean: a carbon pump', *Ocean Climate*, pp. 12–17.

Borkott, H. (1970) 'Geschlechtliche Organisation, Fortpflanzungsverhalten und Ursachen der sexuellen Vermehrung von *Stenostomum sthenum* nov. spec. (Turbellaria, *Catenulida*)', *Z. Morph. Tiere*, **67**, pp. 183–262.

Borowski, C., Giere, O., Krieger, J., Amann, R. and Dubilier, N. (2002) 'New aspects of the symbiosis in the provannid snail *Ifremeria nautilei* from the North Fiji Back Arc Basin', *Cah. Biol. Mar.*, **43**, pp. 321–324.

Boss, K. J. and Turner, R. (1980) 'The giant white clam from the Galapagos Rift *Calyptogena*

magnifica species novum', in *Malacologia*. 20th edn, pp. 161–194.

Bowles, M. W., Mogollón, J. M., Kasten, S., Zabel, M. and Hinrichs, K.-U. (2014) 'Global rates of marine sulfate reduction and implications for sub-sea-floor metabolic activities', *Science*, **344**, pp. 889–891.

Bright, M., Giere, O. (2005) 'Microbial symbiosis in Annelida', in *Symbiosis*. 38th edn, pp. 1–45.

Bright, M. and Bulgheresi, S. (2010) 'A complex journey: transmission of microbial symbionts', *Nature Reviews Microbiology*, **8**, pp. 218–230.

Bright, M., Keckeis, H. and Fisher, C. R. (2000) 'An autoradiographic examination of carbon fixation, transfer and utilization in the *Riftia pachyptila* symbiosis', *Marine biology*, **136**, pp. 621–632.

Bright, M. and Sorgo, A. (2003) 'Ultrastructural reinvestigation of the trophosome in adults of *Riftia pachyptila* (Annelida, Siboglinidae)', *Invertebrate Biology*, **122**, pp. 347–368.

Cavanaugh, C. M. (1983) 'Symbiotic chemoautotrophic bacteria in marine invertebrates from sulphide-rich habitats', *Nature*, **302**, pp. 58–61.

Cavanaugh, C. M., R., L. P., Maki, J. S., Mitchell, R. and Lidstrom, M. E. (1987) 'Symbiosis of methylotrophic bacteria and deep-sea mussels', *Nature*, **325**, pp. 346–348.

Cavanaugh, C. M., Gardiner, S. L., Jones, M. L., Jannasch, H. W. and Waterbury, J. B. (1981) 'Prokaryotic cells in the hydrothermal vent tube wom *Riftia pachyptila* Jones: Possible chemoautotrophic symbionts', *Science*, **213**, pp. 340–342.

Childress, J. J., Fisher, C. R., Brooks, J. M., C, K. I. M., Bidigare, R. and Anderson, A. E. (1986) 'A methanotrophic marine Molluscan (Bivalvia, Mytilidae) symbiosis: mussels fueled by gas', *Science*, **233**, pp. 1306–1308.

Dirks, U., Gruber-Vodicka, H. R., Egger, B. and Ott, J. A. (2012) 'Proliferation pattern during rostrum regeneration of the symbiotic flatworm *Paracatenula galateia*: A pulse-chase-pulse analysis', *Cell and tissue research*, **349**, pp. 517–525.

Dirks, U., Gruber-Vodicka, H. R., Leisch, N., Bulgheresi, S., Egger, B., Ladurner, P. and Ott, J. A. (2012) 'Bacterial symbiosis maintenance in the asexually reproducing and regenerating flatworm *Paracatenula galateia*', *PLoS ONE*, **7**, e34709.

Dirks, U., Gruber-Vodicka, H. R., Leisch, N., Sterrer, W. and Ott, J. A. (2011) 'A new species of symbiotic flatworms, *Paracatenula galateia* sp. nov. (*Platyhelminthes*: *Catenulida*: Retronectidae) from Belize (Central America)', *Marine Biology Research*, **7**, pp. 769–777.

Dmytrenko, O., Russell, S. L., Loo, W. T., Fontanez, K. M., Liao, L., Roeselers, G., Sharma, R., Stewart, F. J., Newton, I. L. G., Woyke, T., Wu, D., Lang, J. M., Eisen, J. A. and Cavanaugh, C. M. (2014) 'The genome of the intracellular bacterium of the coastal bivalve, *Solemya velum*: A blueprint for thriving in and out of symbiosis', *BMC Genomics*, **15**, 924.

Douglas, A. E. (2010) 'The significance of symbiosis', in *The Symbiotic Habit*. Princeton, New

Jersey: Princeton university press, pp. 1–214.

Van Dower, C. L. (2000) *The ecology of deep-sea hydrothermal vents*. Princeton, New Jersey: Princeton university press.

Dubilier, N., Bergin, C. and Lott, C. (2008) 'Symbiotic diversity in marine animals: the art of harnessing chemosynthesis', *Nature Reviews Microbiology*, **6**, pp. 725–740.

Dubilier, N., Blazejak, A. and Ruehland, C. (2006) *Symbioses between bacteria and gutless marine oligochaetes*, *Progress in Molecular and Subcellular Biology*. Edited by J. Overmann. New York, NY, USA: Springer-Verlag Berlin Heidelberg, pp. 251-275.

Dworkin, M. (2012) 'Sergei Winogradsky: a founder of modern microbiology and the first microbial ecologist', *FEMS Microbiology Reviews*, **36**, pp. 364–379.

Edmond, J. M., Von Damm, K. L., McDuff, R. E. and Measures, C. I. (1982) 'Chemistry of hot springs on the East Pacific Rise and their effluent dispersal', *Nature*, **297**, pp. 187–191.

Egger, B., Bachmann, L. and Fromm, B. (2017) 'Atp8 is in the ground pattern of flatworm mitochondrial genomes', *BMC Genomics*, **18**, doi:10.1186/s12864-017-3807-2.

Felbeck, H. (1981) 'Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila* Jones (Vestimentifera)', *Science*, **213**, pp. 336–338.

Feldman, R. A., Shank, T. M., Black, M. B., Baco, A. R., Smith, Craig, R. and Vrijenhoek, R. C. (1998) 'Vestimentiferan on a whale fall', *Biological Bulletin*, **194**, pp. 116–119.

Fenchel, T. M. and Riedl, R. J. (1970) 'The sulfide system: a new biotic community underneath the oxidized layer of marine sand bottoms', *Marine Biology*, **7**, pp. 255–268.

Fiala-Médioni, A., Métivier, C., Herry, A. and Le Pennec, M. (1986) 'Ultrastructure of the gill of the hydrothermal-vent mytilid *Bathymodiolus* sp.', *Marine Biology*, **92**, pp. 65–72.

Fontanez, K. M. and Cavanaugh, C. M. (2014) 'Evidence for horizontal transmission from multilocus phylogeny of deep-sea mussel (Mytilidae) symbionts', *Environmental Microbiology*, **16**, doi:10.1111/1462-2920.12379.

Frank, A. B. (1877) 'Über die biologischen Verhältnisse des Thallus einiger Krustenflechten', *Cohn's Beiträge zur Biologie der Pflanzen*, **2**, pp. 123–200.

Funk, D. J., Wernegreen, J. J. and Moran, N. A. (2001) 'Intraspecific variation in symbiont genomes: Bottlenecks and the aphid-*Buchnera* association', *Genetics*, **157**, pp. 477–489.

Gaill, F., Shillito, B., Ménard, F., Goffinet, G. and Childress, J. J. (1997) 'Rate and process of tube production by the deep-sea hydrothermal vent tubeworm *Riftia pachyptila*', *Marine Ecology Progress Series*, **148**, pp. 135–143.

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', *The ISME journal*, **6**, pp. 766–776.

Giere, O. (1981) 'The gutless marine oligochaete *Phallodrilus leukodermatus* - structural studies on an aberrant tubificid associated with bacteria', *Marine ecology progress series*, **5**, pp. 353–357.

Giere, O. (1985) 'The gutless marine tubificid *Phallodrilus planus*, a flattened Oligochaete with symbiotic bacteria. Results from morphological and ecological studies', *Zoologica Scripta*, **14**, pp. 279–286.

Giere, O. (2009) 'The microscopic motile fauna of aquatic sediments', in *Meiobenthology*. Springer-Verlag Berlin Heidelberg.

Giere, O., Conway, N. M., Gastrock, G. and Schmidt, C. (1991) '"Regulation" of gutless annelid ecology by endosymbiotic bacteria', *Marine ecology progress series*, **68**, pp. 287–299.

Giere, O. and Erséus, C. (2002) 'Taxonomy and new bacterial symbioses of gutless marine Tubificidae (Annelida, Oligochaeta) from the Island of Elba (Italy)', *Org Divers Evol*, **2**, pp. 289–297.

Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., Bibbs, L., Eads, J., Richardson, T. H., Noordewier, M., Rappe, M. S., Short, J. M., Carrington, J. C. and Mathur, E. J. (2005) 'Genome streamlining in a cosmopolitan oceanic bacterium', *Science*, **309**, pp. 1242–1245.

Gómez-Valero, L., Silva, F. J., Simon, J. C. and Latorre, A. (2007) 'Genome reduction of the aphid endosymbiont *Buchnera aphidicola* in a recent evolutionary time scale', *Gene*, **389**, pp. 87–95.

Gonza, F., Latorre, A., Gil, R., Silva, F. J., Zientz, E., Ho, B., Ham, R. C. H. J. Van, Gross, R., Rausell, C. and Kamerbeek, J. (2003) 'The genome sequence of *Blochmannia floridanus*: Comparative analysis of reduced genomes', *Proceedings of the National Academy of Sciences of the United States of America*, **100**, pp. 9388–9393.

Gruber-Vodicka, H. (2011) *Specificity and transmission in two shallow water thiotrophic symbioses*, *PhD Thesis*. Universität Wien.

Gruber-Vodicka, H. R., Dirks, U., Leisch, N., Baranyi, C., Stoecker, K., Bulgheresi, S., Heindl, N. R., Horn, M., Lott, C., Loy, A., Wagner, M. and Ott, J. (2011) '*Paracatenula*, an ancient symbiosis between thiotrophic *Alphaproteobacteria* and catenulid flatworms', *Proceedings of the National Academy of Sciences of the United States of America*, **108**, pp. 12078–12083.

Hand, S. C. (1987) 'Trophosome ultrastructure and the characterization of isolated bacteriocytes from invertebrate-sulfur bacteria symbioses', *The biological bulletin*, **173**, pp. 260–276.

Hentschel, U., Berger, E., Bright, M., Felbeck, H. and Ott, J. A. (1999) 'Metabolism of nitrogen and sulfur in ectosymbiotic bacteria of marine nematodes (Nematoda, Stilbonematinae)', *Marine Ecology Progress Series*, **183**, pp. 149–158.

Hershberg, R. and Petrov, D. A. (2010) 'Evidence that mutation is universally biased towards

AT in bacteria', *PLoS Genetics*, **6**, e1001115.

Hildebrand, F., Meyer, A. and Eyre-Walker, A. (2010) 'Evidence of selection upon genomic GC-content in bacteria', *PLoS Genetics*, **6**, e1001107.

Hurtgen, M. T. (2012) 'The marine sulfur cycle, revisited', *Science*, **337**, pp. 305–306.

Jäckle, O. (2018) 'Mouth- and gutless in tropical marine sediments – Symbiosis makes it possible'*.* In *Scientific partnership for a better future*. Edited by H. G, H. I, and H. A-K. Bremen, Germany: Edition Falkenberg, pp. 25–28.

Jan, C., Petersen, J. M., Werner, J., Teeling, H., Huang, S., Glöckner, F. O., Golyshina, O. V, Dubilier, N., Golyshin, P. N., Jebbar, M. and Cambon-Bonavita, M.-A. (2014) 'The gill chamber epibiosis of deep-sea shrimp *Rimicaris exoculata*: an in-depth metagenomic investigation and discovery of *Zetaproteobacteria*', *Environmental microbiology*, **16**, pp. 2723–2738.

Jannasch, H. W. (1985) 'The chemosynthetic support of life and the microbial diversity at deep-sea hydrothermal vents', *Proceedings of the Royal Society B*, **225**, pp. 277–297.

Jiang, J., Chan, A., Ali, S., Saha, A., Haushalter, K. J., Lam, W. L. M. R., Glasheen, M., Parker, J., Brenner, M., Mahon, S. B., Patel, H. H., Ambasudhan, R., Lipton, S. A., Pilz, R. B. and Boss, G. R. (2016) 'Hydrogen sulfide-mechanisms of toxicity and development of an antidote', *Scientific Reports*, **6**, 6:20831

Jones, M. L. (1981) '*Riftia pachyptila* Jones: Observations on the vestimentiferan worm from the Galapagos Rift', *Science*, **213**, pp. 333–336.

Kádár, E., Davis, S. A. and Lobo-Da-Cunha, A. (2008) 'Cytoenzymatic investigation of intracellular digestion in the symbiont-bearing hydrothermal bivalve *Bathymodiolus azoricus*', *Marine Biology*, **153**, pp. 995–1004.

Kenk, C. and Wilson, B. R. (1985) 'New mussel from hydrothermal vents', in *Malacologia*, pp. 253–271.

Kiers, E. T. and West, S. A. (2015) 'Evolving new organisms *via* symbiosis', *Science*, **348**, pp. 392–394.

Kleiner, M., Petersen, J. M. and Dubilier, N. (2012) 'Convergent and divergent evolution of metabolism in sulfur-oxidizing symbionts and the role of horizontal gene transfer', *Current Opinion in Microbiology*, **15**, pp. 621–631.

Kleiner, M., Wentrup, C., Holler, T., Lavik, G., Harder, J., Lott, C., Littmann, S., Kuypers, M. M. M. and Dubilier, N. (2015) 'Use of carbon monoxide and hydrogen by a bacteria-animal symbiosis from seagrass sediments', *Environmental Microbiology*, **17**, pp. 5023–5035.

Kleiner, M., Wentrup, C., Lott, C., Teeling, H., Wetzel, S., Young, J., Chang, Y.-J., Shah, M., VerBerkmoes, N. C., Zarzycki, J., Fuchs, G., Markert, S., Hempel, K., Voigt, B., Becher, D., Liebeke, M., Lalk, M., Albrecht, D., Hecker, M., Schweder, T. and Dubilier, N. (2012) 'Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use', *Proceedings of the National Academy of*

Sciences of the United States of America, **109**, pp. E1173–E1182.

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', *Nature Microbiology*, **2**, 16193.

Kuwahara, H., Takaki, Y., Shimamura, S., Yoshida, T., Maeda, T., Kunieda, T. and Maruyama, T. (2011) 'Loss of genes for DNA recombination and repair in the reductive genome evolution of thioautotrophic symbionts of *Calyptogena* clams', *BMC Evolutionary Biology*, **11**, doi:10.1186/1471-2148-11-285.

Kuwahara, H., Takaki, Y., Yoshida, T., Shimamura, S., Takishita, K., Reimer, J. D., Kato, C. and Maruyama, T. (2008) 'Reductive genome evolution in chemoautotrophic intracellular symbionts of deep-sea *Calyptogena* clams', *Extremophiles*, **12**, pp. 365–374.

Kuwahara, H., Yoshida, T., Takaki, Y., Shimamura, S., Nishi, S., Harada, M., Matsuyama, K., Takishita, K., Kawato, M., Uematsu, K., Fujiwara, Y., Sato, T., Kato, C., Kitagawa, M., Kato, I. and Maruyama, T. (2007) 'Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*', *Current Biology*, **17**, pp. 881–886.

Larsson, K. and Jondelius, U. (2008) 'Phylogeny of *Catenulida* and support for *Platyhelminthes*', *Organisms Diversity & Evolution*, **8**, pp. 378–387.

Laurent, M. C. Z., Gros, O., Brulport, J.-P., Gaill, F. and Bris, N. Le (2009) 'Sunken wood habitat for thiotrophic symbiosis in mangrove swamps', *Marine Environmental Research*, **67**, pp. 83–88.

Lee, R. W. and Childress, J. J. (1994) 'Assimilation of inorganic nitrogen by marine invertebrates and their chemoautotrophic and metanotrophic symbionts', *Applied and Environmental Microbiology*, **60**, pp. 1852–1858.

Lee, R. W., Thuesen, E. V. and Childress, J. J. (1992) 'Ammonium and free amino acids as nitrogen sources for the chemoautotrophic symbiosis *Solemya reidi* Bernard (Bivalvia: Protobranchia)', *Journal of Experimental Marine Biology and Ecology*, **158**, pp. 75–91.

Leisch, N., Dirks, U., Gruber-Vodicka, H. R., Schmid, M., Sterrer, W. and Ott, J. A. (2011) 'Microanatomy of the trophosome region of *Paracatenula* cf. *polyhymnia* (*Catenulida*, *Platyhelminthes*) and its intracellular symbionts', *Zoomorphology*, **130**, pp. 261–271.

Li, Y., Liles, M. R. and Halanych, K. M. (2018) 'Endosymbiont genomes yield clues of tubeworm success', The *ISME Journal*, **12**, pp. 2785–2795.

Lind, P. A. and Andersson, D. I. (2008) 'Whole-genome mutational biases in bacteria', *Proceedings of the National Academy of Sciences of the United States of America*, **105**, pp. 17878–17883.

Little, A. E. F. (2010) 'Parasitism is a strong force shaping the fungus-growing ant-microbe symbiosis', in Seckbach, J. and Grube, M. (eds) *Symbioses and stress – joint ventures in*

biology. 17th edn. Springer Science+Business Media B.V., pp. 247–264.

Littlewood, D. T. J. (2008) '*Platyhelminth* systematics and the emergence of new characters', *Parasite*, **15**, pp. 333–341.

Lorion, J., Kiel, S., Faure, B., Kawato, M., Ho, S. Y. W., Marshall, B., Tsuchida, S., Miyazaki, J. I. and Fujiwara, Y. (2013) 'Adaptive radiation of chemosymbiotic deep-sea mussels', *Proceedings of the Royal Society B*, **280**, 20131243.

Loy, A., Duller, S., Baranyi, C., Mußmann, M., Ott, J., Sharon, I., Béjà, O., Le Paslier, D., Dahl, C. and Wagner, M. (2009) 'Reverse dissimilatory sulfite reductase as phylogenetic marker for a subgroup of sulfur-oxidizing prokaryotes', *Environmental Microbiology*, **11**, pp. 289–299.

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*, pp. 663–664.

Manzano-Marín, A. and Latorre, A. (2016) 'Snapshots of a shrinking partner: Genome reduction in *Serratia symbiotica*', *Scientific Reports*, **6**, 32590, doi:10.1038/srep32590.

Margulis, L. (1970) *Origin of eukaryotic cells*. 18th edn. New Haven: Yale University press.

Margulis, L. and Fester, R. (1991) 'Symbiosis as a source of evolutionary innovation', *Cambridge, MA, USA: MIT Press*.

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**, pp. 247–250.

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**, pp. 3106–3117.

Martin, B. D. and Schwab, E. (2012) 'Symbiosis: "Living together" in chaos', *Studies in History of Biology*, pp. 7–25.

McCutcheon, J. P., McDonald, B. R. and Moran, N. A. (2009a) 'Convergent evolution of metabolic roles in bacterial co-symbionts of insects', *Proceedings of the National Academy of Sciences of the United States of America*, **106**, pp. 15394–15399.

McCutcheon, J. P., McDonald, B. R. and Moran, N. A. (2009b) 'Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont', *PLoS Genetics*, **5**, e1000565.

McCutcheon, J. P. and Moran, N. A. (2007) 'Parallel genomic evolution and metabolic interdependence in an ancient symbiosis', *Proceedings of the National Academy of Sciences of the United States of America*, **104**, pp. 19392–19397.

McCutcheon, J. P. and Moran, N. A. (2010) 'Functional convergence in reduced genomes of bacterial symbionts spanning 200 my of evolution', *Genome Biology and Evolution*, **2**, pp. 708–718.

McCutcheon, J. P. and Moran, N. A. (2012) 'Extreme genome reduction in symbiotic bacteria', *Nature Reviews Microbiology*, **10**, pp. 13–26.

Di Meo, C. A., Wilbur, A. E., Holben, W. E., Feldman, R. A., Vrijenhoek, R. C. and Cary, S. C. (2000) 'Genetic variation among endosymbionts of widely distributed vestimentiferan tubeworms', *Applied and environmental microbiology*, **66**, pp. 651–658.

Mira, A. and Moran, N. A. (2002) 'Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria', *Microbial Ecology*, **44**, pp. 137–143.

Moraczewski, J. (1977) 'Asexual reproduction and regeneration of *Catenula* (Turbellaria, Archoophora)', *Zoomorphologie*, **88**, pp. 65–80.

Moran, N. A. (1996) 'Accelerated evolution and Muller's rachet in endosymbiotic bacteria', *Proceedings of the National Academy of Sciences of the United States of America*, **93**, pp. 2873–2878.

Moran, N. A. (2002) 'Microbial minimalism: Genome reduction in bacterial pathogens', *Cell*, **108**, pp. 583–586.

Moran, N. A. (2006) 'Symbiosis', *Current Biology*, **16**, pp. R866–R871.

Moran, N. A., McCutcheon, J. P. and Nakabachi, A. (2008) 'Genomics and evolution of heritable bacterial symbionts', *Annual Review of Genetics*, **42**, pp. 165–190.

Moran, N. A. and Plague, G. R. (2004) 'Genomic changes following host restriction in bacteria', *Current Opinion in Genetics and Development*, **14**, pp. 627–633.

Moya, A., Peretó, J., Gil, R. and Latorre, A. (2008) 'Learning how to live together: genomic insights into prokaryote–animal symbioses', *Nature Reviews Genetics*, **9**, pp. 218–229.

Muller, H. J. (1964) 'The relation of recombination to mutational advance', *Mutation Research*, **1**, pp. 2–9.

Newton, I. L. G., Woyke, T., Auchtung, T. A., Dilly, G. F., Dutton, R. J., Fisher, M. C., Fontanez, K. M., Lau, E., Stewart, F. J., Richardson, P. M., Barry, K. W., Saunders, E., Detter, J. C., Wu, D., Eisen, J. A. and Cavanaugh, C. M. (2007) 'The *Calyptogena magnifica* chemoautotrophic symbiont genome', *Science*, **315**, pp. 998–1000.

Ngamniyom, A. and Panyarachun, B. (2016) '*Stenostomum* cf. *leucops* (*Platyhelminthes*) in Thailand: a surface observation using scanning electron microscopy and phylogenetic analysis based on 18S ribosomal DNA sequences', *Sonklanakarin J. Sci. Technol.*, **38**, pp. 41–45.

Nicks, T. and Rahn-Lee, L. (2017) 'Inside out: Archaeal ectosymbionts suggest a second model of reduced-genome evolution', *Frontiers in Microbiology*, **8**, 384.

Nussbaumer, A. D., Fisher, C. R. and Bright, M. (2006) 'Horizontal endosymbiont transmission in hydrothermal vent tubeworms', *Nature*, **441**, pp. 345–348.

Ohta, T. (1973) 'Slightly deleterious mutant substitutions in evolution', *Nature*, **246**, pp. 96–98.

Ohta, T. (1992) 'The nearly neutral theory of molecular evolution', *Annual Review of Ecology and Systematics*, **23**, pp. 263–286.

Ott, J. A., Novak, R., Schiemer, F., Hentschel, U., Nebelsick, M. and Polz, M. (1991) 'Tackling the sulfide gradient: A novel strategy involving marine nematodes and chemoautotrophic ectosymbionts', *Marine Ecology*, **12**, pp. 261–279.

Ott, J., Bright, M. and Bulgheresi, S. (2005) 'Marine microbial thiotrophic ectosymbioses', in Gibson, R. N., Atkinson, R. J. A., M., G. J. D., and Barnes, H. (eds) *In Oceanography and Marine Biology*. 42nd edn. CRC Press, pp. 95–118.

Ott, J., Rieger, G., Rieger, R. and Enderes, F. (1982) 'New mouthless interstitial worms from the sulfide system: Symbiosis with prokaryotes', *Marine Ecology*, **3**, pp. 313–333.

Page, H. M., Fisher, C. R. and Childress, J. J. (1990) 'Role of filter-feeding in the nutritional biology of a deep-sea mussel with methanotrophic symbionts', *Marine Biology*, **104**, pp. 251–257.

Palmberg, I. (1991) 'Differentiation during asexual reproduction and regeneration in a microturbellarian', *Hydrobiologia*, **227**, pp. 1–10.

Pelseneer, P. (1891) 'Contribution à l´étude des Lamellibranches', *Archives de Biologique*, **11**, pp. 147–312.

Le Pennec, M., Beninger, P. G. and Herry, A. (1995) 'Feeding and digestive adaptations of bivalve molluscs to sulphide-rich habitats', *Comparative Biochemistry and Physiology*, **111A**, pp. 183–189.

Le Pennec, M. and Hily, A. (1984) 'Anatomie, structure et ultrastructure de la branchie d' un Mytilidae des sites hydrothermaux du Pacifique oriental', *Oceanologia Acta*, **7**, pp. 517–523.

Perner, M., Gonnella, G., Kurtz, S. and LaRoche, J. (2014) 'Handling temperature bursts reaching 464°C: Different microbial strategies in the sisters peak hydrothermal chimney', *Applied and Environmental Microbiology*, **80**, pp. 4585–4598.

Petersen, J. M., Kemper, A., Gruber-Vodicka, H., Cardini, U., van der Geest, M., Kleiner, M., Bulgheresi, S., Mußmann, M., Herbold, C., Seah, B. K. B., Antony, C. P., Liu, D., Belitz, A. and Weber, M. (2016) 'Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation', *Nature Microbiology*, **2**, 16195.

Petersen, J. M., Wentrup, C., Verna, C., Knittel, K. and Dubilier, N. (2012) 'Origins and evolutionary flexibility of chemosynthetic symbionts from deep-sea animals', *Biological Bulletin*, **223**, pp. 123–137.

Petersen, J. M., Zielinski, F. U., Pape, T., Seifert, R., Moraru, C., Amann, R., Hourdez, S., Girguis, P. R., Wankel, S. D., Barbe, V., Pelletier, E., Fink, D., Borowski, C., Bach, W. and Dubilier, N. (2011) 'Hydrogen is an energy source for hydrothermal vent symbioses', *Nature*, **476**, pp. 176–180.

Pile, A. J. and Young, C. M. (1999) 'Plankton availability and retention efficiencies of cold-seep

symbiotic mussels', *Limnology and Oceanography*, **44**, pp. 1833–1839.

Pimenov, N. V., Savvichev, A. S., Rusanov, I. I., Lein, A. Y. and Ivanov, M. V. (2000) 'Microbiological processes of the carbon and sulfur cycles at cold methane seeps of the North Atlantic', *Mikobiologiia*, **69**, pp. 709–720.

Plague, G. R., Dunbar, H. E., Tran, P. L. and Moran, N. A. (2008) 'Extensive proliferation of transposable elements in heritable bacterial symbionts', *Journal of Bacteriology*, **190**, pp. 777–779.

Ponnudurai, R., Sayavedra, L., Kleiner, M., Heiden, S. E., Thürmer, A., Felbeck, H., Schlüter, R., Sievert, S. M., Daniel, R., Schweder, T. and Markert, S. (2017) 'Genome sequence of the sulfur-oxidizing *Bathymodiolus thermophilus* gill endosymbiont', *Standards in Genomic Sciences*. Standards in Genomic Sciences, **12**, doi:10.1186/s40793-017-0266-y.

Powell, E. (1989) 'Oxygen, sulfide and diffusion: why thiobiotic meiofauna must be sulfide-insensitive first-order respirers', *Journal of Marine Research*, **47**, pp. 887–932.

Raghavan, R., Kelkar, Y. D. and Ochman, H. (2012) 'A selective force favoring increased G + C content in bacterial genes', *PNAS*, **109**, pp. 14504–14507.

Reid, R. G. B. and Bernard, F. R. (1980) 'Gutless bivalves', *Science*, **208**, pp. 609–610.

Rubin-Blum, M., Antony, C. P., Borowski, C., Sayavedra, L., Pape, T., Sahling, H., Bohrmann, G., Kleiner, M., Redmond, M. C., Valentine, D. L. and Dubilier, N. (2017) 'Short-chain alkanes fuel mussel and sponge *Cycloclasticus* symbionts from deep-sea gas and oil seeps', *Nature Microbiology*, **17093**, 2:17093.

Rubin-Blum, M., Antony, C. P., Sayavedra, L., Martínez-Pérez, C., Birgel, D., Peckmann, J., Wu, Y.-C., Cardenas, P., MacDonald, I., Marcon, Y., Sahling, H., Hentschel, U. and Dubilier, N. (Submitted manuscript) 'Fueled by methane: Deep-sea sponges from asphalt seeps gain their nutrition from methane-oxidizing symbionts'.

Ruehland, C., Blazejak, A., Lott, C., Loy, A., Erséus, C. and Dubilier, N. (2008) 'Multiple bacterial symbionts in two species of co-occurring gutless oligochaete worms from Mediterranean sea grass sediments', *Environmental Microbiology*, **10**, pp. 3404–3416.

Russell, S. L., Corbett-Detig, R. B. and Cavanaugh, C. M. (2017) 'Mixed transmission modes and dynamic genome evolution in an obligate animal–bacterial symbiosis', *The ISME Journal*, **11**, pp. 1359–1371.

Sayavedra, L., Kleiner, M., Ponnudurai, R., Wetzel, S., Pelletier, E., Barbe, V., Satoh, N., Shoguchi, E., Fink, D., Breusing, C., Reusch, T. B. H., Rosenstiel, P., Schilhabel, M. B., Becher, D., Schweder, T., Markert, S., Dubilier, N. and Petersen, J. M. (2015) 'Abundant toxin-related genes in the genomes of beneficial symbionts from deep-sea hydrothermal vent mussels', *eLife*, **4**, e07966.

Schmaljohann, R. and Flugel, H. J. (1987) 'Methane-oxidizing bacteria in Pogonophora', *Sarsia*, **72**, pp. 91–98.

Seah, B. K. B., Schwaha, T., Volland, J.-M., Huettel, B., Dubilier, N. and Gruber-Vodicka, H. R. (2017) 'Specificity in diversity: single origin of a widespread ciliate-bacteria symbiosis', *Proceedings of the Royal Society B*, **284**, doi:10.1098/rspb.2017.0764.

Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. and Ishikawa, H. (2000) 'Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS', *Nature*, **407**, pp. 81–86.

Shimamura, S., Kaneko, T., Ozawa, G., Matsumoto, M. N., Koshiishi, T., Takaki, Y., Kato, C., Takai, K., Yoshida, T., Fujikura, K., Barry, J. P. and Maruyama, T. (2017) 'Loss of genes related to Nucleotide Excision Repair (NER) and implications for reductive genome evolution in symbionts of deep-sea vesicomyid clams', *Plos One*, **12**, e0171274.

Sterrer, W. and Rieger, R. (1974) 'Retronectidae-a new cosmopolitan marine family of *Catenulida* (Turbellaria)', in Riser N, M. M. (eds). (ed.) *Biology of the turbellaria*. McGraw-Hill, New York, pp. 63–92.

Stewart, F. J. and Cavanaugh, C. M. (2006) 'Bacterial endosymbioses in *Solemya* (Mollusca: Bivalvia)-model systems for studies of symbiont-host adaptation', *Antonie van Leeuwenhoek*, **90**, pp. 343–360.

Sueoko, N. (1962) 'On the genetic basis of variation and heterogeneity of DNA base composition', *Proceedings of the National Academy of Sciences of the United States of America*, **48**, pp. 582–592.

Tian, R.-M., Zhang, W., Cai, L., Wong, Y.-H., Ding, W. and Qian, P.-Y. (2017) 'Genome reduction and microbe-host interactions drive adaptation of a sulfur-oxidizing bacterium associated with a cold seep sponge', *mSystems*, **2**, e00184-16.

Toft, C. and Andersson, S. G. E. (2010) 'Evolutionary microbial genomics: insights into bacterial host adaptation', *Nature Reviews Genetics*, **11**, pp. 465–475.

Treude, T., Smith, C. R., Wenzhöfer, F., Carney, E., Bernardino, A. F., Hannides, A. K., Krüger, M. and Boetius, A. (2009) 'Biogeochemistry of a deep-sea whale fall: sulfate reduction, sulfide efflux and methanogenesis', *Marine Ecology Progress Series*, **382**, pp. 1–21.

Venton, D. (2012) 'Highlight: Tiny bacterial genome opens a huge mystery: AT mutational bias in *Hodgkinia*', *Genome Biology and Evolution*, **4**, pp. 28–29.

Wanninger, A. (2015) *Evolutionary Developmental Biology of Invertebrates*, doi:10.1007/978-3-7091-1871-9.

Wasmund, K., Mußmann, M. and Loy, A. (2017) 'The life sulfuric: microbial ecology of sulfur cycling in marine sediments', *Environmental Microbiology Reports*, **9**, pp. 323–344.

Wentrup, C., Wendeberg, A., Schimak, M., Borowski, C. and Dubilier, N. (2014) 'Forever competent: deep-sea bivalves are colonized by their chemosynthetic symbionts throughout their lifetime', *Environmental Microbiology*, **16**, pp. 3699–3713.

Williams, K. P., Sobral, B. W. and Dickerman, A. W. (2007) 'A robust species tree for the

Alphaproteobacteria', *Journal of Bacteriology*, **189**, pp. 4578–4586.

Winogradsky, S. (1887) 'Über Schwefelbakterien', *Bot. Zeit*, **45**, pp. 489–610.

Wolf, Y. I. and Koonin, E. V. (2013) 'Genome reduction as the dominant mode of evolution', *BioEssays*, **35**, pp. 829–837.

Won, Y. J., Hallam, S. J., O'Mullan, G. D., Pan, I. L., Buck, K. R. and Vrijenhoek, R. C. (2003) 'Environmental acquisition of thiotrophic endosymbionts by deep-sea mussels of the genus *Bathymodiolus*', *Applied and Environmental Microbiology*, **69**, pp. 6785–6792.

Wood, A. P., Aurikko, J. P. and Kelly, D. P. (2004) 'A challenge for 21st century molecular biology and biochemistry: What are the causes of obligate autotrophy and methanotrophy?', *FEMS Microbiology Reviews*, **28**, pp. 335–352.

Wu, D., Daugherty, S. C., Van Aken, S. E., Pai, G. H., Watkins, K. L., Khouri, H., Tallon, L. J., Zaborsky, J. M., Dunbar, H. E., Tran, P. L., Moran, N. A. and Eisen, J. A. (2006) 'Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters', *PLoS Biology*, **4**, pp. 1079–1092.

Zal, F., Leize, E., Lallier, F. H., Toulmond, A., Van Dorsselaer, A. and Childress, J. J. (1998) 'S-Sulfohemoglobin and disulfide exchange: The mechanisms of sulfide binding by *Riftia pachyptila* hemoglobins', *Proceedings of the National Academy of Sciences of the United States of America*, **95**, pp. 8997–9002.

Zgadzaj, R., Garrido-Oter, R., Jensen, D. B., Koprivova, A., Schulze-Lefert, P. and Radutoiu, S. (2016) 'Root nodule symbiosis in *Lotus japonicus* drives the establishment of distinctive rhizosphere, root, and nodule bacterial communities', *Proceedings of the National Academy of Sciences of the United States of America*, **113**, pp. E7996–E8005.

Zimmermann, J., Wentrup, C., Sadowski, M., Blazejak, A., Gruber-Vodicka, H. R., Kleiner, M., Ott, J. A., Cronholm, B., De Wit, P., Erséus, C. and Dubilier, N. (2016) 'Closely coupled evolutionary history of ecto- and endosymbionts from two distantly related animal phyla', *Molecular Ecology*, **25**, pp. 3203–3223.

Chapter II: Symbiont physiology

A chemosynthetic symbiont with a drastically reduced genome serves as primary energy storage in the marine flatworm *Paracatenula*

Oliver Jäckle[1], Brandon K. B. Seah[1], Mȧlin Tietjen[1], Nikolaus Leisch[1], Manuel Liebeke[1], Manuel Kleiner[2,3], Jasmine S. Berg[1,4], Harald R. Gruber-Vodicka[1]

[1]Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, Germany

[2]Department of Geoscience, University of Calgary, 2500 University Drive Northwest, Alberta T2N 1N4, Canada

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

[4]Institut de Minéralogie, Physique des Matériaux et Cosmochimie, Université Pierre et Marie Curie, 4 Place Jussieu, 75252 Paris Cedex 05, France

Publication status: Manuscript submitted to PNAS

Abstract

Hosts of chemoautotrophic bacteria typically have much higher biomass than their symbionts and consume symbiont cells for nutrition. In contrast to this, chemoautotrophic *Ca*. Riegeria symbionts in mouthless *Paracatenula* flatworms comprise up to half of the biomass of the consortium. Each species of *Paracatenula* harbors a specific *Ca*. Riegeria and the symbionts have been vertically transmitted for at least 500 million years. Such prolonged strict vertical transmission leads to streamlining of symbiont genomes and the retained physiological capacities reveal the services the symbionts provide to their hosts. Here we studied a species of *Paracatenula* from Sant´Andrea, Elba, Italy using genomics, gene expression, imaging analyses as well as targeted and untargeted mass spectrometry. We show that its symbiont, *Ca*. R. standrea has a drastically smaller genome (1.34 Mb) than the symbiont´s free-living relatives (4.29–4.97 Mb), but retains a versatile and energy-efficient metabolism. It encodes and expresses a complete intermediary carbon metabolism and enhanced carbon fixation through anaplerosis, and accumulates massive intracellular inclusions such as sulfur, polyhydroxyalkanoates, and carbohydrates. Compared to symbiotic and free-living chemoautotrophs, *Ca*. R. standrea's versatility in energy storage is unparalleled in chemoautotrophs with such compact genomes. Transmission electron microscopy as well as host and symbiont expression data indicates that *Ca.* R. standrea provisions its host *via* outer membrane vesicle secretion. With its high share of biomass in the symbiosis and large standing stocks of carbon and energy reserves, it has a unique role for bacterial symbionts – serving as the primary energy storage for its animal host.

Significance statement

Animals typically store their primary energy reserves in specialized cells. Here we show that in the small marine flatworm *Paracatenula*, this function is performed by its bacterial chemosynthetic symbiont. The intracellular symbiont occupies half of the biomass in the symbiosis and has a highly reduced genome, but efficiently stocks up and maintains carbon and energy, particularly sugars. The symbiont provides the bulk nutrition by secreting outer membrane vesicles that are digested by the host. Under normal conditions, these vesicles completely support the host demands. This is in contrast to all other nutritional symbioses

where the hosts digests full cells of a small and ideally growing symbiont population without long-term buffering capacity during nutrient limitation.

Introduction

The discovery of dense animal communities at deep-sea hydrothermal vents in the late 1970s led to paradigm shifts in deep-sea ecology and animal physiology. Large macrofauna was shown to be independent of photoautotrophic production and instead lives off primary production performed by chemosynthetic bacterial symbionts (1–3). In the most extreme forms, animals such as the giant tubeworms have lost their mouth and digestive system and instead are nutritionally dependent on their symbionts (4).

Since then, chemosynthetic symbioses have been documented from diverse habitats and in various lineages of marine invertebrates and protists (5, 6). The symbionts are all *Proteobacteria*, typically *Gamma-* or *Epsilonproteobacteria*, that are either vertically transmitted or horizontally acquired from the environment (5–12). Genomes from chemosynthetic symbionts can range from 4.88 Mb to 1.02 Mb, reflecting varying metabolic- and ecological strategies (13–15). Surprisingly, some mouth- and gutless hosts take up their symbionts anew in every generation, although the lack of a digestive system suggests extreme dependence on their symbionts (12, 16). As these symbionts have to thrive in both free-living environmental and symbiotic states, it is difficult to attribute their genomic features to either services they provide to their host, or functions that are necessary for environmental survival or to both.

The smallest genomes of chemoautotrophic symbionts have been observed for the gammaproteobacterial symbionts of vesicomyid clams that are directly transmitted between host generations (13, 14). Such strict vertical transmission leads to substantial and ongoing genome reduction. Reduced genomes of intracellular symbionts reflect the essential set of functions that the symbionts provide to their invertebrate hosts (17–23). Host provisioning by chemosynthetic bacteria is accomplished through a conveyor belt-like turnover of symbiont cells and compared to their hosts, they only have a small share of the total biomass in the consortium (24). The patterns of genome reduction in the symbionts of vesicomyid clams

mirror this mode of nutrition since they only retained limited metabolic and storage capacities and primarily serve as biosynthesis factories (13, 14, 18).

A second host group that lives in obligate symbiosis with vertically transmitted chemosynthetic symbionts is *Paracatenula*, a genus of marine flatworms that can be a dominant member of the meiofauna in shallow-water sediments (9). Unlike the vesicomyid clams that can still filter feed (3, 25), *Paracatenula* lack both a mouth and gut and are completely dependent on intracellular endosymbionts for nutrition (9, 26). All *Paracatenula* host species harbor a species-specific symbiont phylotype (9). These symbionts, called *Candidatus* Riegeria, comprise a third to a half of the total animal volume and are housed in specialized cells, the bacteriocytes, that make up the nutritive trophosome organ (9, 27). *Ca.* Riegeria are the only known alphaproteobacterial chemosynthetic symbionts (9). They are the oldest extant clade of symbiotic chemoautotrophic bacteria, having been passed on in strict vertical transmission for more than 500 million years since the last common ancestor of all *Paracatenula* species (9).

In this study, we focus on a *Paracatenula* species that is highly abundant in the coarse shallow-water sediments of the bay Sant´Andrea on Elba, Italy (Supplementary Note 1). We propose the name *Ca.* Riegeria standrea for its endosymbionts, in reference to the collection site. Based on the genome of *Ca.* R. standrea that we generated using shotgun metagenomics, we analyzed its metabolism, ecophysiology and evolution. Compared to free-living alphaproteobacterial relatives from the *Rhodospirillaceae*, the genome is highly reduced. We reconstructed the physiology of *Ca.* R. standrea by integrating transcriptomics and correlative imaging approaches with highly sensitive mass spectrometry based methods for proteomics and metabolomics. Despite the reduced genome, the symbiont performs energy-efficient sulfide oxidation and couples it to versatile means of carbon fixation and carbon storage. The symbionts likely provision their host *via* outer membrane vesicle (OMV) secretion. With its large share of the biomass in the consortium, the symbiont has a unique role in animal-bacteria associations – serving as the bulk storage of carbon and energy reserves for its animal host.

Results and Discussion

Ca. R. standrea belongs to *Rhodospirillaceae* and has a reduced genome

The complete *Ca.* R. standrea genome was assembled into one contig from the sequenced metagenome of a single *Paracatenula* sp. standrea specimen. The genome had a size of 1.34 Mb, encoding 1344 protein-coding and 56 RNA genes at a coding density of 83.9% and a genomic GC of 51.8% (Supplementary Figure 1). The *Ca.* R. standrea genome encoded pathways for sulfur oxidation based on reverse-acting dissimilatory sulfate reductase (rDSR) and for carbon fixation using the Calvin-Benson-Bassham (CBB) cycle, indicating that *Ca.* R. standrea could acquire energy and carbon *via* thioautotrophy (Supplementary Note 2).

Ca. R. standrea was classified as a member of the family *Rhodospirillaceae* within the class *Alphaproteobacteria*, based on a phylogenetic analysis of 43 conserved single-copy marker genes (Figure 1). It was phylogenetically nested among *Rhodospirillaceae* taxa that have genome sizes of 4.29 to 4.97 Mb, indicating that massive genome reduction has occurred in the lineage leading to *Ca.* R. standrea (Figure 1, Supplementary Table 1). The *Rhodospirillaceae* are remarkably diverse and include photoheterotrophs, chemoheterotrophs but also chemoautotrophs, diazotrophs and even magnetotactic genera (28–31). The taxa most closely related to the *Ca*. R. standrea symbiont included *Rhodospirillum rubrum* [56.19% amino acid identity (AAI)], *Novispirillum itersonii* (56.22% AAI), *Caenispirillum salinarum* (56.99% AAI) and *Magnetospirillum magneticum* (57.01% AAI).
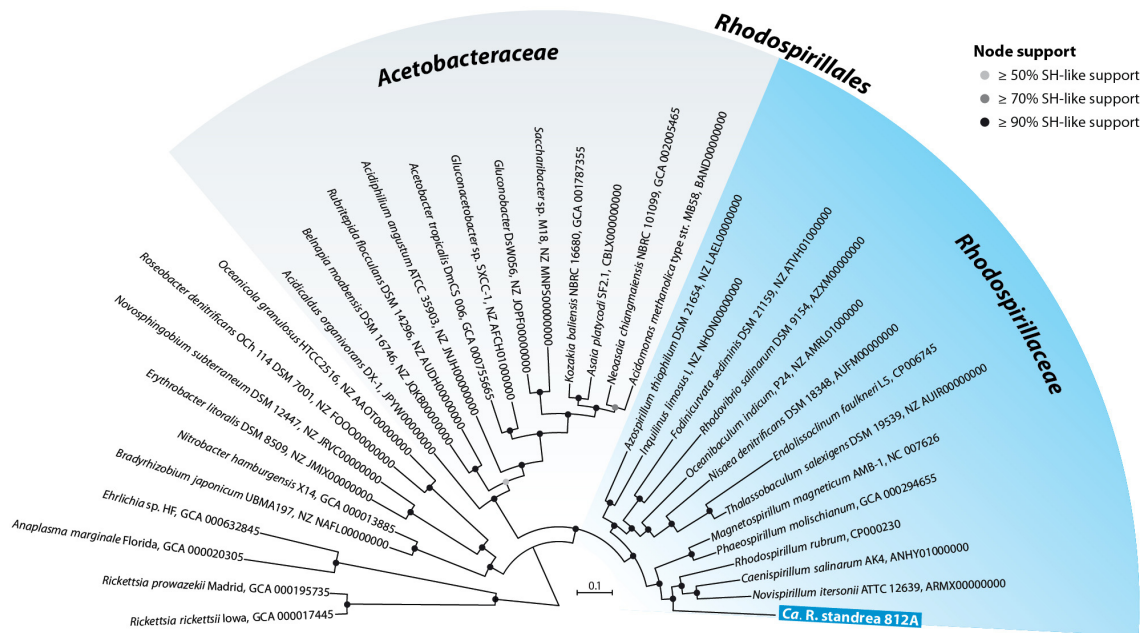
Figure 1. The symbiont *Ca.* R. standrea clusters deeply within *Rhodospirillaceae*. Phylogenetic reconstruction based on a protein alignment of 43 conserved marker genes calculated using FastTree and rooted with *Rickettsiales* as outgroup*.*

Ca. R. standrea is a chemoautotroph with a versatile carbon metabolism despite its reduced genome

To investigate the physiology of *Ca.* R. standrea, we applied a combination of genomics, transcriptomics and imaging analyses with sensitive mass spectrometry protocols that allow proteomic and metabolomic measurements from single microscopic animals. The expression data from specimens sampled directly from the environment covered up to 1288 out of 1400 predicted genes (n = 3 transcriptomes), of which 407 were also identified in the proteomes generated from single *Paracatenula* individuals as well as pools of up to three specimens (n = 8) (Figure 2, Dataset S1–S3).
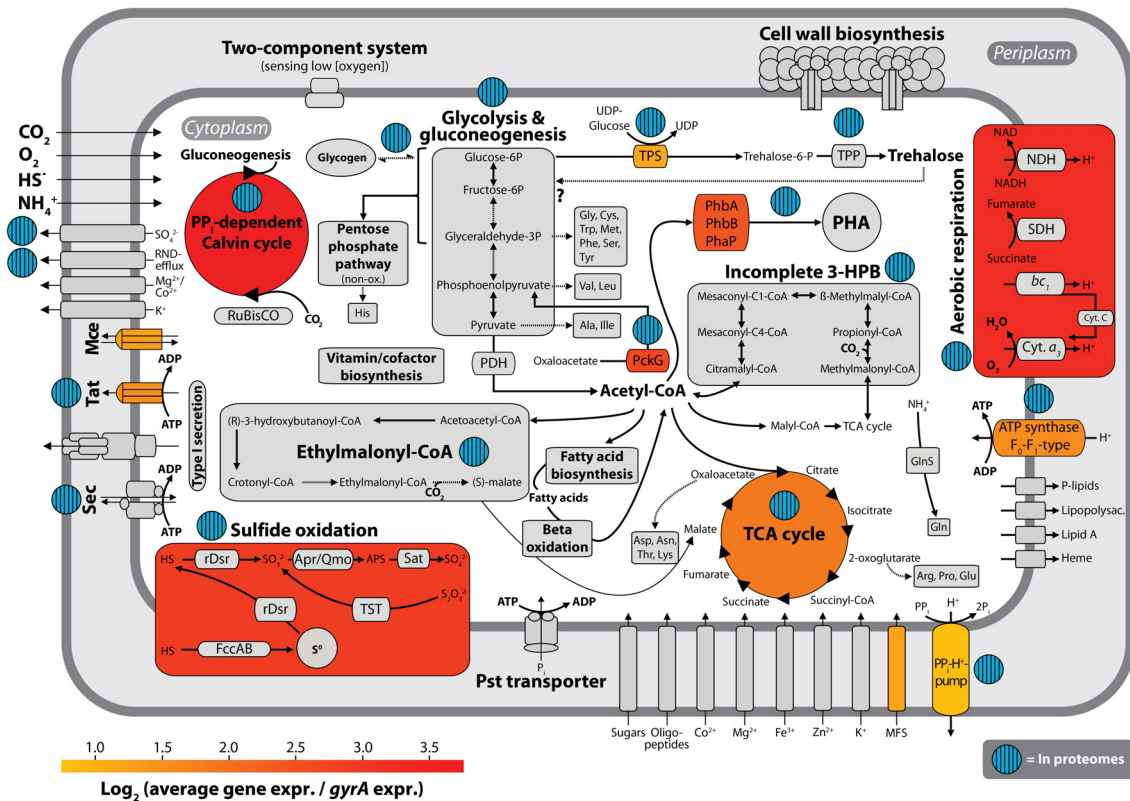
Figure 2. The chemoautotroph *Ca.* R. standrea expresses a versatile metabolism. Metabolic reconstruction of *Ca.* R. standrea, pathways or enzymes are colored by expression levels in the transcriptome (mean of three samples). Only the highest 10% of expressed genes were colored relative to DNA gyrase A (*gyr*A) mean expression. Expressed proteins detected in the proteome samples are indicated. Not all genes of certain pathways were transcribed, and not all transcribed genes were found in the proteomes (Dataset S2, S3). Dotted arrows correspond to indirect synthesis of metabolites. Abbreviations used: **3-HPB**: 3-hydroxypropionate bi-cycle pathway; **Apr**: sulfite reductase; *bc$_1$*: cytochrome C oxidase type *bc$_1$*; **Cyt. a$_3$**: cytochrome C oxidase type *a$_3$*; **Cyt. C**: cytochrome C; **FccAB**: flavocytochrome c sulfide dehydrogenase; **Mce**: mammalian cell entry proteins Mla; **NDH**: NADH-quinone oxidoreductase/dehydrogenase; **PckG**: Phosphoenolpyruvate carboxykinase; **PDH**: pyruvate dehydrogenase complex; **PHA**: polyhydroxyalkanoates; **Pst:** high-affinity phosphate transporter; **rDsr**: reversely operating dissimilatory sulfite reductase; **Sat**: ATP sulfurylase; **SDH**: succinate dehydrogenase; **Sec**: secretion pathway; **TCA**: tricarboxylic acid; **TST**: thiosulfate sulfur transferase (rhodanese); **TPS**: trehalose-P-synthase; **Tat**: twin-arginine translocation.

Genes of the rDSR sulfur oxidation and CBB carbon fixation pathways were highly expressed and underline the thioautotrophic role of *Ca.* R. standrea [sum of 8 CBB cycle proteins 4.88% ± 1.17 normalized spectral abundance factor (NSAF), sum of 14 rDSR pathway proteins 7.48% ± 0.95 NSAF, n = 8 proteomes] (Figure 2, Supplementary Figure 2, 3, Supplementary Note 2). CBB cycle genes including the RuBisCO subunits *cbbL* and *cbbS* were among the top 10% of most highly represented genes in transcriptomes as well as proteomes (Supplementary Figure 2, 3).

Microautoradiography corroborated the gene expression data, as signals for carbon fixation were limited to the symbiont-bearing trophosome region. Without addition of an external

electron donor supply, the symbionts fixed 9.48 ± 10.2 µmol C g^{-1} h^{-1} (mean $\pm$ standard deviation, max.: 33 µmol C g^{-1} h^{-1}) (Supplementary Figure 4–6, Dataset S4). These rates were in the range determined for other sulfur-oxidizing symbionts of e.g. *Ifremeria nautilei* (0.7 µmol C g^{-1} h^{-1}) and *Riftia pachyptila* (27 µmol C g^{-1} h^{-1}) (32). The calculated fixation rates would be sufficient to sustain a doubling of the length of a *Paracatenula* sp. standrea specimen in 10 to 34 days (Dataset S4). Such growth rates together with frequent asexual reproduction (33, 34) could explain their high abundances at the sampling site where they occurred in densities of 233–5400 individuals per m^3 of sediment (mean 1809 $\pm$ 1483, n = 40).

The CBB cycle encoded in *Ca.* R. standrea was missing fructose-1,6-bisphosphatase and sedoheptulose-1,7-bisphosphatase, which could be replaced by a reversible and PP$_i$-dependent 6-phosphofructokinase (35). *Ca*. R. standrea encoded two 6-phosphofructokinases (PFK), an ATP-dependent ATP-PFK and a PP$_i$-PFK. The PP$_i$-PFK had a ten-fold higher expression [mean transcripts per million (TPM) 5073 $\pm$ 780, n = 3] than the ATP-PFK (mean TPM 504 $\pm$ 128, n = 3) and was among the top 10% of highest expressed genes (Supplementary Figure 2, 3). Together with the PP$_i$-energized proton pump (PP$_i$-H$^+$-pump), which also was among the 10% most expressed genes (mean TPM 1686 $\pm$ 163, n = 3), the PP$_i$-dependent variant of the CBB cycle can save up to 31.5% of the ATP invested in carbon fixation (35). We tested for a possible lateral gene transfer of the PP$_i$-PFK and PP$_i$-H$^+$-pump genes to *Ca*. R. standrea as the PP$_i$-dependent CBB cycle has also been found in gammaproteobacterial chemosynthetic symbionts, including the symbionts of gutless oligochaetes, vestimentiferans tubeworms and the bivalve *Solemya* (35–37). In contrast to the *Gammaproteobacteria*, the PP$_i$-PFK and PP$_i$-H$^+$-pump genes in *Ca*. R. standrea did not form an operon (35). Both genes were of alphaproteobacterial origin and clustered with gene sequences of related *Rhodospirillaceae* (Supplementary Figure 7, 8, Supplementary Note 2). Other *Rhodospirillaceae* such as *Rhodospirillum rubrum* only lacked a sedoheptulose-1,7-bisphosphatase, but still encoded a canonical fructose-1,6-bisphosphatase for the last step in the CBB cycle that has the highest energy saving potential (35) (Dataset S1). This suggests that the most energy-efficient variant of PP$_i$-dependent CBB cycle evolved independently in *Ca*. R. standrea, possibly due to the same energy constraints that led to the selection of the most efficient sulfur oxidation pathway across the diversity of thiotrophic symbionts (38) (Supplementary Note 2).

The *Ca.* R. standrea symbiont encoded form IA RuBisCO for carbon fixation rather than the form II RuBisCO that was previously identified by PCR amplification in another *Ca.* Riegeria species, *Ca.* R. galateia (9) (Supplementary Figure 9). The presence of two different RuBisCO forms in the *Ca.* Riegeria clade that both clustered with sequences from *Alphaproteobacteria* suggests that the last common ancestor encoded both forms and that descendants differentially retained only one. Several members of the family *Rhodospirillaceae* are known to encode more than one RuBisCO form per genome, including forms IA and II (39). As form IA is adapted to higher oxygen and lower carbon dioxide levels compared to form II, the retention of form IA in *Ca*. R. standrea could indicate a preference for habitats with higher oxygen concentrations for *Paracatenula* sp. standrea compared to *P. galateia* (39)*.* Niche differentiation *via* different forms of RuBisCO has been observed in free-living thiotrophic bacteria, e.g. in strains of *Ca.* Thiomargarita nelsonii (40), but not within a single clade of chemoautotrophic symbionts (15, 36, 41, 42).

Characteristic of obligate autotrophs such as the vesicomyid clam symbionts is an incomplete TCA cycle that lacks the alpha-ketoglutarate dehydrogenase complex (13, 14). In contrast, *Ca.* R. standrea encoded a complete TCA cycle, which in combination with glycolysis allows it to use sugars and other organic substrates such as fatty acids as carbon and energy sources (Figure 2, Supplementary Figure 10). While these pathways might indicate a heterotrophic lifestyle, the results of our analysis of the transporters suggest that *Ca.* R. standrea uses these pathways for internal cycling of carbon stocks (Supplementary Note 2). We identified seven import transporters in the symbiont's genome with transmembrane domains and found only a very low number of importers of organic compounds (Supplementary Figure 11, Dataset S5). The symbionts can only take up selected peptides, but no sugars or fermentation end products that might be present in the sediments or originate from the hosts metabolism.

Still, TCA cycle genes were highly transcribed, and the corresponding proteins were abundant, constituting a total of 3% ± 0.49 NSAF (sum of 7 proteins, n = 8 proteomes) of all measured proteins in the proteomes (Figure 2, Supplementary Figure 2, 3). We hypothesize that carbon fixation and the TCA cycle do not operate simultaneously in a single cell but are expressed in separate symbiont populations to avoid futile cycling. To replenish intermediates of the TCA cycle and to synthesize and convert acetyl-CoA into biomass, *Ca.* R. standrea expresses the ethylmalonyl-CoA pathway and an incomplete 3-hydroxypropionate bi-cycle (3-HPB) pathway.

The ethylmalonyl-CoA pathway (sum of 5 proteins 2.25% ± 0.34 NSAF, n = 8 proteomes) is typical for *Alphaproteobacteria* (43–45) and has a similar function to the glyoxylate shunt that operates in many *Gammaproteobacteria*, but additionally allows the anaplerotic co-assimilation of carbon dioxide (43) (Figure 2). For three molecules of acetyl-CoA that enter the ethylmalonyl-CoA pathway, two molecules of CO_2 can be fixed at the expense of only one ATP (46). A similarly highly expressed but incomplete 3-HPB pathway (sum of 5 proteins 1.94% ± 0.42 NSAF, n = 8 proteomes) could allow the assimilation of organic compounds such as acetate, propionate, succinate and malate, even though several key reactions that would permit it to function autotrophically were not predicted in *Ca.* R. standrea (35, 47). The elements of the 3-HPB pathway that are encoded represent an anaplerotic pathway, similar to the ethylmalonyl-CoA pathway, and consume 1 ATP for the fixation of one molecule of CO_2 (35, 47). Compared to the 4 $^1/_9$ ATP that would be needed in the PP_i-dependent CBB cycle for the fixation of two molecules CO_2 (35), both the ethylmalonyl-CoA pathway and the 3-HPB pathway would allow for a cheaper maintenance and expansion of carbon stocks in *Ca.* R. standrea. The total proteomic investment in both anaplerotic pathways was as high as for primary carbon fixation *via* the PP_i-dependent CBB cycle (Supplementary Figure 3). This suggests, in combination with the highly expressed TCA cycle, that the two pathways contribute substantially to carbon assimilation in *Ca*. R. standrea. While this has been shown for other mixotrophs (46) or apparent heterotrophs like SAR11, one of the most successful clades of marine bacteria (48), it has not been observed in chemosynthetic symbioses. How much the possible anaplerotic fixation in *Ca*. R. standrea contributes to the total carbon fixation in the symbiosis remains to be shown. It is however tempting to speculate that without an essential benefit to the efficiency of the symbiont, one or both pathways would have been lost as they are not essential to replenish the intermediates of the TCA cycle that they synthetize.

The exact same gene set of the partial 3-HPB pathway described above has also been found in thiotrophic gammaproteobacterial symbionts of gutless oligochaetes (35). Our phylogenetic analyses showed that most of the *Ca*. R. standrea homologs of genes characteristic for the 3-HPB pathway clustered with genes from these two gammaproteobacterial symbiont groups, except for one that clustered with sequences from other *Rhodospirillaceae* (Supplementary Figure 12). For three out the five enzymes, the cluster of sequences from thiotrophic symbionts were the sister clade to homologs from *Rhodospirillaceae*, whereas two of them (Meh and Mct)

had no homologs in *Rhodospirillales*. This extended phylogenetic distribution for several of the enzymes corroborates an origin of these genes for enzymes of the 3-HPB pathway outside of the *Chloroflexi* and points to the *Alphaproteobacteria* as one possible group of origin (49).

Paracatenula sp. standrea shares the Sant´Andrea bay sediments with several groups of chemosynthetic meiofauna such as gutless oligochaetes, stilbonematinae nematodes and the ciliate *Kentrophoros* (6, 50, 51). The most abundant taxon is the gutless oligochaete *Olavius algarvensis* followed by *Paracatenula* sp. standrea. They represent the most contrasting forms of these chemosynthetic symbioses. *O. algarvensis* hosts a mixotrophic consortium of five bacterial symbionts (52) with little signs of genome reduction in any of the symbiont groups (50), while *Paracatenula* relies on a single symbiont with a drastically reduced genome. These contrasting metabolic capabilities likely reflect their different ecological niches. The *Olavius* symbiosis can use several energy sources in addition to sulfur oxidation, such as carbon monoxide, hydrogen and small organic compounds (35, 53), whereas the *Paracatenula-Ca.* Riegeria symbiosis is an energy-efficient specialist with sulfur oxidization as its main energy source.

Specialization and convergence in intracellular thiotrophic symbionts

To test for deviating evolutionary patterns in the alphaproteobacterial *Ca.* R. standrea symbionts we compared its genome to those of 62 bacteria representing free-living relatives of *Ca*. Riegeria, parasites, non-thiotrophic mutualistic symbionts as well as thiotrophic bacteria from other symbiotic and non-symbiotic clades (for a list of genomes see Dataset S6). We analyzed the cellular processes encoded in each their genomes in a non-metric multidimensional scaling (NMDS) ordination of the distribution of each their genes to clusters of orthologous genes (COGs) and to modules of the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Figure 3, Supplementary Note 3).

Apart from *Ca.* R. standrea, the thiotrophic symbiont genomes formed a spectrum that was dependent on their genome sizes, suggesting that COG and KEGG based profiles and genome contents in the other symbionts with small genomes were largely convergent. At the other end of this spectrum, the profiles of symbionts with non-reduced genomes overlapped with free-living thiotrophs such as *Allochromatium vinosum* and *Thiocapsa marina* (Dataset S7, S8).
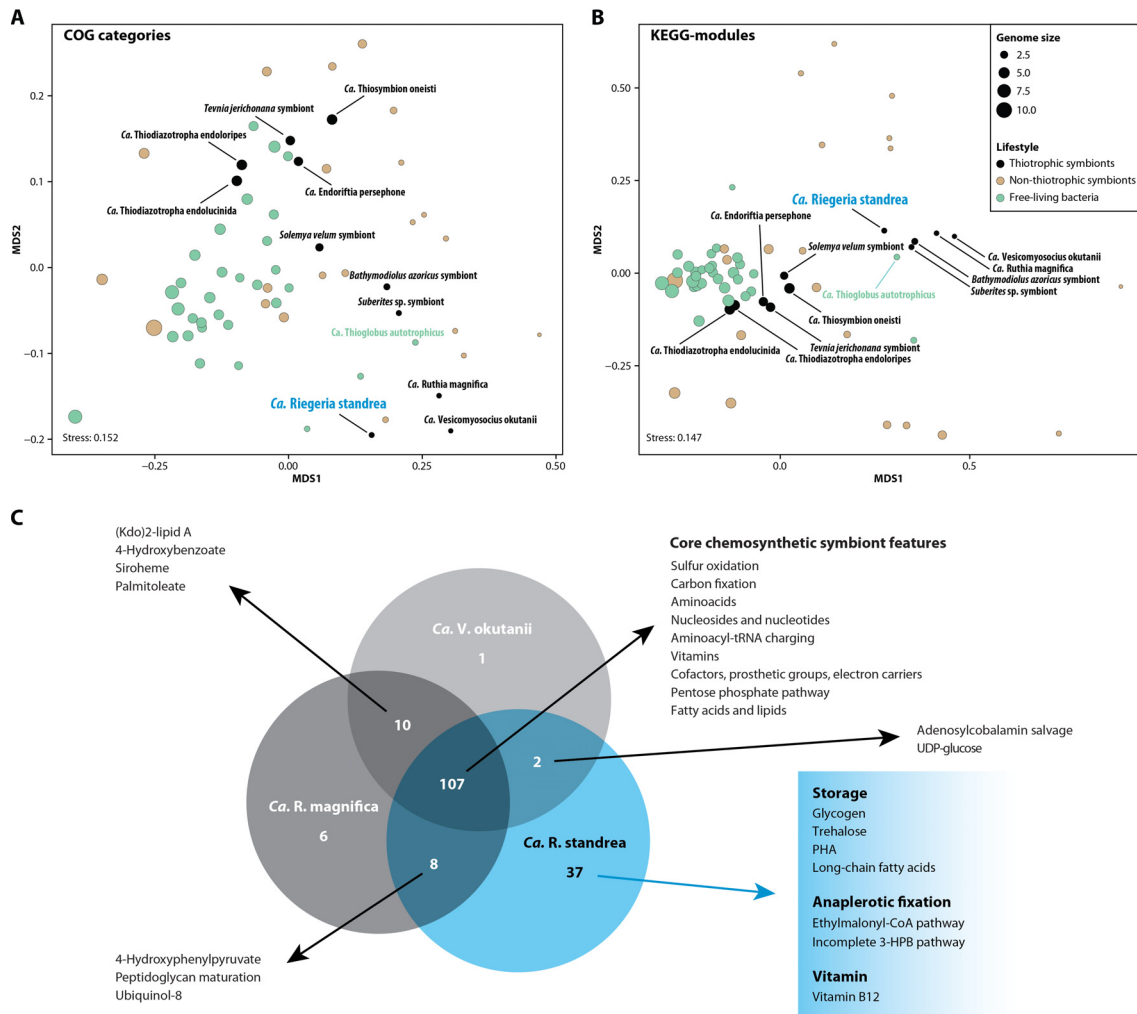
Figure 3. The unique functions in *Ca*. R. standrea allow efficient maintenance of large carbon and energy stocks. *A*, *B*, Functional diversity of thiotrophic symbionts and selected reference bacteria. The diameter of each circle represents the genome sizes. Free-living bacteria include *Alpha-* and *Gammaproteobacteria* and one Cyanobacterium. A list of genomes used and the raw plots can be found in Dataset S6, n = 63. *A*, NMDS plot of COG category distributions with a 2D stress of 0.152. *B*, NMDS plot of KEGG module distributions with a 2D stress of 0.147. *C*, Unique pathways in vertically transmitted thiotrophic symbionts with reduced genomes. Biosynthetic pathways specific for *Ca*. R. standrea or shared between two or more symbionts are highlighted. See details in Dataset S9A.

While *Ca*. R. standrea had the most similar distribution of encoded functions to vesicomyid symbionts, another group of vertically transmitted chemosynthetic symbionts, its profile stood out in the observed NMDS patterns across the categories of COGs and KEGG modules (Figure 3A, Supplementary Figure 13, Supplementary Note 3, Dataset S7, S8, for a list of genomes see Dataset S6). *Ca*. R. standrea deviated from the vesicomyid symbionts with an enrichment of genes coding for carbohydrate transport and metabolism (COG category G), lipid transport and metabolism (COG category I) und coenzyme transport and metabolism (COG category H) (Dataset S7). We could corroborate this pattern in a comparison on the level of pathways as *Ca.* R. standrea encoded 37 unique pathways that allow for flexible carbon fixation and the

versatile storage of carbohydrates (Figure 3B, Supplementary Note 3). In particular, the capabilities to synthesize long-chain fatty acids and PHA as well as trehalose and glycogen underline a possible dual role of the symbionts in the *Paracatenula* symbiosis, where the symbionts provide large storage capacities for the association (Figure 3B, Dataset S9). This is in contrast to all other chemosynthetic symbioses and could be coupled to the high proportion of the total biomass that the *Ca*. Riegeria cells achieve in the *Paracatenula* consortium (9).

While *Ca*. R. standrea has substantial set of unique pathways, it also shared 107 pathways with the thiotrophic symbionts with the most reduced genomes, *Ca.* R. magnifica and *Ca.* V. okutanii. Besides energy conservation through sulfur oxidation, these shared pathways were largely related to biosynthesis and include carbon fixation *via* the CBB cycle and pathways for amino acid, nucleoside, nucleotide, fatty acid and lipid synthesis (Supplementary Note 3, Dataset S9). These pathways represent the core functions of a nutritional chemosynthetic symbiont and reflect the strong selection based on the conserved metabolic needs of their invertebrate hosts. In such nutritional symbioses essential biosynthetic pathways must be retained because their loss would lead to a decrease in fitness or even the extinction of the affected host (13, 14, 17). In contrast, a loss of catabolic pathways might not severely impact the symbionts but might in return even benefit the consortium as it leaves resources accessible to the host.

Ca. R. standrea is a major carbon and energy buffer in the *Paracatenula* symbiosis

Our comparative analyses showed that *Ca.* R. standrea possessed and expressed a surprising number of pathways for a chemoautotrophic symbiont with a reduced genome, many connected to carbon metabolism and storage. The pathways included the synthesis of polyhydroxyalkanoates (PHA), trehalose and glycogen (Supplementary Figure 10, Dataset S9A). We observed a large number of intracellular inclusions in the symbionts in both light and electron microscopy that caused the orange-white appearance of the trophosome region and further highlighted the symbionts storage capabilities (Figure 4 A–C). By combining Raman spectroscopic imaging with sensitive mass spectrometry protocols, we were able to account for most of these storage molecules that make up large proportion of the total *Ca.* R. standrea volume (Supplementary Figure 14, 15).
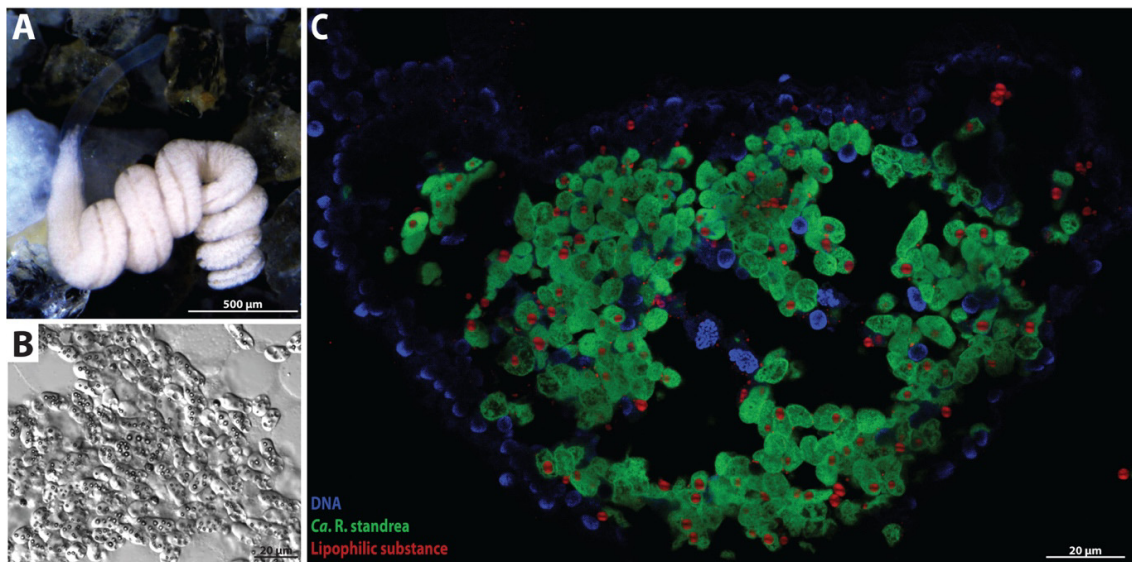
Figure 4. *Ca.* R. standrea are large endosymbiotic bacteria with massive storage inclusions. *A*, Habitus of *Paracatenula* sp. standrea. The white trophosome contains endosymbionts while the anterior and transparent part of the worm (rostrum) is bacteria-free. ***B***, Differential interference contrast image of *Ca.* R. standrea symbionts indicated multiple intracellular inclusions. ***C***, Confocal laser-scanning image of CARD-FISH combined with a lipophilic staining (Nile Red) on a transverse section of a *Paracatenula* specimen in the symbiont-bearing region. Overlay of DAPI signal (blue), Nile Red (red) and probe targeting the symbionts (green, EUB I–III).

Two conspicuous inclusions were detectable in differential interference contrast microscopy. One of them, identified using Raman spectroscopy, was elemental sulfur that appears as light-refractile granules of 1–2 µm in diameter (Figure 4B, Supplementary Figure 14). The second type of inclusions was larger, with sizes of up to 3.5 µm in diameter, apparent in most cells and occupied $13.1 \pm 9.4\%$ of the cell volumes (n = 40 symbiont cells) (Figure 4C). These were likely composed of PHA, as they were selectively stained by the lipophilic dye Nile Red, and both the polymer polyhydroxyvalerate and its monomer hydroxyvalerate could be identified by gas chromatography–mass spectrometry (GC-MS) (Supplementary Figure 15). Furthermore, genes for PHA synthesis and the phasin protein PhaP were highly transcribed, and PhaP was the most abundant protein in six out of eight proteomes (mean $6.67\% \pm 1.83$ NSAF, n = 8 proteomes) (Supplementary Figure 2, 3). Besides storing carbon, PHA might function as an electron sink under anoxic conditions (54–56). The host animal *Paracatenula* likely traverses the oxic and anoxic sediment layers on a regular basis to provide its symbionts with sulfide, a behavior shown for invertebrate meiofauna with thiotrophic symbionts (57, 58). The *Ca.* R. standrea symbiont had the necessary genes for aerobic respiration coupled with ATP synthesis but lacked terminal oxidases for utilization of nitrate or alternative inorganic electron acceptors (Figure 2, Supplementary Note 2). A possible electron sink for the oxidation

of sulfide to elemental sulfur when oxygen is absent could be the synthesis of PHA from acetyl-CoA, which would simultaneously function as both a store of energy and of carbon. The accumulation of PHA under anaerobic conditions has also been shown in other *Alphaproteobacteria* that live in changing redox conditions, such as the free-living *Ca.* Defluviicoccus tetraformis (59). A similar mechanism was shown for *Ca.* Accumulibacter phosphatis (*Betaproteobacteria*) and was also proposed for the unrelated gammaproteobacterial thiotrophic symbionts of *Olavius algarvensis* (54, 55, 60).

Electron dense granules resembling glycogen were identified in both host and symbiont tissue (Supplementary Figure 16). Glycogen is widely used in both bacteria and animals to store energy and carbon and is considered to be a long-term energy reserve under anaerobic conditions (35, 61–63). Proteins for glycogen metabolism were also detected in the symbiont proteomes, although their genes were not highly expressed (Figure 2, Supplementary Figure 2, 3). We detected proteins for both biosynthesis and degradation in the same samples, suggesting a dynamic switching between synthesis and utilization, possibly distributed in sub-populations of the symbionts (Dataset S3).

Unexpectedly, the disaccharide trehalose was by far the most abundant soluble metabolite measured in *Paracatenula* holobionts (Supplementary Figure 15, 17). We could link the trehalose pool to the symbionts as it was not detectable in host tissue without symbionts (anterior rostrum fragments; n = 3, Supplementary Figure 17A). The *Ca.* R. standrea symbiont uses the OtsAB pathway for trehalose synthesis from UDP-glucose and glucose-6-phosphate. The alpha, alpha-trehalose-phosphate synthase was among the 10% most highly expressed genes and was detectable in the proteome (Supplementary Figure 2, 3). Trehalose can be found in a wide spectrum of eukaryotes and bacteria and has multiple biological functions including carbon storage but also the protection from osmotic stress e.g. due to varying salinities (64). *Paracatenula* sp. standrea could experience lower salinities due to freshwater input into Sant´Andrea bay in periods of prolonged precipitation. Changes in salinity during salinity tolerance tests however had no effect on trehalose levels, indicating a storage rather than osmolyte function. Specimens survived 5 hour treatments in seawater of a salinity ranging from 25 to 45‰, but no significant differences in trehalose concentrations could be measured (salinities from 20 to 50‰, n = 5 , two-sided *t*-test, Supplementary Figure 17B).

Based on expression data, chemical measurements and visualizations, *Ca.* R. standrea has an unique ability to accumulate massive amounts of multiple energy and carbon storage compounds. Our chemical characterization and quantification showed that all of these carbon and energy stocks are available with large standing stocks, and that most – elemental sulfur, PHA and trehalose – can be solely attributed to the symbionts. Taken together, our observations imply that the symbiont performs the primary energy storage for the whole animal-microbe symbiosis.

Outer membrane vesicles secreted by the symbiont form the basis of host nutrition

Like many marine invertebrates associated with thiotrophic symbionts, *Paracatenula* lacks a mouth and a gut and depends on its symbionts for essential nutrients. This was reflected in the broad genomic repertoire and the expression of amino acid, vitamin and cofactor biosynthesis pathways in the symbiont (Supplementary Figures 18, 19, Supplementary Note 2).

Active export across the symbiont membrane is limited as we only identified 15 export-related gene products in *Ca.* R. standrea, eight of which had at least one transmembrane (TM) domain based on the identification using the Transporter Classification Database (TCDB) (Supplementary Figure 11, Dataset S5, S6, Supplementary Note 2, n = 63). The main identified transporters were the Tat and Sec secretion systems that could translocate folded and unfolded proteins to the periplasm. A Type I secretion system could export unfolded proteins across the outer membrane and in the hosting vacuoles but no secretion systems for the transfer of proteins into the host bacteriocyte could be detected. Although several transporters such as secretion systems have essential functions conserved across bacteria (65) and might be involved in the nutrition of the host, surprisingly only the Tat secretion system for the translocation of folded proteins was highly expressed (Figure 2, Dataset S2A).

Given that we could not find genomic evidence for substantial export *via* transporters, we were surprised that we also found no evidence for lysosomal digestion of symbiont cells in freshly collected *Paracatenula* based on transmission electron microscopy (TEM) (n = 3). Lysosomal digestion of symbiont cells is readily observed in bacteriocytes of other hosts, e.g. members of the deep-sea mussel genus *Bathymodiolus* as well as in vesicomyid clams (66, 67). In

freshly collected *Paracatenula* specimens host gene expression in such samples indicated lysosome formation and trafficking as well as proteolysis, all likely involved in the digestion of symbiont biomass (Dataset S10). Important enzymes in the lysosomal digestion of bacteria are lysozymes that cleave the bonds of peptidoglycan and lead to cell lysis. Surprisingly, the host did not express lysozymes, despite the expression of several types of digestive enzymes. We could e.g. detect the cysteine protease cathepsin B that has been shown to play a major role in the digestion of gut-bearing flatworms and is also expressed in the symbiotic gutless annelid *Olavius algarvensis* (68, 69) (Dataset S10). Since the symbiont has a peptidoglycan cell wall (Figure 2), the absence of lysozymes suggests that complete symbiont cells are not the main target for the lysosomal digestion expressed in freshly collected *Paracatenula* specimens. This corroborates the ultrastructural observations, but is in contrast to other animals with nutritional bacterial symbionts. Divergent hosts such as the *Bathymodiolus* deep-sea mussels but also insects, e.g. aphids, have been shown to constitutively express lysozymes that play a key role in symbiont digestion and the control of symbiont populations (66, 70).

Our data rather suggests that the main mode of symbiont to host transfer is likely based on the release of outer membrane vesicles (OMVs) and their digestion by the host. OMVs package proteins, lipids and nucleic acids, but despite increased attention into their role in symbiotic interactions they have not been shown to be broadly involved in host nutrition in animal microbe interactions (71). We observed several clusters of small OMVs in close proximity to symbiont cells (Figure 5A–D). Correspondingly, we identified a highly expressed intermembrane lipid transport system (ILTS, also known as mammalian cell entry proteins, *mlaED*) of the symbiont that has been shown to play an important role in the formation of OMVs (72). The two most highly expressed genes of the ILTS (*mlaE* and *mlaD*, mean TPM 2536 ± 682 and 2097 ± 531, respectively, n = 3) were among the top 10% of expressed genes and represent the subunits that form the transport pore in the inner membrane (73) (Figure 2). In contrast, the outer membrane protein of the ILTS, VacJ, had 60x lower expression (VacJ, mean TPM 39 ± 13, n = 3). VacJ maintains outer membrane asymmetry by removing phospholipids from the outer leaflet and supports trafficking of phospholipids from the outer membrane (74). The observed low expression of VacJ has been linked to increased OMV formation in VacJ deficient Gammaproteobacteria (72). Uptake and digestion of symbiont

OMVs *via* phagocytosis and the subsequent lysosomal digestion *via* a phagolysosome would explain the expression for lysosome formation observed in freshly collected samples.
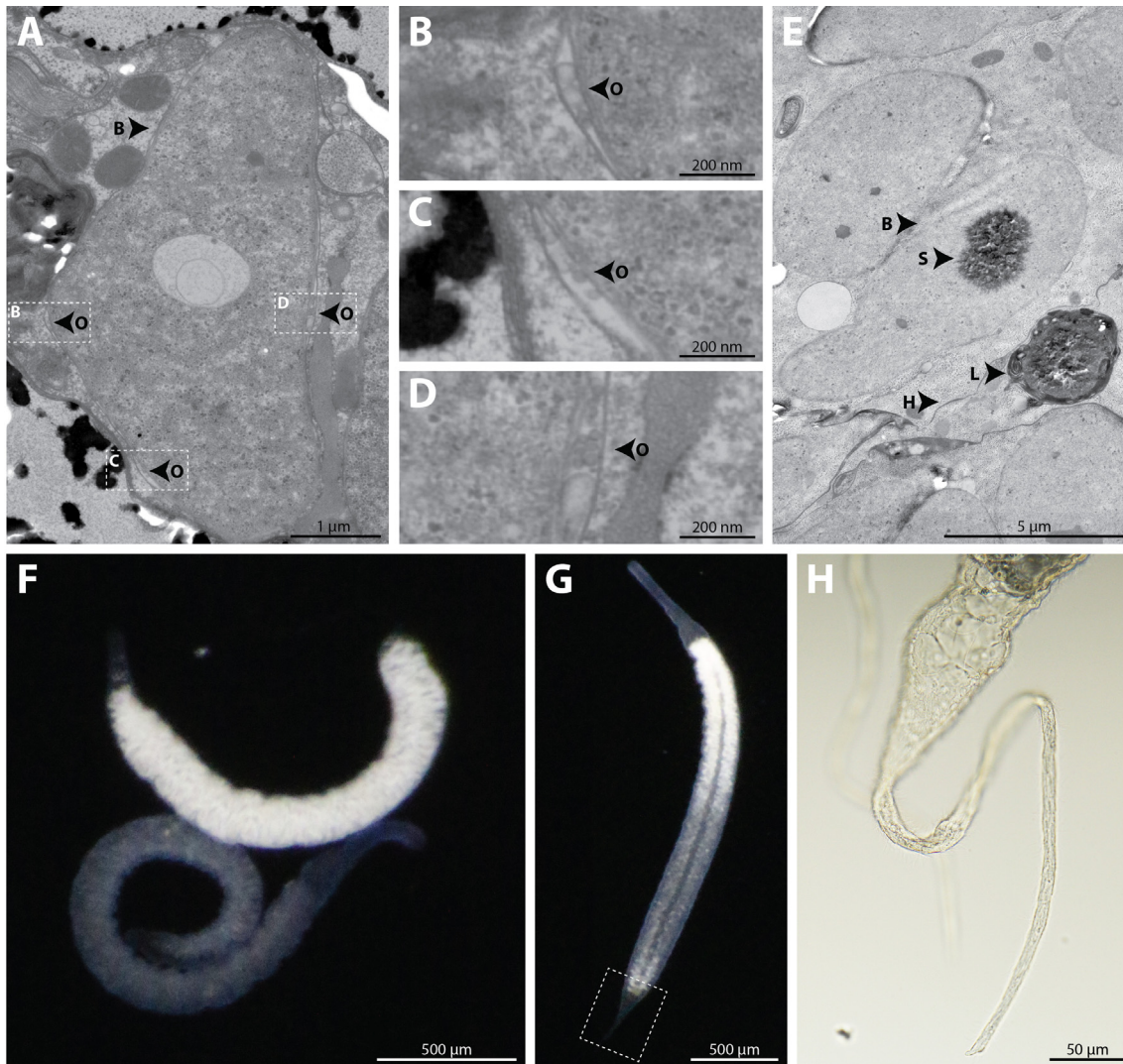

Figure 5. *Ca.* **R. standrea nourishes the host with outer membrane vesicles and is only digested after prolonged starvation. A−D** TEM images of *Paracatenula* sp. standrea. *A*, Overview of a intracellular bacterial symbionts *Ca.* R. standrea (B). *B−D* Represent magnified areas indicated in **A** highlighting the outer membrane vesicles (O). **A**. *E*, A symbiont cell (B), a large storage inclusion (S) and a lysosomal structure (L) in a bacteriocyte surrounded by a host membrane (H). The enclosed structure in the lysosome resembled a PHA granule (Figure 4C). *F*, Two *P.* sp. standrea of pale and white coloration. *G*, A single *P.* sp. standrea with changing coloration within its trophosome. The dotted square indicates a tail-like structure. *H*, Microscopic image of a tail-like structure. No symbionts can be detected in this area.

In addition to the role of OMVs in the nutrition of the *Paracatenula* host, we could also observe digestion of complete symbiont cells, but only in specimens from long-term husbandry experiments (Figure 5E). These experiments often led to a decoloration of the trophosome regions as well as to a formation of an epidermal tail-like structure (Figure 5F–H). Both observations suggest starvation of the host that led to consumption of storage material and symbiont cells, thereby reducing the volume of the trophosome. We detected lysosomal

structures with diameters of ca. 3 µm in such starved specimens after 7 months of husbandry (n = 2) (Figure 5E). We also identified electron-dense structures that resemble the PHA granules of intact cells within lysosomes surrounded by membrane stacks, but no symbionts in ongoing digestion. The observed PHA granules also corresponded to the extracellular position of some of them in the Nile Red-staining (Figure 4C).

Based on (i) the missing transporter-based export of substrates by the symbionts, (ii) the observation of OMV formation and corresponding symbiont expression profiles, (iii) the detection of host expression of lysosomal digestion but no expression of lysozymes and (iv) the absence of digestion of full symbiont cells we propose that digestion of symbiont material in the form of OMVs is the main mode of nutrition in the *Paracatenula* symbiosis. The symbiont builds up and efficiently maintains the primary energy storage for the whole animal-microbe consortium. The host then directly accesses all these stockpiles together with vitamins, amino acids and all other nutritive constituents *via* digestion of symbiont derived OMVs. Only in cases of higher demands or starvation, a subpopulation of its symbionts is digested, in contrast to all other chemoautotrophic symbioses where symbiont cell digestion drives biomass transfer.

Conclusion

After 500 million years of strict vertical transmission and reductive genome evolution, *Ca*. R. standrea has a drastically reduced genome compared to its free-living relatives. It has retained a versatile carbon metabolism and carbohydrate storage, in stark contrast to the similarly reduced vesicomyid clam symbionts that appear to be optimized only for autotrophic biomass production. The versatile function of *Ca.* R. standrea might be connected to its ancestry in *Rhodospirillaceae*, with many of which they share a complete intermediary carbon metabolism as well as the ethylmalonyl-CoA pathway that is common in *Alphaproteobacteria* (43). In addition, *Ca*. R. standrea has several unexpected features, even for a versatile alphaproteobacterial background that are shared with gammaproteobacterial symbionts – an incomplete 3-HPB pathway and a PP_i-dependent CBB cycle. In *Ca*. R. standrea, the majority of the genes involved in these pathways are of alphaproteobacterial origin and not acquired *via* lateral gene transfer. Instead of showing an acquisition as proposed for the gammaproteobacterial symbionts, our analyses extended the phylogenetic distribution of these

energy-efficient and versatile pathways and shed new light on their origins and possible histories of transfer from the metabolically flexible *Alphaproteobacteria*.

The key elements in the physiology of *Ca*. R. standrea such as enhanced modes of carbon fixation, intermediary carbon metabolism and sulfur oxidation are optimized for high energy efficiency. The symbionts can create and maintain versatile carbon pools as they represent the bulk storage organ for the consortium and support its host *via* OMV secretion. This unique integration of storage and transfer into a single and streamlined chemoautotrophic organism with a reduced genome represents a new type of bacteria-animal interaction. *Ca*. R. standrea extends the spectrum of nutritional symbioses that range from single vitamin or amino acid supplementation to full nutrition (17–23), and adds the function of primary energy and carbon storage that in all other animal hosts is performed by the animal itself.

Materials and Methods

Sample collection

Paracatenula sp. standrea specimens were collected between 2013 and 2017 from sediments in the bay off Sant´Andrea, Elba, Italy. Specimens were sampled with 10 liter buckets and extracted by decanting the sediment collected by divers at 6–7 m water depth. Individual specimens were picked manually using glass pipettes and stored in 5.9 ml glass vials filled with seawater (38‰) and natural sediment that was rinsed twice with fresh- and seawater to remove other meiofauna. The abundance of *Paracatenula* per m³ of Sant´Andrea sediment was estimated from specimen counts in 10 L buckets (n = 40) collected during two fieldtrips in 2014 and 2016. *Paracatenula* specimens for nucleic acid extractions and proteomics were fixed in RNAlater (Ambion) and stored at 4 °C. Specimens for metabolomics were fixed in methanol and stored at -20 °C.

Paracatenula worms for transmission electron microscopy (TEM) were kept for up to seven months in 12 ml glass vials filled with washed natural sediment and filtered seawater supplemented with trace elements. The vials had an air-filled headspace, and up to five pellets (Ø 0.5 cm) of seagrass leaves were placed at the bottom of the glass vials to induce a redox gradient due to microbial sulfate reduction around the rotting seagrass. Specimens were visually checked and transferred into freshly prepared glass vials every one to two months.

Light microscopy

Images of live *Paracatenula* were recorded in incident light using a Nikon SMZ25 stereomicroscope (Nikon). To image the bacteria of live organisms, specimens were first transferred onto glass slides and squeezed to bursting by applying pressure with cover slips. Single bacteria were then visualized using transmitted light with differential interference contrast (DIC) on a Zeiss LSM 780 confocal laser microscope (Carl Zeiss AG).

Electron microscopy

Specimens freshly collected from the environment and those kept in glass vials for seven months were fixed overnight with 2.5% (w/v) glutaraldehyde buffered with 1.5x PHEM and 9% w/v sucrose as described by Montanaro and colleagues (75). Samples were then either post-fixed with 1% w/v osmium tetroxide and directly embedded in resin or further processed using a Leica EM ICE high-pressure freezer (Leica Microsystems). Samples subjected to high-pressure freezing were freeze-substituted in acetone with 1% w/v osmium tetroxide and warmed to room temperature for 120 min as described by McDonald and Webb (76). All samples were embedded in Low Viscosity Resin (Agar Scientific) using centrifugation embedding (77). After polymerization for 12 h at 65 °C, 70 nm sections were cut on an ultramicrotome (Leica UC7, Leica Microsystems), mounted on formvar-coated slot grids and contrasted with 0.5% w/v aqueous uranyl acetate (Science Services) for 20 min and with 2% w/v Reynold's lead citrate for 6 min. Sections were imaged at 30 kV with a Quanta FEG 250 scanning electron microscope (FEI Company) equipped with a STEM detector using the xT microscope control software v6.2.6.3123.

CARD-FISH and staining procedures

CARD-FISH was performed on paraffin sections of *Paracatenula* sp. standrea as previously described (78) with minor modifications of the protocol. Specimens were fixed overnight at 4 °C in 2% paraformaldehyde (PFA) (w/v) and dehydrated in an increasing series of ethanol [50%, 80%, 90%, 100% (v/v)]. Dehydrated samples were then infiltrated with Roti-Histol (Carl Roth GmbH) (40 min, then overnight), a 1:1 (v/v) mixture of Roti-Histol/paraffin (50 min, 60 °C) and finally paraffin (3 x 30 min, overnight, 3 x 2 h, 60 °C) and solidified at room temperature for up to one week. 4 μm sections were cut on a microtome (Leica RM2165, Leica

Microsystems) and baked on glass slides for 2 h at 57 °C. *In situ* hybridization was performed for 3 h at 46 °C with the general bacteria probe EUB I–III (79, 80) (8.42 pmol µl^{-1}) mixed in hybridization buffer (1:150 v/v) at a formamide concentration of 30% (v/v). Slides were washed in washing buffer (15 min) at 48 °C followed by washing in 1x PBS (20 min). Sections were incubated in amplification buffer with tyramide fluorochromes (Alexa Fluor 488 tyramide, Life Technologies) at 46 °C for 30 min. Slides were washed in 1x PBS (20 min) and in deionized water. Sections were counterstained with DAPI (1 µg ml^{-1}) and Nile Red (2.5 µg ml^{-1}) for 10 min each, rinsed in deionized water and dried and mounted in a mixture of Vector Shield/Citifluor mounting solution (1:5.5 v/v). Samples were imaged with a Zeiss LSM 780 using a 40x/1.3 oil immersion objective (Carl Zeiss AG). The following excitation and emission spectra were used: tyramide Alexa488, excitation: 488 nm, emission: 490 to 525 nm; DAPI, excitation: 405 nm, emission: 419 to 525 nm; Nile Red, excitation: 561 nm (CY3), emission: 570–623 nm. Histograms were adjusted using the Zen software (2012, Carl Zeiss AG). The relative volume of Nile Red-stained inclusions to the symbionts cell volume was estimated using ImageJ v1.47.

DNA extraction and genome sequencing

DNA was extracted using the DNeasy Blood and Tissue Micro Kit (Qiagen) following the manufacturer´s instructions with two modifications: to increase the overall DNA yield, the proteinase K digestion was performed for four days and the DNA was eluted twice with a total volume of 40 µl each. The Ovation Ultralow Library System V2 (NuGEN) was used for paired-end library preparations following the manufacturer's protocol. 6 million 250 bp paired-end reads were generated with an average insert size of 508 ± 210 bp. Library preparation and sequencing on the Illumina MiSeq platform was performed at the Max Planck Genome Centre in Cologne, Germany (http://mpgc.mpipz.mpg.de/home/).

Metagenome assembly and binning

Prior to assembly, adapters and low-quality reads were removed with bbduk (from the software package BBmap v38.06 (http://sourceforge.net/projects/bbmap/) using a minimum quality value of two and a minimum length of 36. Single reads were excluded from further analyses. To remove the host genomic reads, reads were sorted based on a kmer frequency analysis

using bbnorm (BBmap). Only reads with average kmer frequencies of more than 20 at a kmer length of 31 were kept using bbnorm (BBmap). The assembly was performed with SPADes v3.8 with kmers 21, 33, 55, 77, 99 and 127 (81). Initial binning was done by collecting all contigs linked to the contig that contained the symbiont 16S rRNA sequence using the FASTG files of the SPAdes assembly in Bandage v0.81 (82). Reads were mapped onto the binned contig set and reassembled with SPADes in an iterative approach that was stopped when no more reads could be recruited. Average nucleotide coverage of the final *Ca*. R. standrea assembly was 58x.

Metabolic reconstructions and comparison

The *Ca*. R. standrea genome assembly was annotated by RAST v2 (83) along with the genomes used for comparison (Supplementary Table 1). Annotations of specific genes discussed in detail were verified manually using NCBI PSI-BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (84) against the NCBI nr database. Protein sequences were assigned to Clusters of Orthologous Genes (psiCOGs) and KEGG Orthologs (KOs) with eggNOG (85). Metabolic features were predicted from KOs with the KEGG Mapper Reconstruct Module tool (http://www.genome.jp/kegg/tool/map_module.html) and from RAST annotations with the PathoLogic module of Pathway Tools v19.5 (86). Transporters were annotated with Transporter Classification (TC) numbers by BLASTp v2.2.29 against the Transporter Classification Database (TCDB) (87) (accessed 17 Nov 2017), taking the best-scoring hit with E-values $<10^{-5}$, >30% amino acid identity and >70% coverage of the reference sequences (parameters from Yelton et al. (88)). Hits to TC families or subfamilies that had been manually designated as uptake or export transporters for organic substrates were analyzed.

To compare COG category composition between genomes, the number of genes per COG category was counted. To account for varying genome sizes, compositions were expressed as percentages of COG assignments per genome. To compare KEGG Module composition, the predicted completeness per module per genome was encoded into a binary matrix (complete or missing up to 2 genes: 1, incomplete or absent: 0). The NMDS analyses were performed using the metaMDS function from the R package 'vegan' (http://cran.r-project.org/package=vegan) (Bray-Curtis distances) and visualized with ggplot2. Non-redundant pathways predicted by Pathway Tools were amended for polyhydroxyalkanoate

(PHA) synthesis and the incomplete 3-hydroxypropionate bi-cycle pathway (3-HPB). The presence of pathways was compared between *Ca.* R. standrea, *Ca.* R. magnifica and *Ca.* V. okutanii and represented in a Venn diagram.

RNA extraction and transcriptome analyses

For symbiont transcriptomics, RNA was extracted from single *Paracatenula* sp. standrea specimens that had been fixed in RNAlater directly after collection in the bay of Sant´Andrea, Elba. RNA was extracted with the RNeasy plus micro kit (Qiagen) following the manufacturer´s protocol, and eluted in 11 µl RNase-free water. cDNA was synthesized from total RNA with the Ovation RNASeq System v2 (NuGEN) following the manufacturer's protocol, sheared to 350 bp target size with the Covaris microTUBE system (Covaris) and cleaned up with the Zymo Genomic DNA Clean & Concentrator Kit (Zymo Research). Sequencing libraries were prepared from cDNA with the NEBNext Ultra DNA library preparation kit (New England Biolabs) for Illumina and sequenced on the Illumina HiSeq 2500 platform using 2 x 100 bp paired-end reads. Both library preparation and sequencing were performed at the Max Planck Genome Centre in Cologne. 6 to 7 million single-end 100 bp reads per library were sequenced and transcript counts were generated from raw reads using kallisto v0.44 (89). The expression values [transcripts per million (TPM)] were averaged for the three replicates and normalized to the expression of the housekeeping gene DNA gyrase A (*gyrA*) that has a high expression stability (90).

For host transcriptomics, *Paracatenula* specimens of different physiological states were sequenced in triplicates on the Illumina HiSeq 2500 platform using 2 x 100 bp paired-end reads to a depth of 11–15 million reads. Sequencing adapters were removed and quality trimming was performed as described for the metagenome. Host mRNA reads were filtered out by removing rRNA reads with BBSplit (from the software package BBmap) using the reference databases provided by SortMeRNA v2.1 (91) and the *Ca.* R. standrea genome. A *de novo* transcriptome assembly for the pooled reads of all 12 *Paracatenula* sp. standrea libraries was generated with Trinity v2.5.1 (92). Final assembly cleaning was performed with BLAST searches against the NCBI nr database on GenBank (93) to filter out contaminant sequences using MEGAN v6 (94). To remove duplicated transcripts the assembly was clustered to 97% identity using the centroid-based algorithm of VSEARCH v1.1.3 (95). Assembly quality and

completeness was analyzed with the utility scripts provided by the Trinity package (96) and BUSCO v3 (97). The final host transcriptome consisted of 54,926 predicted genes represented by 63,868 assembled transcripts and contained 79.1% complete or fragmented metazoan BUSCOs. Functional annotation of the assembly was conducted with Trinotate v3.0.1 following the developer´s guidelines (http://trinotate.github.io/).

Protein extraction and proteomic analyses

For proteomics, tryptic digests from eight samples (4x one animal, 4x three animals pooled) were prepared following the filter-aided sample preparation (FASP) protocol (98). In brief, 35 µl of SDT-lysis buffer (4% (w/v) SDS, 100 mM Tris-HCl pH 7.6, 0.1 M DTT) was added to each sample. Samples were heated for lysis to 95 °C for 10 min and the entire lysate (without pelleting) was mixed with 200 µl of UA solution (8 M urea in 0.1 M Tris/HCl pH 8.5) in a 10 kDa MWCO 500 µl centrifugal filter unit (VWR International) and centrifuged at 14,000 x g for 40 min. 200 µl of UA solution was added again and centrifugally filtered at 14,000 x g for 40 min. 100 µl of IAA solution (0.05 M iodoacetamide in UA solution) was added to the filter and incubated at 22 °C for 20 min. The IAA solution was removed by centrifugation and the filter was washed 3x by adding 100 µl of UA solution and then centrifuging. The buffer on the filter was then changed to ABC (50 mM ammonium bicarbonate), by washing the filter 3x with 100 µl of ABC. 2 µg of mass spectrometry grade trypsin (Thermo Scientific Pierce) in 40 µl of ABC was added to the filter and filters were incubated overnight in a wet chamber at 37 °C. The next day, peptides were eluted by centrifugation at 14,000 x g for 20 min, followed by addition of 50 µl of 0.5 M NaCl and another centrifugation. Peptides were not desalted. Approximate peptide concentrations were determined using the Pierce Micro BCA assay (Thermo Scientific Pierce) following the manufacturer's instructions. Peptide concentrations in all samples were below the detection limit of the assay.

Samples were analyzed by 1D-LC-MS/MS. Two wash runs and one blank run were performed between samples to reduce carry over. For each run, 20 µl of peptide solution was loaded onto a 5 mm, 300 µm ID C18 Acclaim PepMap100 pre-column (Thermo Fisher Scientific) using an UltiMate 3000 RSLCnano Liquid Chromatograph (Thermo Fisher Scientific) and desalted on the pre-column. After desalting the peptides, the pre-column was switched in line with a 50 cm x 75 µm analytical EASY-Spray column packed with PepMap RSLC C18, 2 µm material (Thermo

Fisher Scientific) heated to 45 °C. 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 a 260 min gradient as described by Kleiner and colleagues (99) and mass spectra were acquired in the Orbitrap as described by Petersen and colleagues (15). Roughly 110,000 MS/MS spectra were acquired per sample.

For protein identification a database was created using all protein sequences predicted from the *Ca.* R. standrea genome. The cRAP protein sequence database (http://www.thegpm.org/crap/) containing protein sequences of common laboratory contaminants was appended to the database. The final database contained 1460 protein sequences. For protein identification MS/MS spectra were searched against the database using the Sequest HT node in Proteome Discoverer v2.0.0.802 (Thermo Fisher Scientific) as described previously (15). Only proteins identified with medium or high confidence were retained resulting in an overall false discovery rate of < 5%. False discovery rate was controlled using the FidoCT node in Proteome Discoverer. A total of 407 different symbiont proteins were identified.

Phylogenetic analyses

Distance analysis

16S rRNA and *dsrB* sequence identities between the symbionts of two *Paracatenula* species, *Ca.* R. standrea and *Ca.* R. galateia, were determined from Muscle alignments (100). Only overlapping regions were taken into account. The average amino acid identities (AAI) of *Ca.* R. standrea and free-living relatives were calculated with the AAI calculator (http://enve-omics.ce.gatech.edu/aai/).

Conserved marker genes

Publicly available genomes of representatives from different alphaproteobacterial groups (*Rhodospirillales*, *Sphingomonadales*, *Rhodobacteraceae*, *Rhizobiales* and *Rickettsiales*) were downloaded from the NCBI database. 43 conserved marker genes were identified and aligned using the CheckM pipeline v1.0.5 (101). The tree was estimated from the amino acid alignment using maximum likelihood with FastTree v2.1 (102) using the JTT substitution model with 20

rate categories of sites and SH-like support values. The tree was visualized with iTOL v3.5.4 (103).

Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO)

The *Ca.* R. standrea CbbL sequence was used to query the NCBI nr database using BLASTx and the top 100 hits were included in a phylogenetic analysis (84). Additionally, RuBisCO sequences were retrieved from available sequences deposited in the NCBI database based on a RuBisCO matrix analyzed previously (9). OrfPredictor was used to identify protein-coding regions in the nucleic acid sequences and for the translation of the open reading frames in the short *Ca.* R. galateia sequence (104) (http://proteomics.ysu.edu/tools/OrfPredictor.html). Amino acid sequences were aligned using MAFFT with the iterative refinement method G-INS-I (105) and manually inspected in Geneious v11 (106) (http://www.geneious.com). The maximum-likelihood RuBisCO tree was estimated with PhyML (107) using the LG+I+G4 substitution model (108) with 100 bootstrap replicates. The tree was visualized in Geneious and rooted with CbbL.

PP$_i$-dependent phosphofructokinase (PP$_i$-PFK)

The translated PP$_i$-PFK sequence of *Ca.* R. standrea was extracted from the genome and used to query the Swissprot database (20 best hits based on % identity extracted for the PP$_i$-PFK) and the Uniprot database (100 best hits based on % identity extracted) by BLASTp. Hits for ATP-PFK from this BLAST search were removed from the gene database. To cover the translated ATP-PFK sequence of *Ca.* R. standrea, the sequence was used to query the Uniprot database by BLASTp, and the 50 best hits based on % identity were added to the gene database. Sequences without annotations were removed and the gene database was streamlined for analysis, retaining only one representative sequence per genus and gene type. In addition, selected PP$_i$-PFK sequences of thiotrophic symbionts were added from the NCBI database and the superfamily database. The sequences (n = 51) were aligned using MAFFT with the iterative refinement method G-INS-I. The PFK tree was estimated using IQ-TREE v1.5.6 and the LG+G4 evolutionary model with aLRT support values and nonparametric bootstraps (109–111). Tree visualization was performed with iTOL. For the protein motif analysis the ATP-PFK sequence of *E. coli* (accession nr. WP_000591795) was included in the MAFFT G-INS-I alignment and the search for GGDG/D and PG/KTIDXD motifs performed in Geneious.

PP$_i$-energized proton pump (PP$_i$-H$^+$-pump)

The translated PP$_i$-H$^+$-pump sequence of *Ca*. R. standrea was extracted from the genome and used to query the Uniprot database by BLASTp. From the best hits based on % identity, only one representative sequence per genus was added to the gene database. In addition, selected PP$_i$-H$^+$-pump sequences of thiotrophic symbionts were added from the NCBI database. The sequences (n = 32) were aligned using MAFFT with the iterative refinement method G-INS-I. The tree was estimated using IQ-TREE v1.6.7 and the LG+F+I+G4 evolutionary model with aLRT support values and nonparametric bootstraps. Tree visualization was performed with iTOL.

Incomplete 3-hydroxypropionate bi-cycle (3-HPB) pathway

Five key proteins of the 3-HPB pathway (Mch, Mcl, Smt, Mct, Meh) were used for single phylogenetic reconstructions. We used the homologues to proteins of the 3-HPB pathway in *Chloroflexus aurantiacus* as previously described (112). The amino acid sequences were obtained from the UniRef50 clusters containing the *C. aurantiacus* sequences in the UniProt database. The sequences were aligned with the *Ca*. R. standrea homologues using Muscle. The maximum-likelihood trees were estimated with FastTree v2.1.7 applying the same settings as for the conserved marker genes. Tree visualizations were performed with iTOL.

Physiological experiments

All physiology experiments were performed with biological replicates, the numbers are indicated in each experiment. Individual worms served as technical replicates for the carbon fixation experiments. To normalize measurements by the size of each *Paracatenula* specimen, digital images of each specimen were recorded with a Canon EOS 700D camera (Canon) mounted on a Nikon SMZ-745T dissecting microscope (Nikon) and the areas (in mm^2) of each specimen were determined using ImageJ v1.47.

Visualization and quantification of carbon fixation

Carbon fixation was studied with *Paracatenula* specimens using two experimental approaches. The first approach was conducted in two independent experiments. Specimens (2 and 3 individuals per replicate, 3 replicates each) were incubated in 3.9 ml glass vials filled with artificial seawater (ASW) (113) containing ^{13}C (final concentration 2.3 mM) and ^{14}C-labelled bicarbonate (0.02 mM, 160 KBq per vial) with glass beads as sterile sediment replacement

[2:1 ratio of 0.75–1 mm (Roth) and 0.4–0.6 mm beads (Sartorius)]. In the second approach (5 individuals per replicate, 3 replicates), ^{14}C-labelled bicarbonate (0.02 mM, 166 KBq per vial) was used again, but this time in natural Mediterranean seawater and with Sant´Andrea sediment that had been washed three times in freshwater and then was stored in Mediterranean seawater. This second experiment mimicking natural conditions was conducted because the carbon fixation rates were surprisingly low using the artificial seawater and glass beads setup. In the first experiment (ASW + glass beads) one time point was stopped after 6 h by washing in non-labeled seawater and transferring the specimens into 2% PFA (w/v) for 24 h followed by washing in 1x PBS and storage in 1x PBS/EtOH (1:1) (v/v) at 4 °C. In the second incubation experiment (natural seawater + sediment), samples were incubated (pulse) for up to 6 h (5 time points) or transferred to non-labeled seawater after the pulse and kept for 5 d to chase radioactive signals. Incubations were stopped by washing in non-labeled seawater followed by fixation in 2.5% (w/v) glutaraldehyde in 1.5x PHEM with 9% sucrose (w/v). Fixation was done for up to 14 h followed by 3x washing in 4.5x PHEM buffer and samples were stored in 1.5x PHEM with 9% sucrose (w/v) added. Non-incubated specimens as well as dead controls that were fixed for 2 min in 2% PFA and incubated for 6 h under same conditions served as negative controls.

Individual specimen replicates were cut transversely into a posterior half of only the symbiont-containing trophosome region and an anterior half that included both the bacteria-free rostrum and trophosome tissue. Bulk ^{14}C incorporation of the posterior half of each specimen was measured in Ultima Gold scintillation fluid (PerkinElmer) using a Tri-Carb 2900Tr (PerkinElmer) liquid scintillation counter and calculated as decays per minute (DPM) per mm² specimen.

For microautoradiography (MAR), anterior halves of specimens were dehydrated in a graded series of ethanol that was then replaced by mixtures of ethanol and LR-White (London Resin Company) [2:1; 1:1; 1:2 (v/v)] with modifications to the procedure explained elsewhere (16). The mixture was replaced with pure LR-White resin (3x, 30 min each) and incubated overnight at room temperature. Specimens were transferred into gelatin capsules filled with LR-White and polymerized at 50 °C for 5 d. 1 µm sections were cut on an ultramicrotome (Leica EM UC7, Leica Microsystems) and placed on Superfrost Plus glass slides (Thermo Fisher Scientific). Microautoradiography was applied using the ILFORD nuclear emulsion K5 (Ilford) following the manufacturer´s instructions with modifications. The photoemulsion was melted at 40 °C and

diluted with pre-warmed deionized water (1:1 v/v, total volume 7 ml) in a slide holder. Slides were dipped in photoemulsion for 3 sec, held vertically for 10 sec, followed by drying on aluminum ice blocks for 7 min. Slides were stored at 4 °C for up to 10 days (1, 2, 3, 6, 7, 10 days) in a light-tight box containing silica beads. Exposure was stopped by transferring the slides into mixtures of Ilford phenisol developer (1:4 v/v, 4 min), followed by Ilford Ilfostop pro (1:19 v/v, 1 min) and Ilford hypam (1:4 v/v, 4 min) in deionized water. Slides were washed with deionized water for 4 min, post-fixed using 2.5% glutaraldehyde for 1 h at room temperature followed by washing in 0.2% HCl (v/v) for 1 min and rinsing in deionized water. Sections were stained with toluidine blue and washed in deionized water at 37 °C (25 min). Slides were dried and mounted with Permount (Fisher Scientific) for light microscopy. The overview image (Supplementary Figure 4A) was stitched from eight tiles and blended with Adobe Photoshop CS5 allowing only rotation and translation of the individual tiles.

Impact of salinity on trehalose concentrations

Specimens were incubated in 5.9 ml glass vials filled with glass beads as sterile sediment replacement and filtered natural seawater of three different salinities: 25‰ (diluted with deionized water), 45‰ (supplemented with NaCl) and two variants of 38‰ – Mediterranean seawater and seawater first increased to 50‰ salinity using NaCl and then diluted with deionized water to 38‰. Five replicates of single specimens were incubated at 19 °C in darkness. Incubation was stopped after five hours by transferring specimens into 250 µl pre-cooled methanol and stored at -20 °C until the trehalose measurement. For a separate batch of freshly collected specimens, the trophosome and rostrum regions were separated prior to trehalose measurements.

Metabolite measurements

Metabolites were extracted from single (trehalose quantification) and from 25 (bulk metabolome analysis) *Paracatenula* specimens for chemical analysis as previously described by Liebeke and Bundy (114) with a few modifications. Additionally, three *Paracatenula* specimens were sectioned to separately analyze equal sized fragments of rostrum (symbiont-free, anterior) and trophosome (symbiont-hosting, posterior). Samples in tubes were transferred to bead-beater vials [200 µl of 1.0–1.2 mm, SiLibead Typ ZY-S, (Sigmund Lindner GmbH)], the sample tubes were washed with 0.75 ml acetonitrile/methanol/water-solution

mix (AMW) (2:2:1 v/v) and the AMW was also transferred to the bead-beater vials. For extraction blanks and standard samples 250 µl AMW were used. Bead beating was performed 4 x for 30 sec (4 m/sec) using FastPrep (MP Biomedicals), samples were centrifuged and the supernatant was transferred into new tubes. The remaining pellets were again extracted with 1 ml AMW, vortexed, centrifuged and the supernatants combined.

Prior to drying of the extracts, 100 µl ribitol solution [200 mg l^{-1} (1.3 mM)] was added per sample aliquot as an internal standard for GC-MS measurements. Samples were dried for 4 to 7 h at 30 °C and afterwards derivatized with 40 µl methoxyamine hydrochloride solution (0.02% w/v in pyridine) at 37 °C for 90 min followed by the addition of 40 µl N-Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) and further heating at 37 °C for 30 min. Both derivatization steps where performed under constant shaking at 1350 rpm. After quick spin-down, the supernatant was transferred to a small GC-vial for GC-MS analysis.

The GC-MS analysis was performed on an Agilent 8990B gas chromatograph connected to an Agilent 5977A MSD (Agilent Technologies). Samples were injected in splitless mode (1 µl) with an Agilent 7693 autosampler injector and separated using a DB5-MS column (Agilent Technologies) and helium as carrier gas. Further data acquisition and processing details has been described elsewhere (114).

PHA measurements

Paracatenula specimens were dried by centrifugal evaporation for 4 h at 30 °C using the Concentrator plus (Eppendorf). Pellets were re-suspended in 0.5 ml chloroform, vortexed and ultrasonicated (EMAG) for 5 min at 100% efficiency. Suspensions were transferred into GC-vials and 0.1 ml sodium benzoate solution (0.01% w/v) and 0.5 ml of 6% sulfuric acid solution (v/v) were added. The suspensions were boiled for 2 h at 100 °C. After cooling for 5 min at RT, 1 ml of deionized water was added and samples were vortexed. After 30 min of phase separation, the lower phase (nonpolar, chloroform) was transferred into a GC inlet for subsequent GC-MS analysis. To separate PHA monomer derivatives, a different temperature gradient was applied on the otherwise identical GC-MS equipment mentioned above: the gradient started with 50 °C for 1 min, increased to 120 °C at 10 °C/min, further increased to 280 °C at 45 °C/min, held 280 °C for 3 min, increased to 300 °C at 45 °C/min and held 300 °C for 3 min; mass detection was done from 35–400 Da.

Raman spectroscopy

Raman microscopy was performed on live *Paracatenula* sp. standrea specimens using a NTEGRA Spectra confocal spectrometer (NT-MDT Spectrum Instruments) configured with an inverted Olympus IX71 microscope (Olympus Corporation). For spectral analysis, the sample was excited by a solid-state laser at 532 nm and laser power was adjusted to a maximum of 1.5 mW. The pinhole aperture was set to 55 µm, resulting in a spatial resolution of approximately 250–300 µm. Raman scattered light was dispersed with a 150 line · mm^{-1} grating and collected by an electron multiplying charge coupled device (EMCCD) camera (Andor Technology) cooled to -70 °C. Exposure time was set to 0.6 s. Raman spectra were acquired in the spectral range of 0 and 4500 cm^{-1} with a spectral resolution of 0.2 cm^{-1}. Raman spectra were processed for normalization using the software Nova_Px v3.1.0.0 (NT-MDT Spectrum Instruments). For further peak determinations, data were exported to Excel (Microsoft).

Data availability

The assembled and annotated *Ca.* R. standrea genome was deposited at the European Nucleotide Archive (ENA) and is accessible under the project PRJEB26644 and the assembly accession GCA_900576755. Raw total RNA library reads were deposited in ENA under the accession no. PRJEB27565 and the host assembly under the accession no. PRJEB28271.

The mass spectrometry proteomics data and the protein sequence databases have been deposited to the ProteomeXchange Consortium *via* the PRIDE partner repository (115) for the pure culture data with the dataset identifier PXD009856 [**Reviewer access:** log in at http://www.ebi.ac.uk/pride/archive/ with Username: reviewer86761@ebi.ac.uk and Password: s5iurznD].

Code availability

The script that was used to classify the transporter families is available at http://github.com/kbseah/tcdbparse_sqlite.

Competing interests

The authors declare no conflict of interest.

Acknowledgments

Supplementary Material

The Dataset is available on the CD-ROM provided with the thesis.

Supplementary Note 1

Characterization of the *Paracatenula* sp. standrea host

A formal species description would be beyond the scope of this study and will be published elsewhere. So far, the morphotype of *Paracatenula* sp. standrea was exclusively sampled from coarse sediments in the bay off Sant´Andrea on Elba, Italy at waters depths of 5–10 meters in the vicinity of *Posidonia oceanica* meadows.

Paracatenula sp. standrea reach sizes of up to 1.5 cm and are the only *Paracatenula* species found at this sampling site. The specimens had a white-orange coloration and a smooth swimming behavior with the ability of backward movements that are characteristic for *Paracatenula* flatworms (Figure 4). A transparent central line of symbiont-free tissue that has also been observed in other *Paracatenula* species divides the trophosome into two lateral parts (33).

Morphologically, they most closely resembled *Paracatenula urania* (116)*.* The bacteriocytes were interspersed by clusters of bacteria-free and transparent cells that gave the trophosome region a "fluffy" appearance. Each bacteriocyte had an orange inclusion visible in both high magnification dissecting microscopy using a Nikon SMZ25 with incident light and Olympus BX53 compound microscope with transmitted light. These orange inclusions reached sizes of approx. 3 µm and exhibited autofluorescence with the highest intensity at 628 nm with an excitation of 561 nm (Supplementary Figure 20). The function of these inclusions is yet unknown. Non-motile sperm typically described in *Paracatenula* species were so far not detected (33, 116, 117). However, *Paracatenula* sp. standrea individuals occasionally carried clumps of darker colored cells that resembled developing eggs and might be an indication of sexual reproduction in this species (Supplementary Figure 20C).

Supplementary Note 2

Extended physiological description of *Ca.* R. standrea

Sulfur oxidation

The symbionts encoded the reverse-acting dissimilatory sulfite reductase (rDSR) pathway for the oxidation of sulfur to sulfite of which *dsrAB* and *aprAB* were among the highest transcribed and translated genes (Figure 2, Supplementary Figure 2, 3). Thiosulfate might be an important potential electron donor given the fluctuating sulfide concentrations in the sediments of Sant'Andrea Bay, Elba (7, 35, 50, 118). Two genes encoding putative thiosulfate rhodanese for the direct oxidation of thiosulfate to sulfite were encoded and also expressed.

Genes for the sulfur oxidation pathway *via* SOX (thiosulfate) have been found in related *Alphaproteobacteria*, e.g. in *Magnetospirillum* species, that encode genes for SOX and rDSR (119–121). Our genomic analysis also showed that *Caenispirillum salinarum* encoded a complete SOX pathway, but was missing the rDSR (Dataset S1, S11). The *Ca.* R. standrea genome only encoded SoxY and SoxZ that could function in carrying covalently bound reduced sulfur intermediates (122). The selection for the energy-efficient rDSR pathway is convergent to the thiotrophic gammaproteobacterial symbionts of *Olavius algarvensis* and *Riftia pachyptila* (35, 38, 63)*.

Autotrophy

The *Ca.* R. standrea symbiont can fix carbon dioxide using a modified Calvin-Benson-Bassham (CBB) cycle (Figure 2, Supplementary Figure 2, 3, 7–9, Dataset S4). Based on our phylogenetic analysis, we could exclude lateral gene transfer for both genes, the PP_i-dependent phosphofructokinase (PP_i-PFK) and the PP_i-energized proton pump (PP_i-H^+-pump) from *Gammaproteobacteria* as (i) the distinguishing- and active motifs of ATP- and PP_i-PFK sequences were conserved for *Alpha-*, *Beta-* and *Gammaproteobacteria*, and (ii) *Ca.* R. standrea PP_i-PFK and PP_i-H^+-pump gene sequences both clustered with genes from other *Alphaproteobacteria* (Supplementary Figure 7, 8).

The *Ca.* R. standrea symbiont expressed a carbonic anhydrase that converts bicarbonate to carbon dioxide (Dataset S2). This enzyme sustains high concentrations of carbon dioxide as

shown for the chemoautotrophic vesicomyid symbionts and in gills of *Bathymodiolus* mussels and could thus enhance carbon fixation (35, 63, 123).

Aerobic respiration

The aerobic respiration is driven by an electron transport chain consisting of NADH quinone oxidoreductase, succinate dehydrogenase and cytochrome bc_1-type ubiquinol oxidoreductase (cytochrome C1 and cytochrome B) in conjunction with a F_0F_1-type ATP synthase (Figure 2). Several of these genes were among the highest expressed, but only the cytoplasmic ones could be detected in the proteome (Supplementary Figure 2, 3, Dataset S3). The aa_3-type cytochrome c (cytochrome c oxidase polypeptide I–III) oxidase present operates best under atmospheric oxygen concentrations (124, 125).

Amino acid, cofactor and vitamin biosynthesis

In the biosynthesis of 20 amino acids, only the pathway for histidine was missing one step that we could not explain by comparative analyses, while the two other missing steps likely are substituted by yet unknown reactions as they were also missing in related *Rhodospirillales* (Supplementary Figure 18). The *Ca.* R. standrea symbiont expressed the pathways for both essential and non-essential amino acids as well as for vitamins such as cobalamin (vitamin B12), biotin (vitamin H), thiamine (vitamin B1), folate (Vitamin B9), ubiquinone, riboflavin (vitamin B2), pantothenate (vitamin B5), pyridoxine (vitamin B6) and the cofactors heme, nicotinamide adenine dinucleotide (NAD), lipoate, tetrahydrofolate and coenzyme A (Supplementary Figure 19).

Limited import of organic compounds points to internal cycling of carbon stocks

The versatile intermediary carbon metabolism of the *Ca.* R. standrea symbiont, integrating glycolysis, the TCA cycle, the ethylmalonyl-CoA pathway and an incomplete 3-HPB pathway suggests a potential for mixotrophy (Figure 2). Evidence for such heterotrophy in chemoautotrophic symbionts was reported for several of the *Olavius* symbionts, symbionts of the scaly-foot snail, of *Solemya*, of the nematode *Laxus oneistus* and the lucinid bivalve *Loripes lucinalis* (15, 36, 41, 42, 50). These symbionts have diverse types of substrate uptake transporters to fuel a heterotrophic metabolism (36). To assess the potential importance of

net heterotrophy in contrast to internal cycling of carbon stocks in *Ca.* R. standrea, we performed a targeted reanalysis of the transporters using the transporter classification database (TCBD) (87) (Dataset S5A). Our data showed that *Ca.* R. standrea had a limited set of importers (10), similar to vesicomyid clam endosymbionts (2 and 5) (Supplementary Figure 11, for a list of genomes see Dataset S6). The symbiont could import small peptides/opines or nickel, the sugars fructose or mannose and phenylpropionate from its hosts cellular environment. Although this small set of transporters was expressed in its entirety the recycling of metabolites from host fermentation is unlikely – a process suggested e.g. in the *Ca.* Thiosymbion symbionts of *Olavius algarvensis* (35). In summary, it is more likely that *Ca.* R. standrea retained the versatile carbon metabolism for the efficient turnover of internal carbon stocks rather than the import of nutrients from its host environment.

Limited export of organic compounds

We identified fewer export transporters in *Ca.* R. standrea (15 with and without transmembrane domains) than in most other thiotrophic symbionts (up to 85 in *Ca.* Endoriftia persephone), similar to the vesicomyid clam symbionts (18 and 25) (Supplementary Figure 11, Dataset S5).

Supplementary Note 3

COG profile conservation across vertically transmitted chemosynthetic symbionts

Eight COG categories that highlighted important cellular functions were overrepresented in all vertically transmitted thiotrophic symbionts with reduced genomes – *Ca*. R. standrea, a sponge symbiont and the two vesicomyid symbionts (n = 4), such as nucleotide transport and metabolism (category F), coenzyme transport and metabolism (H), translation, ribosomal structure and biogenesis (J) and posttranslational modification, protein turnover, chaperones (O) (Supplementary Figure 13). In contrast, the eight underrepresented COG categories including carbohydrate transport and metabolism (G), cell motility (N), secondary metabolite biosynthesis, transport and catabolism (Q) and defense mechanisms (V). These patterns of enrichment or depletion in reduced genomes were consistent with an intracellular lifestyle and

vertical transmission where they are physically confined to their host, experience no external competition and live in a sheltered and predictable environment.

KEGG profile conservation across vertically transmitted chemosynthetic symbionts

A similar distribution of thiotrophic symbionts with reduced *vs*. non-reduced genomes was also recovered with a NMDS ordination using the predicted metabolic modules of the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Figure 3C, Dataset S7, S8, for a list of genomes see Dataset S6). These modules represent discrete functional units such as glycolysis or the tricarboxylic acid (TCA) cycle and are more specific than the broad COG categories. KEGG modules that were encoded by the majority of the thiotrophic symbionts with non-reduced genomes but absent in the reduced genomes included a broad set of transporters for peptides and sugars (PTS-systems), two-component systems, nitrogen assimilation and fixation and secretion systems.

Pathways for breakdown and assimilation are lost in reduced chemoautotrophic genomes

To characterize the core functions of thiotrophic symbionts with reduced genomes, we identified the shared metabolic pathways based on matches to the BioCyc database (Figure 3C). *Ca*. R. standrea encoded 154 pathways, which was slightly more than the symbionts of vesicomyid clams with similarly reduced genomes (120 / 131 pathways). In contrast, free-living relatives of *Ca.* R. standrea encoded 203 to 251 predicted pathways in genomes that were up to 3.7 times larger. Much of the difference between the free-living relatives and the thiotrophic symbionts with reduced genomes can be explained by the loss of pathways for breakdown and assimilation of organic substrates (*Ca*. R. standrea: 33, vesicomyid clam symbionts: 19 / 22, free-living: 62 to 84) (Supplementary Figure 10A, Dataset S1, S9, S11).

Supplementary Table and Figures

**Supplementary Table 1. Genomes used for comparative analysis.**

| Species	Group	Location	Transmission	Study	Genome source	Size (bp)
Ca. R. standrea	*a-proteob.*	Intracellular	Vertical	Present study	GCA_900576755	1342908
Solemya symbiont	*γ-proteob.*	Intracellular	Mixed mode	(36)	JRAA00000000	2702377
Ca. V. okutanii	*γ-proteobac.*	Intracellular	Vertical	(14)	NC_009465	1022154
Ca. R. magnifica	*γ-proteob.*	Intracellular	Vertical	(13)	NC_008610	1160782
Ca. E. persephone	*γ-proteob.*	Intracellular	Horizontal	(126)	AFOC01000000	3481040
T. crunogena	*γ-proteob.*	Free-living	-	-	NC_007520	2427734
E. coli K12	*γ-proteob.*	Free-living	-	(127)	NC_000913	4641652
M. magneticum	*a-proteob*	Free-living	-	(128)	NC_007626	4967148
C. salinarum	*a-proteob.*	Free-living	-	(129)	ANHY00000000	4952465
R. rubrum	*a-proteob.*	Free-living	-	-	AAAG00000000.2	4403907
N. itersonii	*a-proteob.*	Free-living	-	-	ARMX00000000	4289611
B. aphidicola	*γ-proteob.*	Intracellular	Vertical	(130)	NC_002528	640681
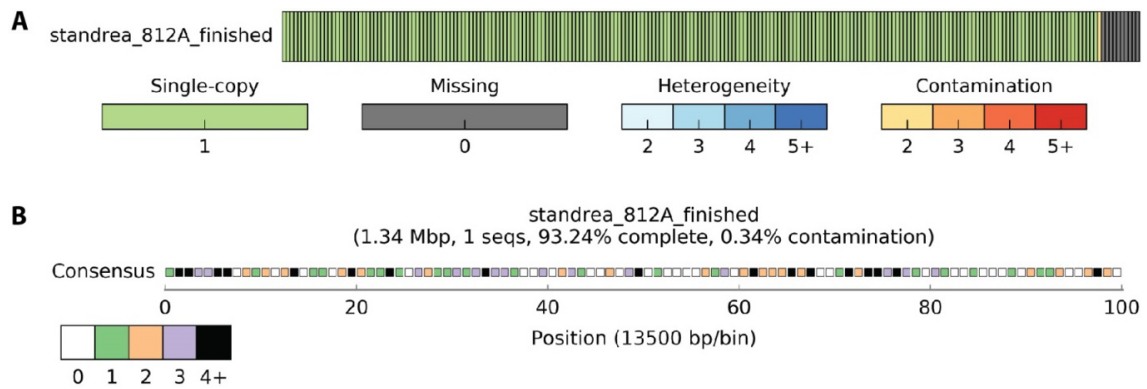
Supplementary Figure 1. Quality assessment of the *Ca.* R. standrea genome bin. The high completeness of the *Ca.* R. standrea genome was evaluated with CheckM using the lineage-specific workflow for *Alphaproteobacteria*. *A*, Visual representation of completeness, contamination and strain heterogeneity within the *Ca.* R. standrea (= standrea_812A_finished) genome bin. Single-copy marker genes are indicated by green bars – missing marker genes in gray. If markers were found multiple times they would be indicated by shades of blue (≥ 90% amino acid identity (AAI) or red (< 90% AAI). *B*, Visual representation of the position of marker genes on the bin sequence (1.34 Mbp). The coloration indicates the number of marker genes per position.
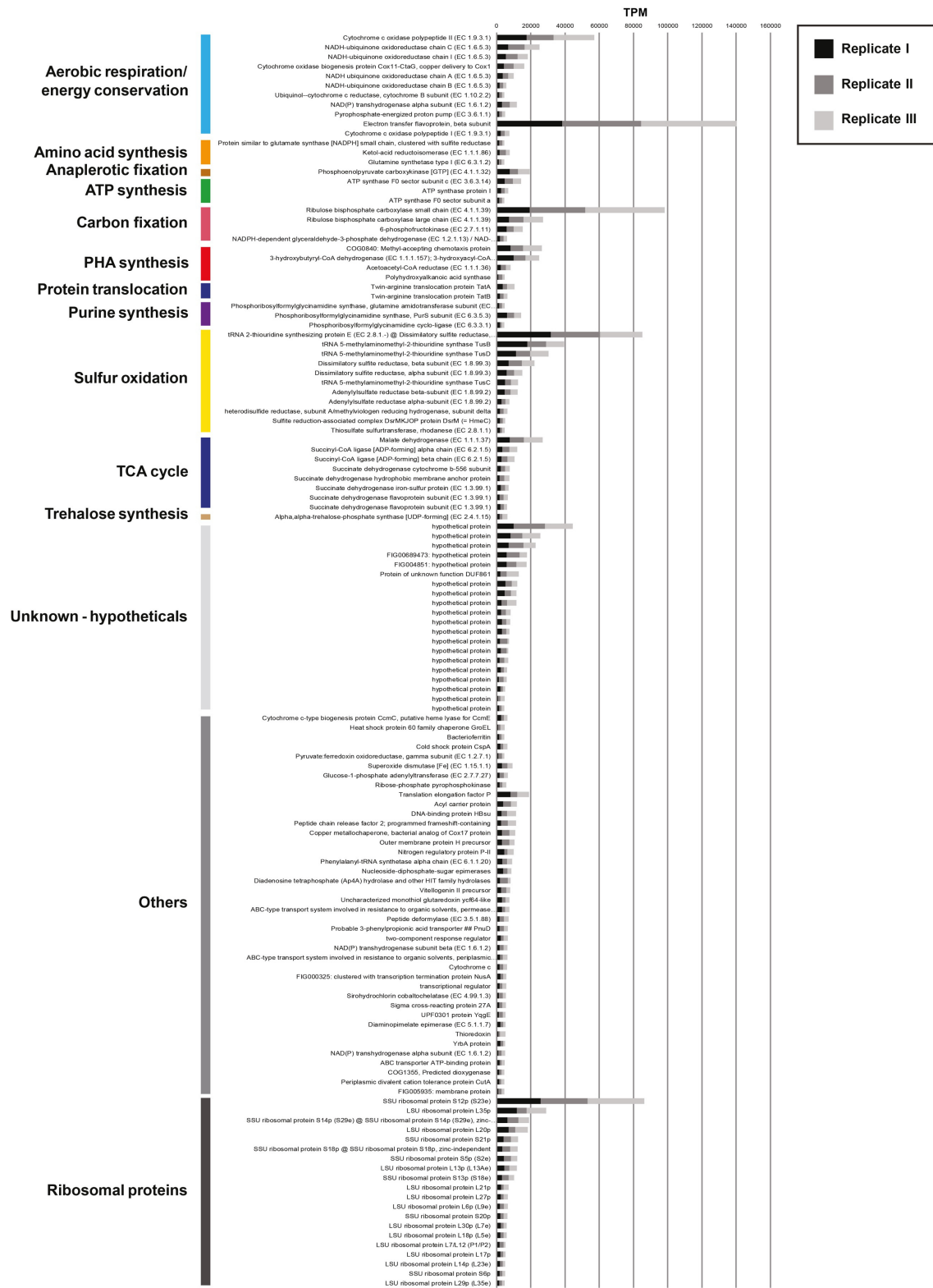
Supplementary Figure 2. Most abundant transcribed genes (top 10%) in *Ca.* R. standrea. The transcription levels were shown as transcripts per million (TPM), genes were sorted by their function and shown for each replicate composed of three pooled specimens.
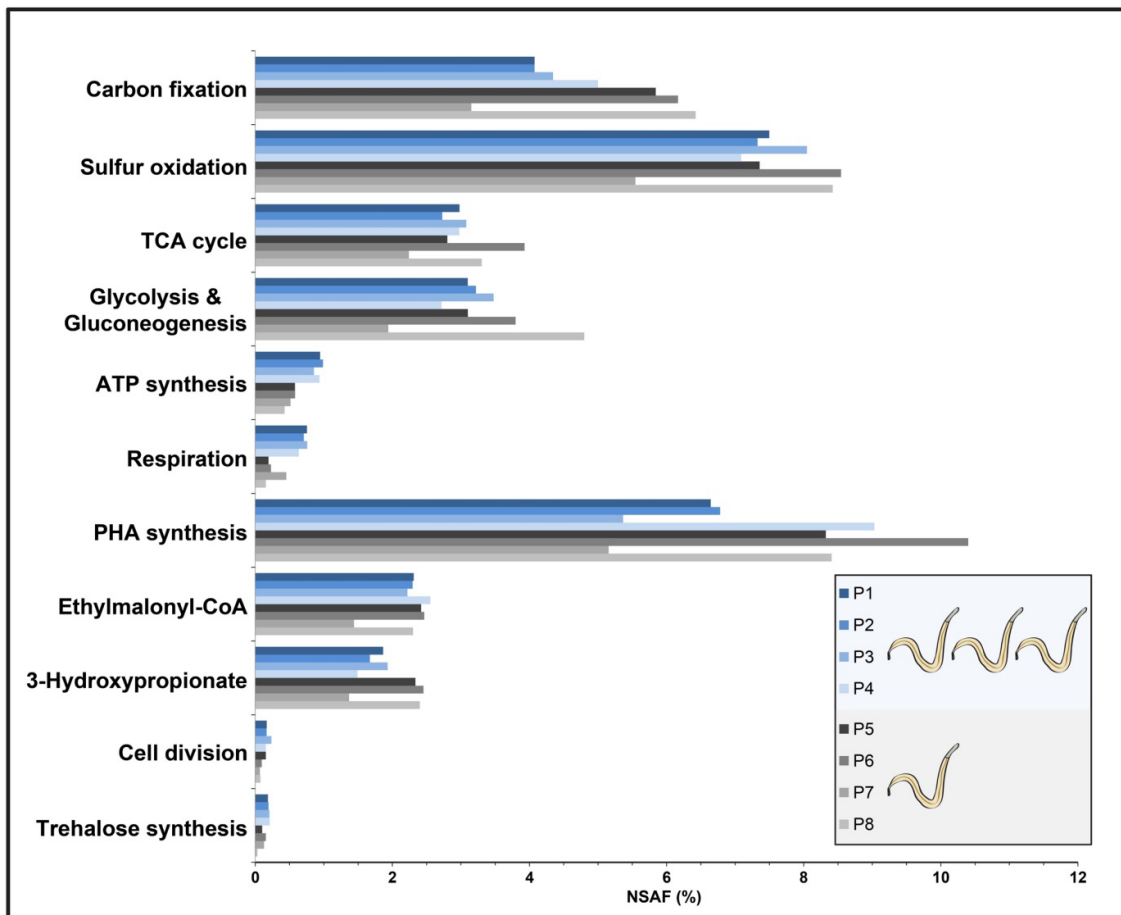
Supplementary Figure 3. Expression of abundant proteins in *Ca*. R. standrea. The bar chart shows the abundance of proteins for selected metabolic pathways identified by proteomics in eight different *Paracatenula* samples (P1–P8). Samples were either a pool of three individuals or single specimens (indicated by the icons). Only proteins that are unique for a pathway were included in this representation (Dataset S3). NSAF (%) = normalized spectral abundance factor.
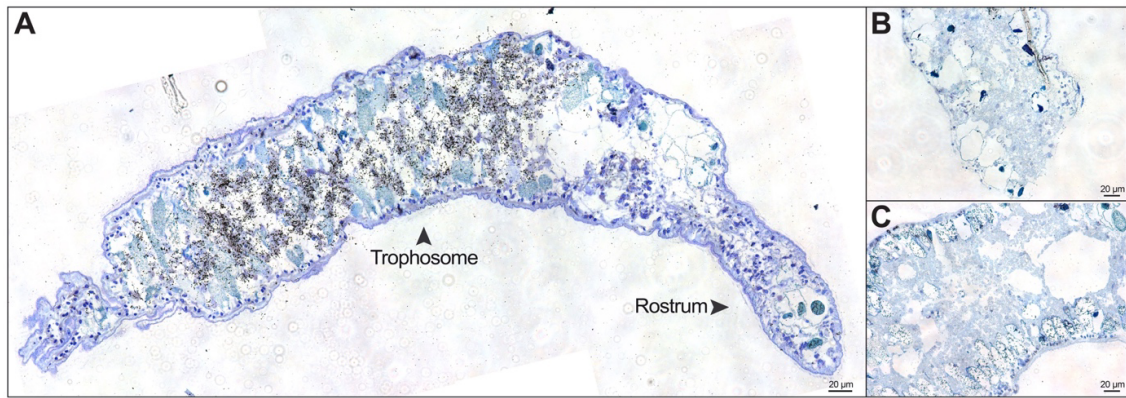
Supplementary Figure 4. Carbon fixation of *Ca*. R. standrea symbionts visualized by microautoradiographs of ^{14}C-bicarbonate incubated *Paracatenula* sp. standrea. *A*, 6 h incubated specimen after 3 d exposure. Signals were present in the trophosome region while the rostrum was signal-free. The image is composed of eight stitched images recorded at 40x magnification. Raw images are shown in Supplementary Figure 5. ***B–C***, Controls after 6 days exposure. ***B***, Dead control. ***C***, Negative control fixed without ^{14}C incubation.
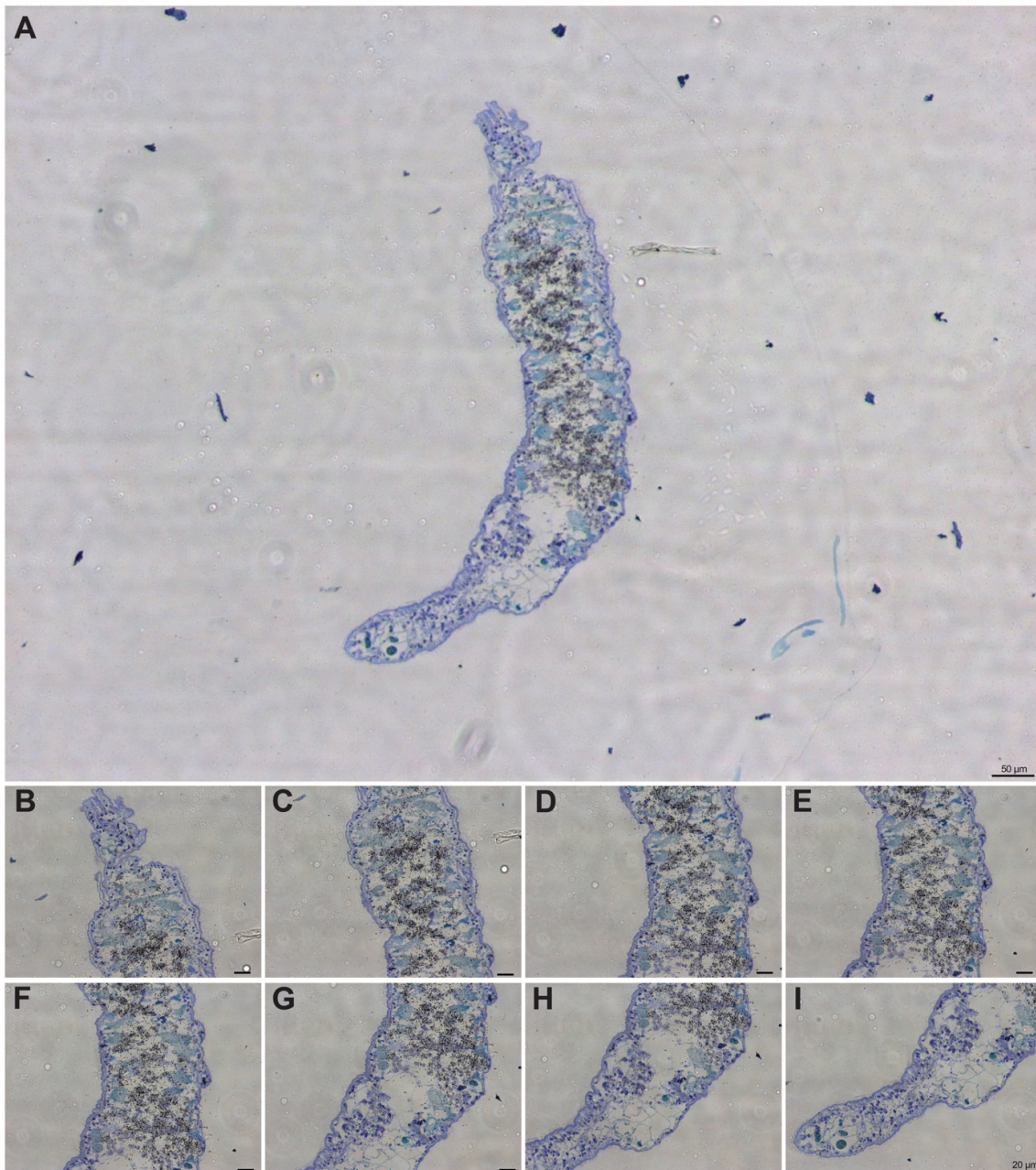
Supplementary Figure 5. Raw image data for Supplementary Figure 4 – Carbon fixation of *Ca*. R. standrea symbionts visualized by microautoradiographs of ^{14}C-bicarbonate incubated *Paracatenula* sp. standrea. 6 h incubated specimen after 3 d exposure. ***A***, Overview image at 10x magnification. ***B–I***, Raw images (40x) for the stitched overview shown in Supplementary Figure 4. ***E*** and ***F*** show the same area but different focal planes. No white balancing of images was performed.
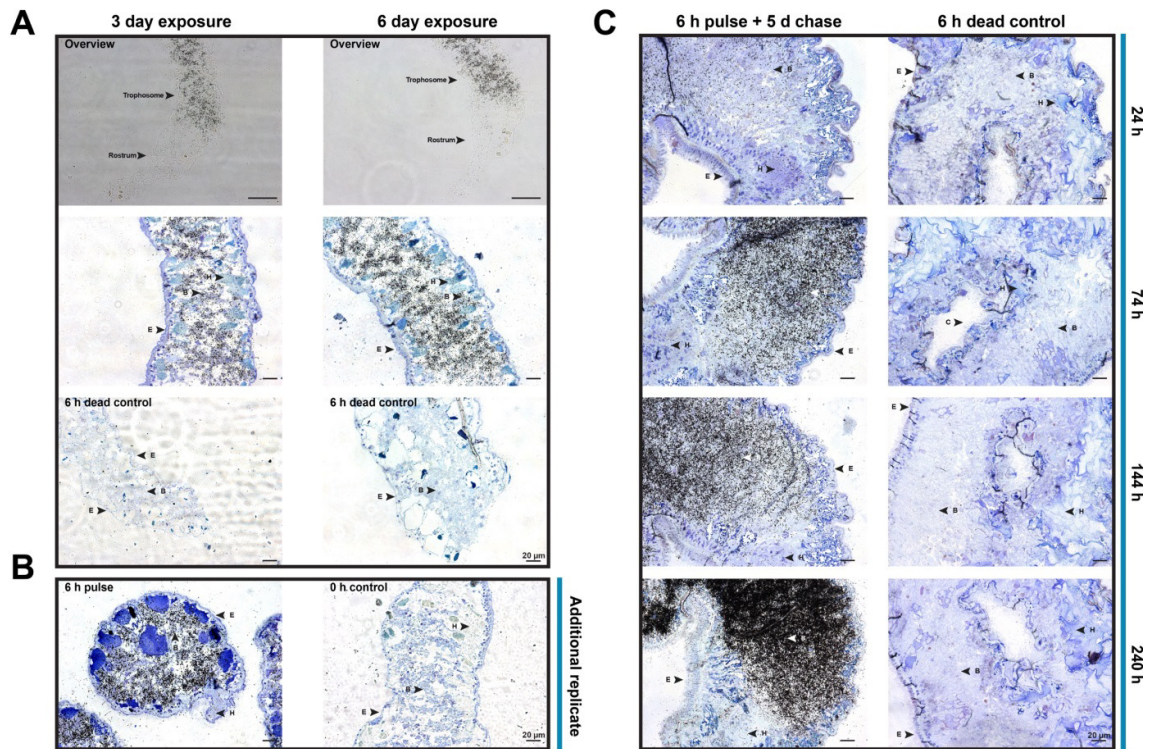
Supplementary Figure 6. Microautoradiographs on sections of ^{14}C-bicarbonate incubated *Paracatenula* sp. standrea individuals to visualize carbon fixation of the symbionts. Specimens were incubated for 6 h in labeled bicarbonate and resin sections were exposed to radiosensitive emulsion (K5, Ilford) for up to 10 days. ***A***, Sections of the same organism as shown in Supplementary Figure 5 after 3 days exposure and 6 days exposure. An increase in signal was observed with time. No signal in host tissue. Controls stayed signal-free. ***B***, Different replicate from a second incubation experiment. Exposure to emulsion for 7 days. No signal in host tissue and 0 h control stayed signal-free. ***C***, Again a different replicate from another experiment. Left panel, specimen was incubated for 6 h followed by a 5 day chase (absence of ^{14}C-HCO$_3^-$) to potentially see transfer into host material. Right panel, 6 h dead control. Exposures of same individual but different sections were done for 24 h to up to 240 h. Instead of artificial seawater we used seawater spiked with ^{14}C-HCO$_3^-$. More signals were observed within shorter time. No signal in host tissue (= no transfer). Dead control stayed signal-free. Abbreviations: E = epidermis; B = bacterial symbionts, H = host; C = cilia.
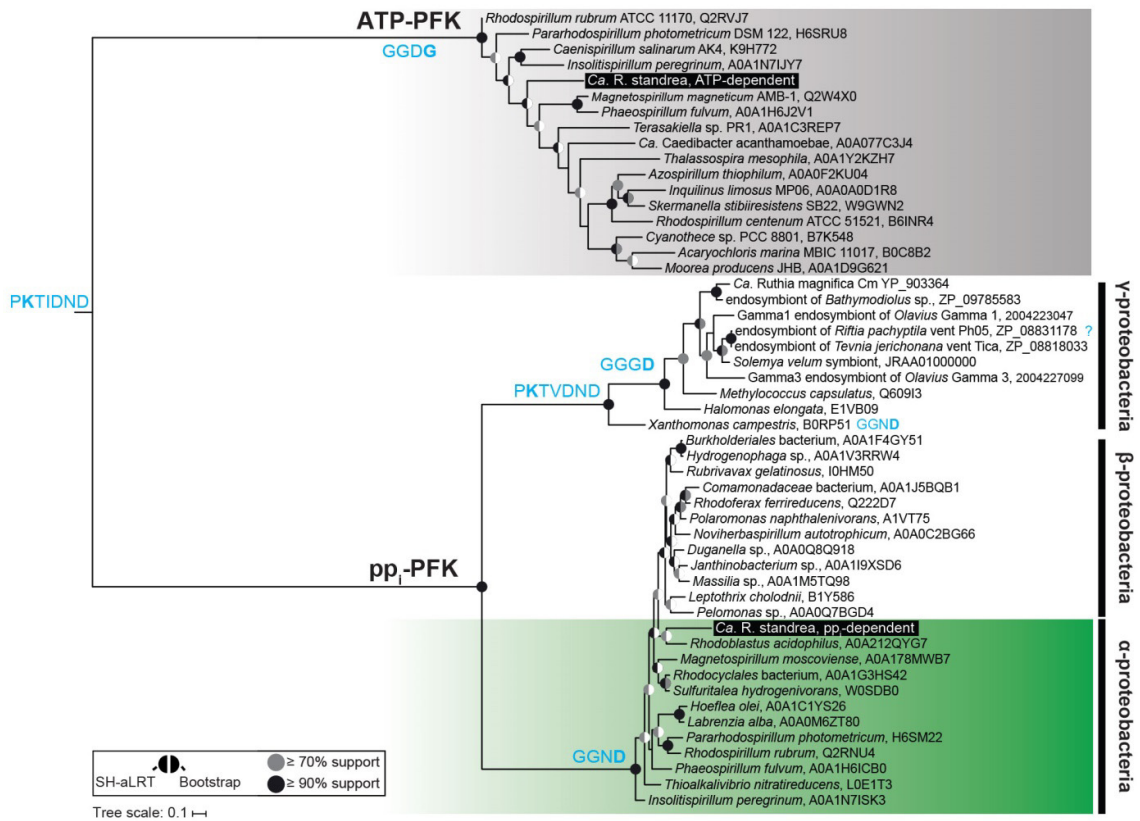
Supplementary Figure 7. Phylogenetic placement of PFK sequences indicates an alphaproteobacterial origin of PP$_i$-PFK in *Ca.* R. standrea. The PP$_i$-PFK sequence of *Ca.* R. standrea clustered with sequences of related *Alphaproteobacteria*. The amino acid sequence alignment was calculated with MAFFT (GIN-S-I) and the tree estimated using maximum likelihood with IQ-TREE with the LG+G4 model. The tree was rooted with the ATP-PFK clade. Respective conserved motifs for ATP-PFK and PP$_i$-PFK shared by groups are shown in light blue. ? = Short sequence missing respective motif. Bootstrap proportions are shown as SH-aLRT support values and nonparametric bootstraps. White half-circles represent support values lower 70%. Scale bar represents 10% estimated sequence divergence.
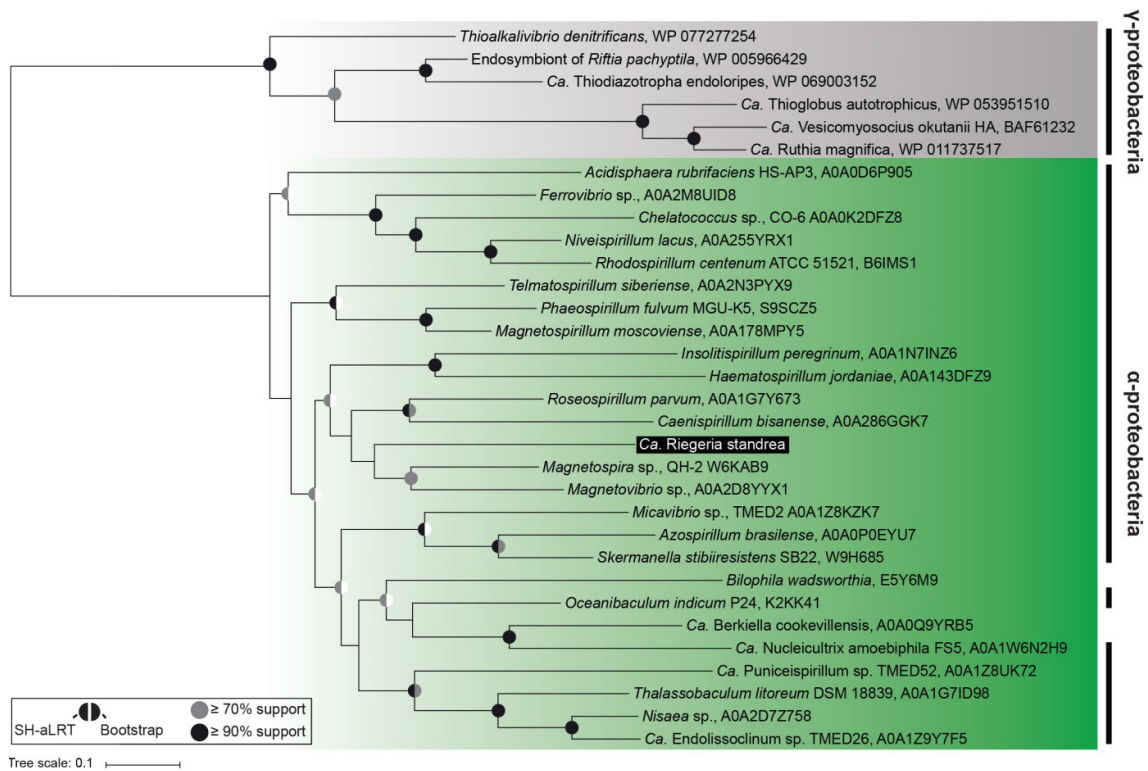
Supplementary Figure 8. Phylogenetic placement of PPi-H+-pump sequences indicates an alphaproteobacterial origin in *Ca*. R. standrea. The PPi-H+-pump sequence of *Ca*. R. standrea clustered with sequences of related *Rhodospirillaceae*. The amino acid sequence alignment was calculated with MAFFT (GIN-S-I) and the tree estimated using maximum likelihood implemented in IQ-TREE with the LG+F+I+G4 model. The tree was rooted with the gammaproteobacterial sequences. Bootstrap proportions are shown as SH-aLRT support values and nonparametric bootstraps. White half-circles represent support values lower 70%. Scale bar represents 10% estimated sequence divergence.
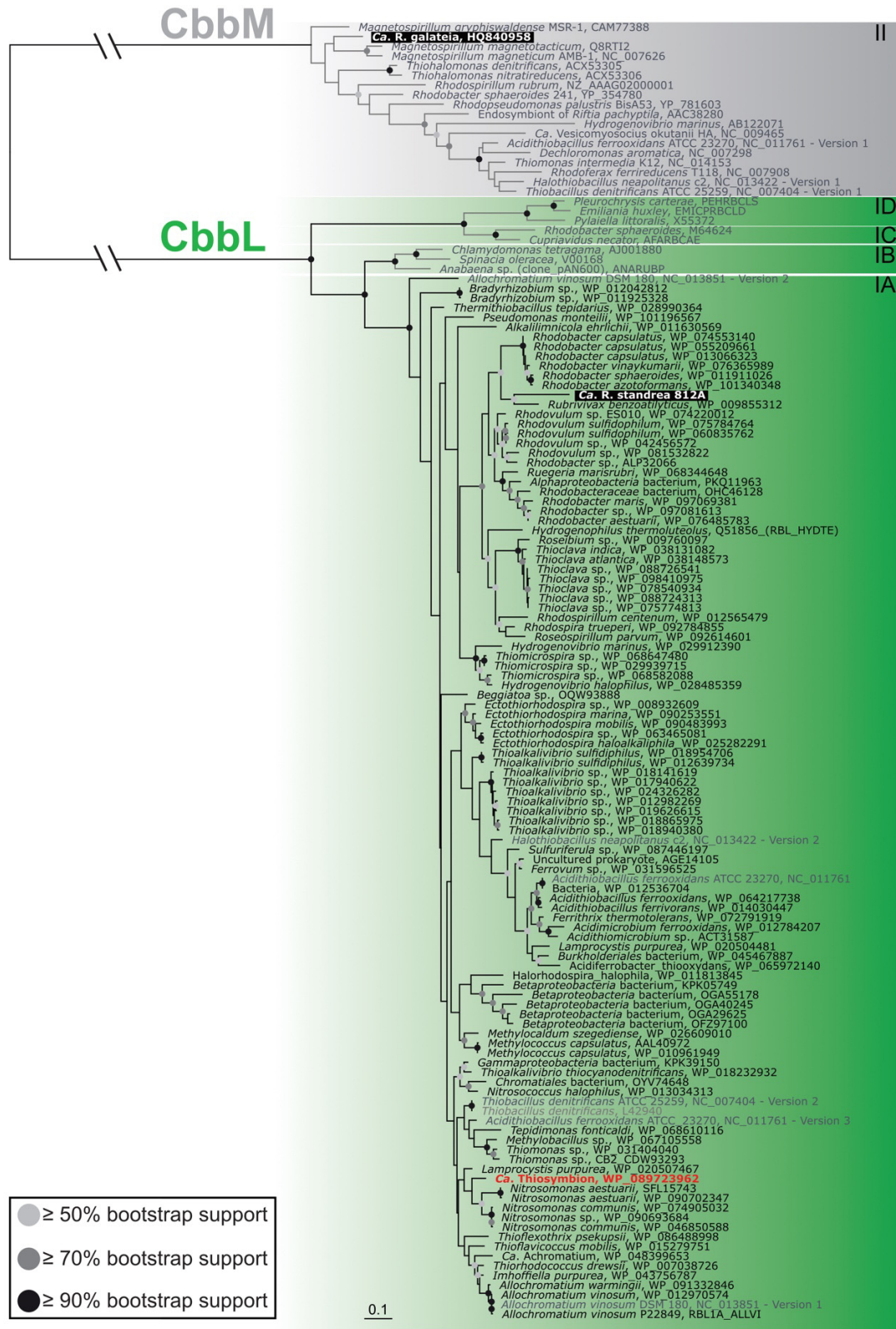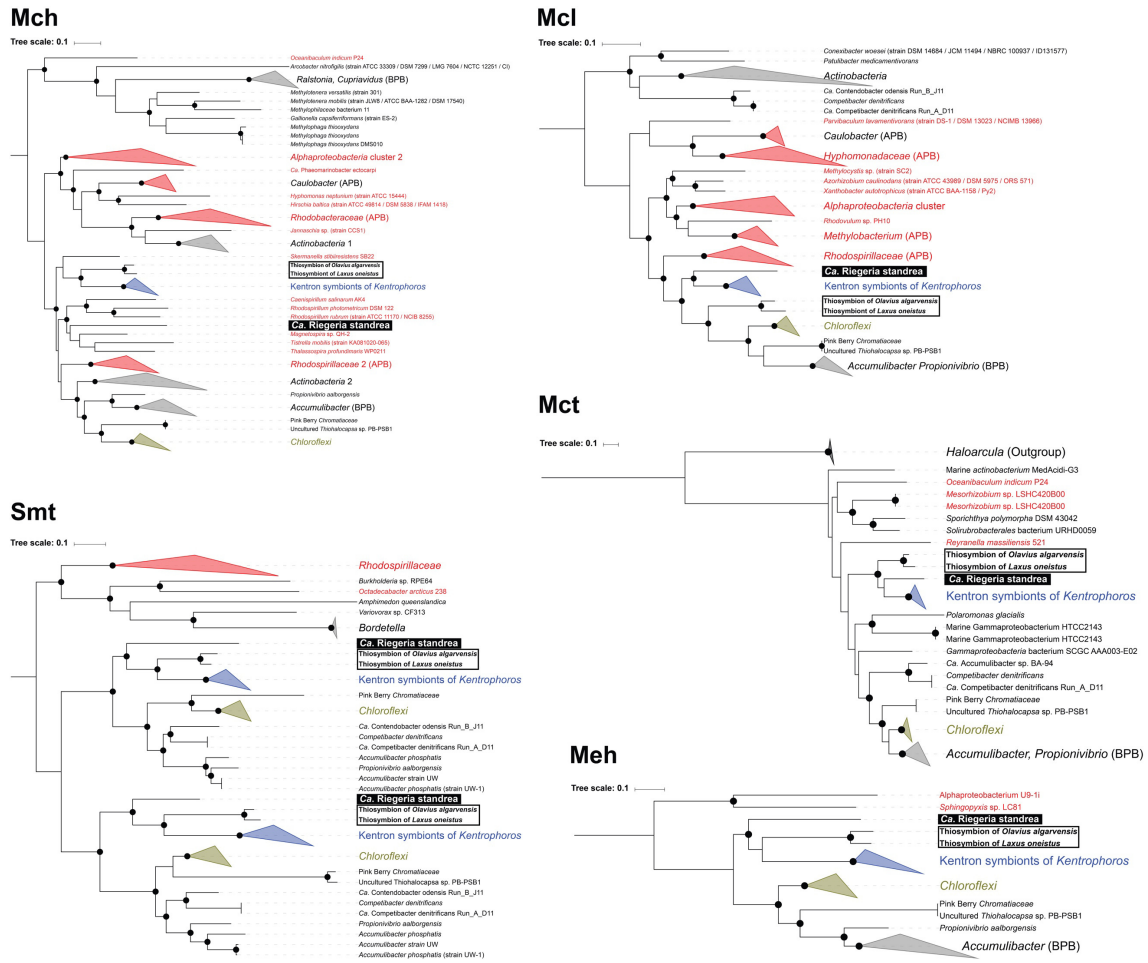
Supplementary Figure 9. Phylogenetic reconstruction of RuBisCO sequences indicates an alphaproteobacterial origin of RuBisCO in *Ca.* R. standrea. The CbbL sequence of *Ca.* R. standrea clustered within the group RuBisCO form I and the sequence of *Ca.* R. galateia in the group RuBisCO form II. The amino acid sequence alignment was calculated with MAFFT (GIN-S-I) and the tree was estimated using maximum likelihood with PhyML with the LG substitution model and 100 BS. The tree was rooted with the CbbL clade. Bootstrap proportions are indicated. Scale bar represents 10% estimated sequence divergence.
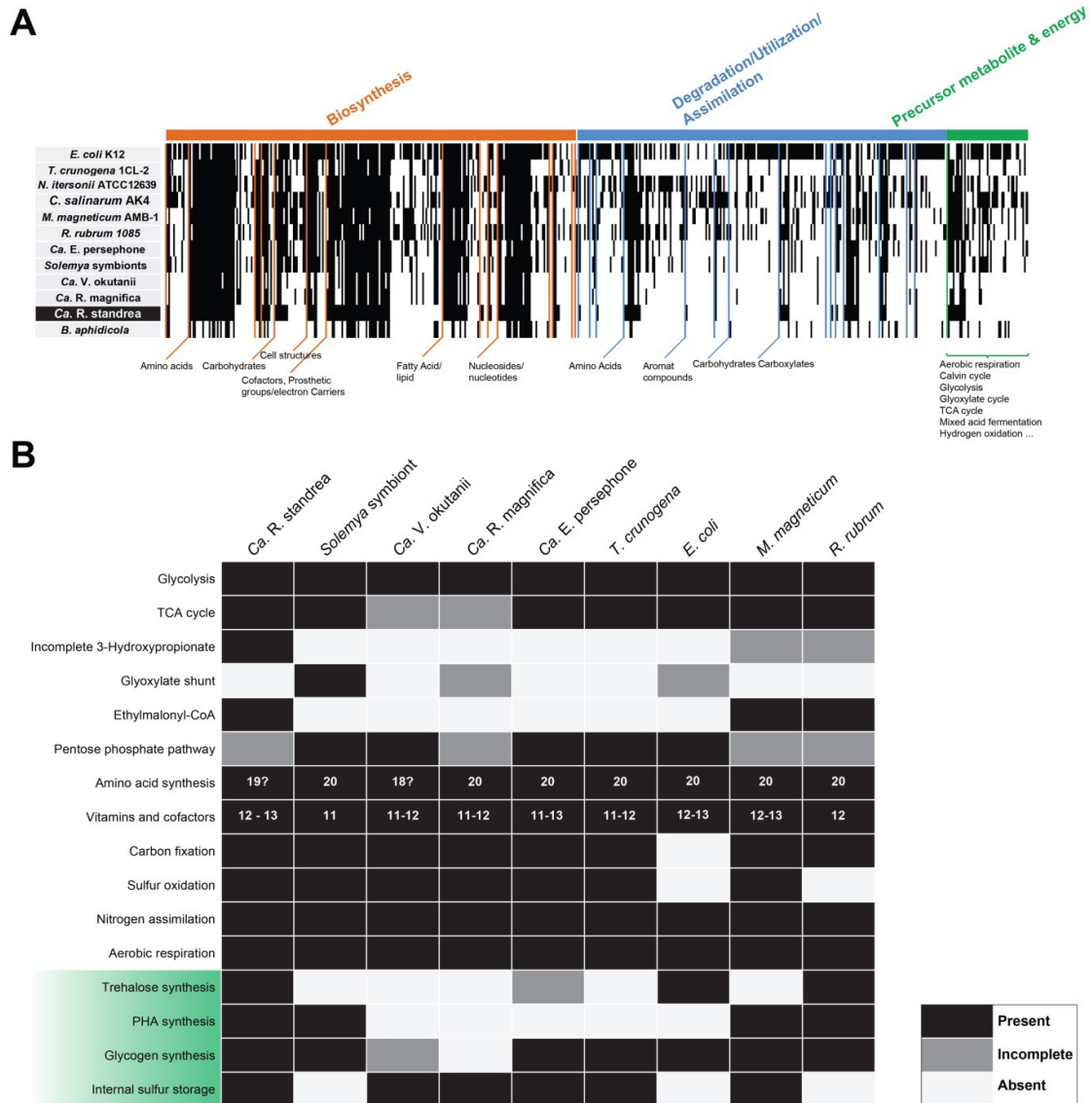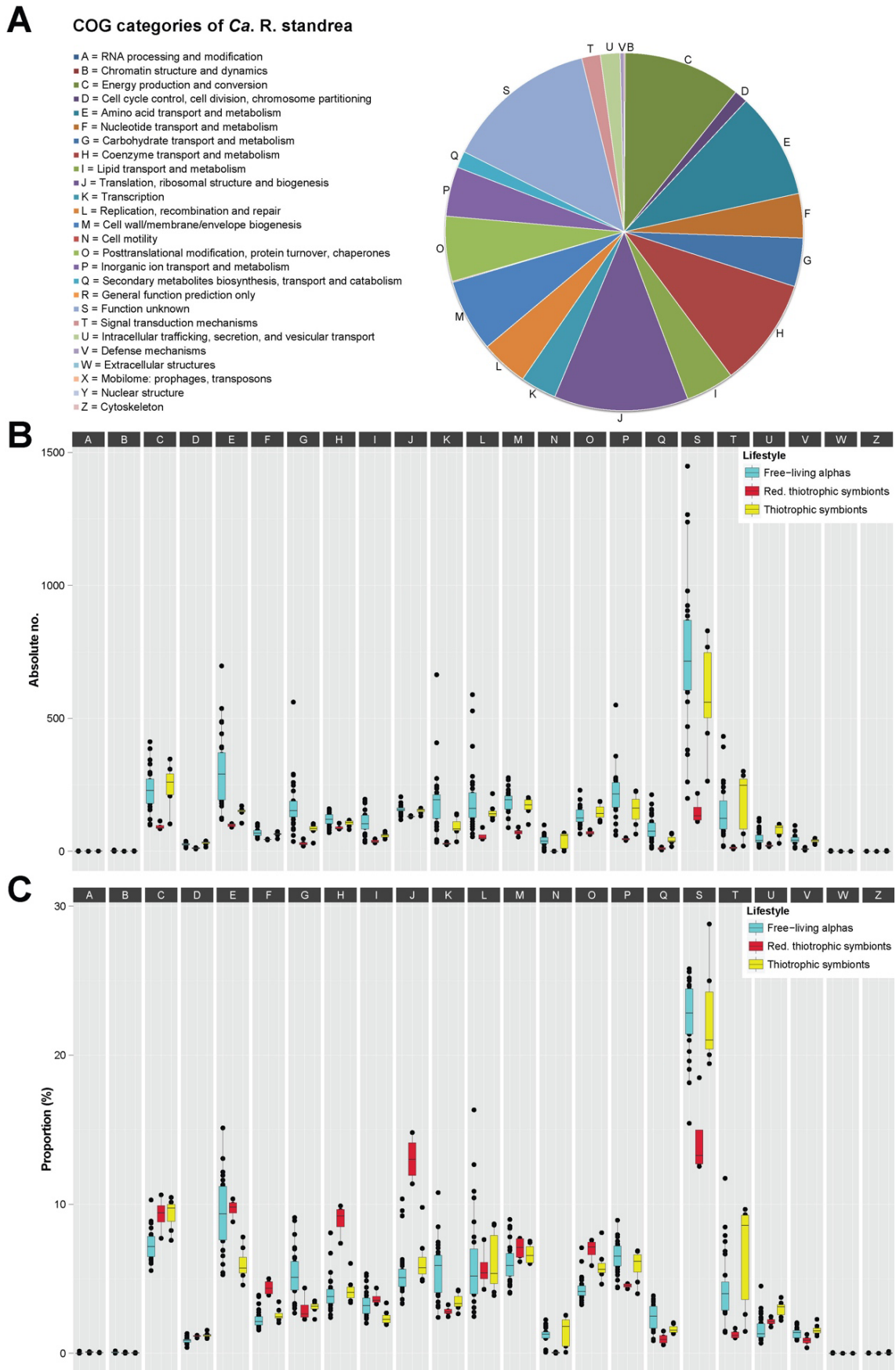
Supplementary Figure 10. Comparison of biosynthetic and catabolic processes of *Ca.* R. standrea, free-living and symbiotic bacteria and thiotrophic symbionts. *A*, A comparative metabolic analysis. Present pathways (based on Pathway Tools/BioCyc classification) were indicated as black lines. ***B***, Occurrence of metabolic features in *Ca.* R. standrea, other thiotrophic symbionts with reduced genomes and free-living bacteria (Dataset S11). Genomes used are summarized in Supplementary Table 1.
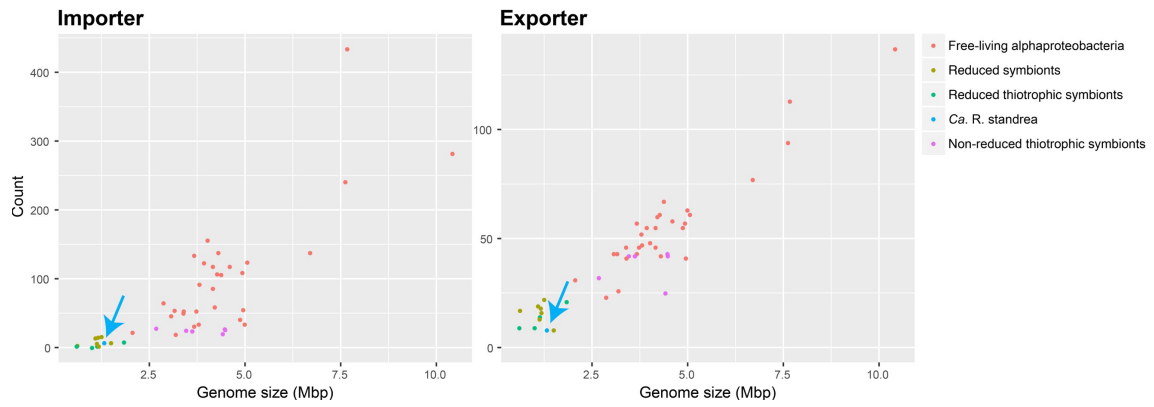
Supplementary Figure 11. Transporters in *Ca.* R. standrea and other organisms. Only hits against the TCDB with at least one transmembrane (TM) domain and ≥ 30% identity are shown. The light blue arrow indicates the counts for *Ca.* R. standrea. The complete list of genomes used can be found in Dataset S6.

Supplementary Figure 12. Phylogenetic trees of representative genes of the incomplete 3-HPB pathway indicate an origin of the *Ca.* R. standrea genes outside of the *Chloroflexi*. Three out of five key genes clustered with sequences of other *Rhodospirillaceae*, and two had no homologs in *Rhodospirillales*. The amino acid alignments were calculated with Muscle and the tree estimated using maximum likelihood. Midpoint rooting was used for each of the five trees. Mch, mesaconyl-C1-CoA hyratase; McI, malyl-CoA/beta-methylmalyl-CoA/citramalyl-CoA (MMC) lyase; Mct, mesaconyl-CoA C1-C4 CoA transferase; Meh, mesaconyl-C4-CoA hydratase; Smt, succinyl-CoA:(S)-malate-CoA transferase. Bootstrap proportions ≥80% are indicated. Scale bar represents 10% estimated sequence divergence.

A

COG categories of *Ca*. R. standrea

- A = RNA processing and modification
- B = Chromatin structure and dynamics
- C = Energy production and conversion
- D = Cell cycle control, cell division, chromosome partitioning
- E = Amino acid transport and metabolism
- F = Nucleotide transport and metabolism
- G = Carbohydrate transport and metabolism
- H = Coenzyme transport and metabolism
- I = Lipid transport and metabolism
- J = Translation, ribosomal structure and biogenesis
- K = Transcription
- L = Replication, recombination and repair
- M = Cell wall/membrane/envelope biogenesis
- N = Cell motility
- O = Posttranslational modification, protein turnover, chaperones
- P = Inorganic ion transport and metabolism
- Q = Secondary metabolites biosynthesis, transport and catabolism
- R = General function prediction only
- S = Function unknown
- T = Signal transduction mechanisms
- U = Intracellular trafficking, secretion, and vesicular transport
- V = Defense mechanisms
- W = Extracellular structures
- X = Mobilome: prophages, transposons
- Y = Nuclear structure
- Z = Cytoskeleton

B

C

Supplementary Figure 13. COG assignments in *Ca.* R. standrea, and a comparative analysis with free-living bacteria, thiotrophic symbionts with reduced and non-reduced genomes. The

complete list of genomes used can be found in Dataset S6. ***A***, Relative abundance of COGs in *Ca.* R. standrea that were assigned using eggNOG-mapper. The list of categories was taken from ftp://ftp.ncbi.nih.gov/pub/COG/COG2014/data/fun2003-2014.tab. ***B−C***, The distribution of COG categories based on absolute (***B***) and relative (***C***) numbers among free-living *Alphaproteobacteria*, thiotrophic symbionts with reduced and non-reduced genomes. The names of COG categories are indicated at the top of the plot. Categories R, X, and Y were not shown as no orthologs were assigned. n = 28 (free-living bacteria); n = 4 (thiotrophic symbionts with reduced genomes that are *Ca.* R. standrea, *Ca.* R. magnifica, *Ca.* V. okutanii and GSub); n = 7 (thiotrophic symbionts with non-reduced genomes).
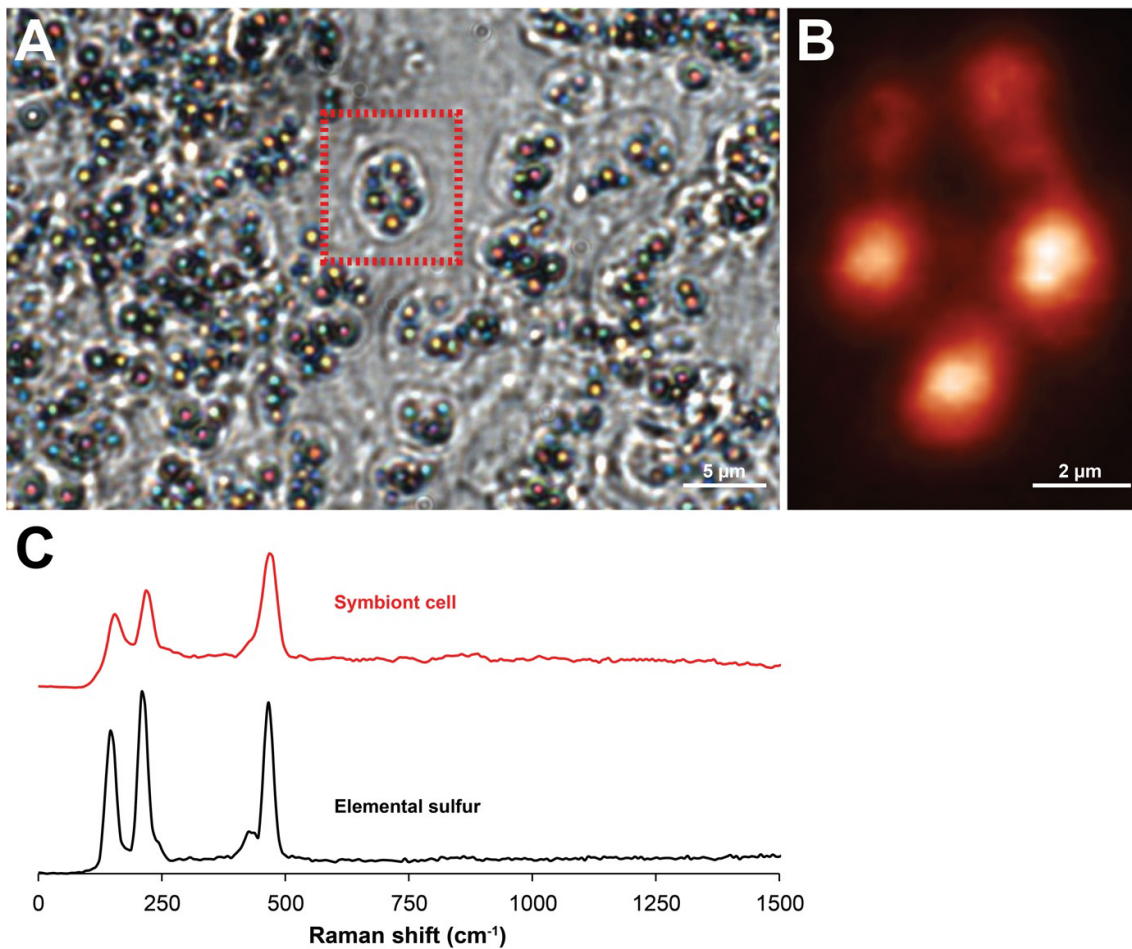
Supplementary Figure 14. Raman analysis of storage compounds in *Ca.* R. standrea. *A*, Optical micrograph of symbionts. Scale bar: 5 µm. The red box indicates the area shown in B. ***B*,** A Raman-map showing the intensity of the 473 cm^{-1} cm peak corresponding to the S-S bond stretch of elemental sulfur. Scale bar: 2 µm. ***C*,** Point spectra of sulfur measured within the cells shown in red compared to an elemental sulfur standard shown in black (Roth 33 GmbH).

102
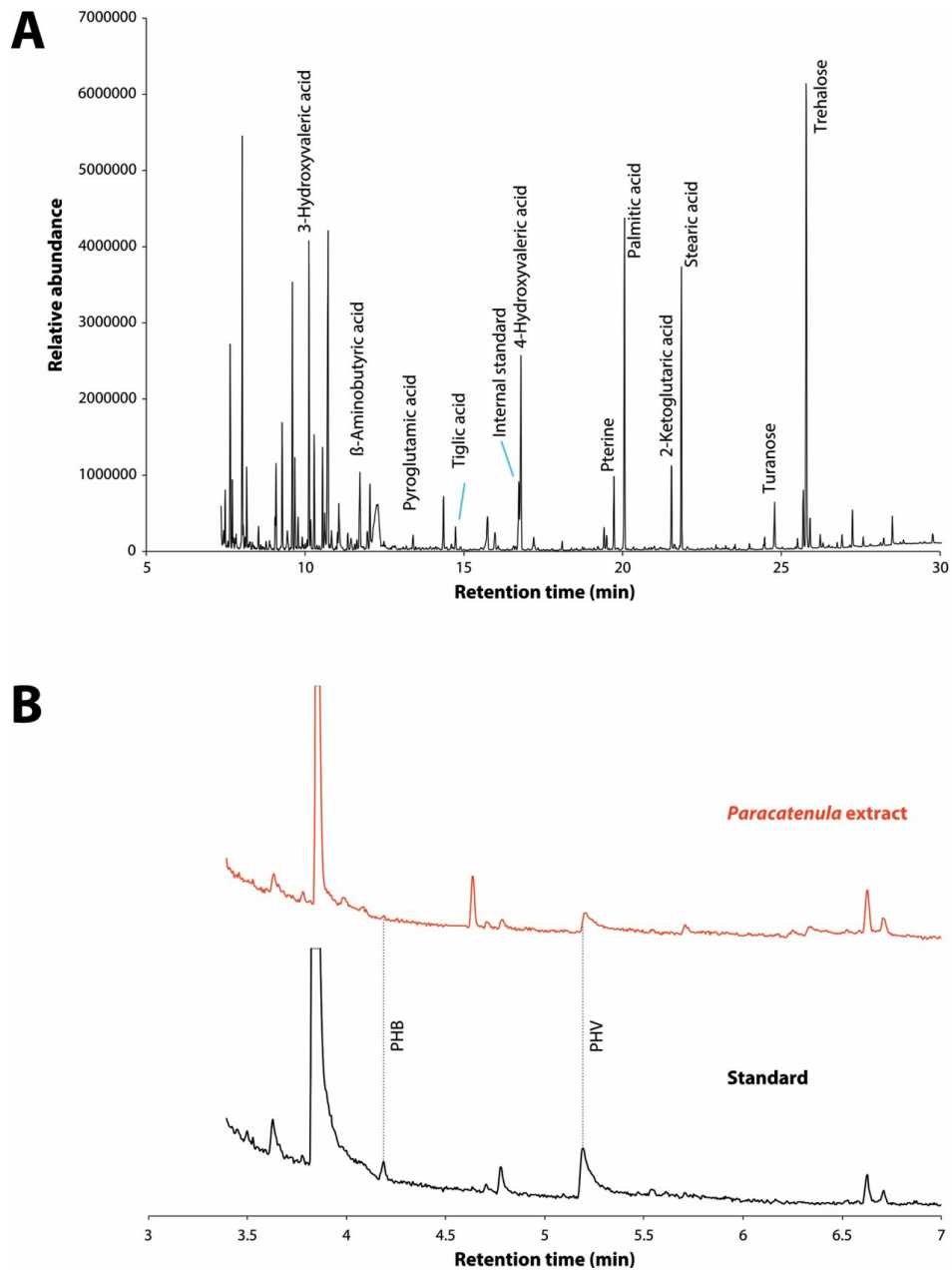
Supplementary Figure 15. Metabolomic analysis of *Paracatenula* sp. standrea specimens. *A*, GC-MS chromatogram of bulk measurement (25 *Paracatenula* specimens) in polar phase. Compounds with >65% confidence identity were indicated. ***B*,** GC-MS chromatograms of Polyhydroxyvalerate (PHV) measured in the *Paracatenula* extract shown in red compared to a Polyhydroxybutyrate (PHB):PHV standard shown in black. Note the absence of the PHB peak in the *Paracatenula* extract.
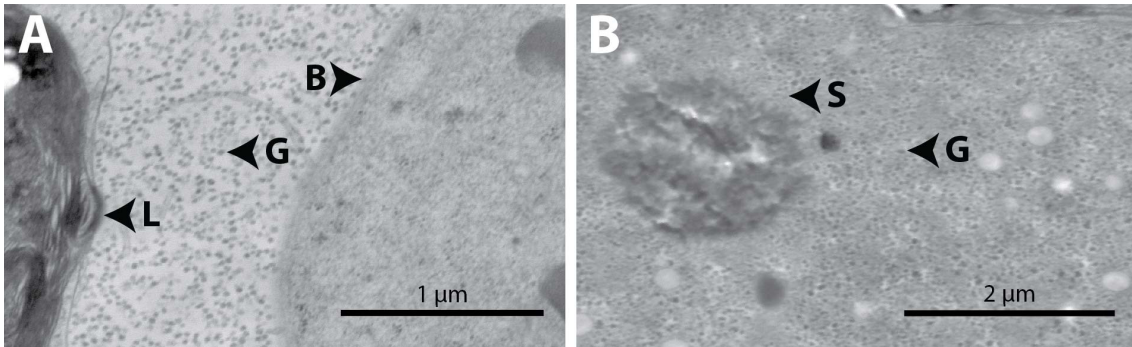
Supplementary Figure 16. Glycogen storages in host and symbiont. *A*, Host tissue and ***B*,** Symbiont cell (B) with glycogen storages (G) next to a lysosome (L). Additionally, large storage inclusions (S) are present in the symbiont. Panel **A** is a magnified area from Figure 5E.
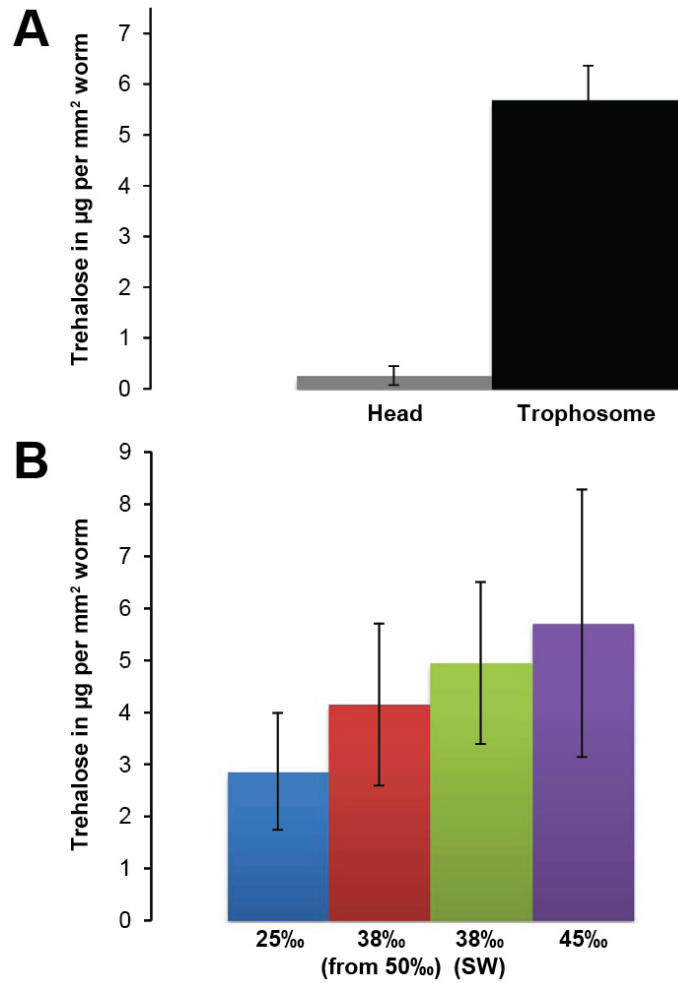
Supplementary Figure 17. Trehalose in *Paracatenula* sp. standrea under changing salinities. *A*, Separation of head (=rostrum, bacteria-free) and trophosome suggested the presence of trehalose as dominant metabolite only present in the trophosome region. n = 3. Error bars are shown as SD. *B*, Trehalose concentration of *Paracatenula* specimens after 5 h in seawater with different salinities. No significant differences between treatments could be detected. n = 5. Error bars are shown as SD.
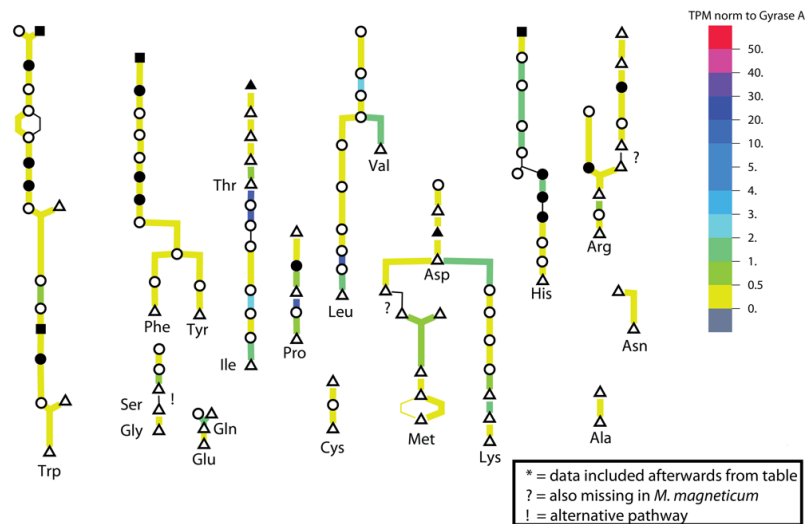
Supplementary Figure 18. Amino acid biosynthesis transcription in *Ca.* R. standrea. TPMs were normalized to the housekeeping gene *gyrA*. Black lines indicate potentially missing enzymes but a comparison with *M. magneticum* and *E. coli* indicated that in most cases these steps were only specific to *E. coli*. *** =** indicates that genes were not assigned based on Pathway Tools but identified using a manual search. **! =** indicates alternative pathways (i.e. for serine synthesis it is the non-phosphorylated pathway using serine hydroxymethyltransferase).
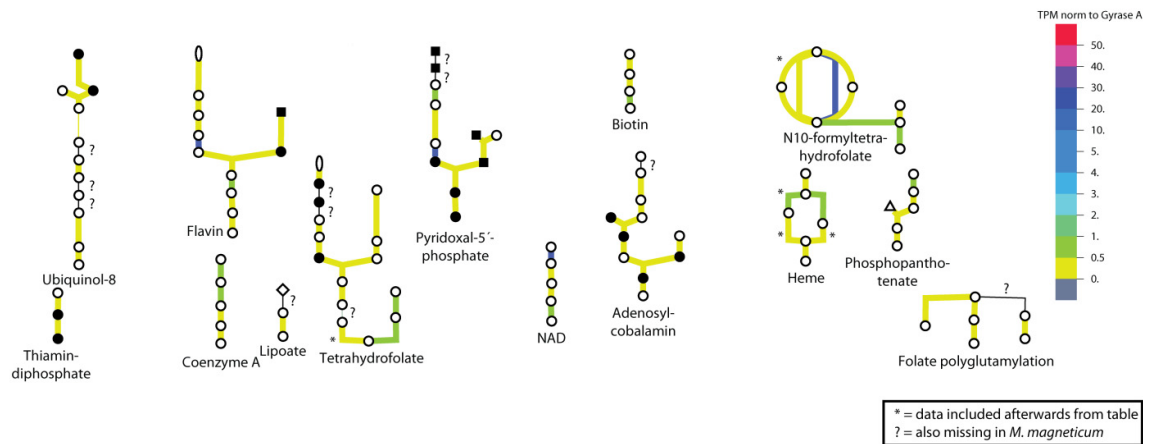
Supplementary Figure 19. Cofactor and vitamin biosynthesis transcription in *Ca.* R. standrea.
TPMs were normalized to the housekeeping gene *gyrA*. Black lines indicate potentially missing enzymes.
? = the comparison with *M. magneticum* and *E. coli* indicated that in most cases these steps were only specific to *E. coli*. *** =** indicates that genes were not assigned based on Pathway Tools but identified using a manual search.

107
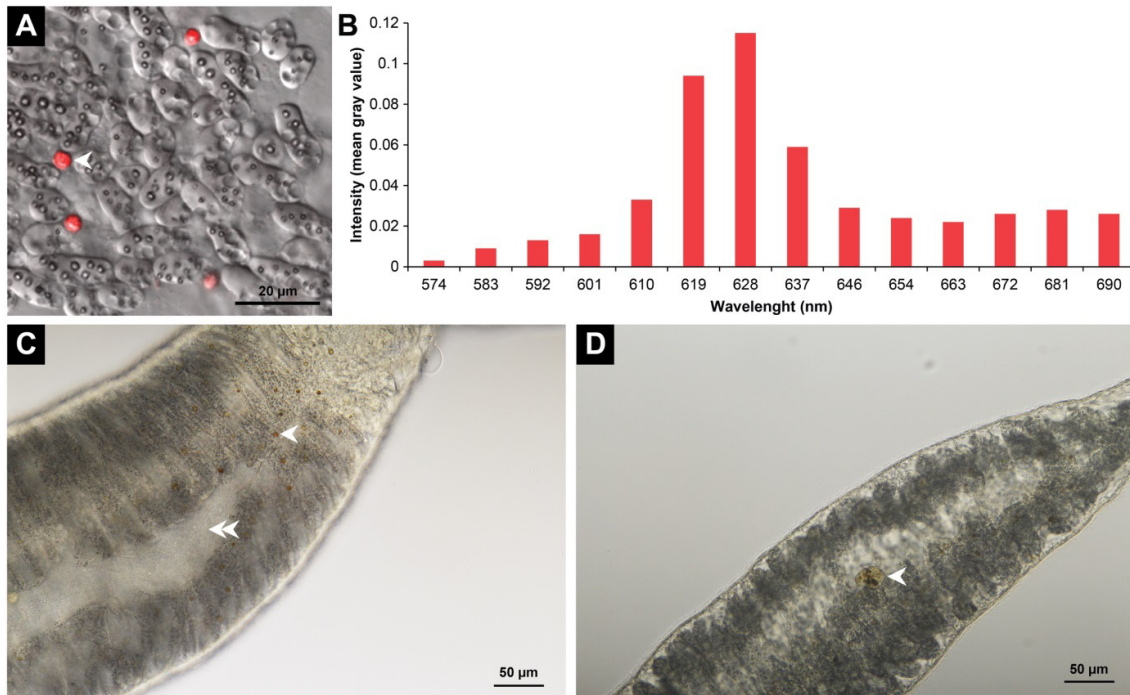
Supplementary Figure 20. Morphology of *Paracatenula* sp. standrea with characteristic inclusions in bacteriocytes. *A*, Confocal laser-scanning microscopy (Zeiss LSM 780 confocal laser microscope) showed the overlay of autofluorescence (excitation: 561 nm, emission. 570 nm–695 nm) with the respective DIC image. Inclusions (arrow) were not detected in symbionts but their surrounding tissue, most likely within bacteriocytes. *B*, Bar plot showing results of the lambda spectra. *C*, Microscopic image of the trophosome region with its orange inclusions (arrow) and the typical midline (double arrow) and *D*, a structure that resembled an egg. Images were taken with a Jenamed microscope (Carl Zeiss AG) with air objectives and a mounted Canon EOS 700D camera.

Bibliography Chapter II

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

2. Cavanaugh CM, Gardiner SL, Jones ML, Jannasch HW, Waterbury JB (1981) Prokaryotic cells in the hydrothermal vent tube wom *Riftia pachyptila* Jones: Possible chemoautotrophic symbionts. *Science* 213:340–342.

3. Cavanaugh CM (1983) Symbiotic chemoautotrophic bacteria in marine invertebrates from sulphide-rich habitats. *Nature* 302:58–61.

4. Jones ML (1981) *Riftia pachyptila* Jones: Observations on the vestimentiferan worm from the Galapagos Rift. *Science* 213:333–336.

5. Dubilier N, Bergin C, Lott C (2008) Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat Rev Microbiol* 6:725–740.

6. Seah BKB, et al. (2017) Specificity in diversity: single origin of a widespread ciliate-bacteria symbiosis. *Proc R Soc B Biol Sci* 284:doi:10.1098/rspb.2017.0764.

7. Dubilier N, et al. (2001) Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. *Nature* 411:298–302.

8. Haddad A, Camacho F, Durand P, Cary SC (1995) Phylogenetic characterization of the epibiotic bacteria associated with the hydrothermal vent polychaete *Alvinella pompejana*. *Appl Environ Microbiol* 61:1679–1687.

9. Gruber-Vodicka HR, et al. (2011) *Paracatenula*, an ancient symbiosis between thiotrophic *Alphaproteobacteria* and catenulid flatworms. *Proc Natl Acad Sci U S A*

108:12078–12083.

10. Duperron S, et al. (2005) Dual symbiosis in a *Bathymodiolus* sp. mussel from a methane seep on the Gabon Continental Margin (Southeast Atlantic): 16S rRNA phylogeny and distribution of the symbionts in gills. *Nature* 71:1694–1700.

11. Assié A, et al. (2016) A specific and widespread association between deep-sea *Bathymodiolus* mussels and a novel family of *Epsilonproteobacteria*. *Environ Microbiol Rep* 8:805–813.

12. Russell SL, Corbett-Detig RB, Cavanaugh CM (2017) Mixed transmission modes and dynamic genome evolution in an obligate animal–bacterial symbiosis. *ISME J* 11:1359–1371.

13. Newton ILG, et al. (2007) The *Calyptogena magnifica* chemoautotrophic symbiont genome. *Science* 315:998–1000.

14. Kuwahara H, et al. (2007) Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*. *Curr Biol* 17:881–886.

15. Petersen JM, et al. (2016) Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation. *Nat Microbiol* 2:16195.

16. Nussbaumer AD, Fisher CR, Bright M (2006) Horizontal endosymbiont transmission in hydrothermal vent tubeworms. *Nature* 441:345–348.

17. McCutcheon JP, Moran NA (2012) Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 10:13–26.

18. Kuwahara H, et al. (2008) Reductive genome evolution in chemoautotrophic intracellular symbionts of deep-sea *Calyptogena* clams. *Extremophiles* 12:365–374.

19. Bennett GM, Moran NA (2013) Small, smaller, smallest: The origins and evolution of ancient dual symbioses in a phloem-feeding insect. *Genome Biol Evol* 5:1675–1688.

20. Moran NA, Bennett GM (2014) The tiniest tiny genomes. *Annu Rev Microbiol* 68:195–215.

21. Anbutsu H, et al. (2017) Small genome symbiont underlies cuticle hardness in beetles. *Proc Natl Acad Sci U S A* 11440:E8382–E8391.

22. Hansen AK, Moran NA (2011) Aphid genome expression reveals host – symbiont cooperation in the production of amino acids. *Proc Natl Acad Sci U S A* 108:2849–2854.

23. 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.

24. Bright M, Sorgo A (2003) Ultrastructural reinvestigation of the trophosome in adults of *Riftia pachyptila* (Annelida, Siboglinidae). *Invertebr Biol* 122:347–368.

25. Le Pennec M, Beninger PG, Herry A (1995) Feeding and digestive adaptations of bivalve molluscs to sulphide-rich habitats. *Comp Biochem Physiol* 111A:183–189.

26. Ott J, Rieger G, Rieger R, Enderes F (1982) New mouthless interstitial worms from the sulfide system: Symbiosis with prokaryotes. *Mar Ecol* 3:313–333.

27. Dirks U, et al. (2012) Bacterial symbiosis maintenance in the asexually reproducing and regenerating flatworm *Paracatenula galateia*. *PLoS One* 7:e34709.

28. Imhoff JF, Hiraishi A, Süling J (2005) Anoxygenic Phototrophic Purple Bacteria. *Bergeys Manual of Systematic Bacteriology*, ed Brenner D. J., Krieg, Staley, J. T., Garrity GM, pp 119–132. 2nd Ed.

29. Lindstrom, E. S., Tove, S. R., Wilson PW (1950) Nitrogen fixation by the green and purple sulfur bacteria. *Science* 112:198–198.

30. Blakemore R (1975) Magnetotactic bacteria. *Science* 190:377–379.

31. Baldani I, et al. (2014) The family *Rhodospirillaceae*. *The Prokaryotes*, ed Rosenberg, E., DeLong, E. F., Stackebrandt, E., Lory, S., Thompson F (Springer, Berlin, Heidelberg), pp 533–618.

32. Childress JJ, Girguis PR (2011) The metabolic demands of endosymbiotic

chemoautotrophic metabolism on host physiological capacities. *J Exp Biol* 214:312–325.

33. Dirks U, Gruber-Vodicka HR, Leisch N, Sterrer W, Ott JA (2011) A new species of symbiotic flatworms, *Paracatenula galateia* sp. nov. (Platyhelminthes: *Catenulida*: Retronectidae) from Belize (Central America). *Mar Biol Res* 7:769–777.

34. Dirks U, Gruber-Vodicka HR, Egger B, Ott JA (2012) Proliferation pattern during rostrum regeneration of the symbiotic flatworm *Paracatenula galateia*: A pulse-chase-pulse analysis. *Cell Tissue Res* 349:517–525.

35. 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 U S A* 109:E1173–E1182.

36. Dmytrenko O, et al. (2014) The genome of the intracellular bacterium of the coastal bivalve, *Solemya velum*: A blueprint for thriving in and out of symbiosis. *BMC Genomics* 15:924.

37. Markert S, et al. (2011) Status quo in physiological proteomics of the uncultured *Riftia pachyptila* endosymbiont. *Proteomics* 11:3106–3117.

38. Klatt JM, Polerecky L (2015) Assessment of the stoichiometry and efficiency of CO_2 fixation coupled to reduced sulfur oxidation. *Front Microbiol* 6:484.

39. Badger MR, Bek EJ (2008) Multiple Rubisco forms in proteobacteria: Their functional significance in relation to CO_2 acquisition by the CBB cycle. *J Exp Bot* 59:1525–1541.

40. Winkel M, et al. (2016) Single-cell sequencing of *Thiomargarita* reveals genomic flexibility for adaptation to dynamic redox conditions. *Front Microbiol* 7:964.

41. 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.

42. Nakagawa S, et al. (2014) Allying with armored snails: the complete genome of gammaproteobacterial endosymbiont. *ISME J* 8:40–51.

43. Alber BE (2011) Biotechnological potential of the ethylmalonyl-CoA pathway. *Appl Microbiol Biotechnol* 89:17–25.

44. Beck DAC, et al. (2015) Multiphyletic origins of methylotrophy in *Alphaproteobacteria*, exemplified by comparative genomics of Lake Washington isolates. *Environ Microbiol* 17:547–554.

45. Erb TJ, et al. (2007) Synthesis of C_5-dicarboxylic acids from C_2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. *Proc Natl Acad Sci U S A* 104:10631–10636.

46. Bill N, et al. (2017) Fixation of CO_2 using the ethylmalonyl-CoA pathway in the photoheterotrophic marine bacterium *Dinoroseobacter shibae*. *Environ Microbiol* 19:2645–2660.

47. Berg IA (2011) Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways. *Appl Environ Microbiol* 77:1925–1936.

48. Moran MA, Miller WL (2007) Resourceful heterotrophs make the most of light in the coastal ocean. *Nat Rev Microbiol* 5:792–800.

49. Shih PM, Ward LM, Fischer WW (2017) Evolution of the 3-hydroxypropionate bicycle and recent transfer of anoxygenic photosynthesis into the Chloroflexi. *Proc Natl Acad Sci U S A* 114:10749–10754.

50. Woyke T, et al. (2006) Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* 443:950–955.

51. Zimmermann J, et al. (2016) Closely coupled evolutionary history of ecto- and endosymbionts from two distantly related animal phyla. *Mol Ecol* 25:3203–3223.

52. Kleiner M, 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.

53. Kleiner M, et al. (2015) Use of carbon monoxide and hydrogen by a bacteria-animal symbiosis from seagrass sediments. *Environ Microbiol*. doi:10.1111/1462-2920.12912.

54. Kleiner M (2012) Metabolism and evolutionary ecology of chemosynthetic symbionts from marine invertebrates. Dissertation (Universität Bremen).

55. Burow LC, Mabbett AN, Blackall LL (2008) Anaerobic glyoxylate cycle activity during simultaneous utilization of glycogen and acetate in uncultured *Accumulibacter* enriched in enhanced biological phosphorus removal communities. *ISME J* 2:1040–1051.

56. Wilkinson JF (1959) The problem of energy-storage compounds in bacteria. *Exp Cell Res* 7:111–130.

57. Ott JA, et al. (1991) Tackling the sulfide gradient: A novel strategy involving marine nematodes and chemoautotrophic ectosymbionts. *Mar Ecol* 12:261–279.

58. Giere O, Conway NM, Gastrock G, Schmidt C (1991) "Regulation" of gutless annelid ecology by endosymbiotic bacteria. *Mar Ecol Prog Ser* 68:287–299.

59. Nobu MK, Tamaki H, Kubota K, Liu W-T (2014) Metagenomic characterization of "*Candidatus* Defluviicoccus tetraformis strain TFO71", a tetrad-forming organism, predominant in an anaerobic-aerobic membrane bioreactor with deteriorated biological phosphorus removal. *Environ Microbiol* 16:2739–2751.

60. Martín HG, et al. (2006) Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nat Biotechnol* 24:1263–1269.

61. Bryant C (1991) Metazoan life without oxygen. *Chapman & Hall, London*.

62. Sorgo A, Gaill F, Lechaire J-P, Arndt C, Bright M (2002) Glycogen storage in the *Riftia pachyptila* trophosome: contribution of host and symbionts. *Mar Ecol Prog Ser* 231:115–120.

63. Ponnudurai R, et al. (2017) Metabolic and physiological interdependencies in the *Bathymodiolus azoricus* symbiosis. *ISME J* 11:463–477.

64. Zhang R, et al. (2011) Mechanistic analysis of trehalose synthase from *Mycobacterium smegmatis*. *J Biol Chem* 286:35601–35609.

65. Green ER, Mecsas J (2015) Bacterial Secretion Systems – An overview. *Am Soc Microbiol* 4:1–32.

66. Fiala-Médioni A, Michalski J-C, Jollès J, Alonso C, Montreuil J (1994) Lysosomic and lysozyme activities in the gill of bivalves from deem hydrothermal vents. *C R Acad Sci Paris, Sci la vie/Life Sci* 317:239–244.

67. 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.

68. 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.

69. Goupil LS, et al. (2016) Cysteine and aspartyl proteases contribute to protein digestion in the gut of freshwater planaria. *PLoS Negl Trop Dis* 10:e0004893.

70. Nakabachi A, et al. (2005) Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, *Buchnera*. *Proc Natl Acad Sci U S A* 102:5477–5482.

71. Lynch JB, Alegado RA (2017) Spheres of hope, packets of doom: The good and bad of outer membrane vesicles in interspecies and ecological dynamics. *J Bacteriol* 199:e00012-17.

72. Roier S, et al. (2016) A novel mechanism for the biogenesis of outer membrane vesicles in Gram-negative bacteria. *Nat Commun* 7:doi: 10.1038/ncomms10515.

73. Ekiert DC, et al. (2017) Architectures of lipid transport systems for the bacterial outer membrane. *Cell* 169:273–285.

74. Abellón-Ruiz J, et al. (2017) Structural basis for maintenance of bacterial outer membrane lipid asymmetry. *Nat Microbiol* 2:1616–1623.

75. Montanaro J, Gruber D, Leisch N (2016) Improved ultrastructure of marine invertebrates using non-toxic buffers. *PeerJ* 4:e1860.

76. McDonald KL, Webb RI (2011) Freeze substitution in 3 hours or less. *J Microsc* 243:227–233.

77. McDonald KL (2014) Rapid embedding methods into epoxy and LR White resins for morphological and immunological analysis of cryofixed biological specimens. *Microsc Microanal* 20:152–163.

78. Blazejak A, Erséus C, Amann R, Dubilier N (2005) Coexistence of bacterial sulfide oxidizers, sulfate reducers, and spirochetes in a gutless worm (Oligochaeta) from the Peru margin. *Appl Environ Microbiol* 71:1553–1561.

79. Daims H, Brühl A, Amann R, Schleifer K-H, Wagner M (1999) The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: Development and evaluation of a more comprehensive probe set. *Syst Appl Microbiol* 22:434–444.

80. Amann RI, et al. (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* 56:1919–1925.

81. Bankevich A, et al. (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477.

82. Wick RR, Schultz MB, Zobel J, Holt KE (2015) Bandage: Interactive visualization of *de novo* genome assemblies. *Bioinformatics* 31:3350–3352.

83. Aziz RK, et al. (2008) The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics* 9:doi:10.1186/1471-2164-9-75.

84. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.

85. Huerta-Cepas J, et al. (2016) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* 44:D286–D293.

86. Karp PD, Paley S, Romero P (2002) The pathway tools software. *Bioinformatics* 18:S225–S232.

87. Saier MH, Tran C V., Barabote RD (2006) TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res* 34:D181–D186.

88. Yelton AP, et al. (2016) Global genetic capacity for mixotrophy in marine picocyanobacteria. *ISME J* 10:2946–2957.

89. Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34:525–527.

90. Rocha DJP, Santos CS, Pacheco LGC (2015) Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis. *Antonie Van Leeuwenhoek* 108:685–693.

91. Kopylova E, Noé L, Touzet H (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28:3211–3217.

92. Grabherr MG, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–654.

93. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2005) GenBank. *Nucleic Acids Res* 33:D34–D38.

94. Huson DH, et al. (2016) MEGAN Community Edition - Interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput Biol* 12:e1004957.

95. Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584.

96.	Haas BJ, et al. (2013) *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8:1494–1512.

97.	Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.

98.	Wiśniewski JR, Zougman A, Nagaraj N, Mann M (2009) Universal sample preparation method for proteome analysis. *Nat Methods* 6:359–362.

99.	Kleiner M, et al. (2017) Assessing species biomass contributions in microbial communities *via* metaproteomics. *Nat Commun* 8:1558.

100.	Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.

101.	Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055.

102.	Price MN, Dehal PS, Arkin AP (2010) FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490.

103.	Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44:W242–W245.

104.	Min XJ, Butler G, Storms R, Tsang A (2005) OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Res* 33:W677–W680.

105.	Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 33:511–518.

106.	Kearse M, et al. (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649.

107.	Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321.

108.	Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* 25:1307–1320.

109.	Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589.

110.	Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274.

111.	Anisimova M, Gascuel O (2012) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* 55:539–552.

112.	Seah BKB, et al. (2018) Sulfur-oxidizing symbionts without canonical genes for autotrophic CO_2 fixation. *Manuscript in preparation*.

113.	Widdel F, Bak F (1992) Gram-negative mesophilic sulfate-reducing bacteria. *The Prokaryotes*, ed A. Balows, H. G. Truper, M. Dworkin, W. Harder K-HS (Springer, New York), pp 3352–3378. IV.

114.	Liebeke M, Bundy JG (2012) Tissue disruption and extraction methods for metabolic profiling of an invertebrate sentinel species. *Metabolomics* 8:819–830.

115.	Vizcaíno JA, et al. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* 44:D447–D456.

116.	Sterrer W, Rieger R (1974) Retronectidae-a new cosmopolitan marine family of Catenulida (Turbellaria). *Biology of the Turbellaria*, ed Riser N MM (eds). (McGraw-Hill, New York), pp 63–92.

117.	Leisch N, et al. (2011) Microanatomy of the trophosome region of *Paracatenula* cf. *polyhymnia* (*Catenulida*, *Platyhelminthes*) and its intracellular symbionts.

Zoomorphology 130:261–271.

118. Scott KM, Cavanaugh CM (2007) CO_2 uptake and fixation by endosymbiotic chemoautotrophs from the bivalve *Solemya velum*. *Appl Environ Microbiol* 73:1174–1179.

119. Dahl C, Friedrich C, Kletzin A (2008) Sulfur oxidation in prokaryotes. *Encycopedia Life Sci*:doi:10.1002/9780470015902.a0021155.

120. Friedrich CG, Bardischewsky F, Rother D, Quentmeier A, Fischer J (2005) Prokaryotic sulfur oxidation. *Curr Opin Microbiol* 8:253–259.

121. Geelhoed JS, et al. (2009) Microbial sulfide oxidation in the oxic-anoxic transition zone of freshwater sediment: involvement of lithoautotrophic *Magnetospirillum* strain J10. *FEMS Microbiol Ecol* 70:54–65.

122. Bamford VA, et al. (2002) Structural basis for the oxidation of thiosulfate by a sulfur cycle enzyme. *EMBO J* 21:5599–5610.

123. Newton ILG, Girguis PR, Cavanaugh CM (2008) Comparative genomics of vesicomyid clam (Bivalvia: Mollusca) chemosynthetic symbionts. *BMC Genomics* 9:585.

124. Garcia-Horsman JA, Barquera B, Rumbley J, Ma J, Gennis RB (1994) The superfamily of heme-copper respiratory oxidases. *J Bacteriol* 176:5587–5600.

125. Preisig O, Zufferey R, Thöny-Meyer L, Appleby C, Hennecke H (1996) A high-affinity cbb3-type cytochrome oxidase terminates the symbiosis-specific respiratory chain of *Bradyrhizobium japonicum*. *J Bacteriol* 178:1532–1538.

126. Gardebrecht A, et al. (2012) Physiological homogeneity among the endosymbionts of *Riftia pachyptila* and *Tevnia jerichonana* revealed by proteogenomics. *ISME J* 6:766–776.

127. Riley M, et al. (2006) *Escherichia coli* K-12: a cooperatively developed annotation snapshot - 2005. *Nucleic Acids Res* 34:1–9.

128. Matsunaga T, et al. (2005) Complete genome sequence of the facultative anaerobic magnetotactic bacterium *Magnetospirillum* sp. strain AMB-1. *DNA Res* 12:157–166.

129. Khatri I, Singh A, Korpole S, Pinnaka AK, Subramanian S (2013) Draft genome sequence of an alphaproteobacterium, *Caenispirillum salinarum* AK4, isolated from a solar saltern. *Genome Announc* 1:e00199-12.

130. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H (2000) Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* 407:81–86.

Chapter III: Genome reduction

Clade-specific patterns of reductive genome evolution in ancient thiotrophic endosymbionts

Oliver Jäckle[1], Brandon K. B. Seah[1], Harald R. Gruber-Vodicka[1]

[1]Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, Germany

OJ and HGV conceived the study. OJ and HGV analyzed biological samples. OJ and HGV collected samples for genomic analysis. HGV prepared libraries and assembled the genomes. OJ, and HGV analyzed genomes. OJ, HGV and KBS performed analyses. OJ and HGV wrote the manuscript with contributions from KBS.

Highlights

- The *Ca.* Riegeria symbionts form a diverse clade, have reduced genomes and a low but stable genomic GC content

- The *Ca.* Riegeria symbionts share most of their metabolic functions

- Phylogenetic subclades of *Ca*. Riegeria have unique metabolic 'fingerprints' of retained or lost functions or genes e.g. involved in cell division and DNA repair

- The three clades show different modes of genome evolution that have previously been interpreted as successive stages of reductive genome evolution
  - For divergent subclades we observe either full gene loss or nucleotide-based erosion that increases pseudogenization
  - The divergent patterns for the loss of recombination and mobile genetic elements are highlighted by a conserved gene order in one clade and completely rearranged gene orders in a second clade of similar phylogenetic distance

- Despite strict vertical transmission of the symbionts, and their intracellular location in a nutritionally dependent host, one of their key metabolic genes encoding RuBisCO form I, has seen replacement by horizontal gene transfer

Significance statement

It is an open question if genomes of closely related endosymbionts follow convergent trajectories of reductive evolution. General predictions of genome reduction processes could yet not be drawn as most established models on underlining processes have been derived from studies on highly specialized insect symbioses. In this study, we compared the members of a diverse clade of vertically transmitted flatworm endosymbionts, which allowed us to look for convergent patterns in the evolution of reduced genome. We showed that each of the symbionts kept species- and lineage-specific metabolic 'fingerprints'. Besides indications for parallel and group-specific reductive genome evolution, we identified contrasting trajectories among the symbiont clade including different degrees of gene fragmentation as well as genome rearrangements. Our results on the flatworm symbioses highlight convergent and stochastic genome evolution processes that are comparable to those in insect symbioses.

Abstract

The marine flatworm *Paracatenula* lacks a mouth and gut, and gains its nutrition from chemosynthetic *Alphaproteobacteria* named *Ca.* Riegeria. These intracellular symbionts are vertically transmitted between host generations, a transmission mode which is known to lead to rampant genome reduction in symbionts. However, given the complete reliance of the flatworm on its symbionts for nutrition, there should be strong selective constraints on symbiont genome reduction.

To trace how reduced genomes evolved in *Ca.* Riegeria, we sequenced 35 symbiont genomes of 23 *Paracatenula* species from the Mediterranean and the Caribbean. The symbiont genomes were highly reduced with sizes between 1.19 to 2.04 Mb and genomic GC contents of 45.1 to 55.2%. Gene losses that affected parts of the DNA repair and cell division machinery were specific for some *Ca.* Riegeria lineages. This lineage-specific gene loss appeared to be stochastic, and ranged from nucleotide-centered erosions to full gene deletions. Genome synteny was lineage-specific and highly variable, with one *Ca.* Riegeria lineage displaying several 100-fold more rearrangements than other lineages of similar phylogenetic distance. Despite the central importance of autotrophic carbon fixation in the symbioses, one of the key metabolic genes for autotrophic carbon fixation encoding RuBisCO form I, has seen replacement through horizontal gene transfer in one *Ca.* Riegeria species. Our comparative analyses revealed that current models of genome reduction processes of vertically transmitted endosymbionts, which are largely based on insect-bacteria symbioses, do not adequately represent the stochastic nature of parallel genome evolution we observed in the *Paracatenula* – *Ca*. Riegeria symbiosis.

Introduction

Genome reduction is a universal process of gene loss, diminution of genome size, and simplifying of genetic information in eukaryotic fungi, bacterial and archaeal genomes (McCutcheon and Moran, 2012; Hauser, 2014; Nicks and Rahn-Lee, 2017). The reduction and loss of genes can occur in microbes from different habitats and ecological roles, including thermophilic archaea, free-living bacteria, insect symbionts and in chemosynthetic symbionts of marine invertebrates (Moran, 2002; Giovannoni *et al.*, 2005; Kuwahara *et al.*, 2007; Newton *et al.*, 2007; McCutcheon and Moran, 2012; Wolf and Koonin, 2013; Nicks and Rahn-Lee, 2017; Tian *et al.*, 2017).

Vertically transmitted bacteria living in obligate intracellular symbiosis with their hosts are particularly affected by extensive genome reduction that often results in tiny genomes down to not more than 0.112 Mb (McCutcheon and Moran, 2012; Bennett and Moran, 2013; Moran and Bennett, 2014; Anbutsu *et al.*, 2017). While genes for essential functions, such as information processing or biosynthesis of key metabolites are retained by the symbionts, those regions not related to the provision of the host often get inactivated or even lost in the processes of reductive genome evolution (Moran, 2002; Morris *et al.*, 2012). For example, insect symbionts that have co-evolved with their hosts for millions of years have evolved highly specialized metabolic interactions which can be as specific as the provision of a single type of amino acid (McCutcheon *et al.*, 2009; McCutcheon and Moran, 2012).

Established models for the dynamic processes of reductive genome evolution are mostly derived from studies on insect symbioses, and postulate various stages in the evolution from a free-living to an obligate intracellular lifestyle (Toft and Andersson, 2010; McCutcheon and Moran, 2012). Early stages are characterized by the acquisition of genes from the environment, rapid modification and gene duplication events (Toft and Andersson, 2010). Mobile elements can also proliferate, which causes recombination events, and the interruption and fragmentation of coding sequences (Toft and Andersson, 2010; McCutcheon and Moran, 2012). Genome reduction is initiated by single-nucleotide mutations which cause pseudogenization, followed by a reduction in length of the affected genes until their complete erosion. Genes affected by erosion are often related to DNA repair and metabolic redundancy (Moya *et al.*, 2008; Toft and Andersson, 2010; Manzano-Marín and Latorre, 2016). In later stages, mobile

elements and the ability to recombine will be lost, causing gene order to be fixed and creating a genome stasis that can persist for over millions of years (Tamas *et al.*, 2002; Manzano-Marín and Latorre, 2014, 2016). In insect symbioses, whether the symbionts finally become reduced into organelle-like organisms is thought to depend on how co-adapted, metabolically integrated, and independent the symbionts and their host are, but it is still under debate if sequence erosion will eventually come to a halt (Tamas *et al.*, 2002; Andersson, 2006; Toft and Andersson, 2010; McCutcheon and Moran, 2012).

Does this progression of stages postulated for insect symbioses also apply to other types of endosymbionts associated with different classes of animal hosts, and does it always follow the same sequence? Most comparative studies thus far have compared reduced obligate endosymbionts with closely related microbes, or based their comparisons on selected genes or categories of genes, such as those associated with DNA repair (Kuwahara *et al.*, 2011; Manzano-Marín and Latorre, 2016; Shimamura *et al.*, 2017). In this study, we compared the genomes of a diverse clade of obligate chemosynthetic symbionts called *Ca*. Riegeria. These bacteria are vertically transmitted, intracellular, and form a nutritional symbiosis with the marine flatworm genus *Paracatenula*, with whom they have been associated for 500–620 million years (Gruber-Vodicka *et al.*, 2011). A recent genomic analyses of one *Ca.* Riegeria species has revealed its high genome reduction (Jäckle *et al.*, 2018). Analysis of the *Ca*. Riegeria clade as a whole would permit us to test whether each lineage follows a similar trajectory of reductive genome evolution, or if it is not as linear as previously supposed.

Here we characterized the genome evolution in 23 *Ca.* Riegeria species covering a broad diversity of *Paracatenula* symbionts. Combining genomics, phylogenetics and functional analyses we could unravel a largely conserved metabolism with lineage-specific functions. We identified varying patterns of gene losses unique for some of the *Ca.* Riegeria lineages. Clade-specific genome rearrangements indicated a dynamic genome structure despite the ancient association. We also observed stabilization processes with early losses of mobile genetic elements only in some clades. Our analyses revealed that current models of genome evolution of vertically transmitted endosymbionts, which are largely based on insect-bacteria symbioses, do not adequately represent the stochastic nature of the processes observed in the *Paracatenula* – *Ca*. Riegeria symbiosis.
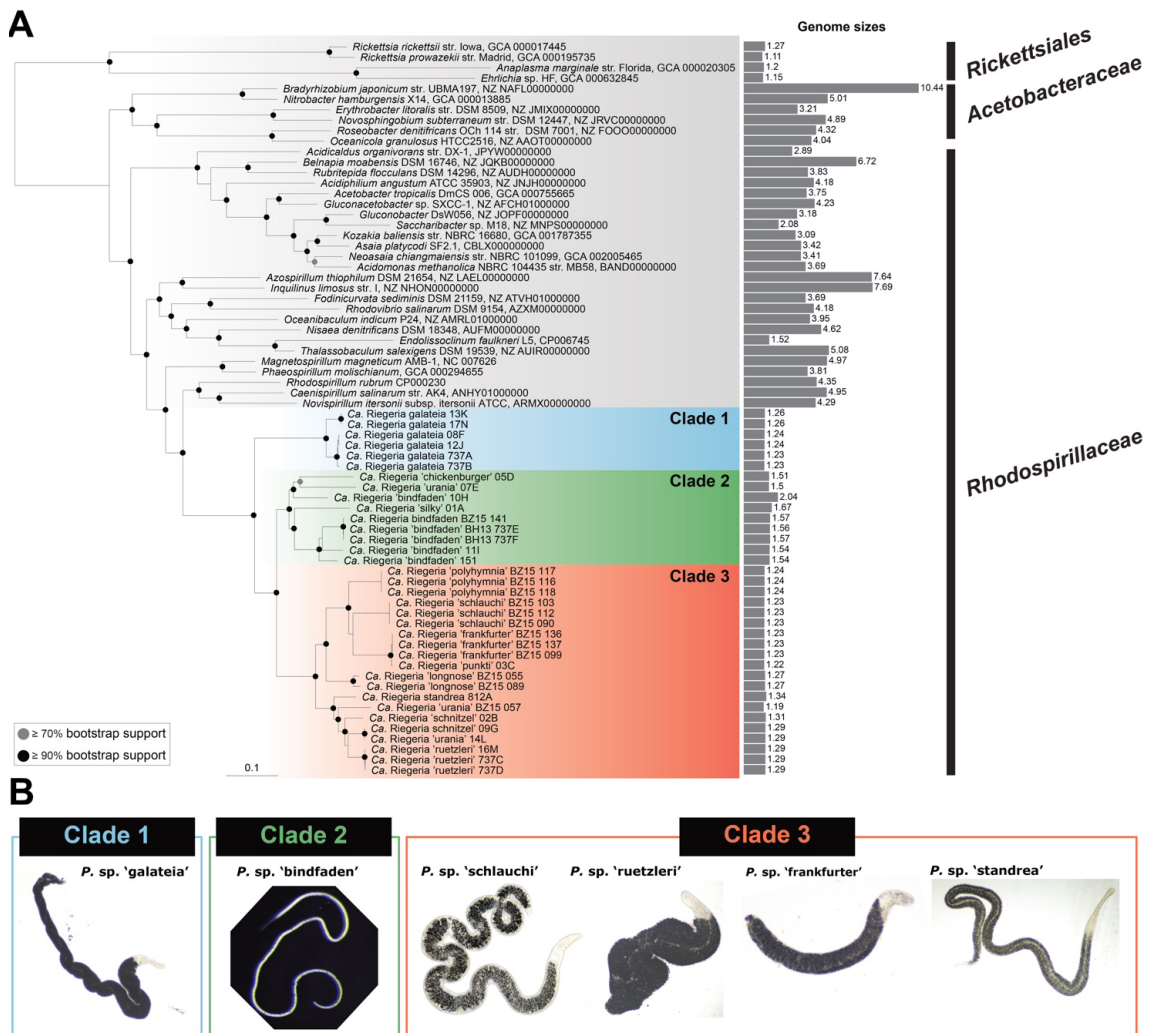
Results and Discussion

Ca. Riegeria symbionts form a phylogenetically diverse clade with reduced genomes and low but stable genomic GC content

We collected 35 single *Paracatenula* specimens from three localities (Belize, Bahamas, Italy) around the world, and assembled the genomes of each their *Ca.* Riegeria symbionts from shotgun metagenomic sequences. The sampled diversity represented 23 species and covered the known diversity of *Paracatenula* hosts (Gruber-Vodicka *et al.*, 2011). The symbiont genome assemblies ranged in size between 1.19–2.04 Mb on 1–249 contigs with coding densities between 69.57–88.23% (mean 82.92% ± 6.17 standard deviation) (Supplementary Table 1). Two genomes could be assembled into closed, circular contigs; estimates of their completeness based on conserved marker genes were 89.64% and 91.89%, indicating genome reduction rather than incomplete genome assemblies. All *Ca*. Riegeria symbionts clustered into a single, well-supported monophyletic group within *Rhodospirillaceae* (*Alphaproteobacteria*) which concurs with previous phylogenetic analyses (Gruber-Vodicka *et al.*, 2011; Jäckle *et al.*, 2018) (Figure 1A). The *Ca.* Riegeria clade covered a large diversity with the lowest amino acid identity (AAI) in the whole clade of 69.93% and could be further divided into at least three subclades that agrees with the morphological classifications of their *Paracatenula* flatworm hosts (Figure 1B, Supplementary Data set, S1). *Ca.* Riegeria symbionts of clade 1 had a lowest AAI of 92.94%, whereas clade 2 and clade 3 were more divergent with lowest within-clade AAIs of 78.36% and 79.07%.

The genomes of *Ca.* Riegeria were both smaller (1.35 Mb ± 0.29) and had lower average genomic GC (gGC) content (45.07–55.19%, mean 50.64% ± 1.69) than their closest relatives in the family *Rhodospirillaceae*. For example, relatives such as *Magnetospirillum magneticum* and *Rhodospirillum rubrum* have genome sizes of up to 4.97 Mb and gGC ranging from 62–69% (Gruber-Vodicka *et al.*, 2011; Jäckle *et al.*, 2018). Vertically transmitted endosymbionts often have lower gGC than their relatives, which has been explained by genetic drift due to population bottlenecks (Moran, 1996; Moran *et al.*, 2009; Roeselers *et al.*, 2010; Gruber-Vodicka *et al.*, 2011). The nucleotide content of genomes is not uniform, with the nucleotide usage in the third codon position being relatively unconstrained because of the redundancy in the genetic code. Non-coding regions of a genome are also under a lower selective pressure

and the comparison between gGC, non-coding regions GC and third codon position can reveal underlying trajectories of gGC evolution masked by gGC alone (Muto and Osawa, 1987). The high GC content (83.23% ± 5.95) in the third codon position in the free-living *Rhodospirillaceae* suggests their genomes are under strong selection towards high gGC (Supplementary Data set, S2). In contrast, the GC contents of non-coding regions (47.26% ± 2.52) and the third codon position (49.22% ± 3.93) in *Ca*. Riegeria were similar to their gGC of 50.64% ± 1.69. This indicates a stationary equilibrium of the *Ca.* Riegeria gGC contents at approximately 50% and suggests an absence of the selective pressure observed in other *Rhodospirillaceae*. The stationary gGC of *Ca.* Riegeria contradicts the mutational bias towards high AT typically observed in insect symbionts (McCutcheon and Moran, 2012).

Figure 1. Phylogenomic tree of the *Ca.* Riegeria clade. *A*, The phylogenetic tree is based on an alignment of 43 conserved genes, calculated using FastTree and rooted with *Rickettsiales.* The placement of the three *Ca.* Riegeria clades was highlighted by a blue, green and orange box, and alphaproteobacterial relatives by a gray box. Genome sizes based on CheckM estimations were included using iTOL. *B*, Micrographs of representative host individuals of each clade. Images are not to scale.

The *Ca.* Riegeria symbionts share most of their metabolic functions and differ in ecologically relevant functions

A large proportion of orthologous genes and metabolic pathways were shared by the diverse *Ca.* Riegeria symbionts suggesting that they retained a broad core metabolism (Figure 2, Supplementary Figure 1, 2). The core genome of all 35 *Ca.* Riegeria symbionts represented 55.3% ± 4.6 of the total gene repertoire of each symbiont and consisted of 721–735 genes. The majority of conserved core genes included those for basic cellular functions such as translation, energy, amino acid and cofactor metabolism – physiological functions that are usually retained in primary endosymbionts with reduced genomes (Wernegreen, 2017) (Supplementary Figure 2). The metabolic similarity of *Ca.* Riegeria symbionts was corroborated by how their genomes clustered together in a non-metric multidimensional scaling (NMDS) analysis. The NMDS was based on the distribution of annotated genes in COG categories, and compared *Ca.* Riegeria with free-living bacteria, parasitic and non-parasitic symbionts, as well as other chemosynthetic symbionts (n = 95, Figure 2B). This suggests functional conservation among *Ca.* Riegeria symbionts, which may be because they need to provide their host's nutrition (Jäckle *et al.*, 2018). Three *Ca.* Riegeria genomes from clade 2 diverged from the main *Ca.* Riegeria cluster, which could be explained by their relative enrichments in COG categories such as T (signal transduction mechanisms) for two *Ca.* Riegeria genomes, and L (replication, recombination and repair) for one genome, compared to other *Ca*. Riegeria. These categories are not directly related to biomass production, but might point to either different habitat preferences, niche adaptations and nutritional needs by the host, or the symbionts being in different stages of reductive genome evolution.

The pan genome of *Ca*. Riegeria consisted of 7235 genes and each subclade or lineage within *Ca*. Riegeria had its own clade-specific genetic features (Figure 2A, C). Furthermore, most of the genes (5636 genes, 78%) were not annotated with a known function, and likely included pseudogenes. Genes or groups of genes that were specific to certain lineages encoded for example RuBisCO form II in clade 1, urea degradation-related genes in both clades 1 and 2 and trehalose synthesis genes in clades 2 and 3 (Figure 2C).

Two different species of *Ca.* Riegeria have already been shown to encode different forms of RuBisCO, the key protein of the Calvin-Benson cycle for carbon fixation, which suggested that different species have dissimilar affinities for carbon dioxide and oxygen in the habitat (Jäckle

et al., 2018). This was thought to originate from the differential loss of one or another copy of the gene encoding RuBisCO from a common ancestor that had encoded both forms of RuBisCO (Jäckle *et al.*, 2018). Unexpectedly, the gene encoding RuBisCO form I present in clade 2 and clade 3 organisms had likely seen horizontal replacement in at least one of the studied symbionts, as its phylogeny showed a disjunction compared to the 16S rRNA genes (Supplementary Figure 3). Horizontal gene transfer (HGT) events were not expected in *Ca.* Riegeria as they have an obligate intracellular lifestyle and such events typically only occur in early stages of adaptation to the within-host environment (Toft and Andersson, 2010). Genes encoding RuBisCO have often been reported to undergo HGT in both free-living and symbiotic *Gammaproteobacteria* (Baxter *et al.*, 2002; Kleiner *et al.*, 2012)*.* We therefore hypothesize that one *Ca.* Riegeria species has taken up a second copy of a RuBisCO form I-encoding gene belonging to a different form of the protein and lost the original.
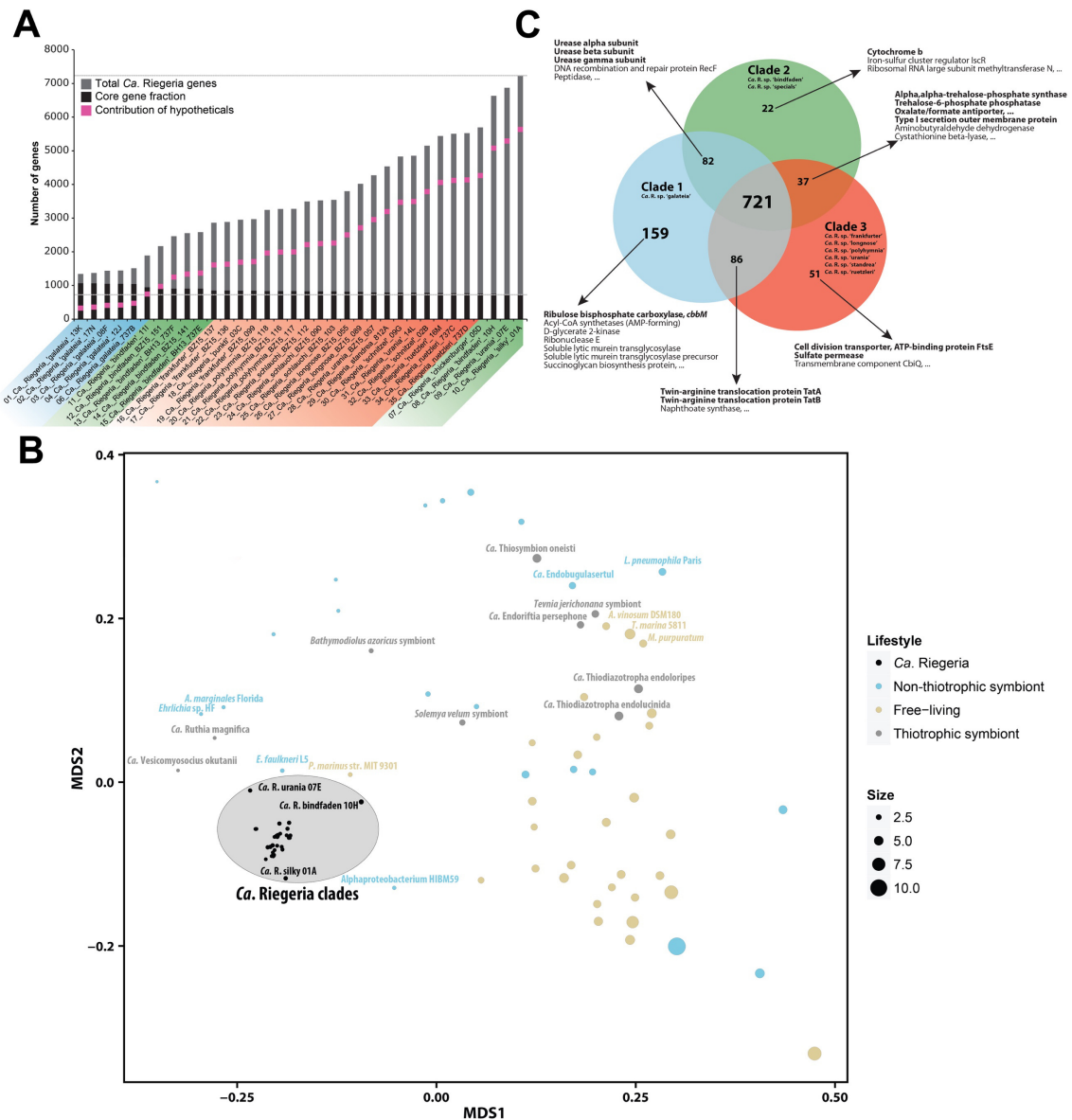
Figure 2. Conserved and unique features in *Ca*. Riegeria symbionts. *A*, *Ca.* Riegeria core genes are shown in black, summing up to a pan genome shown in gray. Core genes include those found in *Ca.* R. sp. galateia 737A and in subsequently analyzed organisms. The pink squares represent the number of hypothetical proteins contributing to the pan genome. ***B*,** The NMDS plot was generated from COG category distributions with a 2D stress of 0.1135. The diameter of circles represent the genome sizes. The *Ca.* Riegeria clade and other reduced thiotrophs are highlighted. ***C*,** Shared and unique genes of *Ca*. Riegeria clades. Genes of interest were highlighted.

Ca. Riegeria harbor lineage-specific nucleotide and gene deletions

Vertically transmitted endosymbionts can gradually lose non-essential genes, leading to highly reduced and specialized genomes (McCutcheon and Moran, 2012). We examined the pattern of gene loss across the three *Ca.* Riegeria clades to determine whether the genes lost are paralleled in each lineage. We could identify clade-specific reductions and losses of *fts* genes and DNA repair genes among the three *Ca.* Riegeria clades (Supplementary Figure 4–6, Figure

3A). Only the genes *ftsH* and *ftsZ* were conserved among the symbionts, showed no degrees of reduction and only one clade 2 symbiont appeared to lack the *ftsZ* gene. Varying degrees of gene reduction, ranging from shortened genes up to drastic gene diminutions, were observed for *ftsQ* and *ftsA* in clade 2 and clade 3 whereas these genes were retained in clade 1 organisms (Figure 3B). Other symbiotic bacteria with reduced genomes have gene diminutions in some *fts* genes, e.g. *ftsK* and *ftsN*, but the encoded proteins were shown to be still functional (Manzano-Marín and Latorre, 2016). Thus, we could not exclude that gene remnants encoding essential protein domains were still present and the enzymes might still be functional.

We identified four DNA repair genes (*unG*, *ntH*, *recA*, *ligA*) that were conserved in all *Ca.* Riegeria symbionts (Supplementary Figure 5, Supplementary Data set, S3). Genes such as *ogT*, *recF*, *recO* that are involved in direct damage reversal and DNA recombination were conserved in at least one clade. The *Ca.* Riegeria symbionts encoded fewer DNA repair genes (8–20, 11.08 ± 2.41) compared to other thiotrophic symbionts such as clam symbionts (11–23, 16.50 ± 4.50), but more than insect symbionts (0–12, 5.75 ± 4.92) (Moran *et al.*, 2008; McCutcheon and Moran, 2012). The set of genes conserved in *Ca.* Riegeria may explain the stationary gGC of *Ca.* Riegeria symbionts, as the loss of DNA repair genes often results in DNA damage such as cytosine deamination and guanine oxidation, which bias a genome towards a lower GC content (Lind and Andersson, 2008; McCutcheon and Moran, 2012). The largest number of DNA repair genes were found in a clade 2 symbiont with the largest known genome among *Ca*. Riegeria with 2.04 Mb, indicative for a less streamlined state (Supplementary Figure 5, 6). The conservation of specific DNA-repair gene repertoires in *Ca.* Riegeria is a usual feature of insect symbioses and suggests convergent processes of genome reduction (Moran *et al.*, 2008; McCutcheon and Moran, 2012).

While the loss of DNA repair genes is common in endosymbionts undergoing genome reduction, the fragmentation and degradation of cell division-related *fts* genes over a diverse clade of symbionts was unknown and unexpected (McCutcheon and Moran, 2012). Although *fts* genes encode proteins that play key roles in bacterial cell division, several bacteria, both free-living and symbiotic, are known to lack some elements such as *ftsZ* and hence must use alternative molecular mechanisms for division (Sayavedra, 2016). In contrast, the most basally branching clade of *Ca*. Riegeria (clade 1) has an intact set of *fts* genes whereas the other clades show

gradual erosion and loss (Figure 3). This suggests that in the latter two clades the genes lost have no essential function for cell division processes or the symbionts encode an alternative mechanism for cell division than clade 1. Lineage-specific absence of *fts* genes is supporting evidence for ongoing progressive and parallel genome reduction in the *Ca.* Riegeria clade.

The presence of distinct stages of ongoing reductive genome evolution in the *Ca.* Riegeria clades is further supported by investigating gene length distributions as indicators for pseudogenes. We identified a dichotomy between organisms of the three different *Ca.* Riegeria clades based on their abundances of short-length genes. Gene length histograms of clade 2 and clade 3 symbionts showed higher abundances of short-length genes (100–200 bp) compared to clade 1 symbionts (Figure 3C). The majority of these genes were annotated as hypothetical proteins and likely represent gene remnants as a result of pseudogenization, and also inflate the total gene count for the *Ca*. Riegeria pan genome (Figure 1A). The accumulation of remnants in these two clades could lead to the assumption that these symbionts were in an early stages of reductive genome reduction. Molecular clock estimates, in combination with their stable gGC contents contradict this assumption, as they suggested that this symbiotic setting is instead ancient and was established hundreds of million years ago (Gruber-Vodicka *et al.*, 2011) (Supplementary Data set, S4). Slow but ongoing processes of reductive genome evolution might be an explanation for the presence of gene fragments in the *Ca.* Riegeria symbionts, or drastic reductions even came to a halt – a scenario that is still under debate to occur in insect symbionts (Andersson, 2006).
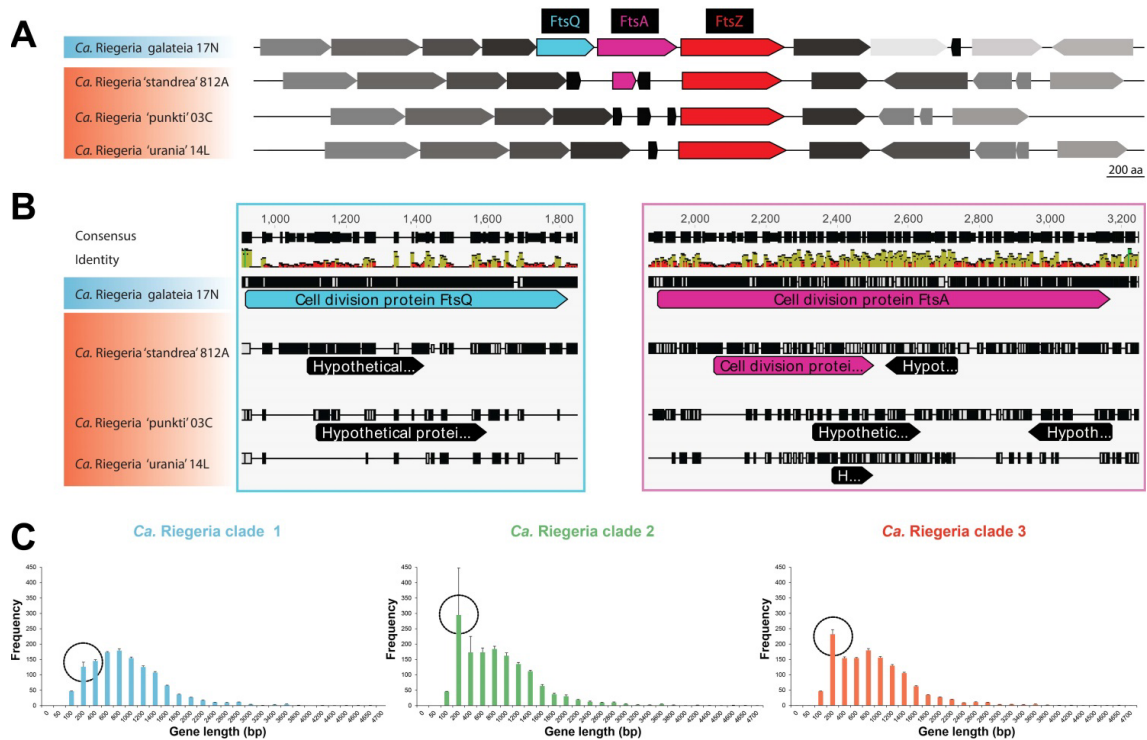
Figure 3. Patterns of genome reduction in *Ca.* Riegeria symbionts. *A*, Schematic representation of the *fts* gene cluster in four *Ca*. Riegeria symbionts showing different degrees of gene reductions. Homologous genes are indicated by their colorations. ***B*,** Zoom-in into the nucleotide alignment of the *ftsQAZ* gene cluster. Gaps indicate nucleotide deletions. Blue and pink squares indicate zoom-ins of the alignment of the *ftsQ* (left) and the *ftsA* gene (right) and illustrate gaps in the *Ca*. Riegeria genetic code. The alignment was done with MUSCLE using a maximum number of iterations of eight. ***C*,** Gene length distributions based on ORF-predictions in the *Ca.* Riegeria clades indicate fragmentations in derived *Ca.* Riegeria clades.

Subclades show varying modes of genome evolution that resemble successive stages of reductive genome evolution in insect symbionts

Generally, bacterial symbionts that live in an obligate intracellular symbiosis and are vertically transmitted show high degrees of genome stability and only few examples are known that do not follow this general trend (Sloan and Moran, 2013). Our analysis of genome architectures within *Ca.* Riegeria allowed us to compare ongoing genome dynamics between clades and to investigate if this general prediction holds true. The degree of genome rearrangements was estimated by counting the number of breakpoints in pairwise genome alignments of organisms from the three symbiont clades.

The genomes of *Ca.* Riegeria from clade 1 and clade 3 were highly syntenic within the respective clades, with few indications of recent genome rearrangements (Figure 4). The genome sequences of clade 1 and clade 3 symbionts had fixed gene orders with 1–13 breakpoints (at in average 95.29% ± 3.58% AAI, and lowest number of breakpoints at

91.96%) and 0–16 breakpoints (at in average 82.54% ± 5.86% AAI, and lowest number of breakpoints at e.g. 78.06%). In contrast to this, genomes within clade 2 had similar average AAI (83.24% ± 6.48%) but less conserved genome architectures, having on average 201 ± 133 breakpoints between each pair of genomes (Figure 4A–D).

High degrees of genome instability were identified when comparing genomes between the three *Ca.* Riegeria clades. Comparing clade 1 and clade 2 organisms resulted in an average AAI of 71.3% and 921 breakpoints, clade 2 and clade 3 in 494 breakpoints at an average AAI of 73.4% and clade 1 and clade 3 in 916 breakpoints at an AAI of 69.5%.

The breakpoint distances in clade 2 are phylogenetically correlated, as the tree based on breakpoint distances had a similar branching pattern as the molecular phylogeny based on conserved marker genes (Figure 4E). This illustrated that within clade 2 we still find ongoing and recent genome rearrangements which are absent within clade 1 and clade 3 organisms, indicating that these organisms have lost genetic mobile elements.

Genome syntenies that vary from being conserved to highly rearranged stand in contrast to the relatively stable genome structures in pairwise comparisons of vertically transmitted symbionts with comparable genetic distances such as *Buchnera*, *Portiera* and chemosynthetic clam symbionts (Supplementary Figure 7) (Tamas *et al.*, 2002; Sloan and Moran, 2013). The architectural conservations in two *Ca.* Riegeria clades was surprising as they all encoded the recombination protein RecA. The lack of the RecA protein is assumed to be key for preserving genome syntenies, although its function was recently suggested for reconsideration what agrees with our findings (Kuwahara *et al.*, 2008; Shimamura *et al.*, 2017).

Genome rearrangements are often a consequence of mobile genetic elements movements such as transposases in early stages of reductive genome (Toft and Andersson, 2010; McCutcheon and Moran, 2012). An alternative to mobile elements is the accumulation of restriction-modification systems that are known to be associated with genome rearrangements (Zheng *et al.*, 2015). We found both, a remnant of a transposase as well as two remnants of type I restriction-modification systems in the largest genome of a clade 2 representative with a genome size of 2.04 Mb. All other *Ca.* Riegeria genomes were lacking these gene remnants, and thus, this *Ca.* Riegeria species is likely in an earlier stage of reductive genome evolution compared to clade 1 and clade 3 organisms. Additional evidence for an earlier stage was given by the presence of five flagellar genes (*motB*, *flbT*, *flaA*, *flip*, *flil*) in this species of which two

were also shared with other clade 2 organisms. These genes showed signatures of a alphaproteobacterial phylogeny and might be indications for a previous free-living state before the establishment of a symbiosis (Supplementary Data set, S5). Not surprisingly, clade 2 organisms had the lowest coding densities (mean: 72.88% ± 2.10) compared to clade 1 (mean: 87.73% ± 0.44) and clade 3 (mean: 86.00% ± 1.31) symbionts. And exactly the clade 2 symbiont with the largest genome had also the lowest coding density of 69.57%, that is lower compared to the average bacterial coding density of 85–90% (Kuo *et al.*, 2009). This is indicative for clade 2 organisms being in an earlier stage of genome reduction compared to the other clades (Kuo *et al.*, 2009).
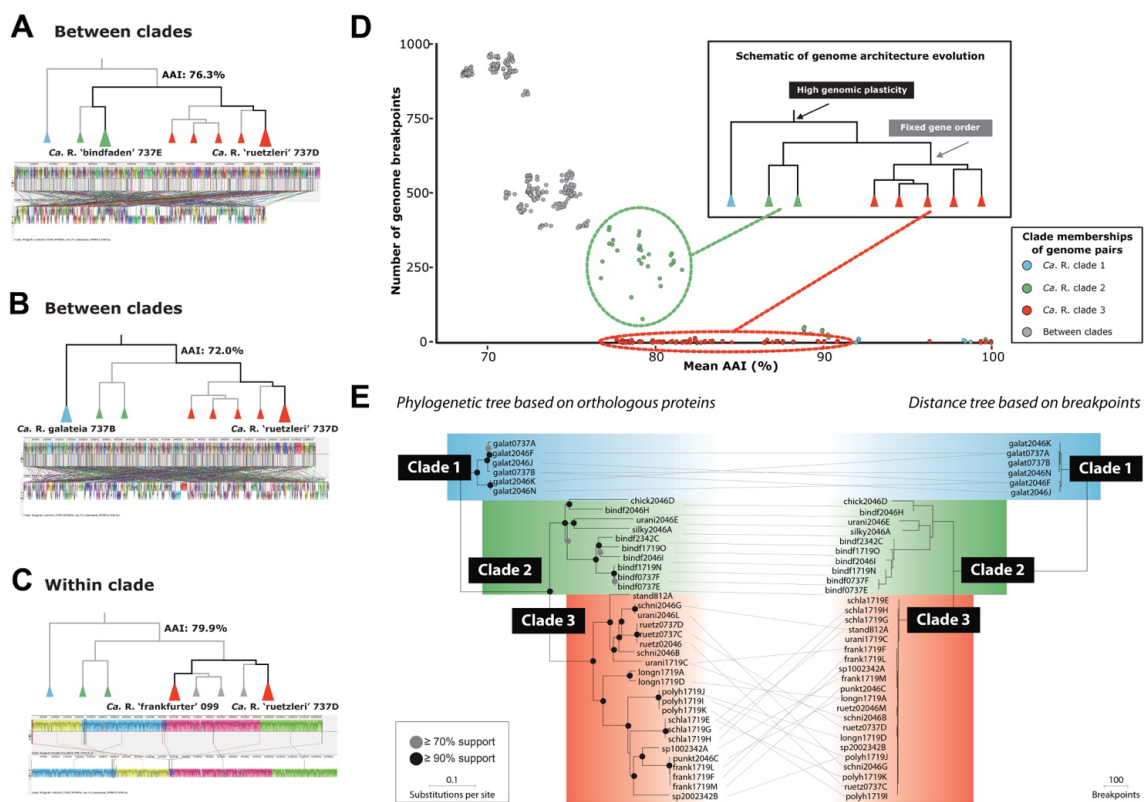
Figure 4. Pairwise whole-genome alignments of *Ca*. Riegeria symbionts. A schematic 16S rRNA gene tree is shown above each pairwise Mauve alignment to highlight the representatives compared. Each colored block represents likely homologous segments. Homologous segments are connected by colored lines. Pairwise comparisons were done with ***A***, *Ca*. R. 'ruetzleri' (clade 3) and 'bindfaden' (clade 2) (identity: 76.3%), ***B***, *Ca*. R. 'ruetzleri' (clade 3) and galatea (clade 1) (identity: 72.0%). ***C***, *Ca*. R. 'ruetzleri' (clade 3) and 'frankfurter' (clade 3) (identity: 79.9%). Identity values are based on amino acid identities (AAI). ***D***, Quantification of synteny breakpoints vs. amino acid identity values. ***E***, Genome syntenies represent phylogenetic distances of *Ca.* Riegeria symbionts. The left side shows a phylogenetic tree that is based on 32 orthologous genes and the right tree is based on a distance matrix from the breakpoint analysis.

For the first time we could provide evidence for the presence of different modes of genome evolution within a single monophyletic group of symbionts (Figure 5). We additionally identified

parallel gene losses among the three clades which agrees with previous studies on insect symbioses (Williams and Wernegreen, 2015; Boscaro *et al.*, 2017; Klasson, 2017; Kinjo *et al.*, 2018). Our study on a diverse symbiont clade allowed to track 'steps along the way', that are usually missing as most symbioses only transitioned once (Klasson, 2017). The different stages among the *Ca.* Riegeria clades have previously been interpreted as successive stages of genome reduction. The *Ca.* Riegeria symbiont genomes ranged from one clade affected by genome rearrangements, pseudogenes and low coding densities to clades with more streamlined genomes, suggesting that these processes are neither linear in *Ca.* Riegeria symbionts, nor directly linked to the age of a symbiotic clade (Supplementary Data set, S4).
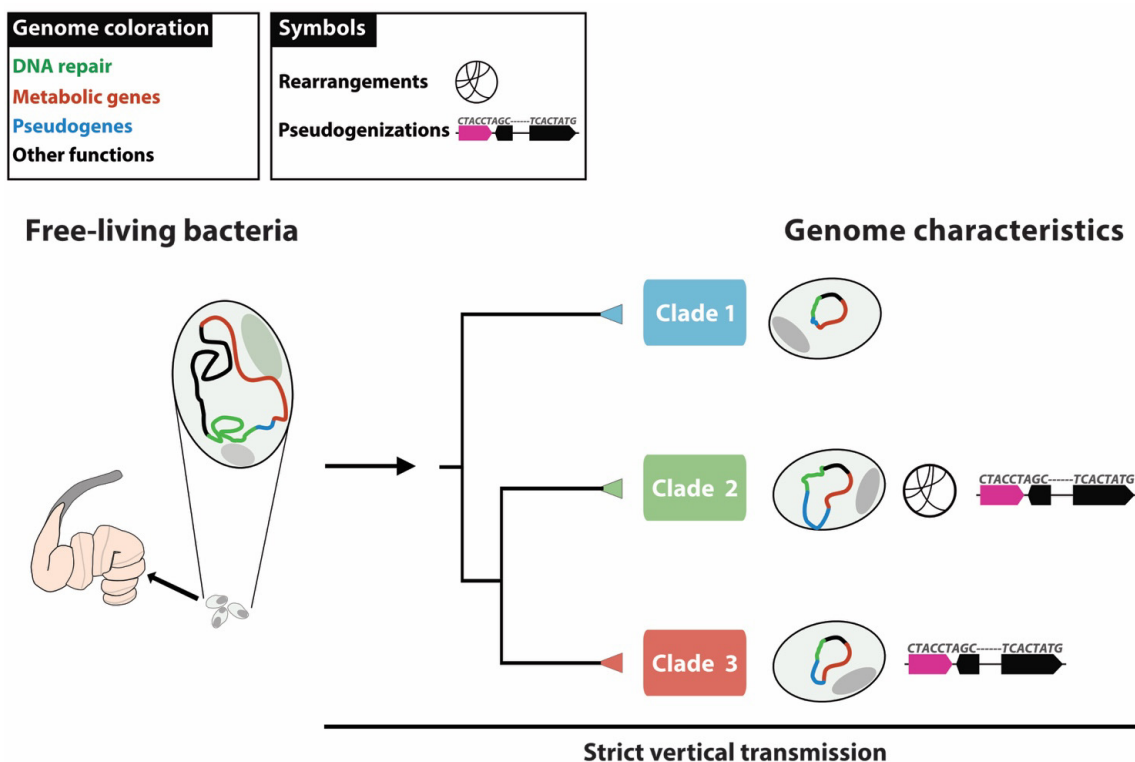
Figure 5. Model of the reductive stages in the three *Ca.* Riegeria clades. Likewise as in insect symbionts, a free-living bacterium that was incorporated is affected by dynamic genome changes. Once taken up as intracellular symbiont, these changes result in losses of DNA repair genes, metabolic genes, the accumulation of pseudogenes and genome rearrangements. Although the *Ca.* Riegeria symbionts from the three clades are ancient and have evolved from a last common ancestor, they show different modes and stages of genome reduction.

Conclusion

Although general mechanisms of convergent reductive genome evolution in insect symbioses are well understood it is still an open question if these processes are predictable or random and if they would apply to symbioses hosted by other animals. One of the challenges is that

such symbiotic systems, with symbionts that are vertically transmitted and have reduced genomes, often are ancient and phylogenetically divergent from their last common ancestor. Additionally, these symbioses were often established once with few known exceptions resulting in no apparent "traces" left to finally result in reduced genomes (Boscaro *et al.*, 2017). A similar symbiotic setup that allows studying such questions is represented by the ancient symbiosis between the flatworm *Paracatenula* and its vertically transmitted symbionts *Ca.* Riegeria.

With our comparative analysis of the diverse *Ca.* Riegeria symbionts we could show that each symbiont left behind species- or even lineage-specific 'footprints' among three distinct *Ca.* Riegeria clades. Besides indications for parallel and group-specific reductive genome evolution we identified atypical traces among the *Ca.* Riegeria clades. These differences included gene-length fragmentations as well as ancient genome rearrangements. We provided evidence for three different stages of genome reduction among the *Ca.* Riegeria symbionts that most likely originated from a shared last common ancestor. These results suggest that the *Ca.* Riegeria symbiont clades undergo reductive genome evolution under different selective pressures and different outcomes. Our results on the flatworm symbioses highlight convergent and comparably stochastic genome reduction processes as suggested for insect symbioses.

Materials and Methods

Sample collection

Paracatenula specimens were collected between 2013 and 2015 from sediments in the bay off Sant´Andrea, Elba, Italy, from Carrie Bow Cay, Belize, and from the Bahamas. Specimens were extracted by decanting the sediment collected by divers at 6–7 m water depth in Elba, and from sediment in 1–3 m water depths in Carrie Bow Cay and the Bahamas. Individual specimens were picked manually using glass pipettes and stored in glass vials filled with seawater and natural sediment that was rinsed with fresh- and seawater to remove other meiofauna. *Paracatenula* specimens for nucleic acid extractions were fixed in RNAlater (Ambion) and stored at 4 °C.

Light microscopy

Live *Paracatenula* specimens were transferred onto glass slides and squeezed by applying pressure to the cover slips. Microscopy was performed using transmitted light with differential interference contrast (DIC) on a modified Zeiss Jenamed microscope (BW-Optik).

DNA extraction and genome sequencing

DNA from single *Paracatenula* specimens was extracted using the DNeasy Blood and Tissue Micro Kit (Qiagen) following the manufacturer´s instructions as explained by Jäckle and colleagues (Jäckle *et al.*, 2018). Between 5 and 20 million 250 bp paired-end reads were generated with an insert size of 300–600 bp. Library preparations and sequencing on the Illumina HiSeq 2500 or 3000 were performed at the Max Planck Genome Centre in Cologne, Germany (http://mpgc.mpipz.mpg.de/home/).

Metagenome assembly and binning

Prior to assembly, adapters and low-quality reads were removed with bbduk (part of BBmap v36.2, https://sourceforge.net/projects/bbmap/) using a minimum quality value of two and a minimum length of 36. Single reads were excluded from further analyses. To remove the host genomic reads, reads were sorted based on a kmer frequency analysis using bbnorm (BBmap). Symbiont genome kmer peaks were identified by plotting the kmer frequency spectra as well as using the bbnorm peak prediction. Only reads with average kmer frequencies of more than $^1/_3$ of the start of the symbiont´s kmer peak were kept using the bbnorm function outhigh. The assembly was performed with SPADes v3.8 with kmer lengths 21, 33, 55, 77, 99 and 127 (Bankevich *et al.*, 2012). Binning was performed by collecting all contigs or scaffolds linked to the contig that contained the symbiont 16S rRNA sequence using the FASTG output files of the SPAdes assembly and the FASTG 'fishing' script available with phyloflash v3 (https://github.com/HRGV/phyloFlash).

Annotation and metabolic comparisons

Assembly sizes, completeness and contamination were estimated using 225 marker genes conserved across *Alphaproteobacteria* using CheckM v1.09 (Parks *et al.*, 2015). 35 *Ca*. Riegeria genome assemblies were annotated by RAST v2 (Aziz *et al.*, 2008) along with the genomes used for comparisons (Supplementary Table 1). Annotations of specific genes discussed in

detail were verified manually using NCBI PSI-BLAST (Altschul *et al.*, 1997) against the NCBI nr database.

Functional annotations and the presence of cell division and DNA repair genes among the different bacterial genomes were manually checked. Gene length patterns in the *Ca.* Riegeria symbionts were studied by exporting the predicted gene lengths from the RAST annotation pipeline (Overbeek *et al.*, 2014) to Geneious v11 (Kearse *et al.*, 2012), transferring into tabular format and a representation as histograms. *fts* gene sequences were aligned with Muscle v3.8.425 (Edgar, 2004) to identify conserved vs. lost gene regions.

Shared and unique orthologous genes of the three defined core genomes of *Ca.* Riegeria clades were analyzed using reciprocal best hits calculated with the EDGAR software (Blom *et al.*, 2009). The core genomes of each clade were defined using reciprocal best BLAST hits with one clade representative set as reference (clade 1: *Ca.* R. galateia 737A; clade 2: *Ca.* R. 'bindfaden' 737E; clade 3: *Ca.* R. 'standrea' 812A) – unique genes were added to the respective pan genomes. The relative proportions of genes contributing to the coregenome was estimated from the total number of genes encoded in each *Ca.* Riegeria organism. Sequences of the core genomes were extracted and transformed into FASTA format and annotated by RAST v2. The Venn diagram was calculated based on reciprocal BLAST hits, manually curated and genes of interest were highlighted.

The core- vs. pan genome analysis was conducted using fractional pan genome calculations implemented in EDGAR. *Ca.* R. galateia 737A was set as reference and further *Ca.* Riegeria genomes were added to the analysis in a stepwise process.

To compare Clusters of Orthologous Genes (COGs) category composition between genomes, protein sequences were assigned to COGs using EggNog (Huerta-Cepas *et al.*, 2016). The numbers of sequences per COG category were tabulated and expressed as percentages of COG assignments per genome to account for varying genome sizes, as described in Jäckle *et al*. (Jäckle *et al.*, 2018). The NMDS analysis was performed using the metaMDS function from the R package 'vegan' (cran.r-project.org/package=vegan) with Bray-Curtis distances, and was visualized with ggplot2.

Phylogenomic analysis

The *Ca.* Riegeria symbiont phylogeny was calculated from a concatenated alignment of 43 conserved marker genes as implemented in the CheckM tree pipeline (Parks *et al.*, 2015). In

brief: publically available genomes of representatives from different alphaproteobacterial groups (*Rhodospirillales*, *Sphingomonadales*, *Rhodobacteracaea*, *Rhizobiales* and *Rickettsiales*) were downloaded from the NCBI database. The amino acid sequences of marker genes of each genome were aligned with HMMER v3.1b2 (Eddy, 2011). The maximum likelihood phylogenomic tree was estimated with FastTree v2.1 (Price *et al.*, 2010), using the JTT substitution model with 20 per-site rate categories and SH-like support values. The annotated tree was visualized with iTOL v3.5.4 (Letunic and Bork, 2016) and edited with Adobe Illustrator CS5.

Phylogenetic distances and gGC analysis

The average amino acid identities (AAI) of *Ca.* Riegeria and free-living relatives were calculated with either the AAI calculator (http://enve-omics.ce.gatech.edu/aai/) or with the tool implemented in the EDGAR software (https://edgar.computational.bio.uni-giessen.de/cgi-bin/edgar_login.cgi). The gGC content of full genomes, non-coding regions, coding regions, and coding regions by codon position were estimated by predicting rRNAs and coding regions using Barrnap (http://www.vicbioinformatics.com/software.barrnap.shtml) and prodigal (Hyatt *et al.*, 2010), and by generating codon usage tables with the EMBOSS tool cusp (Rice *et al.*, 2000) and custom scripts.

Genome rearrangements

To visualize genome synteny for selected genomes, genomes were aligned in pairs with the mauveAligner algorithm (Mauve version snapshot 13 February 2015). The degree of genome synteny and rearrangements for 38 *Ca*. Riegeria genomes (including additional species of clade 2 and 3) was quantified by counting the number of breakpoints between conserved segments in all pairs of genomes. All 703 possible pairs were aligned by their amino-acid translations using the Promer tool from Mummer v3.07, with standard settings (Kurtz *et al.*, 2004). The results were filtered with the command delta-filter to retain only bidirectional best hits (options –q and –r). Coordinates were reported with show-coords, removing alignments that overlap other better-scoring alignments (option -k), in the btab format (option -B). Breakpoints were counted and plotted with scripts available at https://github.com/kbseah/mummer-breakpoints. The definition of breakpoints by Sankoff & Blanchette (Sankoff and Blanchette, 1997) was used, except that breakpoints falling at contig boundaries were regarded as

ambiguous and were not counted, to account for genome assembly fragmentation. The data matrix of breakpoint counts between each genome pair was used as a distance matrix to calculate a neighbor-joining distance tree with BIONJ (Gascuel, 1997). The phylogeny was calculated from a set of 32 orthologous proteins containing conserved markers used by the CheckM pipeline, and which were present in a single copy in each genome. Each protein sequence was aligned with Muscle, and a phylogeny was calculated from the concatenated alignment using RAxML v8.1.3 (Stamatakis, 2014) with CAT-based likelihood and WAG substitution matrix, 10 randomized maximum-parsimony starting trees, and SH-like support values. The breakpoint distance tree was drawn as a tanglegram vs. the phylogenetic tree based on orthologous conserved proteins, using Dendroscope (Huson and Scornavacca, 2012).

Supplementary Table and Figures

The Supplementary Data set is available on the CD-ROM provided with the thesis.

**Supplementary Table 1. Genomes used for comparative analysis.**

| Organism	Size (Mb)	gGC (%)	Contigs	Complete (%)	Contamination (%)
Ca. R. galateia 13K	1.26	50.4	108	94.93	0.00
Ca. R. galateia 17N	1.26	50.4	102	94.93	0.00
Ca. R. galateia 08F	1.24	51.2	139	94.93	0.00
Ca. R. galateia 12J	1.24	51.2	106	94.93	0.00
Ca. R. galateia 737A	1.23	51.3	13	94.93	0.00
Ca. R. galateia 737B	1.23	51.1	12	94.93	0.00
Ca. R. 'chickenburger' 05D	1.51	45.1	18	95.95	0.34
Ca. R. 'urania' 07E	1.50	50.3	39	92.57	0.34
Ca. R. 'bindfaden' 10H	2.04	55.2	249	98.31	0.00
Ca. R. 'silky' 01A	1.67	50.5	33	92.23	0.00
Ca. R. 'bindfaden' 141	1.57	50.8	80	91.89	0.34
Ca. R. 'bindfaden' 737E	1.56	50.8	15	91.89	0.34
Ca. R. 'bindfaden' 737F	1.57	50.8	25	91.89	0.34
Ca. R. 'bindfaden' 11I	1.54	51.2	63	94.59	0.34
Ca. R. 'bindfaden' 151	1.54	50.7	84	92.91	0.34
Ca. R. 'polyhymnia' 117	1.24	50.1	12	90.52	0.34
Ca. R. 'polyhymnia' 116	1.24	50.1	12	90.52	0.34
Ca. R. 'polyhymnia' 118	1.24	50.1	12	90.52	0.34
Ca. R. 'schlauchi' 103	1.23	49.0	3	89.64	0.34
Ca. R. 'schlauchi' 112	1.23	49.0	1	89.64	0.34
Ca. R. 'schlauchi' 090	1.23	49.0	3	89.64	0.34
Ca. R. 'frankfurter' 136	1.23	48.5	9	89.86	0.34
Ca. R. 'frankfurter' 137	1.23	48.5	9	89.86	0.34
Ca. R. 'frankfurter' 099	1.23	48.5	9	89.86	0.34
Ca. R. 'punkti' 03C	1.22	48.5	6	90.54	0.34
Ca. R. 'longnose' 055	1.27	53.6	6	93.24	0.68
Ca. R. 'longnose' 089	1.27	53.3	18	91.89	0.68
Ca. R. 'standrea' 812A	1.34	51.8	1	93.24	0.34
Ca. R. 'urania' 057	1.19	50.7	12	91.22	0.34
Ca. R. 'schnitzel' 02B	1.31	51.8	12	93.92	0.34
Ca. R. 'schnitzel' 09G	1.29	51.5	12	91.22	0.34
Ca. R. 'urania' 14L	1.29	51.5	12	91.22	0.34
Ca. R. ''ruetzleri' 16M	1.29	51.9	6	91.89	0.34
Ca. R. 'ruetzleri' 737C	1.29	51.9	4	91.89	0.34
Ca. R. 'ruetzleri' 737D	1.29	51.9	1	91.89	0.34
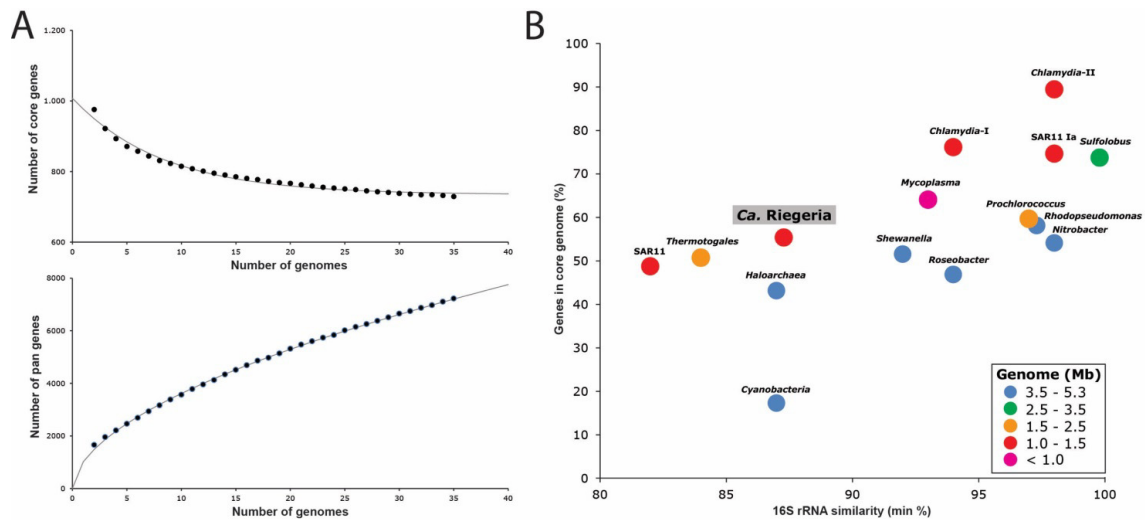
Supplementary Figure 1. Coregenome and pan genome development and genome conservation in *Ca.* Riegeria symbionts. *A,* Coregenome and pan genome development plots of 35 different *Ca.* Riegeria symbionts. ***B***, Comparison of minimal 16S rRNA gene similarity, core genome conservation and average genome sizes. The different coloration indicates the average genome sizes. Data taken and modified from Grote *et al.*, 2012.

COG categories of *Ca*. R. core

- A = RNA processing and modification
- B = Chromatin structure and dynamics
- C = Energy production and conversion
- D = Cell cycle control, cell division, chromosome partitioning
- E = Amino acid transport and metabolism
- F = Nucleotide transport and metabolism
- G = Carbohydrate transport and metabolism
- H = Coenzyme transport and metabolism
- I = Lipid transport and metabolism
- J = Translation, ribosomal structure and biogenesis
- K = Transcription
- L = Replication, recombination and repair
- M = Cell wall/membrane/envelope biogenesis
- N = Cell motility
- O = Posttranslational modification, protein turnover, chaperones
- P = Inorganic ion transport and metabolism
- Q = Secondary metabolites biosynthesis, transport and catabolism
- R = General function prediction only
- S = Function unknown
- T = Signal transduction mechanisms
- U = Intracellular trafficking, secretion, and vesicular transport
- V = Defense mechanisms
- W = Extracellular structures
- X = Mobilome: prophages, transposons
- Y = Nuclear structure
- Z = Cytoskeleton

Supplementary Figure 2. COG assignments in *Ca.* Riegeria core. Relative abundance of COGs in *Ca.* Riegeria core that consists of 735 genes that were assigned using EggNog-mapper. The list of categories was taken from ftp://ftp.ncbi.nih.gov/pub/COG/COG2014/data/fun2003-2014.tab. The relative abundances are similar to *Ca.* R. standrea (Jäckle *et al.*, 2018).
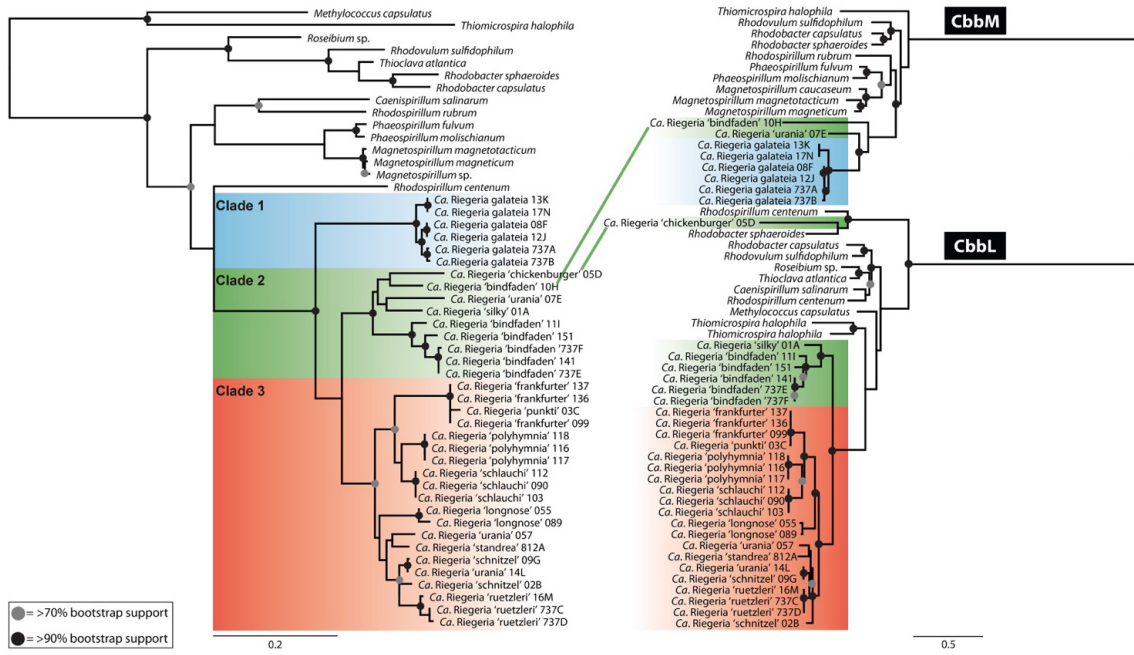
Supplementary Figure 3. Comparison of 16S rRNA and CbbM/L phylogenies shows congruency. The 16S rRNA tree shown is based on MAFFT-qinsi alignment and was calculated with PhyML (100 bootstraps). Scale bar: 20% estimated sequence divergence. The RuBisCO gene tree is based on MUSCLE nucleotide alignments. The tree shown was calculated using PhyML (100 bootstraps) and rooted based on the CbbM-clade. Scale bar: 50% estimated sequence divergence.
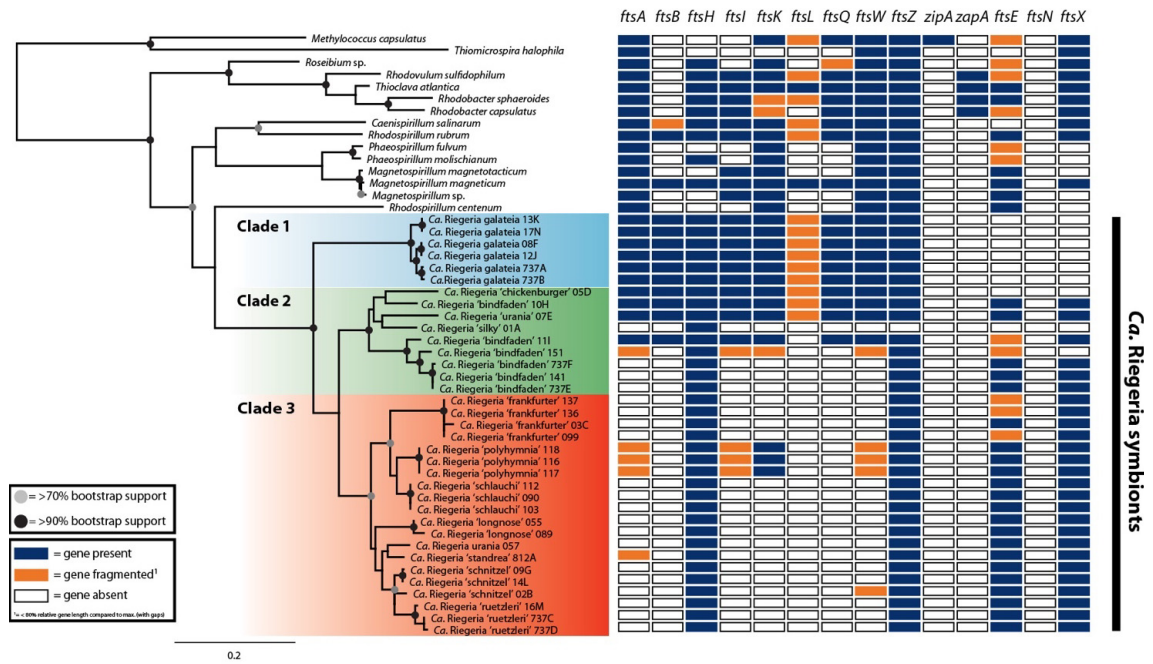
Supplementary Figure 4. *Fts*-genes in free-living *Alphaproteobacteria* and *Ca.* Riegeria symbionts. The phylogenetic tree using 16S rRNA genes is based on MAFFT nucleotide alignment incorporating predicted secondary structure information and was estimated under the HKY85 substitution model using PhyML with 100 bootstraps. Present genes are shown in blue, absent genes in white and those fragmented (below 80% of length when compared to standard length) in orange.
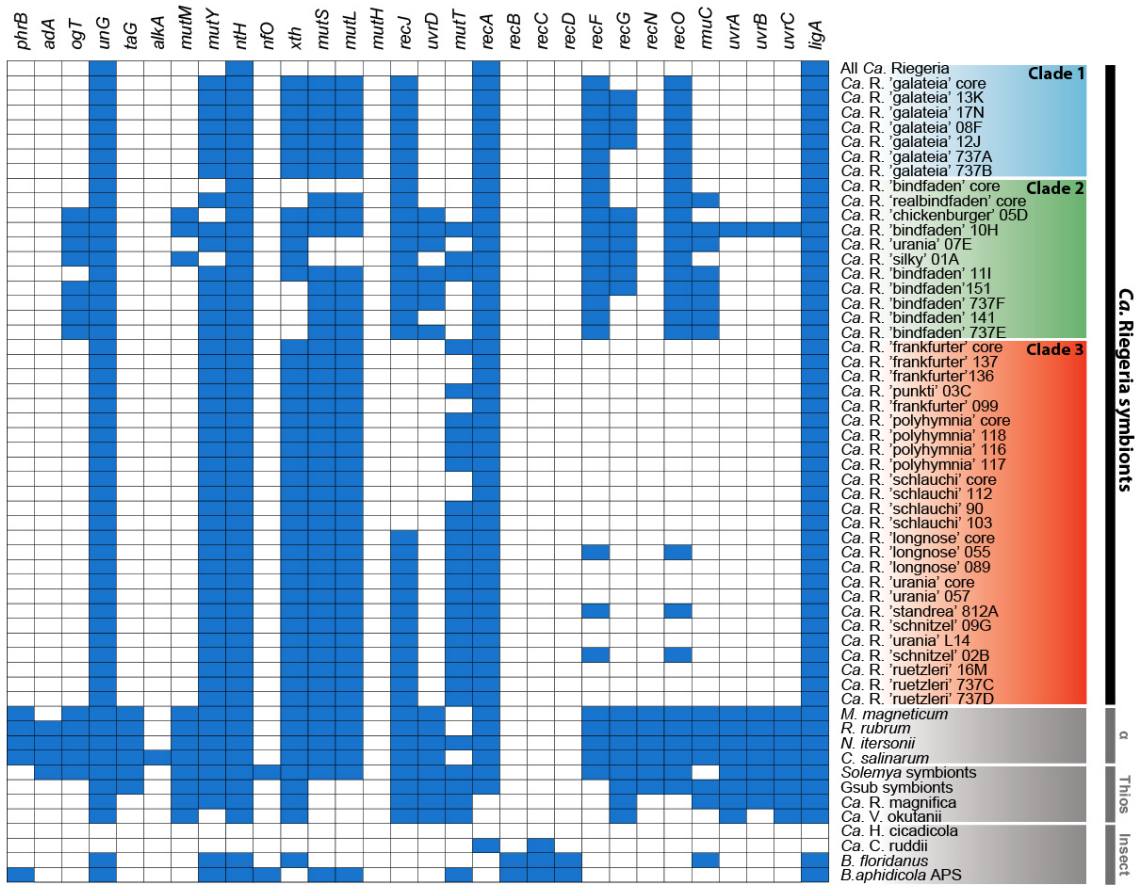
Supplementary Figure 5. DNA-repair gene reduction in *Ca.* Riegeria symbionts, free-living *Alphaproteobacteria*, reduced symbiotic thioautotrophs and highly reduced insect symbionts. Present genes were shown in blue, absent genes in white.

141
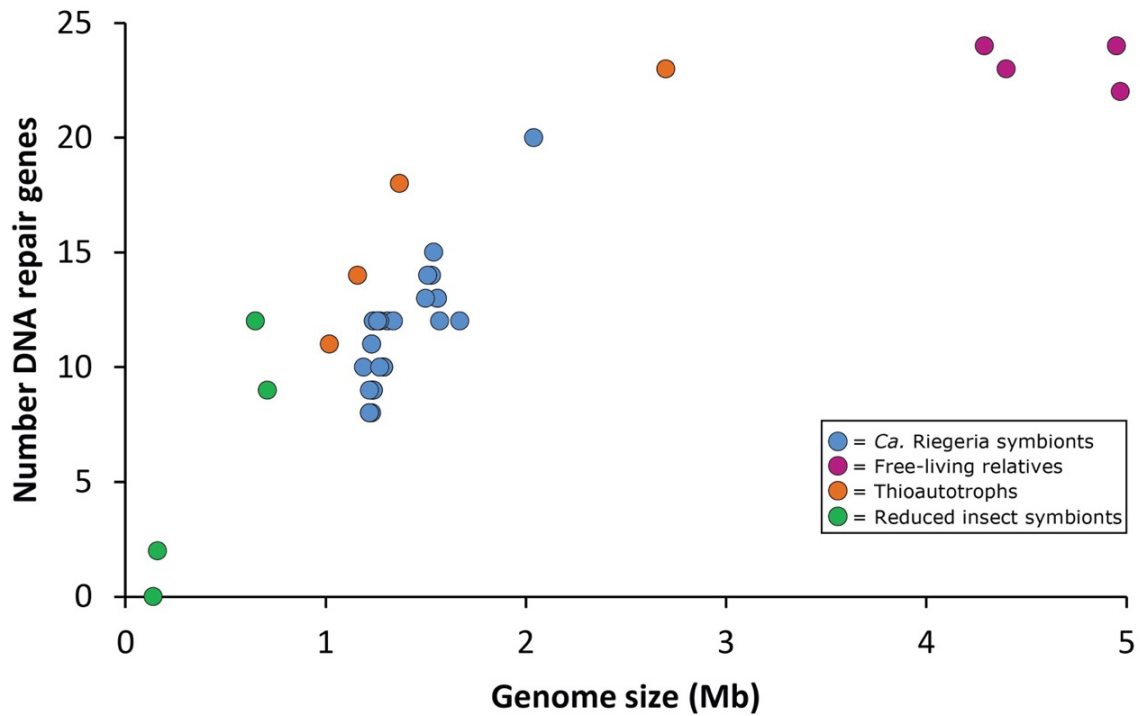
Supplementary Figure 6. Number of DNA-repair genes plotted against genome sizes. The comparison includes *Ca.* Riegeria symbionts, free-living relatives (*Caenispirillum salinarium*, *Novispirillum itersonii*, *Rhodospirillum rubrum* and *Magnetospirillum magneticum*), thioautotrophic symbionts (*Ca.* V. okutanii*, Ca.* R. magnifica, Gsub and *Solemya* symbionts) and reduced insect symbionts (*Buchnera aphidicola* APS, *Blochmannia floridanus, Ca.* Carsonella ruddii and *Ca.* Hodgkinia cicadicola).

A Whitefly symbionts, 78.5% AAI
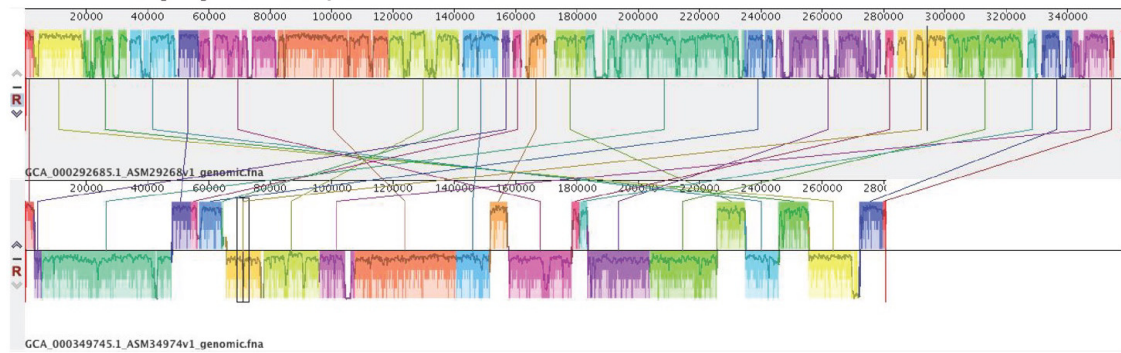
B Aphid symbionts, 73.4% AAI
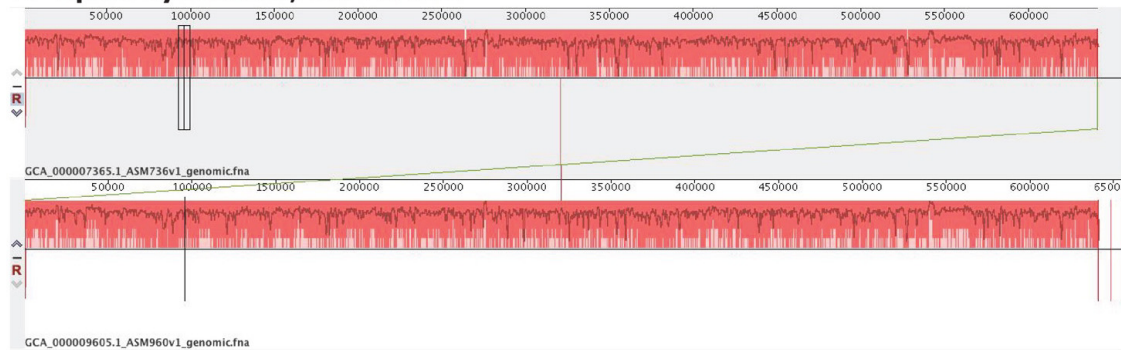
C Clam symbionts, 82.3% AAI
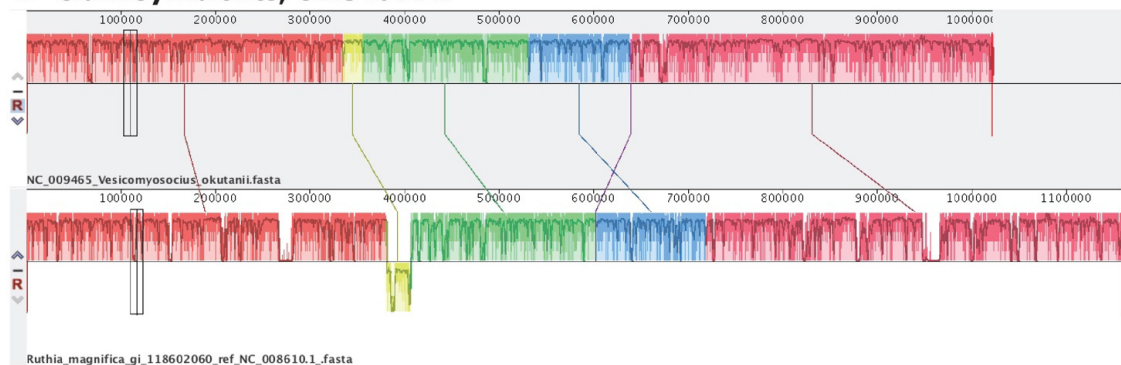
Supplementary Figure 7. Pairwise whole-genome alignments of vertically transmitted closely related symbionts. A, The whitefly symbionts *Ca.* P. aleyrodidarum BT-B and TV, **B**, The aphid symbionts *Buchnera aphidicola* SG and APS and **C**, The clam symbionts *Ca.* V. okutanii and *Ca.* R. magnifica. AAI values are indicated.

Competing interests

The authors declare no conflict of interest.

Acknowledgments

Bibliography Chapter III

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic Acids Research*, **25**, pp. 3389–3402.

Anbutsu, H., Moriyama, M., Nikoh, N., Hosokawa, T., Futahashi, R., Tanahashi, M., Meng, X.-Y., Kuriwada, T., Mori, N., Oshima, K., Hattori, M., Fujie, M., Satoh, N., Maeda, T., Shigenobu, S., Koga, R. and Fukatsu, T. (2017) 'Small genome symbiont underlies cuticle hardness in beetles', *Proceedings of the National Academy of Sciences of the United States of America*, **11440**, pp. E8382–E8391.

Andersson, S. G. E. (2006) 'The bacterial world gets smaller', *Science*, **314**, pp. 259–260.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A. and Zagnitko, O. (2008) 'The RAST Server: Rapid annotations using subsystems technology', *BMC genomics*, **9**, doi:10.1186/1471-2164-9-75.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A. and Pevzner, P. A. (2012) 'SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing', *Journal of Computational Biology*, **19**, pp. 455–477.

Baxter, N. J., Hirt, R. P., Bodrossy, L., Kovacs, K. L., Embley, M. T., Prosser, J. I. and Murrell, C. J. (2002) 'The ribulose-1,5-bisphosphate carboxylase/oxygenase gene cluster of *Methylococcus capsulatus* (Bath)', *Archives of Microbiology*, **177**, pp. 279–289.

Bennett, G. M. and Moran, N. A. (2013) 'Small, smaller, smallest: The origins and evolution of ancient dual symbioses in a phloem-feeding insect', *Genome Biology and Evolution*, **5**, pp. 1675–1688.

Blom, J., Albaum, S. P., Doppmeier, D., Pühler, A., Vorhölter, F.-J., Zakrzewski, M. and Goesmann, A. (2009) 'EDGAR: a software framework for the comparative analysis of prokaryotic genomes.', *BMC bioinformatics*, **10**, 154.

Boscaro, V., Kolisko, M., Felletti, M., Vannini, C., Lynn, D. H. and Keeling, P. J. (2017) 'Parallel genome reduction in symbionts descended from closely related free-living bacteria', *Nature Ecology & Evolution*, **1**, pp. 1160–1167.

Eddy, S. R. (2011) 'Accelerated profile HMM searches', *PLoS Computational Biology*, **7**.

Edgar, R. C. (2004) 'MUSCLE: Multiple sequence alignment with high accuracy and high throughput', *Nucleic Acids Research*, **32**, pp. 1792–1797.

Gascuel, O. (1997) 'BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data', *Molecular Biology and Evolution*, **14**, pp. 685–695.

Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., Bibbs, L., Eads, J., Richardson, T. H., Noordewier, M., Rappe, M. S., Short, J. M., Carrington, J. C. and Mathur, E. J. (2005) 'Genome streamlining in a cosmopolitan oceanic bacterium', *Science*, **309**, pp. 1242–1245.

Grote, J., Thrash, J. C. and Huggett, M. J. (2012) 'Streamlining and Core Genome Conservation among Highly Divergent Members of the SAR11 Clade', **3**, pp. 1–13.

Gruber-Vodicka, H. R., Dirks, U., Leisch, N., Baranyi, C., Stoecker, K., Bulgheresi, S., Heindl, N. R., Horn, M., Lott, C., Loy, A., Wagner, M. and Ott, J. (2011) '*Paracatenula*, an ancient symbiosis between thiotrophic *Alphaproteobacteria* and catenulid flatworms', *Proceedings of the National Academy of Sciences of the United States of America*, **108**, pp. 12078–12083.

Hauser, P. M. (2014) 'Genomic Insights into the fungal pathogens of the genus *Pneumocystis*: Obligate biotrophs of humans and other mammals', *PLoS Pathogens*, **10**, e1004425.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., Rattei, T., Mende, D. R., Sunagawa, S., Kuhn, M., Jensen, L. J., Von Mering, C. and Bork, P. (2016) 'eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences', *Nucleic Acids Research*, **44**, pp. D286–D293.

Huson, D. H. and Scornavacca, C. (2012) 'Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks', *Systematic Biology*, **61**, pp. 1061–1067.

Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W. and Hauser, L. J. (2010) 'Prodigal: prokaryotic gene recognition and translation initiation site identification.', *BMC bioinformatics*, **11**, 119.

Jäckle, O., Seah, B. K. B., Tietjen, M., Leisch, N., Liebeke, M., Kleiner, M., Berg, J. S. and Gruber-Vodicka, H. R. (2018) 'A chemosynthetic symbiont with a drastically reduced genome serves as primary energy storage in the marine flatworm *Paracatenula*', *Submitted manuscript*.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. and Drummond, A. (2012) 'Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data', *Bioinformatics*, **28**, pp. 1647–1649.

Kinjo, Y., Bourguignon, T., Tong, K. J., Kuwahara, H., Lim, S. J., Yoon, K. B., Shigenobu, S.,

Park, Y. C., Nalepa, C. A., Hongoh, Y., Ohkuma, M., Lo, N. and Tokuda, G. (2018) 'Parallel and gradual genome erosion in the *Blattabacterium* endosymbionts of *Mastotermes darwiniensis* and *Cryptocercus* wood roaches', *Genome Biology and Evolution*, **10**, pp. 1622–1630.

Klasson, L. (2017) 'The unpredictable road to reduction', *Nature Ecology and Evolution*, **1**, pp. 1062–1063.

Kleiner, M., Petersen, J. M. and Dubilier, N. (2012) 'Convergent and divergent evolution of metabolism in sulfur-oxidizing symbionts and the role of horizontal gene transfer', *Current Opinion in Microbiology*, **15**, pp. 621–631.

Kuo, C., Moran, N. A. and Ochman, H. (2009) 'The consequences of genetic drift for bacterial genome complexity', *Genome Research*, **19**, pp. 1450–1454.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C. and Salzberg, S. L. (2004) 'Versatile and open software for comparing large genomes.', *Genome Biology*, **5**, R12.

Kuwahara, H., Takaki, Y., Shimamura, S., Yoshida, T., Maeda, T., Kunieda, T. and Maruyama, T. (2011) 'Loss of genes for DNA recombination and repair in the reductive genome evolution of thioautotrophic symbionts of *Calyptogena* clams', *BMC Evolutionary Biology*, **11**, doi:10.1186/1471-2148-11-285.

Kuwahara, H., Takaki, Y., Yoshida, T., Shimamura, S., Takishita, K., Reimer, J. D., Kato, C. and Maruyama, T. (2008) 'Reductive genome evolution in chemoautotrophic intracellular symbionts of deep-sea *Calyptogena* clams', *Extremophiles*, **12**, pp. 365–374.

Kuwahara, H., Yoshida, T., Takaki, Y., Shimamura, S., Nishi, S., Harada, M., Matsuyama, K., Takishita, K., Kawato, M., Uematsu, K., Fujiwara, Y., Sato, T., Kato, C., Kitagawa, M., Kato, I. and Maruyama, T. (2007) 'Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*', *Current Biology*, **17**, pp. 881–886.

Letunic, I. and Bork, P. (2016) 'Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees', *Nucleic Acids Research*, **44**, pp. W242–W245.

Lind, P. A. and Andersson, D. I. (2008) 'Whole-genome mutational biases in bacteria', *Proceedings of the National Academy of Sciences of the United States of America*, **105**, pp. 17878–17883.

Manzano-Marín, A. and Latorre, A. (2014) 'Settling down: the genome of *Serratia symbiotica* from the aphid *Cinara tujafilina* zooms in on the process of accommodation to a cooperative intracellular life', *Genome Biology and Evolution*, **6**, pp. 1683–1698.

Manzano-Marín, A. and Latorre, A. (2016) 'Snapshots of a shrinking partner: Genome reduction in *Serratia symbiotica*', *Scientific Reports*, **6**, 32590.

McCutcheon, J. P., McDonald, B. R. and Moran, N. A. (2009) 'Convergent evolution of metabolic roles in bacterial co-symbionts of insects', *Proceedings of the National Academy of Sciences of the United States of America*, **106**, pp. 15394–15399.

McCutcheon, J. P. and Moran, N. A. (2012) 'Extreme genome reduction in symbiotic bacteria', *Nature Reviews Microbiology*, **10**, pp. 13–26.

Moran, N. A. (1996) 'Accelerated evolution and Muller's rachet in endosymbiotic bacteria', *Proceedings of the National Academy of Sciences of the United States of America*, **93**, pp. 2873–2878.

Moran, N. A. (2002) 'Microbial minimalism: Genome reduction in bacterial pathogens', *Cell*, **108**, pp. 583–586.

Moran, N. A. and Bennett, G. M. (2014) 'The tiniest tiny genomes', *Annual Review of Microbiology*, **68**, pp. 195–215.

Moran, N. A., McCutcheon, J. P. and Nakabachi, A. (2008) 'Genomics and evolution of heritable bacterial symbionts', *Annual Review of Genetics*, **42**, pp. 165–190.

Moran, N. A., McLaughlin, H. J. and Sorek, R. (2009) 'The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria', *Science*, **323**, pp. 379–382.

Morris, J. J., Lenski, R. E. and Zinser, E. R. (2012) 'The Black Queen Hypothesis: Evolution of dependencies through adaptative gene loss', *Mbio*, **3**, pp. 1–7.

Moya, A., Peretó, J., Gil, R. and Latorre, A. (2008) 'Learning how to live together: genomic insights into prokaryote–animal symbioses', *Nature Reviews Genetics*, **9**, pp. 218–229.

Muto, A. and Osawa, S. (1987) 'The guanine and cytosine content of genomic DNA and bacterial evolution', *PNAS*, **84**, pp. 166–169.

Newton, I. L. G., Woyke, T., Auchtung, T. A., Dilly, G. F., Dutton, R. J., Fisher, M. C., Fontanez, K. M., Lau, E., Stewart, F. J., Richardson, P. M., Barry, K. W., Saunders, E., Detter, J. C., Wu, D., Eisen, J. A. and Cavanaugh, C. M. (2007) 'The *Calyptogena magnifica* chemoautotrophic symbiont genome', *Science*, **315**, pp. 998–1000.

Nicks, T. and Rahn-Lee, L. (2017) 'Inside out: Archaeal ectosymbionts suggest a second model of reduced-genome evolution', *Frontiers in Microbiology*, **8**, 384.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F. and Stevens, R. (2014) 'The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST)', *Nucleic Acids Research*, **42**, pp. D206–D214.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. and Tyson, G. W. (2015) 'CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes', *Genome research*, **25**, pp. 1043–1055.

Price, M. N., Dehal, P. S. and Arkin, A. P. (2010) 'FastTree 2 - Approximately maximum-likelihood trees for large alignments', *PLoS ONE*, **5**, e9490.

Rice, P., Longden, L. and Bleasby, A. (2000) 'EMBOSS: The European Molecular Biology Open Software Suite', *Trends in Genetics*, **16**, pp. 276–277.

Roeselers, G., Newton, I. L. G., Woyke, T., Auchtung, T. A., Dilly, G. F., Dutton, R. J., Fisher, M. C., Fontanez, K. M., Lau, E., Stewart, F. J., Richardson, P. M., Barry, K. W., Saunders, E., Detter, J. C., Wu, D., Eisen, J. A. and Cavanaugh, C. M. (2010) 'Complete genome sequence of *Candidatus* Ruthia magnifica', *Standards in Genomic Sciences*, **3**, pp. 163–173.

Sankoff, D. and Blanchette, M. (1997) 'The median problem for breakpoints in comparative genomics', in *Computing and Combinatorics, Proceedings of COCOON '97, Lecture Notes in Computer Science*. Springer-Verlag New York, pp. 251–263.

Sayavedra, L. (2016) *Host-symbiont interactions and metabolism of chemosynthetic symbiosis in deep-sea Bathymodiolus mussel*, *PhD Thesis*. Universität Bremen.

Shimamura, S., Kaneko, T., Ozawa, G., Matsumoto, M. N., Koshiishi, T., Takaki, Y., Kato, C., Takai, K., Yoshida, T., Fujikura, K., Barry, J. P. and Maruyama, T. (2017) 'Loss of genes related to Nucleotide Excision Repair (NER) and implications for reductive genome evolution in symbionts of deep-sea vesicomyid clams', *Plos One*, **12**, e0171274.

Sloan, D. B. and Moran, N. A. (2013) 'The evolution of genomic instability in the obligate endosymbionts of whiteflies', *Genome Biology and Evolution*, **5**, pp. 783–793.

Stamatakis, A. (2014) 'RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies', *Bioinformatics*, **30**, pp. 1312–1313.

Tamas, I., Klasson, L., Canbäck, B., Näslund, A. K., Eriksson, A.-S., Wernegreen, J. J., Sandström, J. P., Moran, N. A. and Andersson, S. G. E. (2002) '50 Million Years of Genomic Stasis in Endosymbiotic Bacteria.', *Science (New York, N.Y.)*, **296**, pp. 2376–2379.

Tian, R.-M., Zhang, W., Cai, L., Wong, Y.-H., Ding, W. and Qian, P.-Y. (2017) 'Genome Reduction and microbe-host interactions drive adaptation of a sulfur-oxidizing bacterium associated with a cold seep sponge', *mSystems*, **2**, e00184-16.

Toft, C. and Andersson, S. G. E. (2010) 'Evolutionary microbial genomics: insights into bacterial host adaptation', *Nature Reviews Genetics*, **11**, pp. 465–475.

Wernegreen, J. J. (2017) 'In it for the long haul: evolutionary consequences of persistent endosymbiosis', *Current Opinion in Genetics and Development*. Elsevier Ltd, **47**, pp. 83–90.

Williams, L. E. and Wernegreen, J. J. (2015) 'Genome evolution in an ancient bacteria-ant symbiosis: parallel gene loss among *Blochmannia* spanning the origin of the ant tribe Camponotini', *PeerJ*, **3**, e881.

Wolf, Y. I. and Koonin, E. V. (2013) 'Genome reduction as the dominant mode of evolution', *BioEssays*, **35**, pp. 829–837.

Zheng, H., Dietrich, C., Hongoh, Y. and Brune, A. (2015) 'Restriction-modification systems as mobile genetic elements in the evolution of an intracellular symbiont', *Molecular Biology and Evolution*, **33**, pp. 721–725.

Chapter IV: Tissue regeneration

The role of endosymbionts in rostrum regeneration of the marine flatworm *Paracatenula* sp. standrea

Oliver Jäckle[1], Målin Tietjen[1], Harald R. Gruber-Vodicka[1]

[1]Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, Germany

Publication status: Manuscript draft

HGV conceived the study. OJ and HGV analyzed biological samples. HGV collected samples for transcriptomic analysis and prepared transcriptomic libraries. MT performed differential gene expression analysis. OJ analyzed the results and generated the figures and tables. OJ and HGV wrote the manuscript with contributions from MT.

Highlights

- *Paracatenula* flatworms have the ability to regenerate a rostrum within two weeks

- They harbor nutritional symbionts that provide nutrition and serve as primary energy storage

- The symbiont population changes gene expression in response to fragmentation and regeneration

- Symbiont populations in *Paracatenula* individuals regenerating a rostrum showed differential expression in 168 of 1344 genes

- Most upregulated genes were involved in cellular processes such as translation, transcription and energy conservation

- Some genes were involved in the biosynthesis of amino acids and the vitamin thiamine

- The large-scale change of gene expression suggests an active function of the symbionts in rostrum regeneration

Abstract

Most flatworm species have the ability to regenerate parts of their body within hours up to a few days. Marine flatworms of the genus *Paracatenula* are mouthless and live in endosymbiosis with nutritional symbionts. The symbionts are housed in the trophosome region that is posterior to the brain while the anterior located rostrum is symbiont-free. Their symbionts provide both bulk nutrition and primary energy storage for their host. In *Paracatenula*, asexual reproduction appears to be the predominant mode of reproduction that naturally includes tissue regenerating capabilities. Within a few days, *Paracatenula* can perform wound healing and regrowth of the its rostrum within two weeks after induced sectioning.

Here we provide evidence that *Ca.* R. standrea symbionts are involved in tissue regeneration processes in the flatworm *Paracatenula* sp. standrea. The symbionts change their expression in 168 out of 1344 protein-coding genes. The majority of differentially expressed genes were upregulated and related to basic cellular processes such as translation, transcription as well as carbon and energy metabolism. This upregulation suggests a boost of their metabolism under such conditions. Additionally, some genes were involved in the biosynthesis of amino acids and the vitamin thiamine, which reflect a potential importance of the symbionts in nutrient supplementation. We conclude that the transcriptome response of *Ca.* R. standrea to tissue regeneration shows an active function of the symbionts in complex host processes as rostrum regeneration.

Introduction

The regeneration of body parts is widespread in several species of Platyhelminthes (Egger *et al.*, 2007). Underlying processes of tissue regeneration were well studied in model systems such as *Schmidtea mediterraneae*, *Dugesia japonica* and *Macrostomum lignano* and recently investigated in the symbiotic flatworm *Paracatenula* (Newmark and Sánchez Alvarado, 2002; Agata, 2003; Saló, 2006; Gentile *et al.*, 2011; Dirks, Gruber-Vodicka, Egger, *et al.*, 2012). The ability for regeneration tissues varies as some flatworms are capable to regrowing into a complete animal from tiny flatworm pieces while others can only regenerate the posterior fragment when brain and pharynx are present (Wanninger, 2015). In general, regenerative abilities are more pronounced in taxa that contain species with asexual reproduction as regeneration is essential for their survival (Egger *et al.*, 2007).

Stem cells, so-called neoblasts, confer the remarkable ability of tissue regeneration in flatworms (Newmark and Sánchez Alvarado, 2000; Bode *et al.*, 2006; Wanninger, 2015). These cells are scattered throughout the animal body but are absent in the head tip and pharynx (Wanninger, 2015). Neoblasts can differentiate into all cell types and were shown to have critical functions for regeneration in e.g. *Tricladida* and basally branching *Catenulida* (Moraczewski, 1977; Palmberg, 1990; Wagner *et al.*, 2011). The regenerative abilities of some flatworms involve the *de novo* synthesis of a functional central nervous system (Cebrià *et al.*, 2007; Agata and Umesono, 2008; Umesono *et al.*, 2011; Fraguas *et al.*, 2012).

The mouthless flatworms genus *Paracatenula* reproduce by asexual fragmentation (Dirks, Gruber-Vodicka, Leisch, *et al.*, 2012). They belong to the order *Catenulida* that are known for their excellent regeneration capabilities (Egger *et al.*, 2007). The species *Paracatenula galateia* can regrow the region anterior of the brain (rostrum) within two weeks after sectioning (Dirks, Gruber-Vodicka, Egger, *et al.*, 2012). It has been shown, that regeneration can only occur in the presence of a sufficient trophosome fragment size because tissue-regenerating neoblasts are restricted to this region. (Dirks, Gruber-Vodicka, Egger, *et al.*, 2012). Such neoblasts were identified as the source of bacteriocytes – the symbiont-containing cells – and all other somatic cell types (Dirks, Gruber-Vodicka, Leisch, *et al.*, 2012). In tissue regeneration, the neoblasts migrate through the worms body to the wounded regenerating part (Dirks, Gruber-Vodicka, Leisch, *et al.*, 2012). Within two days after rostrum sectioning, the cutting wound is closed by

surrounding epidermal cells and longitudinal nerves end blindly in the wound area. In this stage, neoblasts start to accumulate in the wound area to initiate complex regeneration processes. With the start of the rostrum regeneration clusters of proliferating neoblasts (blastema) are formed, and within a week of regeneration, the rostrum appears in its characteristic shape (Dirks, Gruber-Vodicka, Egger, *et al.*, 2012).

What has mostly been ignored in processes of *Paracatenula* regeneration is their symbiotic counterpart, the bacterial symbionts *Ca.* Riegeria, that can comprise up to half of the biomass (Gruber-Vodicka *et al.*, 2011). The mutualistic symbionts of *Paracatenula* sp. standrea were previously shown to provide both nutrition as well as primary energy storage for their mouthless host (Jäckle *et al*., 2018). The *Ca.* Riegeria symbionts likely interact with their host during tissue regeneration.

A function of a bacterial microbiome in tissue regeneration was recently illustrated in planarian flatworms (Arnold *et al.*, 2016). While healthy individuals of *Schmidtea mediterraneae* showed a higher abundance of a certain type of bacteria (*Bacteriodetes vs. Proteobacteria*), an experimental-induced shift of bacteria towards the *Proteobacteria* caused tissue degeneration, presumably because signaling pathways responsible for complex tissue regeneration were affected.

In this study, we investigated what functions the bacterial symbionts perform in trophosome regions undergoing rostrum regeneration. *Paracatenula* individuals performed rostrum regeneration for five days. Transcriptomic sequencing showed that the symbionts changed their gene expression in 12.5% of all genes present. The majority of genes (121 out of 168) were significantly upregulated and of these, most were related to information processing and energy and carbon metabolism. Our results provide first evidence for potential supporting function of endosymbionts in tissue regeneration in Platyhelminthes.

Results and Discussion

Rostrum regeneration in *Paracatenula* sp. standrea occurs within two weeks

The sectioning of *Paracatenula* sp. standrea specimens in an anterior (with rostrum) and a posterior (without rostrum) fragment was performed so that both parts still consisted of trophosome tissue (Figure 1A). The anterior part continued its smooth swimming behavior and

the posterior part moved in circles, similar to what has been observed in sectioned *Paracatenula galateia* individuals (Dirks, Gruber-Vodicka, Egger, *et al.*, 2012; Jäckle *et al.*, 2018). The progress of tissue regeneration was photodocumented for the posterior 'trophosome region' for the next 11 days, as the anterior 'rostrum region' just performs wound healing. Already within five days after rostrum sectioning, wound closure and regeneration had started, as a forming blastema could be observed at the tip of the wound (Figure 1B). As the regeneration process was initiated but not completed after five days, this timepoint was chosen for the transcriptomics experiment. Only six days later, the regenerated rostrum had already formed its characteristic shape and will just continue growing in length reaching ~0.4 mm for a few days (Figure 1C). Our observations on *Paracatenula* sp. standrea resemble previous results on *Paracatenula galateia* and suggest that the tissue regeneration patterns are characteristic for the genus *Paracatenula*.
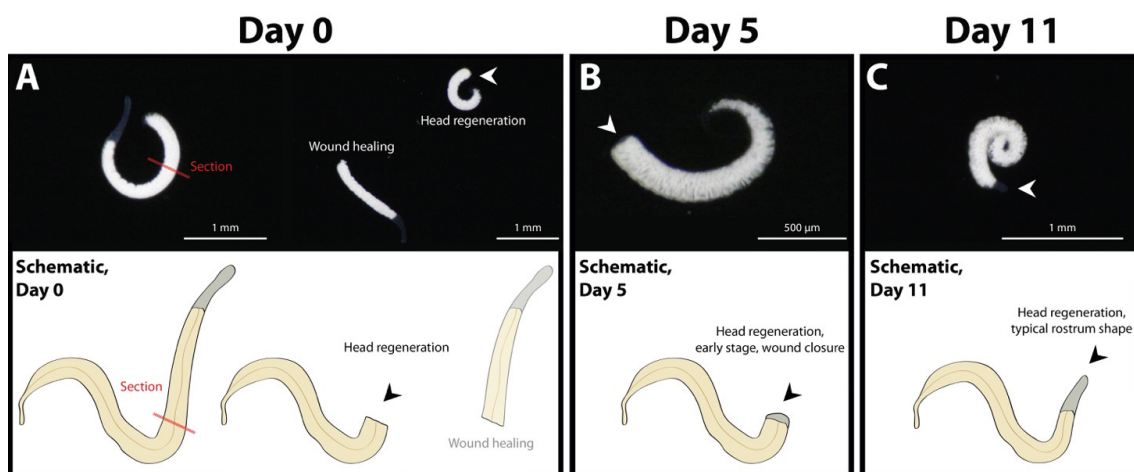
Figure 1. Head regeneration in *Paracatenula* sp. standrea observed over 11 days. Representative microscopic images and schematics illustrating early and later stages of head regeneration in a single specimen. ***A***, *Paracatenula* specimen before and just after rostrum sectioning. ***B***, After 5 days of regeneration; healing and head regeneration is in process. ***C***, Late stage of head regeneration after 11 days; the rostrum has its typical shape and will continue growing in length. Gene expression of *Paracatenula* symbionts was investigated after 5 days.

Differential gene expression of *Ca*. R. standrea in the rostrum-regenerating trophosome region

To study potential functions of symbionts adjacent to the regenerating area, we investigated gene expression changes in specimens five days after sectioning (Figure 1B). The mRNA was sequenced from individuals that performed rostrum regeneration for five days, wound-healing individuals and compared to control individuals kept for five days under the exact same conditions and one specimen from each glass vial (n = 3) was sampled for transcriptomics. No

differential expression was observed for individuals that performed wound healing compared to the control samples. In individuals that performed rostrum regeneration we identified 168 significant differentially expressed genes in *Ca.* R. standrea symbionts accounting for 12.5% of all genes affected. Of the 1344 protein-coding genes, 1336 genes were expressed, being indicative for covering almost all genes in the transcriptomics. Of these 168 differentially expressed genes, 47 were downregulated and 121 were upregulated, with significant expression differences of ≥ 1.23-fold at false-discovery rate (FDR) < 0.05 (Figure 2A). To categorize the cellular functions of up- and downregulated genes in *Ca.* R. standrea symbionts, the differentially expressed genes were assigned to Clusters of Orthologous Genes (COGs) (Supplementary Table 1, Figure 2B). Of all differentially expressed genes, 83.9% could be assigned to COG categories, and of those that could not be assigned, 39.3% were annotated as hypothetical proteins.
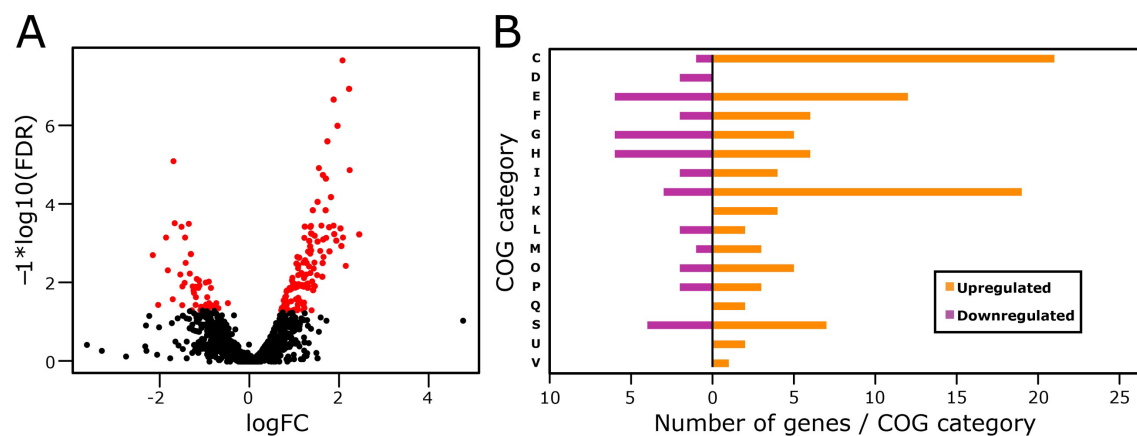
Figure 2. Differential expression of *Ca*. R. standrea genes in head-regenerated trophosome region. *A*, Volcano plot. Differentially expressed genes (FDR < 0.05) and ≥ 1.23-fold expression difference are shown in red. 168 genes were differentially expressed. *B*, Categorization of the differentially expressed genes based on Clusters of Orthologous Genes (COGs). 141 out of 168 genes could be grouped into COGs. The list of COGs was taken from ftp://ftp.ncbi.nih.gov/pub/COG/COG2014/data/fun2003-2014.tab. A detailed list of genes categorized in COGs, there functional categories and functional processes is listed in Supplementary Table 1.

Upregulation of information processing and energy metabolism in *Ca.* R. standrea

Most upregulated genes in *Ca.* R. standrea symbionts that were assigned COGs were classified to function in the production and conversion of energy (category C) (Figure 2B). Upregulated genes in this category included those for the tricarboxylic acid cycle (TCA), suggesting a potential role of utilizing sugars and other organics. Of the eight enzyme-encoding genes of the TCA cycle, five were upregulated (Figure 3, Supplementary Data set, S1). Those genes

included the 2-oxoglutarate dehydrogenase (E_3), succinyl-CoA ligase, the succinate dehydrogenase complex and the fumarate hydratase. The primary source of acetyl-CoA utilized in the TCA cycle originates from the breakdown of sugars by glycolysis that is present in *Ca.* R. standrea (Jäckle *et al.*, 2018). Only two genes involved in glycolysis were upregulated that included the glucokinase, phosphorylating glucose to retain it in the bacterial cell, and a bifunctional PP_i-dependent phosphofructokinase, while the fructose-bisphosphate aldolase for formation of two C3 molecules was downregulated (Supplementary Data set, S1). At this point a clear functioning of energy conservation cannot be claimed. Alternatively, the TCA cycle could operate as the central carbon metabolism producing a variety of metabolic intermediates serving for e.g. amino acid synthesis and vitamin synthesis (Figure 3B). We identified upregulated genes for the synthesis of the amino acids glutamine, lysine and serine, as well as the vitamin thiamine and the cofactor NAD^+ although not all genes of the respective pathways were affected. (Supplementary Data set, S1).

Pathways for thioautotrophy such as thiosulfate oxidation and the Calvin Benson-Bassham (CBB) cycle were upregulated under regeneration conditions (Figure 3, Supplementary Data set, S1). Genes related to sulfur oxidation included a thiosulfate sulfurtransferase (rhodanese) for the oxidation of thiosulfate to sulfite (Figure 3A). Sulfite could then be further oxidized to adenosine-5'-phosphosulfate by the upregulated adenylylsulfate reductase (alpha and beta subunit). The upregulated sulfur oxidation starting from thiosulfate was surprising, as the symbionts of freshly collected individuals typically express a reverse-acting dissimilatory sulfite reductase for the oxidation of other types of sulfur species (Jäckle *et al.*, 2018) that was not differentially expressed. Thiosulfate often represents a small dynamic pool of reduced sulfur that can serve as electron donor in chemosynthetic symbiont but nothing is known about its concentrations in the habitat of *Paracatenula* (Giere *et al.*, 1988; Jørgensen, 1990; Dubilier *et al.*, 2006).

To fix carbon dioxide into biomass, the *Ca.* R. standrea symbionts use a modified PP_i-dependent version of the CBB cycle (Jäckle *et al.*, 2018) of which four genes fulfilling five enzymatic reactions were found to be upregulated (Figure 3). The upregulated genes included both the small and large RuBisCO subunit (*cbbL* and *cbbM*), the multifunctional PP_i-dependent phosphofructokinase (PP_i-PFK) and the PP_i-energized proton pump (PP_i-H^+-pump). The fixed carbon could then be stored as trehalose or polyhydroxyalkanoates (PHA) granules, as genes

encoding the alpha,alpha-trehalose-phosphate synthase as well as a phasin and the acetoacetyl-CoA reductase were also upregulated (Figure 3, Supplementary Data set, S1).

To avoid futile cycling, it was recently suggested that both the CBB cycle and the TCA cycle do not operate at the same time but rather in subpopulations of the *Ca.* R. standrea symbionts (Jäckle *et al.*, 2018) (Figure 3). We cannot exclude that we investigated gene expressions of symbiont populations in varying microniches and therefore different metabolic activities. Another possibility would be that a simultaneous cycling could be a short-term response of the symbionts to increase their net C productivity and facilitate the host nutritional needs. To fix 2 molecules of carbon dioxide 4 $\frac{1}{9}$ ATP are needed (Kleiner *et al.*, 2012) while the complete oxidation of 2 molecules of acetyl-CoA in the TCA cycle (4 CO_2) and redox chain results in 30 ATP. Thus, the oxidation of acetyl-CoA yields additional energy (10 $\frac{8}{9}$ ATP) that can then be used for the fixation of carbon dioxide and could increase the net carbon pool in *Ca.* R. standrea. However, as only the succinyl dehydrogenase genes but no further downstream reactions in the respiratory chain and the ATP-synthase were affected in their expression, the functioning of the TCA cycle for the synthesis of intermediate instead of energy conservation is more likely than for energy conservation.

A large proportion of upregulated genes in *Ca.* R. standrea from the rostrum-regenerating trophosome region were assigned to function in translation, ribosomal structure and biogenesis (category J) (Figure 2B, Figure 3A, Supplementary Data set, S1). While the cell division gene *ftsZ* was significantly downregulated and indicated reduced proliferation in the trophosome, the upregulation of protein biosynthesis genes suggests a general increase in the physiological performance of the symbionts. The biosynthesis genes included those encoding tRNAs for the attachment of the amino acids proline, phenylalanine, and lysine. Additionally, these genes included those encoding eleven ribosomal proteins and the translation elongation factors Tu, G, P that are involved in the elongation cycle of protein biosynthesis (Supplementary Data set, S1). Together, these results point to an upregulation of functionality and biomass production, combined with increased carbon and energy densities in non-growing cells, a hypothesis that still has to be proven.

The most prominent downregulated genes included those involved in the formation of two cofactors (Supplementary Data set, S1). Four iron-sulfur cluster proteins (*sufBCD* and *sufS*) were affected that were located on one operon in the *Ca.* R. standrea genome (Figure 3A).

Iron-sulfur cluster proteins play vital roles in e.g. regulation of enzyme activity, cofactor biogenesis and gene expression regulation (Wollers *et al.*, 2010; Xu and Møller, 2011). Also, these proteins were shown to function as a scaffold-complex receiving and transferring sulfur to other proteins for the assembly of Fe-S clusters (Wollers *et al.*, 2010; Xu and Møller, 2011). The assembly of Fe-S clusters typically occurs under conditions when sufficient amounts of bioavailable iron and sulfide are present (Ali and Nozaki, 2013). We hypothesize that the downregulation of iron-sulfur cluster genes indicates a critical need of free sulfide, not being used for the utilization of iron-sulfur proteins but rather for its oxidation to generate energy. A lack of free sulfide would agree with our finding that the rDSR pathway typically expressed in *Ca.* R. standrea was not affected in its expression (Figure 3, Supplementary Data set, S1). Additionally, genes encoding an adenosylcobyric acid synthase and an adenosylcobinamide-phosphate guanylyltransferase were downregulated, indicative for energy saving for the synthesis of this complex vitamin.
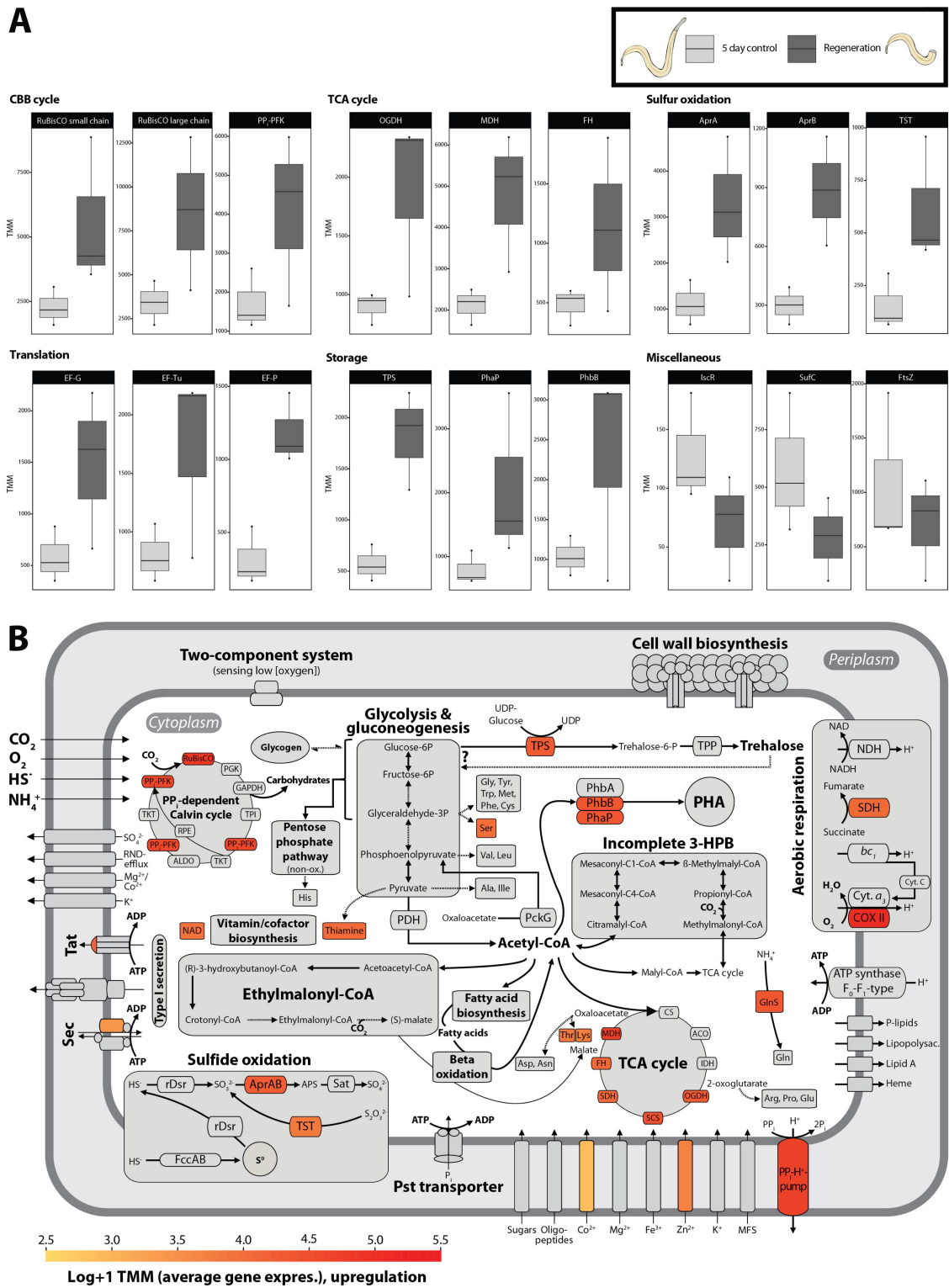
Figure 3. Upregulated energy metabolism of *Ca.* R. standrea symbionts in rostrum-regenerated trophosome region. *A*, A selection of differentially expressed genes covering several metabolic functions. A complete list of genes is shown in Supplementary Data set, S1. ***B*,** Metabolic scheme highlighting upregulated reactions. The scheme was adapted from a previous study on the *Paracatenula* symbiosis (Jäckle *et al.*, 2018). Pathways or enzymes are colored by their expression levels in the transcriptome (mean of three samples). Only upregulated genes are shown in the illustration. Not all genes of certain pathways were differentially transcribed. Dotted arrows correspond to indirect synthesis of metabolites.

159

The importance of the *Ca.* R. standrea symbionts in head regulation processes is suggested by their upregulation of genes related to the central carbon metabolism, storage compounds and transcription and translation (Figure 4). The reported genes showed significant differential expression, and although an increase in expression was observed for the housekeeping genes *gyrA* and *recA* this change was not significant (Supplementary Data set, S1). The significant upregulation of certain metabolic processes allows the production of versatile metabolites such as carbon and energy storages as well as amino acids and vitamins that can serve the host during nutrient-demanding regeneration processes. The transfer of nutrients likely occurs *via* leaking or digestion as it has been recently suggested in *Ca.* R. standrea under natural conditions (Jäckle *et al.*, 2018).
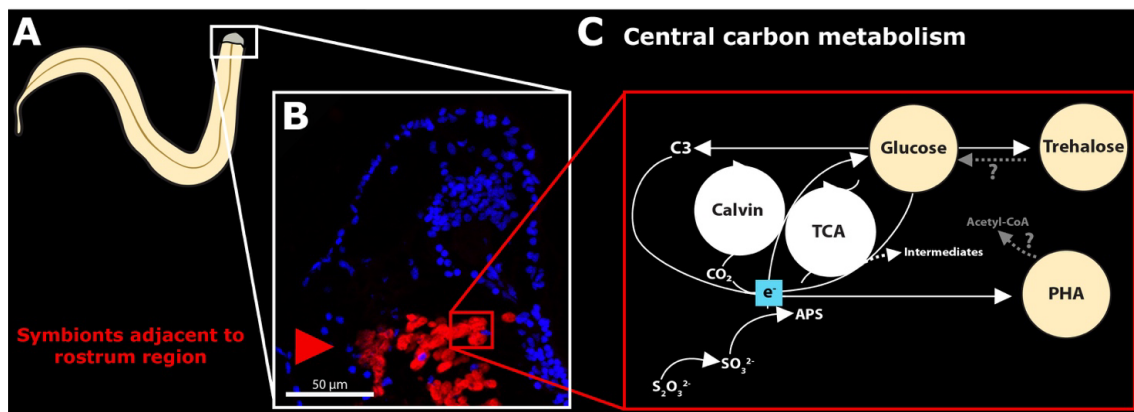
Figure 4. Model summarizing the functions of *Ca.* R. standrea in rostrum regeneration. *A*, Schematic of *Paracatenula* after five days rostrum regeneration and ***B*,** A CARD-FISH-image [EUB I–III, (Amann *et al.*, 1990; Daims *et al.*, 1999)] of the rostrum-trophosome area. The FISH image was adapted from Kiers et al., 2015 (Kiers and West, 2015). ***C*,** Major steps of the central carbon metabolism that were affected in their gene expression. The white arrows indicate substrate and electron flows, the gray arrows indicate reactions which enzymatic reactions are yet. Dotted arrows indicate indirect reactions.

Conclusion

The regeneration of a rostrum is a fascinating process, especially if it occurs in mouthless flatworms that are completely dependent on their symbionts for nutrition. While some species of Platyhelminthes lack the ability for tissue regeneration, flatworms of the genus *Paracatenula* can regenerate a complete rostrum within two weeks. Regeneration is a natural process because *Paracatenula* reproduce by asexual fission.

Our study shows that the *Ca.* R. standrea symbionts are involved in the processes of rostrum regeneration. As a response to sectioning and tissue regeneration, the symbionts change their expression profile in mostly nutrition-related genes within five days of regeneration. Tissue

regeneration represents a natural scenario as *Paracatenula* flatworms proliferate by paratomy. Regeneration experiments, however, have shown that rostrum fragments have limits in their ability of regeneration (Dirks, Gruber-Vodicka, Egger, *et al.*, 2012). The role of endosymbionts in these regeneration processes by the provision of e.g. carbon and energy resources is likely the explanation for previously observed absence of posterior growth in the related species *Paracatenula galateia* (Dirks, Gruber-Vodicka, Egger, *et al.*, 2012). As the symbionts are essential for the host, the number and density of symbionts in the trophosome will be key for regeneration. Furthermore, physiochemical conditions such as oxygen and sulfide availability will have great influence on the pace of tissue regeneration, as these influence the metabolic performance of the symbionts. It was shown in regenerating planarians that the presence of morphogen gradients along the body is essential for tissue development (Adell *et al.*, 2010). It is still unclear if and to what extend the symbionts might be involved in regulating such signaling molecules, either in their production, transfer or depletion.

It remains unresolved if the responses in expression are consistent in symbiont populations of different regions of the trophosome. If we assume an underlying cellular communication between the host and the symbionts the signal could either be consistent for all symbiont cells, or be increased in regions close to the regenerating fragment. Single-cell transcriptomics could reveal if the symbiont´s gene expression change over the distance to the wound area, likely with largest responses close to it.

A deeper understanding of host-symbiont interaction especially on a cellular level, as well as the investigation of host changes of expression during rostrum regeneration will be of great importance to understand interdependent symbiotic associations such as the *Paracatenula-Ca.* Riegeria symbiosis.

Materials and Methods

Sample collection and sectioning experiment

Paracatenula sp. standrea specimens were collected in 2013 and 2016 from sediments in the bay off Sant´Andrea, Elba, Italy. Specimens were extracted as previously described (Jäckle *et al.*, 2018). A razor blade was used to section the living individuals transversally posterior to the brain region in way there was still trophosome that allowed the rostrum to regenerate

(Figure 1A). The anterior and posterior fragments were placed to 2 ml glass vials filled with washed natural sediment and filtered seawater without the addition of external energy or carbon sources. The vials had an air-filled headspace. The experiment was performed with biological replicates (n = 3 each). Untreated individuals kept for 5 days under identical conditions served as comparison group. Additionally, those fragments that performed wound healing as well as non-incubated controls were prepared but not further analyzed in this study. *Paracatenula* specimens for nucleic acid extractions were fixed in RNAlater (Ambion) and stored at 4 °C.

To document the different stages of rostrum regeneration in *Paracatenula* sp. standrea, the steps described above were repeated, and digital images of each individual were recorded with a Canon EOS 700D camera (Canon) mounted on a Nikon SMZ-745T dissecting microscope (Nikon).

RNAseq and differential expression

For symbiont transcriptomics, RNA was extracted from single *Paracatenula* sp. standrea specimens. RNA was extracted with the RNeasy plus micro kit (Qiagen) following the manufacturer´s protocol, and eluted in 11 µl RNase-free water. The cDNA was synthesized from total RNA with the Ovation RNASeq System v2 (NuGEN) following the manufacturer's protocol, sheared to target sizes of 300 bp with the Covaris microTUBE system (Covaris) and cleaned up with the Zymo Genomic DNA Clean & Concentrator Kit (Zymo Research). Sequencing libraries were prepared from cDNA with the NEBNext Ultra DNA library preparation kit (New England Biolabs) for Illumina and sequenced on the Illumina HiSeq 2500 platform using 2 x 100 bp paired end reads. Both library preparation and sequencing were performed at the Max Planck Genome Centre in Cologne. 6 to 7 million single-end 100 bp reads per library were sequenced and to quantify gene expression, the reads were mapped against the *Ca.* R. standrea genome (accession no. GCA_900576755) with BBMap v38.06 and quantified with featureCounts v1.5.0-p1 (Liao *et al.*, 2014). Differential gene expression of *Ca.* R. standrea between rostrum-sectioned individuals and the 5 day controls was analyzed with the accessory scripts implemented in Trinity v2.5.1 (Grabherr *et al.*, 2011) using the Bioconductor package edgeR (Robinson *et al.*, 2010). Significance was accepted at a False-Discovery-Rate (FDR) of < 0.05. For normalization of each sequenced library a normalization factor was estimated based on the trimmed mean of M-values (TMM) implemented in the edgeR package. The

differentially expressed genes were assigned to Clusters of Orthologous Genes with eggNOG (Huerta-Cepas *et al.*, 2016).

Competing interests

The authors declare no conflict of interest.

Acknowledgments

Supplementary Table

The Supplementary Data set is available on the CD-ROM provided with the thesis.

Supplementary Table 1. Differentially expressed functional categories and processes based on COG categories.

Functional category	Functional process	Abbrev.	Upreg.	Downreg.
Information storage and processing				
	Translation, ribosomal structure and biogenesis	J	19	3
	Transcription	K	4	0
	Replication, recombination and repair	L	2	2
Cellular processes and signaling				
	Cell cycle control, cell division, chromosome partitioning	D	0	2
	Defense mechanisms	V	1	0
	Signal transduction mechanisms	T	0	0
	Cell wall/membrane/envelope biogenesis	M	3	1
	Cell motility	N	0	0
	Intracellular trafficking, secretion, and vesicular transport	U	2	0
	Posttranslational modification, protein turnover, chaperons	O	5	2
Metabolism				
	Energy production and conversion	C	21	1
	Carbohydrate transport and metabolism	G	5	6
	Amino acid transport and metabolism	E	12	6
	Nucleotide transport and metabolism	F	6	2
	Coenzyme transport and metabolism	H	6	6
	Lipid transport and metabolism	I	4	2
	Inorganic ion transport and metabolism	P	3	2
	Secondary metabolites biosynthesis, transport and catabolism	Q	2	0
Other functions				
	General function prediction only	R	0	0
	Function unknown	S	7	4
	Mobilome: prophages, transposons	X	0	0
	Extracellular structures	W	0	0
	Chromatin structure and dynamics	B	0	0
	Cytoskeleton	Z	0	0
	Nuclear structure	Y	0	0
	Total		**102**	**39**

Bibliography Chapter IV

Adell, T., Cebrià, F. and Saló, E. (2010) 'Gradients in planarian regeneration and homeostasis', *Cold spring harbor perspectives in biology*, 2:a000505.

Agata, K. (2003) 'Regeneration and gene regulation in planarians', *Current Opinion in Genetics and Development*, **13**, pp. 492–496.

Agata, K. and Umesono, Y. (2008) 'Brain regeneration from pluripotent stem cells in planarian', *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, pp. 2071–2078.

Ali, V. and Nozaki, T. (2013) 'Iron-Sulphur clusters, their biosynthesis, and biological functions in protozoan parasites', *Advances in Parasitology*, **83**, pp. 1–92.

Amann, R. I., Binder, B. J., Olson, R. J., Chisholm, S. W., Devereux, R. and Stahl, D. A. (1990) 'Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations', *Applied and environmental microbiology*, **56**, pp. 1919–1925.

Arnold, C. P., Merryman, M. S., Harris-Arnold, A., McKinney, S. A., Seidel, C. W., Loethen, S., Proctor, K. N., Guo, L. and Sánchez Alvarado, A. (2016) 'Pathogenic shifts in endogenous microbiota impede tissue regeneration *via* distinct activation of TAK1/MKK/p38', *eLife*, **5**, e16793.

Bode, A., Salvenmoser, W., Nimeth, K., Mahlknecht, M., Adamski, Z., Rieger, R. M., Peter, R. and Ladurner, P. (2006) 'Immunogold-labeled S-phase neoblasts, total neoblast number, their distribution, and evidence for arrested neoblasts in *Macrostomum lignano* (Platyhelminthes, Rhabditophora)', *Cell and Tissue Research*, **325**, pp. 577–587.

Cebrià, F., Guo, T., Jopek, J. and Newmark, P. A. (2007) 'Regeneration and maintenance of the planarian midline is regulated by a *slit* orthologue', *Developmental Biology*, **307**, pp. 394–406.

Daims, H., Brühl, A., Amann, R., Schleifer, K.-H. and Wagner, M. (1999) 'The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: Development and evaluation of a more comprehensive probe set', *Systematic and applied microbiology*, **22**, pp. 434–444.

Dirks, U., Gruber-Vodicka, H. R., Egger, B. and Ott, J. A. (2012) 'Proliferation pattern during rostrum regeneration of the symbiotic flatworm *Paracatenula galateia*: A pulse-chase-pulse analysis', *Cell and tissue research*, **349**, pp. 517–525.

Dirks, U., Gruber-Vodicka, H. R., Leisch, N., Bulgheresi, S., Egger, B., Ladurner, P. and Ott, J. A. (2012) 'Bacterial symbiosis maintenance in the asexually reproducing and regenerating flatworm *Paracatenula galateia*', *PLoS ONE*, **7**, e34709.

Dubilier, N., Blazejak, A. and Ruehland, C. (2006) *Symbioses between bacteria and gutless marine oligochaetes*, *Progress in Molecular and Subcellular Biology*. Edited by J. Overmann. New York, NY, USA: Springer-Verlag Berlin Heidelberg.

Egger, B., Gschwentner, R. and Rieger, R. (2007) 'Free-living flatworms under the knife: Past and present', *Development Genes and Evolution*, **217**, pp. 89–104.

Fraguas, S., Barberán, S., Ibarra, B., Stöger, L. and Cebrià, F. (2012) 'Regeneration of neuronal cell types in *Schmidtea mediterranea*: An immunohistochemical and expression study', *International Journal of Developmental Biology*, **56**, pp. 143–153.

Gentile, L., Cebria, F. and Bartscherer, K. (2011) 'The planarian flatworm: an in vivo model for stem cell biology and nervous system regeneration', *Disease Models & Mechanisms*, **4**, pp. 12–19.

Giere, O., Wirsen, C. O., Schmidt, C. and Jannasch, H. W. (1988) 'Contrasting effects of sulfide

and thiosulfate on symbiotic CO_2-assimilation of *Phallodrilus leukodermatus* (Annelida)', *Marine Biology*, **97**, pp. 413–419.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., 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. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. and Regev, A. (2011) 'Full-length transcriptome assembly from RNA-Seq data without a reference genome', *Nature Biotechnology*, **29**, pp. 644–654.

Gruber-Vodicka, H. R., Dirks, U., Leisch, N., Baranyi, C., Stoecker, K., Bulgheresi, S., Heindl, N. R., Horn, M., Lott, C., Loy, A., Wagner, M. and Ott, J. (2011) '*Paracatenula*, an ancient symbiosis between thiotrophic *Alphaproteobacteria* and catenulid flatworms', *Proceedings of the National Academy of Sciences of the United States of America*, **108**, pp. 12078–12083.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., Rattei, T., Mende, D. R., Sunagawa, S., Kuhn, M., Jensen, L. J., Von Mering, C. and Bork, P. (2016) 'eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences', *Nucleic Acids Research*, **44**, pp. D286–D293.

Jäckle, O., Seah, B. K. B., Tietjen, M., Leisch, N., Liebeke, M., Kleiner, M., Berg, J. S. and Gruber-Vodicka, H. R. (2018) 'A chemosynthetic symbiont with a drastically reduced genome serves as primary energy storage in the marine flatworm *Paracatenula*', *Submitted manuscript*.

Jørgensen, B. B. (1990) 'A thiosulfate shunt in the sulfur cycle of marine sediments', *Science*, **249**, pp. 152–154.

Kiers, E. T. and West, S. A. (2015) 'Evolving new organisms *via* symbiosis', *Science*, **348**, pp. 392–394.

Kleiner, M., Wentrup, C., Lott, C., Teeling, H., Wetzel, S., Young, J., Chang, Y.-J., Shah, M., VerBerkmoes, N. C., Zarzycki, J., Fuchs, G., Markert, S., Hempel, K., Voigt, B., Becher, D., Liebeke, M., Lalk, M., Albrecht, D., Hecker, M., Schweder, T. and Dubilier, N. (2012) 'Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use', *Proceedings of the National Academy of Sciences of the United States of America*, **109**, pp. E1173–E1182.

Liao, Y., Smyth, G. K. and Shi, W. (2014) 'featureCounts: an efficient general purpose program for assigning sequence reads to genomic features', *Bioinformatics*, **30**, pp. 923–930.

Moraczewski, J. (1977) 'Asexual reproduction and regeneration of *Catenula* (Turbellaria, Archoophora)', *Zoomorphologie*, **88**, pp. 65–80.

Newmark, P. A. and Sánchez Alvarado, A. (2000) 'Bromodeoxyuridine specifically labels the regenerative stem cells of planarians', *Developmental Biology*, **220**, pp. 142–153.

Newmark, P. A. and Sánchez Alvarado, A. (2002) 'Not your father's planarian: A classic model enters the era of functional genomics', *Nature Reviews Genetics*, **3**, pp. 210–219.

Palmberg, I. (1990) 'Stem cells in microturbellarians - An autoradiographic and

immunocytochemical study', *Protoplasma*, **158**, pp. 109–120.

Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010) 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data', *Bioinformatics*, **26**, pp. 139–140.

Saló, E. (2006) 'The power of regeneration and the stem-cell kingdom: freshwater planarians (Platyhelminthes)', *BioEssays*, **28**, pp. 546–559.

Umesono, Y., Tasaki, J., Nishimura, K., Inoue, T. and Agata, K. (2011) 'Regeneration in an evolutionarily primitive brain - the planarian *Dugesia japonica* model', *European Journal of Neuroscience*, **34**, pp. 863–869.

Wagner, D. E., Wang, I. E. and Reddien, P. W. (2011) 'Clonogenic neoblasts are pluripotent adult stem cells that underlie planarian regeneration', *Science*, **332**, pp. 811–816.

Wanninger, A. (2015) *Evolutionary Developmental Biology of Invertebrates*. Springer-Verlag Wien 2015.

Wollers, S., Layer, G., Garcia-Serres, R., Signor, L., Clemancey, M., Latour, J.-M., Fontecave, M. and De Choudens, S. O. (2010) 'Iron-sulfur (Fe-S) cluster assembly: The SufBCD complex is a new type of Fe-S scaffold with a flavin redox cofactor', *Journal of Biological Chemistry*, **285**, pp. 23331–23341.

Xu, X. M. and Møller, S. G. (2011) 'Iron–Sulfur Clusters: biogenesis, molecular mechanisms, and their functional significance', *Antioxidants & Redox Signaling*, **15**, pp. 271–307.

Chapter V: General discussion and perspectives

A mutualistic symbiosis between microorganisms and the marine flatworm of the genus *Paracatenula* has been proposed already 36 years ago (Ott *et al.*, 1982). Since then, few studies have been performed on the mouthless *Paracatenula* flatworms. These studies mainly focused on morphological characterizations and molecular developmental aspects such as tissue and rostrum regeneration (Dirks *et al.*, 2011; Leisch *et al.*, 2011; Dirks, Gruber-Vodicka, Egger, *et al.*, 2012; Dirks, Gruber-Vodicka, Leisch, *et al.*, 2012). Only one study examined their unusual symbionts, the first known chemoautotrophic symbionts of the class *Alphaproteobacteria* and their role in this ancient symbiosis (Gruber-Vodicka *et al.*, 2011). This study revealed the phylogenetic placement of the symbiont and their co-speciation with the *Paracatenula* host for hundreds of millions of years. Additionally, key genes for thioautotrophy of the symbionts were characterized and it was reported that the symbionts store sulfur. However, the genetic repertoire and function in the symbiosis with *Paracatenula*, their ecophysiology as well as their evolution during several million years have not been analyzed and thus, left a knowledge gap which has been closed by the studies reported here.

The data presented here contribute substantially to the knowledge in the fields of microbial ecology, symbiosis and developmental biology by elucidating the essential functions of chemosynthetic symbionts in mouthless host animals (Chapter 2), their evolutionary history and processes of genome evolution (Chapter 3) and their requirement during rostrum (anterior region with brain) regeneration of the host after its sectioning (Chapter 4).

What are essential functions of chemosynthetic symbionts in mouth- and gutless animals?

"What do *Paracatenula* flatworms feed on?" and "How do they get access to nutrients?" were the first questions I addressed in my studies. Chemosynthetic symbioses are characterized by the fixation of carbon and oxidation of reduced compounds such as sulfur (Kleiner *et al.*, 2012). However, it has been unknown which additional contributions of the chemosynthetic symbionts are required to nurture the host organisms. Tackling such a research question in chemosynthetic symbioses has been so far quite challenging since no suitable experimental system for these analysis could be established for several reasons. One reason is the mode of horizontal transmission, in which the host organism takes up the symbionts newly from the

environment in every generation. Examples of this mode of symbiont acquisition are the bathymodiolin mussel *Bathymodiolus* and the vestimentiferan tubeworm *Riftia* (Nussbaumer *et al.*, 2006; Russell *et al.*, 2017). It implies that the symbionts taken up are able to survive in a free-living stage in the environment without a host, and in symbiosis. A free-living stage of a symbiont requires a flexible metabolic and genetic makeup that is reflected in the wide range of genome sizes up to 4.5 Mb (Kuwahara *et al.*, 2007; Newton *et al.*, 2007; Petersen *et al.*, 2016). Thus, the essential set of core functions that have to be performed by the bacteria living in symbiosis with their host animal cannot be determined.

Instead, a second way to answer this question would be to investigate conserved functions of symbionts that underwent genome streamlining over generations and time, supposedly down to a critical core (Kuwahara *et al.*, 2007; Newton *et al.*, 2007). In contrast to studying *Paracatenula*, most other systems would not allow unambiguous results. In theory, the vertically transmitted symbionts of the vesicomyid clams would be an ideal system to explore the host-chemosynthetic symbiont interdependence, because they have reduced genomes. Thus, they would be a suitable study system to investigate which genes encoding metabolic pathways are maintained. Nevertheless, similar to *Bathymodiolus* mussels, the vesicomyid clams still filter feed (Cavanaugh, 1983; Le Pennec *et al.*, 1995). They have reduced or rudimentary guts, feeding appendages and their stomach was found to be filled with various particles. This suggests an at least partially functioning digestive system, which could mask the essential functions that are provided by a chemosynthetic symbiont.

I have taken advantage of the symbiosis of *Paracatenula* sp. standrea and its chemosynthetic symbionts *Ca.* Riegeria standrea from Elba (Italy) to study the physiology of their symbionts (Chapter 2) to explore the above outlined questions. This system has the advantage that *Paracatenula* lacks mouth and gut and transmits its symbionts vertically (Sterrer and Rieger, 1974; Gruber-Vodicka *et al.*, 2011). Also, the symbionts have reduced genome sizes, suggesting that they have evolutionarily co-adapted with their host over millions of years. My analyses shows that *Ca.* R. standrea, in addition to its dual function as biosynthesis factory and storage organ (see below and Chapter 2), shares a large part of its metabolic functions with other reduced chemosynthetic symbionts such as the symbionts of vesicomyid clams. Considering the environments the hosts live in, i.e. the vesicomyid clams in the deep sea and *Paracatenula* in shallow-water sediments, these common functions of their genome repertoires

seemed surprising at a first. Unexpectedly, our findings represent an example of convergent evolution and show that essential functions in chemosynthetic symbioses are independent of phylogeny and the environment, and rather reflect the nutritional needs of the host organism.

The data indicate that the *Paracatenula* symbiosis is mainly driven by energy-efficient carbon fixation *via* a PP_i-dependent Calvin-Benson-Bassham (CBB) cycle and sulfur oxidation. Core functions shared with other reduced thioautotrophs included genes encoding enzymes for the synthesis of amino acids, nucleosides and nucleotides, vitamins, cofactors and fatty acids. The exchange of these nutrients *via* transporters is not a critical process. Similar to the symbionts of the vesicomyid clams, *Ca.* R. standrea contains only few genes encoding substrate transporters when compared to other chemosynthetic symbionts with non-reduced genomes. Instead, the synthesized nutrients are more likely to be provided to the host through outer membrane vesicles (OMVs) or lysosomal digestion of symbionts – the exact mechanisms need to be addressed in future studies.

The *Ca.* R. standrea symbionts have a unique and powerful set of additional metabolic tools compared to other chemosynthetic symbionts with reduced genomes. These features include genes for enzymes that constitute a complete intermediary carbon metabolism with different ways of incorporating carbon into biomass. The corresponding genes were missing in other reduced thiotrophic symbionts (Kuwahara *et al.*, 2007; Newton *et al.*, 2007). Additionally and in contrast to the vesicomyid symbionts, *Ca.* R. standrea has the capability of storing carbon by at least three different means, i.e. glycogen, trehalose and polyhydroxyalkanoates (PHA). These storage compounds allow a wider metabolic flexibility. To which dimensions this occurs still needs to be investigated.

What might be the importance of storage compounds in *Ca.* Riegeria?

Chemosynthetic meiofaunal residents face challenges in their natural environment as oxygen and sulfide are both spatially distributed (Fenchel and Riedl, 1970). Storage compounds generally represent physiological buffers that are either produced and accumulated or utilized, depending on the availability of carbon and energy sources (Wilkinson, 1959). The *Ca.* R. standrea symbiont can build up a variety of storage compounds of which sulfur and PHA are the most prominent components. These storages can be easily identified as large inclusions (1

to 3 µm in diameter) in the bacterial cells. During the cultivation process of *Paracatenula*, the flatworms constantly migrated through the redoxcline and rested in the anoxic sediment layer for longer times (own observation). This leads to the speculation that under these anoxic conditions, environmental sulfide diffuses into the bacterial cells and is being oxidized, forming elemental sulfur granules (Figure 1). As oxygen is absent, the formation of PHA could serve as an electron sink for sulfide oxidation as previously proposed for *Thiosymbion,* the gammaproteobacterial thiotrophic symbionts of gutless oligochaetes (Kleiner, 2012). Under oxic conditions, stored sulfur could be oxidized to sulfate, and by this, provide energy for carbon fixation. The fixed carbon is either stored as glycogen or as trehalose. Together with these additional storage compounds, the *Ca.* Riegeria symbionts could shift their carbon intermediates and adapt to the altered environmental redox conditions they are exposed to.
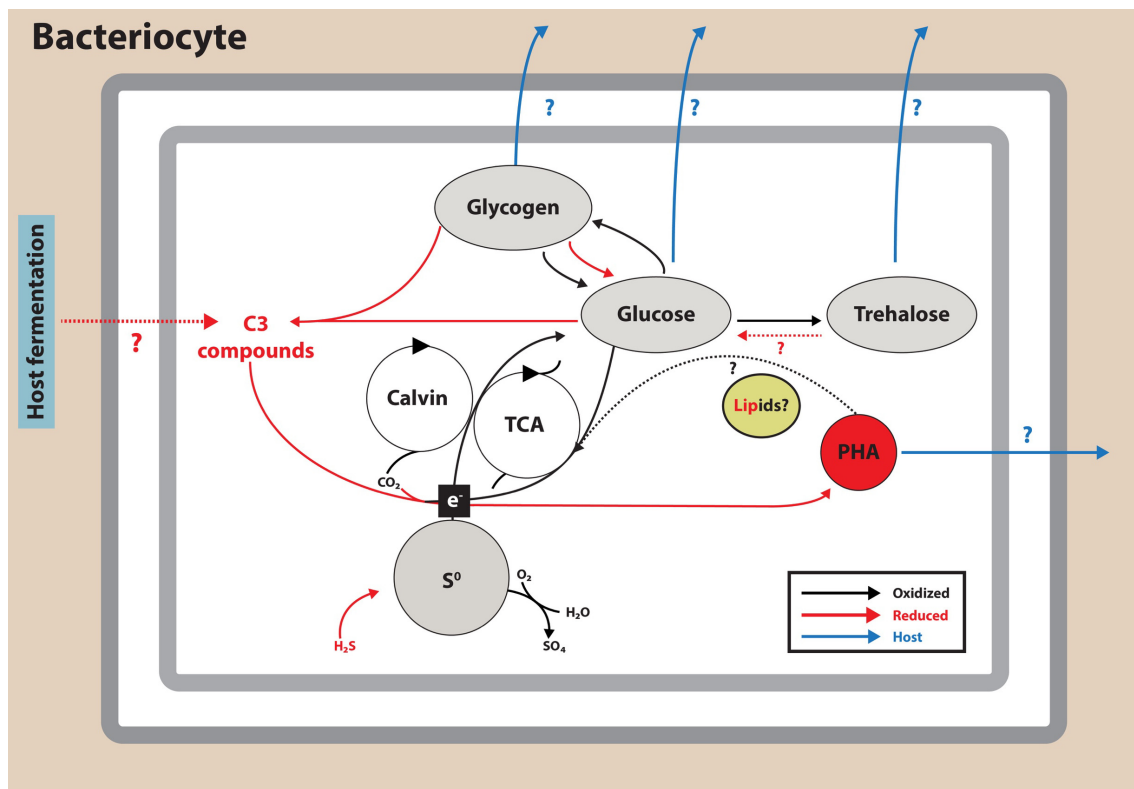
Figure 1. Energy buffering in *Ca.* Riegeria and possible exchange with its host. A schematic model of interactions and metabolism under oxidized (black) and reduced (red) conditions. A selection of possible symbiont carbon substrates that could be utilized by the host are indicated in blue. The figure was adapted from Chapter 4.

It still remains unclear what additional functions these storage compounds might play in the *Paracatenula* symbiosis. It can be assumed that both trehalose and PHA serve as nutrition. While the symbionts lack enzymes for the degradation of trehalose to form glucose monomers,

the host encodes a trehalose synthesis enzyme (TreS, own preliminary data) with bidirectional function previously demonstrated in *C. glutamicum* (Carpinelli *et al.*, 2006). Alternatively or in addition, trehalose could simply function as a protective from osmotic stress (Zhang *et al.*, 2011). However, we could yet not confirm the function and both explanations are not exclusive (Chapter 2).

My results are consistent with the proposal that the *Paracatenula* symbionts, including their storage content, will be digested by lysosomes under starvation conditions (Chapter 2). Symbiont digestion has also been suggested to occur in chemosynthetic symbioses such as the clams *Lucina*, *Bathymodiolus* and *Calyptogena* (Fiala-Médioni *et al.*, 1986; Liberge *et al.*, 2001; Newton *et al.*, 2007). In fresh *Paracatenula*, however, lysosomes were not abundant but host gene expression of transcripts related to lysosome formation, trafficking and proteolysis were indicative for ongoing digestion procedures. Besides the rare occasion of digestion, we could find evidence for OMV formation in the symbiont membrane based on symbiont expression data and TEM. As OMVs are known sources of for example carbon and nitrogen we propose that they represent the main mode of how *Paracatenula* gets its nutrition from the symbionts (Schwechheimer and Kuehn, 2015; Lynch and Alegado, 2017).

In one of the lysosomes of a starving *Paracatenula*, an electron-dense structure that resembled the PHA granules that were typically detected inside intact symbiont cells could be found using TEM (Chapter 2). It is noteworthy that no gene encoding a PHA depolymerase for the degradation of these granules could be found in the genome of *Ca.* R. standrea. It is, however, unlikely that these energy and carbon rich compounds represent only remnants of symbiont digestion in an energy-efficient and optimized symbiosis. The degradation of PHA might therefore either occur *via* unknown enzymatic reactions by the symbionts or by enzymes encoded by the host. This way, the substrate could deliver both energy and carbon. Such a scenario has not been demonstrated yet. However, bioinformatic analysis of the *Paracatenula* gene sequences predicts four kinds of PHA depolymerases in the host transcriptome (preliminary data). It remains to be analyzed whether these sequences represent contaminations with bacterial DNA or whether they are truly transcripts encoded by the host. This could be done by for example phylogenetic analyses and by the visualization of gene expression using mRNA *in situ* hybridization to localize the transcripts within the host.

Furthermore, it remains to be shown whether the predicted enzymes are functional according to the bioinformatic prediction.

In addition to having a possible function as carbon and energy storage, there is increasing evidence for additional physiological roles of the polymer PHA. These functions include the facilitated uptake of DNA by the formation of complexes with calcium and polyphosphate (Madison and Huisman, 1999) which have been described to participate in stress response (Pham *et al.*, 2004) and in the retention of efficient energy metabolism (Escapa *et al.*, 2012). Typically, high intracellular charges are characterized by the presence of acetyl-CoA, NAD(P)H and ATP which all boost the formation of PHA (Haywood *et al.*, 1988; Uchino *et al.*, 2007; Narancic *et al.*, 2016). Moreover, the presence of stored PHA [specifically polyhydroxybutyrate (PHB)] has been reported to stimulate the assimilation of acetate into biomass instead of further into PHB in the bacterium *Rhodospirillum rubrum* (Narancic *et al.*, 2016). Additionally, an ongoing assimilation of PHB has been shown to cause a decreased activity of RuBisCO under anaerobic conditions, contradicting the typical known carbon fixation activities under such conditions (Narancic *et al.*, 2016). Altogether, these findings on the PHA metabolism in *R. rubrum* are consistent with our results on the *Ca.* Riegeria symbionts and could indicate an additional function of PHA for the following reasons:

(i) PHA formation in *Ca.* R. standrea is likely to occur under anaerobic conditions to form sulfur granules from sulfide for its complete oxidation *via* the rDSR pathway (Figure 1),

(ii) *Ca.* R. standrea expresses a RuBisCO form IA that is adapted to higher oxygen and lower carbon dioxide concentrations (Chapter 2). Thus, carbon fixation rather occurs under oxic conditions,

(iii) there is no unambiguous evidence yet for the degradation of PHA granules from neither the symbionts nor the host.

Nevertheless, further studies are needed to address possible additional functions of the diverse storage compounds in the *Paracatenula* symbiosis.

Is the *Paracatenula-Ca.* Riegeria symbiosis a "dead-end street"?

Paracatenula has co-evolved with its endosymbionts for hundreds of millions of years (Gruber-Vodicka *et al.*, 2011) which resulted in a significant genome reduction in its symbionts (Chapter 2). Such a close and obligate long-term symbioses of two organism often results in highly specialized interactions between the two partners (e.g. McCutcheon *et al.*, 2009; McCutcheon and Moran, 2012; Bennett and Moran, 2013). In bacterial insect symbionts, the processes of reductive genome evolution frequently cause tiny symbiont genomes down to sizes below 0.2 Mb (McCutcheon and Moran, 2012; Bennett and Moran, 2013; Moran and Bennett, 2014; Anbutsu *et al.*, 2017). Because the genome sizes usually correlate with the number of coding sequences, these organisms end up with ten times lower gene numbers (~140) as also observed with the *Ca.* Riegeria symbionts (Lynch, 2006; Bennett and Moran, 2013) (Chapter 2, 3). Symbionts of the leafhopper *Macrosteles quadrilineatus* called *Ca.* Nasuia bacteria encode the enzymatic pathway of two amino acids and components for DNA replication, transcription and translation while genes e.g. for ATP synthesis were lost in streamlining processes (Bennett and Moran, 2013). In the case of the *Nardonella*, symbionts that are associated with weevils, only genes for the synthesis of the amino acid tyrosine were retained in its highly reduced genome (Anbutsu *et al.*, 2017). In both examples, the symbionts co-evolved with their host for 100–260 million years (Bennett and Moran, 2013; Anbutsu *et al.*, 2017). During this process, both symbiont genomes underwent modifications and reductions, down to a point that they fail to support cellular life outside their hosts or without a different co-existing symbiont (McCutcheon and Moran, 2012). In the case of the leafhopper, a co-primary symbiont exists with *Ca.* Nasuia which complements the insufficient function of the other symbiont (Bennett and Moran, 2013). Such a scenario has not been reported for the weevil symbionts, but it may potentially be established by the infection through a secondary facultative symbiont or by gut associates (Anbutsu *et al.*, 2017). The metabolic complementation is thought to provide essential functions to overcome a collapsing scenario resulting in reduced fitness of the host (Andersson, 2006; Pérez-Brocal *et al.*, 2006).

According to the genome streamlining theory, smaller genomes provide more adaptive advantages than larger genomes as every function that loses its beneficial effect brings a cost (Ochman and Moran, 2001; Giovannoni *et al.*, 2005). This concept was extended by the Black

Queen Hypothesis, addressing the evolution of dependencies in a given population that is based on the loss of costly functions (= loss of genes), and by the principle of "common and public goods" (Morris *et al.*, 2012; Mas *et al.*, 2016). In insect symbiont communities, functions can thus be lost in one symbiont genotype and complemented by a second symbiont. *Ca.* Riegeria represents the "Black Queen genotype" and the symbiont must provide most if not all nutritional functions for the host. Its genome is under large selection pressure to retain various metabolic functions to maintain its symbiotic partnership with *Paracatenula*. Also, an ongoing gene loss towards functional organelles is rather unlikely in *Ca.* Riegeria for the following reasons:

(i) The symbionts have evolved with their host for a long period of supposedly 500 million years and they retained a broad versatile metabolic toolkit (Chapter 2, 3). Despite this long period of co-evolution, the *Ca.* Riegeria symbionts have not yet undergone an extensive genome reduction as observed with insect symbionts during much shorter periods of co-evolution. Also, *Ca.* Riegeria shows no signs of losing symbiotic capacities (Chapter 3).

(ii) The *Ca.* Riegeria symbionts have to provide more extensive functions than the insect symbionts since their host is unable to feed. In other words, *Paracatenula* relies nutritionally mostly or completely on its symbionts, whereas the insect symbionts often only supplement the diet of the host with vitamins or amino acids (McCutcheon and Moran, 2012). The continuous additional loss of metabolic functions in *Ca.* Riegeria would therefore most likely affect the fitness of the host in a negative way.

(iii) An uptake of a secondary symbiont by *Paracatenula* that takes over lost functions as in insect symbionts is unlikely, since there is no intensive exchange with their environment. Additionally, *Paracatenula* is less likely than freely movable terrestrial insects to move large distances which could prevent significant changes of the habitat.

What is the role of symbionts in tissue regeneration?

In animals, bacteria are known to support wound healing and tissue regenerating processes (Poutahidis *et al.*, 2013; Arnold *et al.*, 2016; Lukic *et al.*, 2017). For example in mammalian wounds, the provision of a lactic acid bacterium such as *Lactobacillus reuteri* as probiotics has been noted to enhance the healing by upregulating the synthesis of the hormone oxytocin and by initiating downstream reactions (Poutahidis *et al.*, 2013). Furthermore, a shift in the healthy

microbiome of the flatworm *Schmidtea mediterraneae* has been shown to strongly affect tissue regeneration and results in negative effects (Arnold *et al.*, 2016). However, the exact mechanisms and interactions underlying this phenomenon are not yet understood.

Ca. R. standrea displays nutritional functions in *Paracatenula* flatworms under environmental conditions (Chapter 2). When the flatworms perform their most fascinating feature, i.e. the regeneration of body parts, 168 of their genes were found to be up- or down-regulated (Chapter 4). Most of the upregulated genes were related to energy and carbon metabolism as well as transcription and translation. Our findings suggest an increased biomass production by the symbionts in regenerating *Paracatenula* flatworms. The simultaneous operation of carbon fixation *via* the PP_i-dependent CBB cycle and the utilization of organic compounds through the tricarboxcylic acid (TCA) cycle might indicate a need for the synthesis of intermediary metabolites required for the synthesis of amino acid, vitamins or cofactors. Yet, it is uncertain what metabolites produced by the symbionts are essential for the regeneration process or whether a cocktail of nutrients is needed. In addition of satisfying the nutritional needs of the host, the findings suggest underlying cell signaling and communication pathways as described for the *Schmidtea mediterranea* flatworm (Arnold *et al.*, 2016).

Perspective

Understanding genome reduction in *Ca.* Riegeria clades

The presence of different genome reduction stages is one of the most puzzling observations in the *Ca.* Riegeria clades (Chapter 3). Occasions of parallel genome reduction processes have been reported within symbiont clades which result in a different genetic makeup but without drastic differences in the overall genome architecture such as genome rearrangements (Williams and Wernegreen, 2015; Boscaro *et al.*, 2017; Kinjo *et al.*, 2018). It can be hypothesized that different genes retained in the *Ca.* Riegeria clades might be the reason for such variations, but finding an answer to this question by experimental evidence would compare to the search of a needle in a haystack.

To address the evolutionary events that have occurred in the three *Ca.* Riegeria clades, it is critical to reconstruct the last common ancestral (LCA) genome of *Ca.* Riegeria. Ancestral genome reconstructions are often applied to understand what traits in a genome are ancestral and which genomic constellation was present within the LCA genome. Such an analysis

provides information concerning the genes which were derived when the organisms have transitioned and adapted to new or changing environments (Schluter *et al.*, 1997; Larsson *et al.*, 2011; Latysheva *et al.*, 2012; Oyserman *et al.*, 2016). Also, studies exploring gene gain and gene loss situations can be performed to illuminate the influence of horizontal gene transfer (HGT) events that is known to contribute to the enrichment of metabolic tools (Pál *et al.*, 2005). HGT events were indeed likely to have occurred during early stages of the *Paracatenula-Ca.* Riegeria symbiosis, when lateral exchanges with the environment were still feasible.

Reconstructing the LCA genome of *Ca.* Riegeria symbionts would further help to illuminate which metabolic frameworks were gained from the environment and maintained to establish *Ca.* Riegeria as a successful and most suitable partner within the *Paracatenula* symbiosis. Such a detailed analysis would help to better understand the presence of three stages of genome reduction in the *Ca.* Riegeria clades as well as their trajectories (Chapter 3).

Single-nucleotide polymorphism (SNPs) and evolutionary pressures in *Ca.* Riegeria symbionts

The processes of genomic changes and reductive genome evolution in intracellular symbionts are mainly driven by transmission bottlenecks, mutations and genetic drift (McCutcheon and Moran, 2012). However, during these processes, gene loss is not completely random and genes that encode essential functions are usually conserved (Moya *et al.*, 2008; McCutcheon and Moran, 2012). Until now, the rate of evolution of the diverse *Ca.* Riegeria symbionts is unknown, although it can be assumed that their genes also accumulate slightly deleterious substitutions (Chapter 3). Single-nucleotide polymorphisms (SNPs) among bacteria within a population as well as between populations can be used as sensitive measure to detect small genomic changes and to possibly take them as indicators of directional selection. The investigation of microdiversity at the level of nucleotides and single genes can help to enlighten the evolutionary history and selective pressures during the evolution of symbiont populations (Syvänen, 2001). In *Ca*. Riegeria symbionts, SNP analyses can establish the mutation rates of single genes among the different symbiont species. For instance, such an analysis can be performed on symbionts of individuals of the offspring from a single *Paracatenula* sp. standrea species that was cultured under defined conditions in the laboratory over a time period of more than three years.

Alternatively, SNP analyses can also be applied to detect evolutionary pressures of single genes among closely related *Ca.* Riegeria species, e.g. those that belong to the same clade, or between the clades, to identify fixed differences in their genomes. The number, frequency and the type of SNPs will allow to identify candidate genes which are subject to a negative selection pressure ("purifying selection"), neutral selection pressure ("non-effective") or a positive selection pressure ("adaptive selection"). The SNP type then determines whether the mutation leads to a change in the encoded amino acid sequence ("non-synonymous", N) of the protein product or whether the substitution translates into the same amino acid ("synonymous", S). The ratio of these SNP types can be used to calculate ratios such as dN/dS values between species and pN/pS values between populations that are indicative for either rapid (high ratio) or slow protein evolution for example caused by deleterious mutations. These values are indicative of the direction of selection applying, for example, McDonald-Kreitman-based tests (Vos *et al.*, 2013). Furthermore, SNP frequencies indicate loci representing local adaptation across populations (Pfeifer and Lercher, 2018). Taken together, these analyses would elucidate if *Ca.* Riegeria genes are under selective pressure and might be modified or even lost in the ongoing processes of genome evolution.

Function of symbiont subpopulations and the host in tissue regeneration

Symbionts showed significant differences in the expression patterns five days after host rostrum sectioning (Chapter 4). These differences suggest that metabolites produced by the symbionts participate or are even essential for the regeneration process in *Paracatenula*. However, due to limitations in the experimental setup, which only allows to scratch at the surface of the regeneration phenomenon, some questions remained unanswered.

The results shown here only indicate that the symbionts are essential for rostrum regeneration and that there are changes in the expression profiles of a subset of symbiont genes. However, these results do not reveal (i) possible differential gene activities of distinct symbiont subpopulations along the body axis of the flatworm as well as (ii) different temporal activities of the symbionts or symbiont populations during the temporal progression of the tissue regeneration process. Thus, further experiments are required to explore the importance of the symbionts and to determine the function of the differentially expressed genes which contribute to the metabolically demanding regeneration process.

In the presented study, I have analyzed the differentially expressed symbiont transcripts five days after sectioning and in the process of rostrum regeneration. However, since the regeneration processes start immediately after sectioning, it is possible that major steps and expression changes have been missed. Thus, I propose a continuous sampling of individuals from early stages up to eleven days when the rostrum regeneration is mostly completed. Without such a data collection, it is difficult to assign gene functions and to separate them according to cause and consequences. Also, it appears challenging to determine the function of the genes, and their contribution to possible steps of the regeneration process. Finally, it could also be possible that anterior and posterior trophosome regions contain subsets of the symbiont population with different gene activities in general, or in response to sectioning and during regeneration. Despite these uncertainties, it remains interesting that after five days, which is about half the time until the rostrum is completely regenerated, the expression patterns of the symbionts have changed.

Since symbiont-free flatworms in the absence of a trophosome are also able to regenerate an entire flatworm, it is unlikely that the symbionts contribute any "regeneration factors" or morphogens to the host development. Therefore, it is more likely that the regeneration process of *Paracatenula* requires both energy supply and substrates for biosynthesis which are supplied by their symbionts. The observed differential expression of metabolic pathway components by the symbionts during the regeneration process of *Paracatenula* is consistent with this conclusion. Future studies addressing the mentioned gaps should therefore aim towards the temporal and spatial profiles of the differentially expressed genes and establishing their function. A correlative study concerning the expression profiles can be performed with the currently available methodology, whereas the assessment of the gene functions has to await the development of new tools which make gene functions even more accessible. As *Paracatenula* can be kept in the laboratory its paves the way into this direction, provided that the genome of the symbionts or the host can be genetically manipulated by applying for example RNA interference knockdown or knock-out experiments.

One aspect that needs further investigation is the contribution of the host to its rostrum regeneration. In fact, bioinformatic tools allow the assembly of host transcriptomes (Chapter 2) and the investigation of the genes that are differentially expressed. These tools also allow the comparison with other flatworm transcriptomes which are not known to carry nutritional

symbionts such as planarian flatworms of the genus *Schmidtea* (Levin *et al.*, 2016; Stückemann *et al.*, 2017) or *Catenula*, a close relative of *Paracatenula* (Ngamniyom and Panyarachun, 2016). Underlying signaling pathways that are involved in other tissue-regenerating flatworms might be affected in the presence and absence of a trophosome region as found in *Paracatenula*. Additionally, such a comparison is of particular interest since the upregulation of nutrition-related genes, such as transporters and degrading enzymes, might give hints to key roles of metabolites which are provided by the *Paracatenula* symbionts.

Bibliography Discussion

Anbutsu, H., Moriyama, M., Nikoh, N., Hosokawa, T., Futahashi, R., Tanahashi, M., Meng, X.-Y., Kuriwada, T., Mori, N., Oshima, K., Hattori, M., Fujie, M., Satoh, N., Maeda, T., Shigenobu, S., Koga, R. and Fukatsu, T. (2017) 'Small genome symbiont underlies cuticle hardness in beetles', *Proceedings of the National Academy of Sciences of the United States of America*, **11440**, pp. E8382–E8391.

Andersson, S. G. E. (2006) 'The bacterial world gets smaller', *Science*, **314**, pp. 259–260.

Arnold, C. P., Merryman, M. S., Harris-Arnold, A., McKinney, S. A., Seidel, C. W., Loethen, S., Proctor, K. N., Guo, L. and Sánchez Alvarado, A. (2016) 'Pathogenic shifts in endogenous microbiota impede tissue regeneration *via* distinct activation of TAK1/MKK/p38', *eLife*, **5**, e16793.

Bennett, G. M. and Moran, N. A. (2013) 'Small, smaller, smallest: The origins and evolution of ancient dual symbioses in a phloem-feeding insect', *Genome Biology and Evolution*, **5**, pp. 1675–1688.

Boscaro, V., Kolisko, M., Felletti, M., Vannini, C., Lynn, D. H. and Keeling, P. J. (2017) 'Parallel genome reduction in symbionts descended from closely related free-living bacteria', *Nature Ecology & Evolution*, **1**, pp. 1160–1167.

Carpinelli, J., Krämer, R. and Agosin, E. (2006) 'Metabolic engineering of *Corynebacterium glutamicum* for trehalose overproduction: Role of the TreYZ trehalose biosynthetic pathway', *Applied and Environmental Microbiology*, **72**, pp. 1949–1955.

Cavanaugh, C. M. (1983) 'Symbiotic chemoautotrophic bacteria in marine invertebrates from sulphide-rich habitats', *Nature*, **302**, pp. 58–61.

Dirks, U., Gruber-Vodicka, H. R., Egger, B. and Ott, J. A. (2012) 'Proliferation pattern during rostrum regeneration of the symbiotic flatworm *Paracatenula galateia*: A pulse-chase-pulse analysis', *Cell and tissue research*, **349**, pp. 517–525.

Dirks, U., Gruber-Vodicka, H. R., Leisch, N., Bulgheresi, S., Egger, B., Ladurner, P. and Ott, J. A. (2012) 'Bacterial symbiosis maintenance in the asexually reproducing and regenerating flatworm *Paracatenula galateia*', *PLoS ONE*, **7**, e34709.

Dirks, U., Gruber-Vodicka, H. R., Leisch, N., Sterrer, W. and Ott, J. A. (2011) 'A new species of symbiotic flatworms, *Paracatenula galateia* sp. nov. (*Platyhelminthes*: *Catenulida*: Retronectidae) from Belize (Central America)', *Marine Biology Research*, **7**, pp. 769–777.

Escapa, I. F., García, J. L., Bühler, B., Blank, L. M. and Prieto, M. A. (2012) 'The polyhydroxyalkanoate metabolism controls carbon and energy spillage in *Pseudomonas putida*', *Environmental Microbiology*, **14**, pp. 1049–1063.

Fenchel, T. M. and Riedl, R. J. (1970) 'The sulfide system: a new biotic community underneath the oxidized layer of marine sand bottoms', *Marine Biology*, **7**, pp. 255–268.

Fiala-Médioni, A., Métivier, C., Herry, A. and Le Pennec, M. (1986) 'Ultrastructure of the gill of the hydrothermal-vent mytilid *Bathymodiolus* sp.', *Marine Biology*, **92**, pp. 65–72.

Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., Bibbs, L., Eads, J., Richardson, T. H., Noordewier, M., Rappe, M. S., Short, J. M., Carrington, J. C. and Mathur, E. J. (2005) 'Genome streamlining in a cosmopolitan oceanic bacterium', *Science*, **309**, pp. 1242–1245.

Gruber-Vodicka, H. R., Dirks, U., Leisch, N., Baranyi, C., Stoecker, K., Bulgheresi, S., Heindl, N. R., Horn, M., Lott, C., Loy, A., Wagner, M. and Ott, J. (2011) '*Paracatenula*, an ancient symbiosis between thiotrophic *Alphaproteobacteria* and catenulid flatworms', *Proceedings of the National Academy of Sciences of the United States of America*, **108**, pp. 12078–12083.

Haywood, G. W., Anderson, A. J., Chu, L. and Dawes, E. A. (1988) 'The role of NADH- and NADPH-linked acetoacetyl-CoA reductases in the poly-3-hydroxybutyrate synthesizing organism Alcaligenes eutrophus', *FEMS Microbiology Letters*, **52**, pp. 259–264.

Kinjo, Y., Bourguignon, T., Tong, K. J., Kuwahara, H., Lim, S. J., Yoon, K. B., Shigenobu, S., Park, Y. C., Nalepa, C. A., Hongoh, Y., Ohkuma, M., Lo, N. and Tokuda, G. (2018) 'Parallel and gradual genome erosion in the *Blattabacterium* endosymbionts of *Mastotermes darwiniensis* and *Cryptocercus* wood roaches', *Genome Biology and Evolution*, **10**, pp. 1622–1630.

Kleiner, M. (2012) *Metabolism and evolutionary ecology of chemosynthetic symbionts from marine invertebrates*, *PhD Thesis*. Universität Bremen.

Kleiner, M., Petersen, J. M. and Dubilier, N. (2012) 'Convergent and divergent evolution of metabolism in sulfur-oxidizing symbionts and the role of horizontal gene transfer', *Current Opinion in Microbiology*, **15**, pp. 621–631.

Kuwahara, H., Yoshida, T., Takaki, Y., Shimamura, S., Nishi, S., Harada, M., Matsuyama, K., Takishita, K., Kawato, M., Uematsu, K., Fujiwara, Y., Sato, T., Kato, C., Kitagawa, M., Kato, I. and Maruyama, T. (2007) 'Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*', *Current Biology*, **17**, pp. 881–886.

Larsson, J., Nylander, J. A. A. and Bergman, B. (2011) 'Genome fluctuations in cyanobacteria reflect evolutionary, developmental and adaptive traits', *BMC Evolutionary Biology*, **11**, 187.

Latysheva, N., Junker, V. L., Palmer, W. J., Codd, G. A. and Barker, D. (2012) 'The evolution

of nitrogen fixation in cyanobacteria', *Bioinformatics*, **28**, pp. 603–606.

Leisch, N., Dirks, U., Gruber-Vodicka, H. R., Schmid, M., Sterrer, W. and Ott, J. A. (2011) 'Microanatomy of the trophosome region of *Paracatenula* cf. *polyhymnia* (*Catenulida*, *Platyhelminthes*) and its intracellular symbionts', *Zoomorphology*, **130**, pp. 261–271.

Levin, M., Anavy, L., Cole, A. G., Winter, E., Mostov, N., Khair, S., Senderovich, N., Kovalev, E., Silver, D. H., Feder, M., Fernandez-Valverde, S. L., Nakanishi, N., Simmons, D., Simakov, O., Larsson, T., Liu, S.-Y., Jerafi-Vider, A., Yaniv, K., Ryan, J. F., Martindale, M. Q., Rink, J. C., Arendt, D., Degnan, S. M., Degnan, B. M., Hashimshony, T. and Yanai, I. (2016) 'The mid-developmental transition and the evolution of animal body plans', *Nature*, **531**, pp. 637–641.

Liberge, M., Gros, O. and Frenkiel, L. (2001) 'Lysosomes and sulfide-oxidizing bodies in the bacteriocytes of *Lucina pectinata*, a cytochemical and microanalysis approach', *Marine Biology*, **139**, pp. 401–409.

Lukic, J., Chen, V., Strahinic, I., Begovic, J., Lev-Tov, H., Davis, S. C., Tomic-Canic, M. and Pastar, I. (2017) 'Probiotics or Pro-healers the role of beneficial bacteria in tissue repair', *Wound Repair Regeneration*, **25**, pp. 912–922.

Lynch, J. B. and Alegado, R. A. (2017) 'Spheres of hope, packets of doom: The good and bad of outer membrane vesicles in interspecies and ecological dynamics', *Journal of Bacteriology*, **199**, e00012-17.

Lynch, M. (2006) 'Streamlining and simplification of microbial genome architecture', *Annual Review of Microbiology*, **60**, pp. 327–349.

Madison, L. L. and Huisman, G. W. (1999) 'Metabolic engineering of poly(3-hydroxyalkanoates): from DNA to plastic', *Microbiology and Molecular Biology Reviews*, **63**, pp. 21–53.

Mas, A., Jamshidi, S., Lagadeuc, Y., Eveillard, D. and Vandenkoornhuyse, P. (2016) 'Beyond the Black Queen Hypothesis', *The ISME Journal*, **10**, pp. 2085–2091.

McCutcheon, J. P., McDonald, B. R. and Moran, N. A. (2009) 'Convergent evolution of metabolic roles in bacterial co-symbionts of insects', *Proceedings of the National Academy of Sciences of the United States of America*, **106**, pp. 15394–15399.

McCutcheon, J. P. and Moran, N. A. (2012) 'Extreme genome reduction in symbiotic bacteria', *Nature Reviews Microbiology*, **10**, pp. 13–26.

Moran, N. A. and Bennett, G. M. (2014) 'The tiniest tiny genomes', *Annual Review of Microbiology*, **68**, pp. 195–215.

Morris, J. J., Lenski, R. E. and Zinser, E. R. (2012) 'The Black Queen Hypothesis: Evolution of dependencies through adaptative gene loss', *Mbio*, **3**, pp. 1–7.

Moya, A., Peretó, J., Gil, R. and Latorre, A. (2008) 'Learning how to live together: genomic insights into prokaryote–animal symbioses', *Nature Reviews Genetics*, **9**, pp. 218–229.

Narancic, T., Scollica, E., Kenny, S. T., Gibbons, H., Carr, E., Brennan, L., Cagney, G., Wynne,

K., Murphy, C., Raberg, M., Heinrich, D., Steinbüchel, A. and O'Connor, K. E. (2016) 'Understanding the physiological roles of polyhydroxybutyrate (PHB) in *Rhodospirillum rubrum* S1 under aerobic chemoheterotrophic conditions', *Applied Microbiology and Biotechnology*, **100**, pp. 8901–8912.

Newton, I. L. G., Woyke, T., Auchtung, T. A., Dilly, G. F., Dutton, R. J., Fisher, M. C., Fontanez, K. M., Lau, E., Stewart, F. J., Richardson, P. M., Barry, K. W., Saunders, E., Detter, J. C., Wu, D., Eisen, J. A. and Cavanaugh, C. M. (2007) 'The *Calyptogena magnifica* chemoautotrophic symbiont genome', *Science*, **315**, pp. 998–1000.

Ngamniyom, A. and Panyarachun, B. (2016) '*Stenostomum* cf. *leucops* (*Platyhelminthes*) in Thailand: a surface observation using scanning electron microscopy and phylogenetic analysis based on 18S ribosomal DNA sequences', *Songklanakarin Journal of Science and Technology*, **38**, pp. 41–45.

Nussbaumer, A. D., Fisher, C. R. and Bright, M. (2006) 'Horizontal endosymbiont transmission in hydrothermal vent tubeworms', *Nature*, **441**, pp. 345–348.

Ochman, H. and Moran, N. A. (2001) 'Genes lost and genes found: Evolution of bacterial pathogenesis', *Ecology and evolution of infection*, **292**, pp. 1096–1098.

Ott, J., Rieger, G., Rieger, R. and Enderes, F. (1982) 'New mouthless interstitial worms from the sulfide system: Symbiosis with prokaryotes', *Marine Ecology*, **3**, pp. 313–333.

Oyserman, B. O., Moya, F., Lawson, C. E., Garcia, A. L., Vogt, M., Heffernen, M., Noguera, D. R. and McMahon, K. D. (2016) 'Ancestral genome reconstruction identifies the evolutionary basis for trait acquisition in polyphosphate accumulating bacteria', *The ISME Journal*, **10**, pp. 2931–2945.

Pál, C., Papp, B. and Lercher, M. J. (2005) 'Adaptive evolution of bacterial metabolic networks by horizontal gene transfer', *Nature Genetics*, **37**, pp. 1372–1375.

Le Pennec, M., Beninger, P. G. and Herry, A. (1995) 'Feeding and digestive adaptations of bivalve molluscs to sulphide-rich habitats', *Comparative Biochemistry and Physiology*, **111A**, pp. 183–189.

Pérez-Brocal, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J. M., Silva, F. J., Moya, A. and Latorre, A. (2006) 'A small microbial genome: the end of a long symbiotic relationship?', *Science*, **314**, pp. 312–313.

Petersen, J. M., Kemper, A., Gruber-Vodicka, H., Cardini, U., van der Geest, M., Kleiner, M., Bulgheresi, S., Mußmann, M., Herbold, C., Seah, B. K. B., Antony, C. P., Liu, D., Belitz, A. and Weber, M. (2016) 'Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation', *Nature Microbiology*, **2**, 16195.

Pfeifer, B. and Lercher, M. J. (2018) 'BlockFeST: Bayesian calculation of region-specific F_{ST} to detect local adaptation', *Bioinformatics*, **34**, pp. 3205–3207.

Pham, T. H., Webb, J. S. and Rehm, B. H. A. (2004) 'The role of polyhydroxyalkanoate

biosynthesis by *Pseudomonas aeruginosa* in rhamnolipid and alginate production as well as stress tolerance and biofilm formation', *Microbiology*, **150**, pp. 3405–3413.

Poutahidis, T., Kearney, S. M., Levkovich, T., Qi, P., Varian, B. J., Lakritz, J. R., Ibrahim, Y. M., Chatzigiagkos, A., Alm, E. J. and Erdman, S. E. (2013) 'Microbial symbionts accelerate wound healing *via* the neuropeptide hormone oxytocin', *PLoS ONE*, **8**, e78898.

Russell, S. L., Corbett-Detig, R. B. and Cavanaugh, C. M. (2017) 'Mixed transmission modes and dynamic genome evolution in an obligate animal–bacterial symbiosis', *The ISME Journal*, **11**, pp. 1359–1371.

Schluter, D., Price, T., Mooers, A. Ø. and Ludwig, D. (1997) 'Likelihood of ancestor states in adaptive radiation', *Evolution*, **51**, pp. 1699–1711.

Schwechheimer, C. and Kuehn, M. J. (2015) 'Outer-membrane vesicles from gram-negative bacteria: biogenesis and functions', *Nature Reviews Microbiology*, **13**, pp. 605–619.

Sterrer, W. and Rieger, R. (1974) 'Retronectidae-a new cosmopolitan marine family of Catenulida (Turbellaria)', in Riser N, M. M. (eds). (ed.) *Biology of the turbellaria*. McGraw-Hill, New York, pp. 63–92.

Stückemann, T., Cleland, J. P., Werner, S., Thi-Kim Vu, H., Bayersdorf, R., Liu, S.-Y., Friedrich, B., Jülicher, F. and Rink, J. C. (2017) 'Antagonistic self-organizing patterning systems control maintenance and regeneration of the anteroposterior axis in Planarians', *Developmental Cell*, **40**, pp. 248–263.

Syvänen, A. C. (2001) 'Accessing genetic variation: Genotyping single nucleotide polymorphisms', *Nature Reviews Genetics*, **2**, pp. 930–942.

Uchino, K., Saito, T., Gebauer, B. and Jendrossek, D. (2007) 'Isolated poly(3-hydroxybutyrate) (PHB) granules are complex bacterial organelles catalyzing formation of PHB from acetyl coenzyme A (CoA) and degradation of PHB to acetyl-CoA', *Journal of Bacteriology*, **189**, pp. 8250–8256.

Vos, M., te Beek, T. a H., van Driel, M. a., Huynen, M. a., Eyre-Walker, A. and van Passel, M. W. J. (2013) 'ODoSE: A webserver for genome-wide calculation of adaptive divergence in prokaryotes', *PLoS ONE*, **8**, pp. 8–11.

Wilkinson, J. F. (1959) 'The problem of energy-storage compounds in bacteria', *Experimental Cell Research*, **7**, pp. 111–130.

Williams, L. E. and Wernegreen, J. J. (2015) 'Genome evolution in an ancient bacteria-ant symbiosis: parallel gene loss among *Blochmannia* spanning the origin of the ant tribe Camponotini', *PeerJ*, **3**, e881.

Zhang, R., Pan, Y. T., He, S., Lam, M., Brayer, G. D., Elbein, A. D. and Withers, S. G. (2011) 'Mechanistic analysis of trehalose synthase from *Mycobacterium smegmatis*', *Journal of Biological Chemistry*, **286**, pp. 35601–35609.

Acknowledgments

I wish to thank the following people that contributed to my doctoral thesis:

Prof. Nicole Dubilier, for giving me the opportunity to perform my studies on the *Paracatenula* system, for always giving critical and constructive feedback and promoting me to tease out the best from my studies, be it in written format or in presentations.

Harald Gruber-Vodicka, for giving me the chance to work with you on the fascinating *Paracatenula* symbiosis, for all the suggestions you were giving me over the years we have spent together, especially on how to explain things in simple ways, for all the fruitful discussions we had (on weekends, in late evenings, …) and for being such a great advisor, mentor and (table) tennis partner.

Prof. Andreas Schramm, for kindly agreeing to be part of my examination commission and taking the time to evaluate the work I have done.

Prof. Karlheinz Altendorf, for being part of my thesis committee and examination commission, for all the discussion we had, for all helpful non-scientific conversations and for being an inspiring lecturer during mar*mic*.

Prof. Michael Friedrich, for kindly agreeing to participate in my examination commission, and the enjoyable lectures during mar*mic*.

Rebecca Ansorge and **Caroline Zeidler**, for being part of my examination commission.

Gunter Wegener and Manuel Liebeke, for being part of my thesis committee and all experimental advice.

The staff at the Hydra Institute on Elba, especially **Miriam Weber**, **Christian Lott**, **Hannah Kuhfuß**, **Matthias Schneider** for support during all the fieldwork that have started in Summer 2014.

The staff at the Carrie Bow Cay field station, especially **Martha**, **Zelda** and **Craig** during our fieldtrip in Summer 2015.

Maria Jung and **Greta Sondej** for all the support in the worm cultivations, the patient, concise and conscientious work you were doing. Without your support it would not have been possible to keep the worms "happy" and alive for now more than three years!

Antony for the discussions on physiology and for being a great colleague. Thank you **Anna** and **Maggie** for your help with RStudio.

Thank you to all TAs from the **Symbiosis Department**, especially **Martina Meyer** and **Wiebke Ruschmeier** for their support in worm cultivations, media preparations and "*Paracatenula* hunting".

I would like to thank the mar*mic* class 2016, you guys made the time fantastic! Special thanks to my ectosymbiont and friend **Burak**, **Karen**, **Tanja**, **Josi** and **Nadine**. A big thank you goes to **Rebecca**, for your friendship, positivity and all the great suggestions you could give me especially in the last phase of my thesis and for proof-reading part of my discussion.

Thank you **Brandon** for your friendship and at the same time a great colleague and fantastic collaborator, who could give me the best advice. Thanks for all our scientific discussions, Vitamalz sessions and jogs, and for the enjoyable time we always had during our Elba fieldtrips.

Thanks to the **best office** I could imagine, thanks **Janine**, **Merle** and **Juliane** for making every working day so enjoyable, for great times and discussions. **Janine** and **Merle**, thanks for proof-reading parts of thesis. **Juliane**, thanks for the great times we spent during our fieldtrips, for all the scientific discussions we had and the numerous questions you answered concerning bioinformatic issues.

Ein besonderer Dank geht an meine **Familie**. Ich danke euch, **Mama** und **Papa** besonders für eure moralische Unterstützung innerhalb dieser Zeit, dass ihr mir den Rücken gestärkt wann immer es möglich war, und besonders **Papa** für all deine hilfreichen Ratschläge. Danke **Daniel** und **Kathi**, dass ich immer auf euch zählen konnte, und danke **Chipsy**, dass du mich immer mit wedelndem Schwänzchen im Kreis rennend begrüßt und zum Lächeln gebracht hast! Ich möchte auch meinem **Opa Heinz** und meiner **Oma Otte** danken - ihr wart immer für mich da wenn ich euch brauchte!

Målin, die wohl meine größte Unterstützung innerhalb der Doktorarbeit gewesen ist. Danke, dass du mir (wissenschaftliche) Ratschläge gegeben hast, auch wenn es manchmal mitten in der Nacht oder das Erste am Morgen gewesen ist. Danke für deine emotionale Stütze, dass du immer aufpasst, dass ich nicht vom Fleisch falle ('Nimmersatt') und für die leckeren Kuchen als Nervennahrung. Danke, dass du jeden Tag so schön machst!

Personal contribution to each manuscript

Manuscript 1:

A chemosynthetic symbiont with a drastically reduced genome serves as primary energy storage in the marine flatworm *Paracatenula*

Experimental concept and design: 50%

Experimental work and/or acquisition of data: 70%

Data analysis and interpretation: 60%

Preparation of figures and tables: 90%

Drafting the manuscript: 70%

Manuscript 2:

Clade-specific patterns of reductive genome evolution in ancient thiotrophic endosymbionts

Experimental concept and design: 50%

Experimental work and/or acquisition of data: 70%

Data analysis and interpretation: 70%

Preparation of figures and tables: 90%

Drafting the manuscript: 70%

Manuscript 3:

The role of endosymbionts in rostrum regeneration of the marine flatworm *Paracatenula* sp. standrea

Experimental concept and design: 30%

Experimental work and/or acquisition of data: 50%

Data analysis and interpretation: 80%

Preparation of figures and tables: 90%

Drafting the manuscript: 90%

Appendix

Mouthless in marine sediments – Symbiosis makes it possible

Oliver Jäckle

Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, Germany

Book chapter

Mouthless in Marine Sediments – Symbiosis Makes it Possible

Oliver Jäckle

When symbiosis researchers from Bremen think about the Caribbean Sea, it is the hidden diversity in these sediments that comes to mind. Only with a microscope is it possible to detect the various organisms of the "meiofauna" which move between sand grains. Almost all animal phyla are represented here, and some are even endemic to this environment. Thousands of tiny worms with an intense white coloration can be discovered in a single litre of sediment. This sediment smells of rotten eggs, indicating the presence of hydrogen sulphide, which also explains

An aerial view of the island Carrie Bow Cay with its Smithsonian Institution research station (light blue). The island is surrounded by different habitats that include algae, seagrass, and coral reef sediments
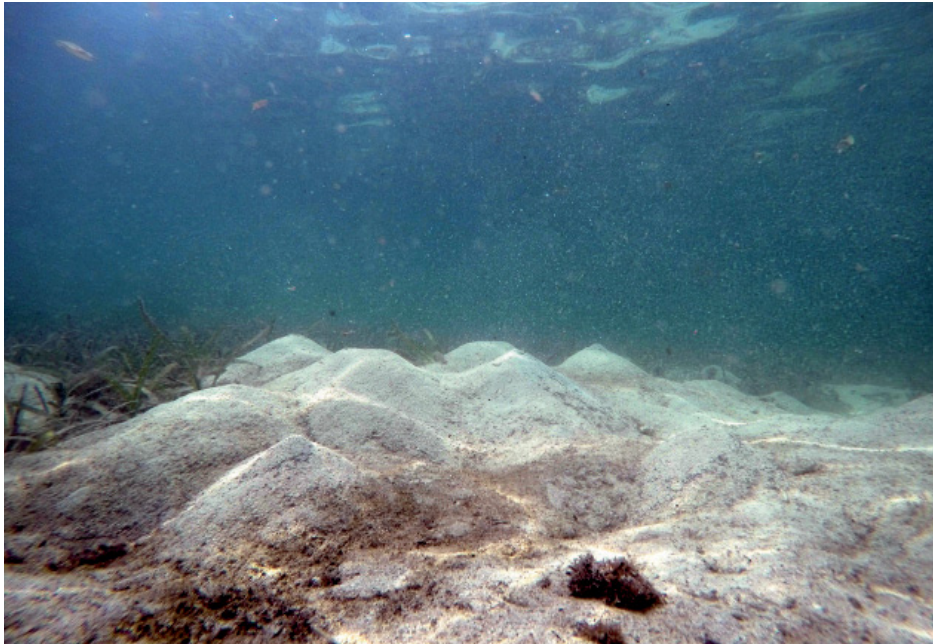
Seagrass meadow adjacent to sediment hills with appreciable amounts of chemosynthetic symbioses

the white coloration of the animals: the worms are densely colonised by bacteria that are capable of storing sulphur in the form of small, spherical granules. Sulphur serves as an energy source for the bacteria in the energy-gaining process of chemosynthesis, whereby sulphide is oxidised in the presence of oxygen. Since oxygen and sulphide do not share the same spatial distribution in the sediment, free-living bacteria may not encounter both reduced and oxidised conditions. Perhaps for this reason, some of the free-living bacteria formed symbiotic relationships with worms a long time ago to use them as elevators to travel through the sediment's redox gradient. To compensate for this "shuttle service", the bacteria deliver energy and essential nutrients to the host – an interdependent benefit called mutualistic symbiosis.

All organic material that ends up in these sulphidic sediments originates in mangroves, dense algae, seaweed, and seagrass meadows, or on the land. Bacteria that are present in these sediments colonise the surfaces and degrade the plant debris. When the oxygen in the sediments

is respired, anaerobic processes take over and sulphate is chemically transformed to what we experience as malodorous sulphide.

To explore the fascinating world of chemosynthetic symbioses, a team of researchers led by Nicole Dubilier at the Max Planck Institute for Marine Microbiology regularly head out on sampling trips. The Carrie Bow Cay field station operated by the US Smithsonian Institution provides one of our favorite sampling spots, located 24 km off the shore of the Belizean mainland on the Meso-American Barrier reef. Its sediments provide ideal habitats for chemosynthetic symbioses. The diverse symbioses have evolved convergently in protists, and at least eight animal phyla – including Mollusca, Nematoda, and Platyhelminthes – with bacteria form diverse phylogenetic groups. The evolution of such tight interdependent symbioses between bacteria and their hosts often resulted in complete losses of mouth, gut, and digestive systems in the host animals.

A small selection of the diversity of white meiofaunal worms including annelids, nematodes and a flatworm that host chemosynthetic bacterial symbionts in Belizean sediments
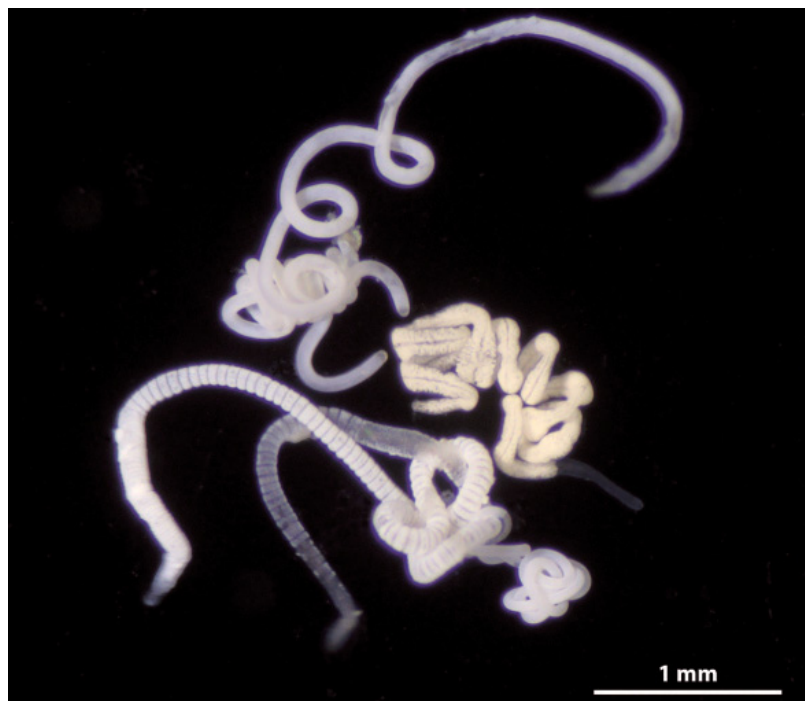
Microscopic image of *Paracatenula sp. "schlauchi".* Bacteria with storage compounds are shown in black

Flatworms of the genus *Paracatenula* lack a mouth, gut, and a digestive system and instead live in an ancient symbiosis with their bacteria that make up half of their body volume. This lack means that these worms depend entirely on their symbionts for nutrition. Conversely, their symbiotic bacteria gain a stable and sheltered environment and a shuttle service. Based on morphology and molecular analyses, seven different *Paracatenula* species were identified in Belize. Surprisingly, each of them hosts its own, species-specific symbiont. Recent data even suggests that this symbiosis was established more than 500 million years ago, i.e. long before the dinosaurs became extinct and the human species evolved among the primates.

So the next time you're at the beach and suddenly sense the odour of rotten eggs, just remember: you might be stepping on sediments that are full of chemosynthetic symbiosis miracles!

Name: Oliver Jäckle   Ort, Datum: Bremen, den 7.11.2018

Anschrift: Bürgermeister-Deichmann-Straße 63, 28217 Bremen

<u>ERKLÄRUNG</u>

.

Hiermit erkläre ich, dass ich die Doktorarbeit mit dem Titel:

Evolution and physiology of the *Paracatenula* symbiosis

selbstständig verfasst und geschrieben habe und außer den angegebenen Quellen keine weiteren Hilfsmittel verwendet habe.

Ebenfalls erkläre ich hiermit, dass es sich bei den von mir abgegebenen Arbeiten um drei identische Exemplare handelt.

_____

Oliver Jäckle