Deep Se(a)quencing:

A study of deep sea ectosymbioses using next generation sequencing

Dissertation

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

- Dr. rer. nat. -

dem Fachbereich Biologie/Chemie der

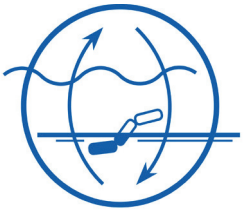
Universität Bremen

vorgelegt von

Adrien Assié

Bremen

April 2016



Die vorliegende Arbeit wurde von Januar 2012 bis April 2016 in der Abteilung Symbiose am Max-Planck-Institut für marine Mikrobiologie in Bremen angefertigt

1. Gutachterin: Dr. Jillian Petersen
2. Gutachter: Dr. Sebastien Duperron

Tag des Promotionskolloquiums: 6. Jun. 2016

“Jacques Cousteau could never get this low”

ODB – Wu Tang Clan

Cover credits - Front: ©IFREMER, Back : Dr. N.Leisch

Summary

Deep-sea hydrothermal vent fields and cold seeps are oases for deep-sea life in an otherwise nutrient-poor environment. They release energy-rich inorganic compounds that sustain rich microbial and invertebrate communities on the basis of bacterial chemosynthesis. Many endemic invertebrate species have established symbiotic relationships with chemosynthetic bacteria and thrive in these habitats. Symbioses occur in many forms: in endosymbioses the bacteria are located within a host cell or tissue, whereas in ectosymbioses the bacteria colonize their host's body surfaces such as epithelia. Deep-sea research is challenging with isolated and remote study sites requiring extensive logistical operations for sample collection. However, the rise of next-generation sequencing has allowed the gathering of large datasets from small samples and thus allows in-depth exploration of symbiotic systems. This Ph.D. thesis was focused on the investigation of two deep-sea epibiotic systems using next-generation sequencing methods.

The first project investigates Epsilonproteobacteria that occur on several deep sea mussels of the subfamily Bathymodiolinae. Previous work with 16S rRNA clone libraries had suggested that the mussels may host epsilonproteobacterial symbionts in addition to the well-known gammaproteobacterial endosymbionts. First I analyzed the localization of the epsilonproteobacterial sequences within the mussel's gill tissue using microscopy and investigated their diversity and phylogeny using 16S rRNA sequencing methods. I was able to show an epibiotic association of the Epsilonproteobacteria and determined that seven out of the twelve mussel species studied were associated with closely-related Epsilonproteobacteria. The phylogenetic reconstruction of the 16S rRNA sequences suggested that the epibionts belong to a new family of Epsilonproteobacteria.

In the second part of this project, I aimed to determine the nature of the association and metabolic potential of the epibionts, using metagenome and metatranscriptome analysis of two different bathymodiolin species. Based on genomic data, I was able to reconstruct their inorganic carbon fixation pathway, which was unexpectedly predicted to occur through the Calvin Benson Bassham (CBB) cycle. To date every other chemoautotrophic Epsilonproteobacteria has been described to fix inorganic carbon using the reverse tricarboxylic acid (rTCA) cycle. These epibionts acquired the CBB cycle from two separate horizontal gene transfer (HGT) events and lost the rTCA cycle. The key gene of the CBB, coding for 1,5-ribulose bisphosphate carboxylase, may have been acquired from a relative of the bathymodiolin

gammaproteobacterial endosymbionts, whereas all the other CBB genes originate from an unknown Betaproteobacteria. I then discussed the implication of such HGTs and hypothesized that the epibionts are commensal or mutualistic, because most pathogens are not autotrophic.

The third part of this project was a comparative analysis of the genomic data of the two epsilonproteobacterial epibionts. My phylogenomic analysis using multigene phylogeny showed that these two epibionts were two different species. I described the genetic potential of these epibionts and presented their reconstructed metabolism. This shows small metabolic differences between both bacterial draft genomes, and an overall array of different metabolic and genetic tools available that gives them a metabolic versatility to adapt to the environment.

My second project investigated the ectosymbiotic bacterial populations associated with the deep-sea shrimp *Rimicaris hybisae*. The *R. hybisae* shrimp was discovered in 2010 along with two hydrothermal vent fields, Von Damm and Piccard, located on the Mid-Cayman Spreading Center. These two hydrothermal vents had very different environmental conditions and offered a unique setting to study the influence of the environment on ectosymbiotic populations. Von Damm, an ultramafic vent with high concentration of methane and low concentration of hydrogen sulfide, is located at 2500 m depth. Piccard, a basaltic vent field with low concentration of methane but high concentration of hydrogen sulfide, is located at 5000 m depth and is the deepest hydrothermal vent field found so far. I compared the different symbiotic and free-living populations using amplicon libraries of a variable region of the 16S rRNA sequence. I showed that the ectosymbiotic populations associated with *R. hybisae* are significantly different between the two hydrothermal vent fields and are more similar to their respective free-living bacterial communities. I hypothesize that the *R. hybisae* shrimp are taking up ectosymbionts from their environment, because they are probably the best-adapted to local environmental conditions.

Zusammenfassung

In der nahrungsarmen Tiefsee sind die heißen Hydrothermalquellen und die kalten Quellen Oasen für Lebewesen. Diese Quellen geben energiereiche anorganische Verbindungen ab, die auf der Basis von Chemosynthese Lebensgemeinschaften zwischen Mikroben und Invertebraten erhalten. Viele endemische Arten von Invertebraten besiedeln diese ungewöhnlochen Habitate und leben in Symbiose mit verschiedenen Bakterien. Es sind verschiedene Symbioseformen zu unterscheiden: Endosymbionten sind Mikroben die im Gewebe oder in den Zellen des Wirts leben, während Ektosymbionten die Oberflächen des Wirts kolonisieren. Die Forschung an solchen chemosynthetischen Symbiosen und die Probennahme in der Tiefsee ist aufgrund der isolierten und abgelegenen Orte logistisch sehr aufwendig. Dank dem Fortschritt im Bereich der ‚Next Generation-Sequenzier-Technologien' ist es jedoch nun möglich große Datenmengen von einer kleinen Anzahl von Proben zu generieren und dadurch symbiotische Assoziationen tiefgründig zu erforschen. Der Schwerpunkt dieser Doktorarbeit liegt auf der Erforschung zweier ektosymbiotischer Systeme, der Tiefseemuscheln der Gattung *Bathymodiolus* und den Tiefseegarnelen *Rymicaris* mittels Next-Generation-Sequenziermethoden

Das erste Projekt untersuchte 16S rRNA Klonbibliotheken von mehreren Muschelarten der Gattung *Bathymodiolus*, welche die Präsenz von bisher nicht-beschriebenen Epsilonproteobakterien zusätzlich zu den bereits gut charakterisierten gammaproteobakteriellen Endosymbionten näher legten. Dieses Projekt war in drei Teile unterteilt: Der erste Teil erforschte die Lokalisierung der Epsilonproteobakterien in den Muscheln und analysierte ihre Diversität und Phylogenie mittels 16S rRNA Sequenziermethoden und Mikroskopie. Ich konnte zeigen, dass die Epsilonproteobakterien Ektosymbionten sind und habe festgestellt, dass 7 von 12 verschiedene Muschelarten mit nah verwandten Epsilonproteobakterien assoziiert sind. Die phylogenetische Rekonstruktion der 16S rRNA Sequenzen wies darauf hin, dass diese Ektosymbionten eine neue Familie innerhalb der Epsilonproteobakterien bilden. Um die Natur dieser Symbiose und das Stoffwechselpotential der Ektosymbionten aufzuklären habe ich im zweiten und dritten Teil dieses Projekts Metagenom- und Metatranskriptomanalysen von zwei *Bathymodiolus* Arten durchgeführt. Wider Erwarten verfügen die Ektosymbionten über den Calvin-Benson-Bassham-Zyklus (CBB-Zyklus) zur Fixierung von anorganischem Kohlenstoff. Dessen tiefgehende Analyse befindet sich im zweiten Teil dieses Projektes. Bisher wurde in allen beschriebenen chemoautotrophen Epsilonproteobakterien nur der

reduktive Citratzyklus (rTCA-Zyklus) gefunden. Mit dieser Arbeit beschreibe ich die Aufnahme von Genen des CBB-Zyklus mittels zweier, unabhängigen horizontalen Gentransfer-Prozessen (HGT) und den darauffolgenden Verlust des rTCA Zyklus in der *Bathymodiolus* Ektosymbionten. Die Schlüsselgene des CBB-Zyklus, die für die Ribulose-1,5-bisphosphat-Carboxylase/ Oxygenase kodieren, wurden von Verwandten des gammaproteobakteriellen Endosymbionten der Bathymodiolinae aufgenommen, während die anderen CBB Gene von unbekannten Betaproteobakterien stammen. Ich diskutiere die Folgerungen solcher HGTs und stelle eine Hypothese bezüglich der kommensalistischen bzw. mutualistischen Interaktionen der Ektosymbionten auf, da bisher nur wenige autotrophe Parasiten beschrieben sind. Der dritte Teil dieses Projektes beschäftigte sich mit der Analyse der Metagenomdaten der zwei epsilonproteobakteriellen Ektosymbionten. Meine phylogenomische Analyse mittels ‚Multi-Gen' Phylogenie zeigte, dass diese zwei Ektosymbionten zweier unterschiedlichen Arten angehören. Ich beschreibe das genetische Potential der zwei Bakterien und rekonstruierte ihren Metabolismus. Die Analyse zeigt metabolische Unterschiede zwischen den beiden verschiedenen Ektosymbionten und eine Reihe unterschiedlicher metabolischer und genetischer Werkzeuge die es ihnen ermöglicht ihren Metabolismus an die Umwelt anzupassen.

Im zweiten Projekt dieser Doktorarbeit beschäftigte ich mich mit den Ektosymbiontengemeinschaften die mit der Tiefseegarnele *Rimicaris hybisae* assoziiert sind. Die *R. hybisae* Garnele wurde 2010 an den zwei Hydrothermalquellenfelder Von Damm und Piccard, beide im Kaimangraben, entdeckt. Diese beiden Hydothermalquellen haben sehr unterschiedliche Umweltbedingungen und ermöglichen es daher den Einfluss der Umwelt auf die Ektosymbiontenpopulation zu untersuchen. Von Damm, eine ultramafische Quelle mit einer hohen Konzentration an Methan und niedriger Konzentration an Wasserstoff, liegt in 2500 m Tiefe. Piccard, ein Basalt Quellenfeld mit niedriger Konzentration an Methan aber hoher Konzentration an Wasserstoff liegt bei 5000 m Tiefe und ist das tiefste Hydrothermalquellenfeld das bisher beschrieben wurde. Ich habe die unterschiedlichen symbiotischen und freilebenden Populationen mittels Amplikon-Bibliothek der variablen Region der 16S rRNA Sequenz verglichen. Ich konnte zeigen, dass sich die Ektosymbiontenpopulationen die mit *R. hybisae* assoziiert sind sich signifikant zwischen den zwei Hydrothermalquellenfeldern unterscheiden und ähnlicher zu den jeweiligen freilebenden Bakterienpopulationen in ihrer Umgebung sind. Ich stelle die Hypothese auf, dass die *R. hybisae* Garnelen ihre Ektosymbionten aus ihrer Umgebung da diese bereits an die lokalen Umweltbedingungen angepasst sind.

Table of Contents

Chapter 1 **Introduction**

1.1 The Deep

The deep sea is the largest habitat on Earth, representing about 85% of the area of the Earth (Figure 1.1) covered by water. It is characterized by low temperatures, high pressure, a lack of light and limited food resources. Because of its remoteness, the deep-sea realm is the least known and researched ecosystem on the planet. No zonation scheme of the ocean has been universally accepted but usually the deep sea is regarded as the environment below 200 m depth and is most commonly divided into the bathyal (200-4,000 m), abyssal (4,000-6,000 m) and hadal (6,000-10,000 m) zones, with corresponding pelagic and benthic zones (Nybakken and Bertness, 2005). Most of the "deep-pelagic" zone is structured only by the physical properties of water. However, where pelagic zones intersect continental landmasses, seamounts or mid oceanic ridges can change the local and regional circulation patterns, which can have an effect on the distribution of deep-sea organisms in the area (Sutton *et al*., 2007).

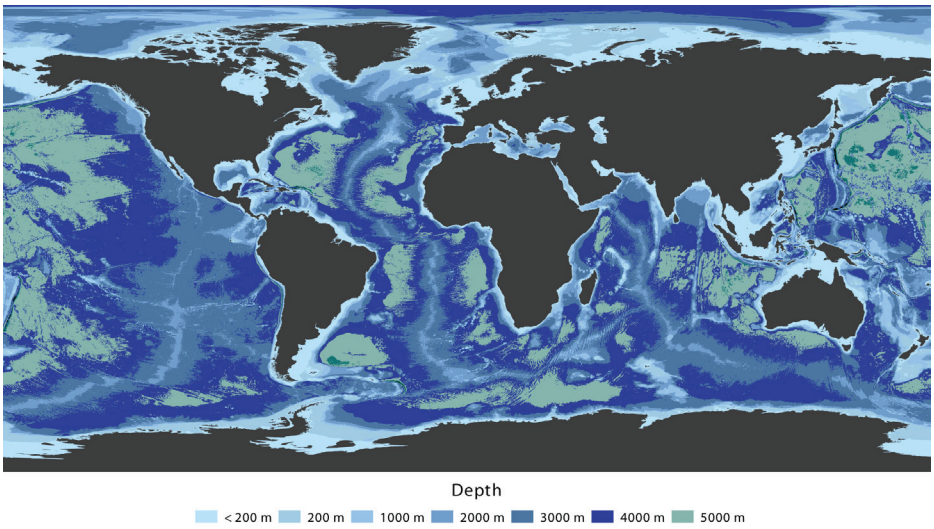


Figure 1.1 World's ocean depth map. Made with Natural Earth, map data sets.

Most of the deep-sea benthic zone consists of abyssal plains, where the principal source of nutrition is sinking particulate matter from the pelagic zone (Bergstad *et al*., 2008; Sutton *et al*., 2009). Exceptions exist in the form of localized organic material input, such as whale or wood falls. Initially thought to be a desert-like environment, due to the low abundance of megafauna, the abyssal plains have been shown to shelter a rich and heterogeneous biodiversity, with hundreds of species of macrofaunal invertebrates, nematode worms, harpacticoid copepods and foraminiferan protozoa in a typical square meter of sediment (Glover *et al*. 2001; Lambshead*et al*., 2002; Smith and Demopoulos, 2003; Durden *et al*., 2015, Rex and Etter. 2010).

Mid-oceanic ridges are poorly studied habitats in the deep sea, yet they are thought to significantly affect the flow of ocean currents across them and are believed to be important as unique habitats, supporting diverse species and ecosystems (Read *et al*., 2010). Their importance to the fauna inhabiting the deep sea is due to the wide range of depths and varying substrate types they offer and the larger availability of food compared to adjacent abyssal plains. However, mid-oceanic ridges lack the terrigenous material sedimentation observed on continental shelves and also rely mostly on surface primary productivity and food falls for nutritional input (Eppley and Peterson, 1979).

Deep-sea hydrothermal vents and cold seeps occur on continental plate margins and are also important habitats. The local release of reduced compounds fuels dense megafaunal communities, which offer a striking contrast to the surrounding abyssal plains (Figure 1.2). These communities are mainly dominated by invertebrates and rely on chemosynthetic primary production (Vrijenhoek, 2010).

1.2 Hydrothermal vent

1.2.1 Geologic Characteristics

Hydrothermal vents usually occur on mid-oceanic ridges and back-arc basins are fuelled by tectonic and volcanic activity at convergence zone

(Figure 1.3). Oceanic ridges are high-relief linear oceanic features that occur at the boundaries of tectonic plates or spreading centers. As two tectonic plates spread from each other, mantle rocks rise from the depth to form a new ocean crust. Oceanic ridges are thus characterized by a fairly thin oceanic crust and are heavily affected by the underlying mantle activity (Lin *et al*., 2000; Nicolas, 1995). These regions account for more than 75% of volcanism in the ocean worldwide. In contrast, back-arc basins occur on subduction zones, where tectonic plates sink beneath each other (Hannington *et al*., 2005).

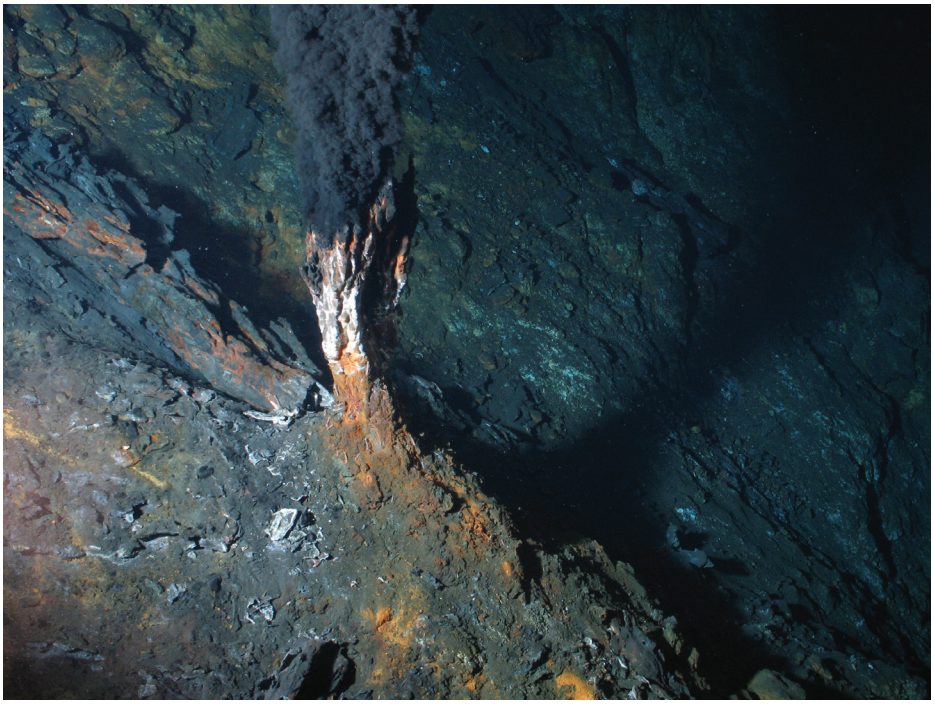


Figure 1.2 Picture of a hydrothermal vent chimney taken at the Rainbow vent fields on the north mid-Atlantic ridge. ©IFREMER

The speed at which two tectonic plates move away from each other is directly linked to the volcanic activity of the spreading center. The total length of spreading centers worldwide is estimated to be around 75,000 km (Figure 1.3). They do not all spread at the same speed and the different classes have been defined as ultraslow (less than 200 mm/yr), slow

(20- 50 mm/yr), intermediate (50—90 mm/yr), fast (90-130 mm/yr) and super-fast (130- 170 mm/yr) (Ramirez-Lloodra *et al*., 2007). Fast and super-fast spreading centers, such as the East Pacific Rise ridges, have a less pronounced relief and a thinner crust. This results in more exposed volcanic activity, fuelling many hydrothermal vent sites that have a homogeneous geological structure. On the other hand, ultraslow and slow spreading centers such as the Mid-Cayman Spreading Center (MCSC) and the Mid-Atlantic Ridge (MAR), have a more pronounced relief and impressive rift and canyon structures (Rosencrantz *et al*., 1988). They therefore have fewer hydrothermal vents but in more heterogeneous geological settings.

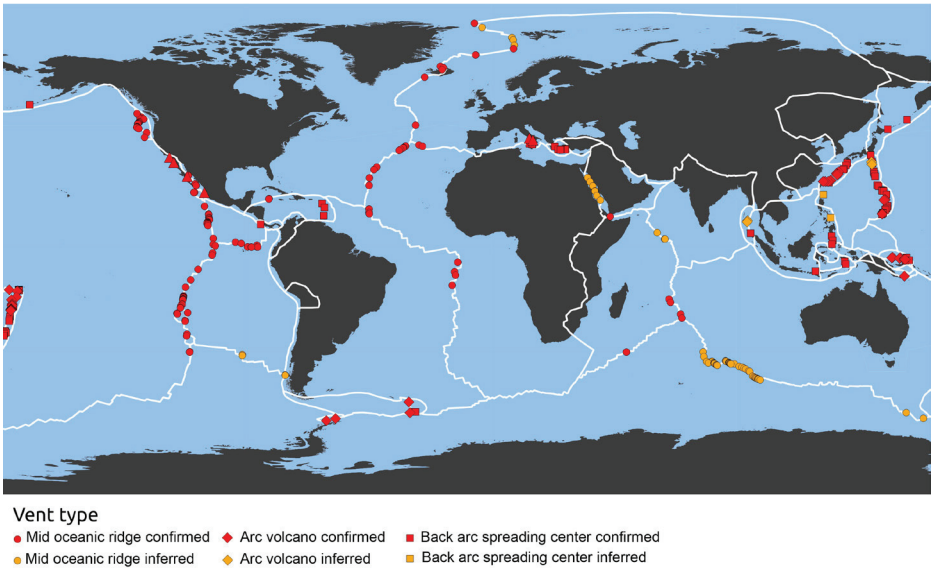


Figure 1.3 World map of the different known and predicted hydrothermal vents. Produced with the Interrige vent data of Version 3.3 on 16 Sept. 2015. Credits:S. Beaulieu

Back-arc basin hydrothermal vents are episodic events and are mainly located at the convergence zones of two tectonic plates. When the overlapping tectonic slab cracks under the collision pressure, magma and hot rock infiltrate the new opening, creating a new oceanic crust. In contrast to spreading centers, the geological settings of back-arc basins are often heterogeneous (Stern *et al*., 1990; Sibuet *et al*., 1998).

1.2.2 Vent fluid chemistry

Hydrothermal vent activity and chemical characteristics are directly linked to tectonic and volcanic activity. After a tectonic event, the newly created crust solidifies, creating fissures and cracks when the lava cools down. Ridges, where the crust is thin, or the fracture zones, where deep cracks occur, are areas often associated with magma chambers or pockets located close to the seabed. Deep-sea waters infiltrating the sea floor through the fissures or cracks get progressively heated as they come closer to the heat source. This heated water is then driven upward by convection, creating a water flow that fuels hydrothermal venting (Reviewed in Van Dover, 2000). The chemical composition of the seawater is modified as it travels towards the heat source and rises back up by convection. As the water infiltrates the crust, it interacts with the surrounding minerals and becomes progressively anoxic. Additionally, the seawater is depleted of minerals and hydroxyl ions by the heat, which induces a drop in pH. However, when the seawater comes close to the magma cell, the surrounding rock composition releases various chemicals into the water flow through phase separation processes. Hydrothermal vents can be classified into different categories based on the geological structure of the vent base (Hannington *et al*., 2005; Tivey, 2007).

Hydrothermal vents with a basaltic base are usually found close to the spreading center and are influenced by the magma activity. The vent fluids are typically enriched in hydrogen and hydrogen sulfide but poor in methane and have an acidic pH. On the other hand, hydrothermal vents located further away in the crust within ultramafic rocks (such as dunites, peridotites and pyroxenites) are influenced by a serpentinization reaction. This is a process of oxidation of iron and manganese minerals in ultramafic rocks, which occurs close to a heat source and generates fluids depleted in hydrogen sulfide but rich in methane and with a high pH (Figure 1.4 - Charlou *et al*., 2000; Petersen *et al*., 2009; Perner *et al*., 2013; Tivet 2007).

Black smokers are vents in which hot water (around 300-400°C) erupts from the cold deep-sea floor, causing the metal and sulfidic compounds

present in the fluids to precipitate immediately and produce the black colour of the fluids. White smokers, on the other hand, have a porous base and water bursts out at a lower temperature (100-150°C). They are rich in calcium sulfate and silicate, which causes the white colour of the end fluids.

Hydrothermal vent sites are ephemeral events and their longevity is also correlated with the underlying magma activity. Tectonic shifts can remove the heat source, thus interrupting the hydrothermal vent flow and the source of primary production, leading to local megafauna extinction. Many vents are short-lived, particularly on fast spreading centers but others can occur for tens, to hundreds, to thousands of years (Fruh-Green, 2003; Lilley *et al.*, 2003).

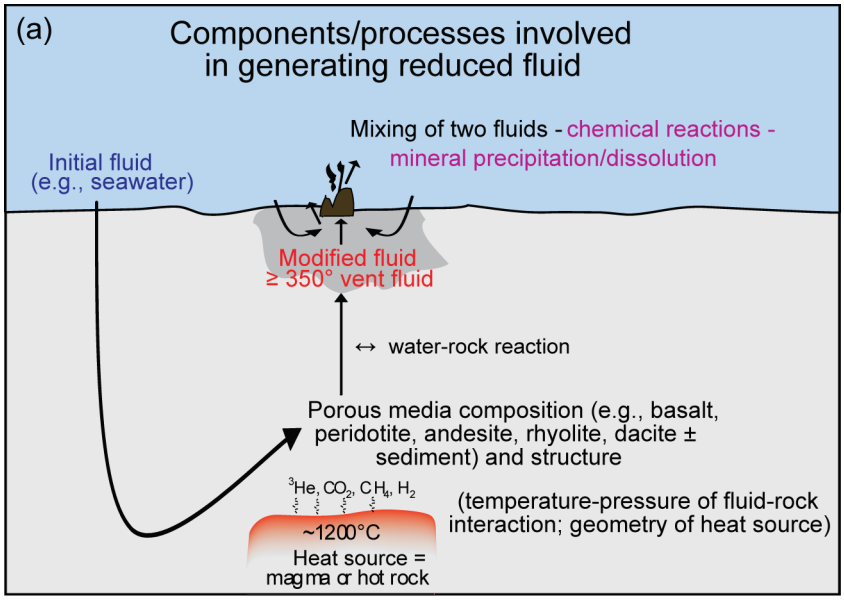


Figure 1.4 Schematic drawing of a hydrothermal system within oceanic crust showing the different components and processes that can affect the composition of the fluid that vents at the seafloor. Adapted from Tivet *et al.*, 2007.

1.2.3 Examples

1.2.3.1 Mid-Atlantic Ridge

The largest topographical feature in the Atlantic is the Mid-Atlantic Ridge (MAR) extending between the junction with the Gakkel Ridge northeast of Greenland southward to the Bouvet Triple Junction in the South Atlantic; the Charlie Gibbs fracture zone divides the ocean basin into north MAR and south MAR (Longhurst, 1998). The MAR is a slow spreading center hosting a heterogeneous set of hydrothermal vent fields. Both basaltic and ultramafic-hosted vents are present along the ridge and their vent fluids exhibit various geological and chemical profiles, summarized in Table 1.1.

1.2.3.2 Mid-Cayman Spreading Center

The Mid-Cayman Spreading Center is one of the smallest oceanic ridges and is located south of the Cayman Islands. This ultraslow spreading center is 420 km long. In 2010, an expedition discovered two hydrothermal vent fields, Von Damm and Piccard, which are located at 2500 m and 5000 m depth, respectively (German *et al*., 2010). The Piccard vent field is the deepest discovered hydrothermal vent field to date.

The two hydrothermal vent fields have very different geological settings. The Von Damm vent field is located on top of a seamount with an ultramafic, talc rich base (Hodgkinson *et al*., 2015), while Piccard is located close to the ridge and is exclusively basaltic (Kinsey and German, 2013). This difference influences the hydrothermal vent fluids from the two vent fields, which have a distinct chemical signature (Table 1.2- Reveillaud *et al*., 2015).

1.3 Cold seeps

1.3.1 Characteristics

The term cold seeps describe another type of geological activity occurring in the deep sea. Similar to hydrothermal vents, these are also often associated with dense benthic invertebrate communities. Many seeps

Table 1.1 Physical and chemical characteristics of the end fluids of some of the MAR hydrothermal vents. *: And reference therein.

Vent	Type	Depth (m)	T (°C)	pH	H_2S (mM)	CH_4 (mM)	H_2 (mM)	Reference
Menez Gwen	Basaltic	850	245	4.2	1.6	0.52	0.038	Charlou *et al*, 2000
Lucky Strike	Basaltic	1700	300	3.65	2.7	1.7	0.207	Charlou *et al*, 2000
Rainbow	Ultramafic	2300	365	2.5	1	2.8	16	Schmidt *et al* 2007*
Logatchev	Ultramafic	3000	350	3.9	2.5	3.5	19	Schmidt *et al* 2007
Lilliput	Basaltic	1500	2.5	6.4	0.053	6.1	0.8	Perner *et al* 2011
Wideawake	Basaltic	3000	16	7	0.033		1.5e-6	Perner *et al* 2013

Table 1.2 Physical and chemical characteristics of the end fluids of the MCSC hydrothermal vents. From Reeves *et al*. 2014 and McDermott *et al*. 2015.

Vent	Type	Depth (m)	T (°C)	pH	H_2S (mM)	CH_4 (mM)	H_2 (mM)
Von Damm	Ultramafic	2500	226	5.6	3.2	2.84	19.2
Piccard	Basaltic	5000	397	3.2	12	0.123	20.7

have been identified to date and are found distributed across deep (7000 m) and shallow (200 m) waters, as well as on active (e.g. Japan subduction zone – Kojima *et al*., 1995) or passive (e.g. Gulf of Mexico – Cordes *et al*., 2007) continental margins (Sibuet *et al*., 1998). Cold seeps have a different origin to hydrothermal vents. On active continental margins, accumulated organic compounds or sediments on the underlying slab are compressed and pushed towards the mantle. The accumulation pockets can become geothermally altered, producing methane or hydrogen sulfide in solid or gaseous form (Reviewed in Van Dover, 2000). On passive margins, the seepage originates from the accumulation of organic matter or sedimentation on continental slopes, which is slowly transformed by geothermal or biotic processes into oil or gases. In both cases, the tectonic plate friction causes the hydrocarbon pockets to crack and leak through the sea floor into the water-column above (Reviewed in Van Dover, 2000).

Different types of cold seeps can be distinguished, depending on the nature of the compounds accumulated in the Earth's crust. Seep fluids and gas can be of various origins, including asphalt or hydrocarbonate deposits, oil, methane or hydrogen sulfide seepage, as well as mud volcanoes or brine pools (Reviewed in Van Dover, 2000).

1.3.2 Example: Gulf of Mexico

The Gulf of Mexico (GoM) is one of the most geologically active continental passive margins. During the Jurassic era, the connection between the GoM and the Atlantic closed (Pindel *et al*, 1985), which led to a large evaporation event and the formation of salt deposits. These deposits were then covered by successive sediment loadings over an extended period. The early accumulation of sediment trapped pockets of hydrocarbon from the Mesozoic rock located below. The continuous accumulation of sediments ultimately deformed the salt layer into mobile pillars and salt domes, causing them to pierce the dense layer of sediment and rock above them, thus creating an escape conduit for the trapped hydrocarbon source below. As the oil migrated

upward, it got altered by the surrounding rock and generated different hydrocarbons, such as petroleum, as well as methane and hydrogen sulfite gases. When a single block of salt reached the sub-seafloor, it created brine pools. As a result, many hydrocarbon and saline seeps have been discovered in the past three decades. Particularly on the northern part of the GoM, the Louisiana slopes (Brooks *et al*., 1990; Cordes *et al*., 2009, 2010)

1.4 Chemosynthesis

1.4.1 Microbial communities at vents and seeps

Hydrothermal vents and cold seeps are an important source of reduced compounds. The sudden outburst of high concentrations of chemical compounds can sustain rich and diverse microbial communities (Dubilier *et al*, 2008). Chemolithoautotrophs are capable of harnessing reduced compounds such as hydrogen and hydrogen sulfides and fixing CO2, while chemoorganoheterotrophs are capable of oxidizing and fixing methane (Jannasch 1995). These microorganisms are the primary producers in such environments, responsible for transferring carbon to higher trophic levels of a very dense food web.

Two main primary production processes exist on Earth, photosynthesis and chemosynthesis. It was thought for a long time that primary production was derived exclusively through photosynthesis, by which plant algae and some microorganisms are capable of using the sun's energy to fix molecules of CO2. Chemosynthesis was described at the end of the 19th century by the Russian Microbiologist Sergei Winogradsky (Winogradsky, 1887 and Jannasch, 1995). Although the process was well studied, it was not considered to play a major role in the Earth's primary production until the discovery of hydrothermal vent systems almost 100 years later. Chemosynthesis is the process used by Bacteria and Archaea to harness energy produced via oxidation of reduced compounds, such as hydrogen, hydrogen sulfide and methane, and most of the time later fix carbon from inorganic sources such as CO2 or methane (Jannasch, 1995).

1.4.2 Inorganic carbon fixation

Various metabolic pathways exist that allow microorganisms to fix inorganic carbon. These are key processes in the carbon cycle, because primary producers are able to take inorganic carbon from the environment and redistribute it to upper trophic levels (Kelley *et al*., 2002). The most widespread pathway on Earth is the Calvin-Benson-Bassham (CBB) cycle, since it is associated with photosynthesis and is present in plants, algae and photosynthetic microorganisms. Additional to the CBB cycle, five other inorganic carbon fixation pathways exist (Displayed on Figure 1.5). These are: the reductive tricarboxylic acid (rTCA) cycle, the reductive acetyl-CoA,or Wood-Ljungdahl (WL) pathway, the 3-hydroxypropionate (3-HP) bicycle, as well as the 3-hydroxypropionate/4- hydroxybutyrate (3-HP/4-HB) and dicarboxylate/4-hydroxybutyrate (DC/4-HB) cycles (reviewed in Hügler and Sievert, 2011). All five pathways are associated with different chemosynthetic organisms.

One of the key enzyme to the CBB cycle is the 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) enzyme. This enzyme catalyzes the oxygenation of ribulose 1,5-bisphosphate, which generates two molecules of 3-phosphoglycerate, intermediate between the CBB cycle and glycolysis. Four types of RuBisCO have been described so far (forms I-IV) and although form IV does not catalyze the same reaction as the others, they all share the same phylogenetic origin, which probably predates the bacterial and archaeal divergence. RuBisCO form I is widespread among higher plants, eukaryotic algae, Cyanobacteria and Proteobacteria. Form II RuBisCO is found in various types of Proteobacteria and in one group of eukaryotes, a branch of dinoflagellate, whereas form III has been found only in Archaea. Finally, form IV is distributed among Proteobacteria, Bacteroidetes, non-methanogenic Archaea and a few eukaryotes (Reviewed in Tabita *et al*. 2007).

The rTCA cycle can be summarized as an oxidative TCA cycle going in reverse. The TCA cycle generates ATP by oxidizing Acetyl -CoA to CO2, while the rTCA cycle consumes energy generated by chemosynthesis to fix

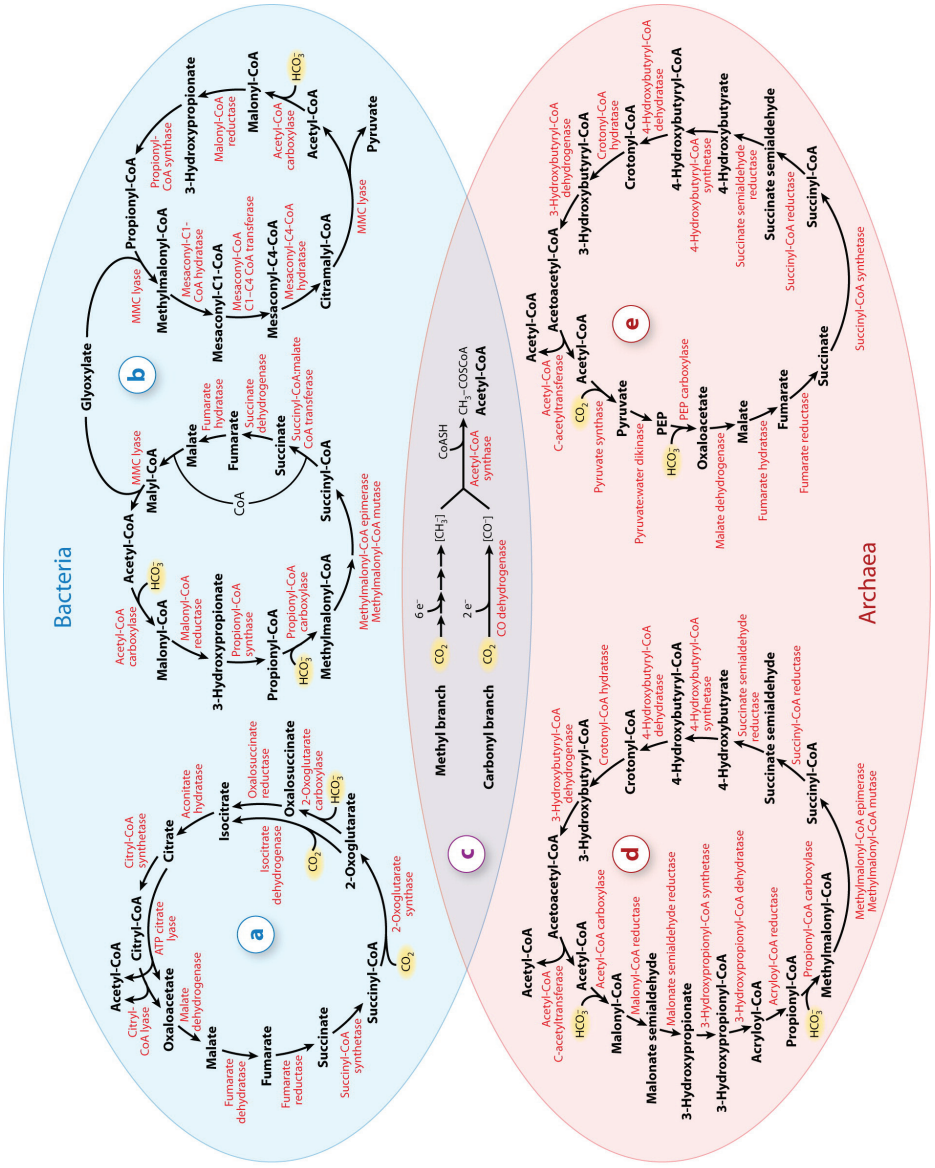
CO2 molecules and generates Acetyl CoA. Many of the enzymes involved in the pathway (malate dehydrogenase, fumarate hydratase, succinyl- CoA synthetase, isocitrate dehydrogenase, aconitate hydratase) can actually perform the enzymatic reactions in both directions. Nevertheless, three enzymes are necessary to perform the complete cycle: ATP citrate lyase, ferredoxin:pyruvate kinase and 2-oxoglutarate oxidoreductase (Aoshima, 2007).

The WL pathway is the only non-cyclic carbon fixation pathway. Two enzymes reduce a CO2 molecule to CH3- and CO- separately and later a third enzyme combines the products into a molecule of acetyl-CoA (Aoshima, 2007).

As the name suggests, the 3-HP bicycle is based on two cycles. The first cycle fixes two bicarbonate molecules and generates glyoxylate, which is then disproportionated with propionyl-COA in the second cycle to produce acetyl-CoA and pyruvate (Herter *et al*. 2002, Zarzycki *et al*. 2009).

The 3-HP/4-HB cycle is a variant of the 3-HP bicycle in which the reactions occur in one cycle. The 3-HP part shares similarities with the previous cycle. The most characteristic enzyme of the 4-HB part of the cycle is 4-hydroxybutyryl-CoA dehydratase, transforming 4-hydroxybutyryl-CoA to crotonyl-CoA (Berg *et al*., 2007; 2010).

Finally, DC/4-HB cycles are the latest described inorganic fixation pathways. These pathways are an alternative to the rTCA cycle. Three enzymes are present: pyruvate synthase, pyruvate:water dikinase and PEP carboxylase. These enzymes convert 2-oxoglutarate to oxalosuccinate, incorporating a molecule of bicarbonate, and then generate citryl-CoA. The other steps of the cycle are performed by regular TCA cycle enzymes (Hubert *et al*., 2008).



Hügler M, Sievert SM. 2011.
Annu. Rev. Mar. Sci. 3:261–89

Figure 1.5 Schematic of the different inorganic carbon fixation cycles. (*a*) reductive tricarboxylic acid cycle, (*b*) 3-hydroxypropionate bicycle, (*c*) reductive acetyl-CoA pathway, (*d*) 3-hydroxypropionate/4-hydroxybutyrate cycle, and (*e*) dicarboxylate/4-hydroxybutyrate cycle. In red are the names of enzymes. Figure reproduced with permission from Hügler and Sievert, 2011.

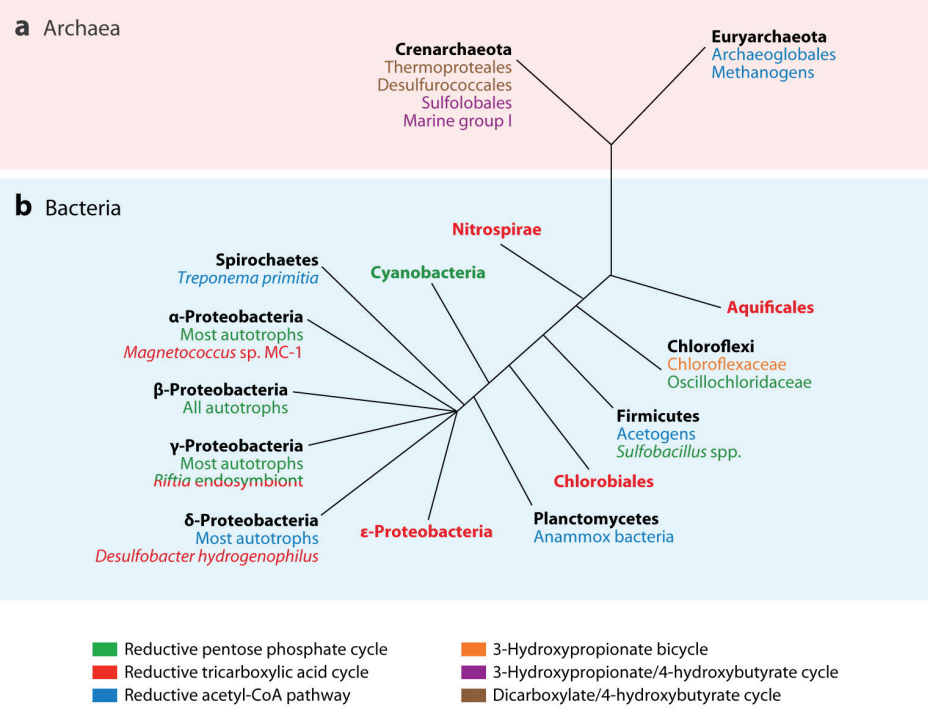
Current knowledge of chemosynthetic communities at hydrothermal vents suggests that the CBB cycle and the rTCA cycle are the most common pathways of carbon fixation (Nakagawa and Takai, 2008; Sievert *et al*., 2008). The rTCA cycle has been found in microorganisms inhabiting areas where the temperature fluctuates between 20 and 90°C, whereas other cycles, such as the CBB cycle, are more commonly found within regions below 20°C. The WL-pathway or DC/4-HB, however, are more present in areas with temperatures above 90°C (Byrne *et al*., 2009; F. P. Wang *et al*., 2009;S. Wang *et al*., 2009).

This distribution could partially be explained by the respective enzymes' tolerance to oxygen. CBB cycle enzymes are not sensitive to oxygen and can be found in cooler, more oxygenated waters (Berg, 2011). In warmer areas, usually near the mixing zone, where the vent fluids reach the sea floor, hot temperature is often linked with lower oxygen concentration. Here, the rTCA and other cycles that have ferredoxin-based enzymes sensitive to oxidation can be found in micro- to strictly an-aerobic microorganisms (Ragsdale, 2003; Imlay, 2006). These cycles prevail over the CBB cycle in such areas and they are much more energy efficient. The CBB cycle has a high-energy demand of seven ATP molecules to generate one molecule of pyruvate, whereas the rTCA cycle only needs two molecules of ATP to generate one pyruvate. The WL-pathway, 3-HP, 3-HP/4-HB and DC/4-HB need 1, 7, 9 and 5 ATP molecules, respectively, to generate one pyruvate (reviewed in Hügler and Sievert, 2011).

Cold seeps have a similar chemistry to hydrothermal vents; however, they do not have such extreme temperature gradients. Because the temperature does not affect the chemosynthesis process, the elevated levels of dissolved inorganic carbon and hydrogen sulfide found in the environment are enough to sustain chemosynthetic communities (Sibuet *et al*., 1998, Jannasch, 1995). The temperature and the local oxygen concentration might influence the enzymatic efficiency of the different carbon fixation cycles and shape the community distribution at hydrothermal vent or seeps.

1.4.3 Phylogenetic distribution

The different inorganic fixation pathways are distributed differently within the bacteria and Archaea kingdoms. The CBB cycle has a limited phylogenetic distribution, despite being the most widespread cycle on Earth. The cycle probably evolved from Cyanobacteria before being incorporated into Eukaryotes via symbiotic assimilation of the chloroplast. It is also found in some Alpha-, Beta- and Gammaproteobacteria, as well as some Chloroflexi and iron-oxidizing Firmicutes (reviewed in Tabita *et al*., 2007 and Hügler and Sievert, 2011).



Hügler M, Sievert SM. 2011.
Annu. Rev. Mar. Sci. 3:261–89

Figure 1.6 Schematic phylogenetic tree showing the distribution of the different carbon fixation pathways among major phylogenetic lineages in (a) Archaea and (b) Bacteria. Figure reproduced with permission from Hügler and Sievert, 2011.

Autotrophs from the Chlorobiales, Epsilonproteobacteria and Aquificales have been reported to perform anaerobic photosynthesis through the rTCA cycle. Because the rTCA cycle enzymes are oxygen sensitive, this cycle is often associated with microaerobic or anaerobic bacteria (Campbell and Cary, 2004). A notable exception is the gammaproteobacterial endosymbiont of the deep sea tube worm, *Riftia pachyptila*. Studies have shown that the endosymbionts possess and express both CBB and rTCA cycle genes. This is the only known example to date in which a Gammaproteobacteria has the potential to use the rTCA cycle. The presence of both cycles is thought to allow the bacteria metabolic versatility and enables them to adapt quickly to sudden environmental changes, such as a drop or a rise in oxygen concentration (Markert *et al*., 2007; Robidart *et al*., 2011).

The WL cycle was initially described within acetogenic bacteria, mostly Clostridiales, but has since been described in autotrophic sulfate-reducing bacteria and archaea as well as in methanogenic archaea, and potentially also in Planctomycetes, carrying out anaerobic oxidation of ammonium (Aoshima, 2007).

The other inorganic fixation pathways have a much more limited phylogenetic distribution. The 3-HP bicycle was originally described to be present in Chloroflexaceae (Herter *et al*., 2002) and the 3-HP/4-HB and DC/4-HB cycles have so far only been found in Crenarchaeota (Berg, *et al*., 2010).

1.5 Symbiosis

1.5.1 Definition of symbiosis

The word symbiosis (from the greek “syn”= with and “bios”=life) was originally defined by the German mycologist, Albert Bernhard Frank, in 1877 and described two species living on or in one another, in a way that is not simply coexistence. In 1879, another German mycologist, Anton de Barry, elaborated the concept further and outlined three criteria for symbiosis: two entities must live together, they must be intimate (in physical contact) and

they must be dissimilar (different species).

Today the definition of symbiosis includes three different types of associations classified on the basis of the cost/benefit for the organisms involved. When both partners profit from the association, it is defined as a mutualistic association. When one organism profits from the association to the detriment of the other, it is defined as parasitism. Finally, the term commensalism is used when the organisms do not have any obvious cost or benefit from the association (Martin and Schwab, 2012).

However, these categories are not rigid and the balance between cost and benefit is influenced by many factors (Leung and Poulin, 2008). For example, the pea aphid Acyrthosiphon pisum has established an obligate symbiosis with the bacterial endosymbiont, Buchnera aphidicola (Douglas, 1998) in which the symbiont complements its host's diet with amino acids. This symbiotic system can also harbor secondary symbionts, Serratia symbiotica and Hamiltonella defensa, which confer increased protection to the host against parasitoid wasps (Oliver *et al*., 2005). The presence of the secondary symbiont, however, greatly decreases the host's fertility in comparison with individuals without the secondary symbionts. When exposed to a parasitoid population, the aphids infected with secondary symbiont would have greater probability of surviving than non-infected individuals, although this advantage becomes a burden when the parasitoid threat disappears (Oliver *et al*., 2006). These associations can be seen as hovering between mutualism and parasitism.

Symbioses must be stable from one generation to the next and different strategies exist for the transmission of the symbiont from the parents to the offspring (Figure 1.6). The two modes of transmission are horizontal or lateral transmission and vertical transmission (Bright and Bulgheresi, 2010). Horizontal transmission is defined by a recolonization of the host by the symbiont from the environment, most of the time implying a free-living stage for the symbionts. The squid-vibrio symbiosis is an example of such a transmission mode, in which specific Vibrio fisheri strains from the

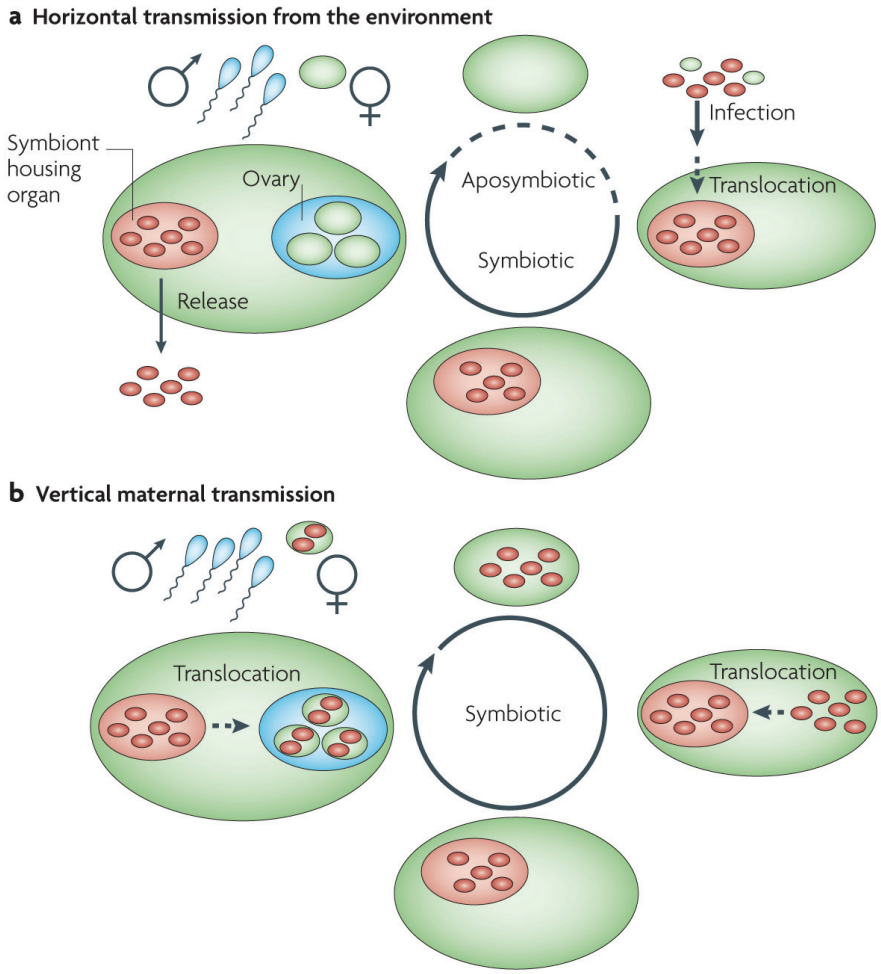


Figure 1.7 Schematic depicting the different mode of symbionts transmission to the next generation of offspring. (a) Horizontal transmission where the symbionts have to undergo a living phase outside its host and (b) vertical transmission where the symbionts are transmitted to the next generation via the host gametes. Figure reproduced with permission from Bright and Bulgheresi, 2010.

surrounding sea-water recolonize the light organ of juvenile squid (Nyholm and McFall-Ngai, 2004). Vertical transmission occurs when symbiotic bacteria are directly transferred to the next generation via the host gametes. The Alphaproteobacteria Wolbachia is a widespread parasite or symbiont of arthropods; these bacteria are manipulators of reproduction and are transmitted to the next generation via gametes (Zug and Hammerstein, 2015).

1.5.2 Where the interaction takes place

Symbiotic relationship occurs in many different forms and in every degree of specificity, from loose ectosymbiosis to intimate endosymbiosis. An ectosymbiosis is defined as when the symbiont colonizes an organ or a body part directly in contact with the environment, whereas endosymbiosis is defined as a symbiont hosted within the tissue or the host cells (Martin and Schwab, 2013).

Ectosymbioses range from plastic or optional to very tight associations. On one side of the spectrum, organisms such as deep sea Kiwa crabs farm chemosynthetic bacteria on their ventral setae for nutrition. In this example, the ectosymbionts are thought to be opportunistic surface colonizers and are not specific to the host; the population varies among individuals (Goffredi, 2010). On the other side of the spectrum, some ectosymbioses are extremely specific, such as with stilbonematine nematodes, of which the many different species of nematodes each host its own species of chemosynthetic ectosymbiont (Zimmermann *et al*., 2016).

Endosymbioses share the same variation of specificity; intracellular vertically transmitted endosymbioses are a perfect example of highly intimate association. Many examples show the co-evolution of such systems, such as the flatworm Paracatenula (Gruber-Vodicka *et al*., 2011) and the aphid *Brachycaudus* (Jousselin *et al.,* 2009). However, other associations are less dependent, and vertically transmitted endosymbioses exist without co-evolution between host and symbiont, such as is the case with the fly Bemisia

and one of its secondary endosymbionts (Ahmed *et al.,* 2013).

In strict vertically transmitted associations genome reduction can occur (Nakabachi *et al.,* 2006; He *et al.*, 2015), genome erosion happens when part of a genome becomes useless, for example a metabolic pathway required for a compound that can be directly imported from the host environment. Most frequently, pathways with several enzymatic steps and higher energy requirements, including amino acid biosynthesis pathways, disappear first (Toft and Andersson, 2010; McCutcheon and Moran, 2011). The evolutionary advantage of very strict mutualistic associations has often been discussed and it appears to be an evolutionary dead end for the endosymbiont. This is because the endosymbiont genomes become progressively smaller and can ultimately be reduced to an organelle in the best case, or become inefficient in the worst case (Bennett and Moran, 2015). Although ectosymbiosis has been proposed as the first step toward endosymbiosis (Smith, 1979), it has the advantage of keeping symbionts genetically fit, and prone to switching if the environmental conditions are no longer suitable.

1.5.3 Relevance of symbiosis

The relevance of symbiotic associations has been established in various fields, from being a driving force in evolution to having economic impacts. Symbiosis is a key factor in evolution, as can be seen in the endosymbiosis theory of mitochondria and chloroplasts; one of the most important symbiotic associations that enabled the evolution of Eukaryotes (Margulis, 1975, 1993; Szathmáry and Smith, 1995). Ancestors of Alphaproteobacteria and Cyanobacteria established tight endosymbiotic associations with primordial eukaryotic cells. These energy based symbioses were the key to the evolutionary success of Eukaryotes. They generated an explosion of diversity, the evolution of multicellular organisms, such as plants, animals and fungi, and allowed their radiation across new habitats (Keeling, 2010; Douglas, 2014). Symbiosis is ubiquitous and occurs among organisms of all three domains of life (Moya *et al*., 2008; McFall-Ngai *et al*., 2013) and it

is thought that many, if not all multicellular eukaryotes are in symbiotic associations (Zilber-Rosenberg and Rosenberg, 2008; Gordon *et al*., 2013; Singh *et al*., 2013).

Nowadays, symbiosis is seen as a key player in human health. The human body is colonized by a wide array of microorganisms, such as Aacteria, Archaea and Fungi (Turnbaugh *et al*., 2007). Microbial communities close to mucosal surfaces, such as the gastrointestinal tract, have a direct influence on their host's health (Eloe-Fadrosh and Rasko, 2013). These communities live in a homeostatic relationship with the host's cells and mammals require their gut to be colonized by mutualistic bacteria, which participate in the synthesis of vitamins and cofactors and also modulate the host's immune system to enhance its protection against pathogens. Disruption of this balance can lead to disease states, in which local communities can induce obesity or inflammatory bowel disease (Li *et al*., 2008; Chow *et al*., 2010). Understanding the network interaction of the many players of the gut and their impact on different aspects of the human health, therefore, holds great promise and modulation of bacterial communities is a new target of non-invasive treatments. The development of "microbiome therapeutics" is still in its early stages but is a part of the future of human health treatment (Cani and Delzenne, 2011; Wallace and Redinbo, 2013; Korem *et al*., 2015). Understanding the mechanism behind communication between beneficial microbial communities and host health in one system could help understand how others work.

Finally, symbiosis has an economic impact in particular in agricultural systems (Abbott and Lumley, 2014). Arbuscular mycorrhizal associations can greatly improve plant nutrition and growth, by enhancing the nutrient (Smith and Smith, 2012) or water (Manoharan *et al*., 2010) uptake in depleted soil, building soil structure (Rillig and Mummey, 2006) and increasing the plant's resistance to parasites or disease (Maffei *et al*., 2014).

1.5.4 Chemosynthetic symbioses

Hydrothermal vents were first discovered in 1977 and were found to host unexpectedly dense invertebrate communities. These were initially thought to be filter feeders but subsequent studies suggested otherwise (Lonsdale, 1977). Geological analyses of the fluid composition showed a release of reduced compounds in high concentration, which were capable of sustaining chemosynthetic bacterial communities (Jannasch *et al*., 1995). At the same time, anatomical and ultrastructural analysis showed that many invertebrate species either lacked or had a reduced gastrointestinal track, refuting the filter-feeding hypothesis (Kenk and Wilson, 1985). However, these invertebrates had specialized anatomical structures enabling them to host dense communities of bacteria (Fiala-Médioni *et al*., 1986). A new paradigm emerged, stating that the bacteria might be actively contributing to the invertebrate host diet. This was confirmed with a series of microscopic, enzymatic and isotopic experiments performed on *Riftia* tube worm in the early 1980's (Cavagnaugh *et al* 1981, Felbeck 1981 and Rau 1981).

The discovery of chemosynthetic symbioses at hydrothermal vents sparked a scientific exploration to find similar systems elsewhere. It has since been shown that chemoautotrophic bacteria are not restricted only to hydrothermal vents. A wide range of chemosynthetic symbioses have been described at cold seeps, wood and whale falls, reducing sediments on continental margins or shallow water coasts, as well as mangroves and swamps. All of these habitats generate reduced compounds, either by the abiotic reaction of water and rock or biotic degradation of biomass (Dubilier *et al*, 2008).

Chemosynthetic symbioses occur with at least seven different animal phyla (Ciliophora, Annelida, Platyhelminthes, Arthropoda, Mollusca, Nematoda and Porifera) and Archaea. The symbiotic bacteria have been described as being both ecto- and endo- symbiotic and are mainly distributed within Gamma- and Epsilonproteobacteria (Dubilier *et al*., 2008).

The most abundant chemosynthetic symbionts are sulfur-oxidizing bacteria, capable of harnessing energy from the oxidation of hydrogen sulfide or other sulfidic compounds (e.g. sulfate or thiosulfate) fuelling inorganic carbon fixation through the CBB or rTCA cycle. Methane oxidizing bacteria are the second most abundant chemosynthetic bacteria and are capable of using methane as both energy and carbon source. Finally, it has also been shown that many chemosynthetic bacteria can also rely on hydrogen as an energy source (Dubilier *et al*., 2008; Petersen *et al*., 2011).

1.6 Deep Sea Fauna

1.6.1 Overview

Since the discovery of the first hydrothermal vent fields, numerous new invertebrate species associated with hydrothermal vents and colds seeps have been described. From the initial inventory in 1983, enumerating 22 different taxa (Hessler and Smithey, 1983), a total of 712 species had been described at hydrothermal vents by 2005. Of these, 508 are thought to be endemic to hydrothermal vents, 35 are present at cold seep and vent sites and the others occur in other deep sea environments (Wolff, 2005).

Invertebrate communities inhabiting hydrothermal vents and cold seeps have had to adapt to unique geological and chemical features, such as high concentrations of reduced compounds and heavy metals, as well as fluctuating levels of oxygen and limited food availability. This has induced a high level of specialization and has driven many species to be unique to such biotopes (Hessler and Smithey, 1983; Wolff, 2005). Hydrothermal vents are often dominated by one or a few species belonging to Mollusca, Arthopoda and Polychaeta, which have been estimated to represent respectively 36.1%, 34.3%, and 18.1% of all animals in the habitat (Wolff, 2005). Mollusca are mainly represented by bathymodiolin mussels, with more than 20 species described so far. Crustaceans are represented mainly by Copepoda, with more than 80 species described, *Rimicaris* shrimps forming dense swarms on chimney columns and other Decapoda, with 20 crab species described.

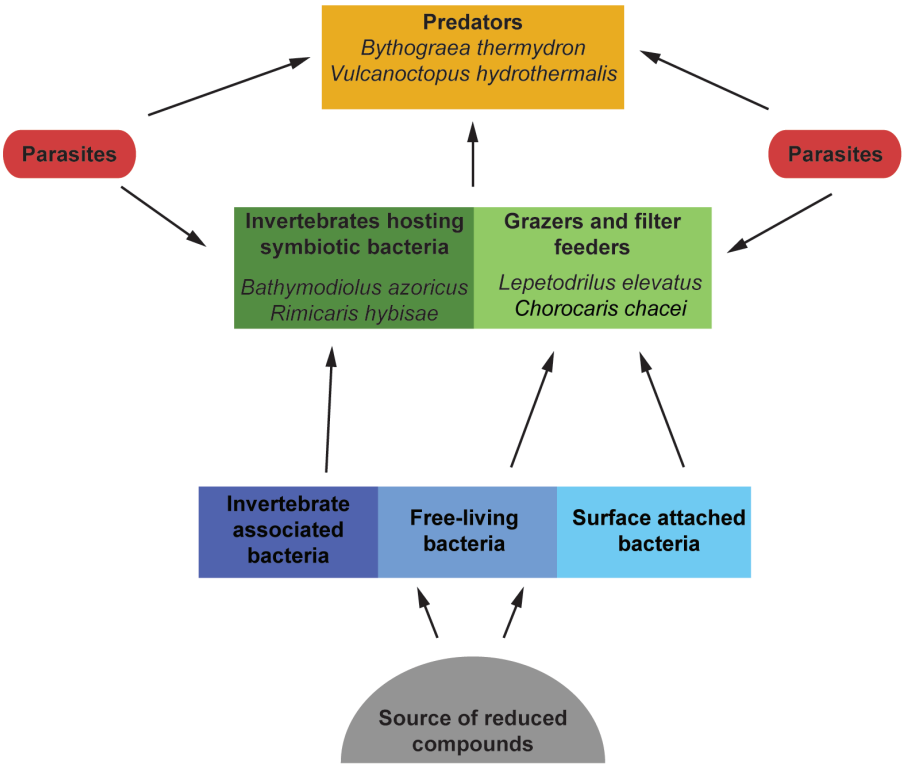


Figure 1.8 Schematic representing hydrothermal vent or cold seeps food web structure. Inspired by Galkin *et al.*, 2016

The polychaetes include all the Polynoid worms, with more than 45 species, Siboglinidae with 16 species and Alvinellidae with 13 species. (Reviewed in Galkin, 2016)

The food web of cold seeps and hydrothermal vents relies on chemosynthetic primary production (Figure 1.8). Bacterial mats and biofilms attract grazers such as *Alvinella pompejana* worms or *Chorocaris* shrimps. The dense biomass and vent activity also generate large amounts of particulate organic matter that can sustain filter-feeding animals, such as the *Neolepas rapanuii* barnacle (Jones, 1993).

However, the highest organism densities are generated by animals associated with chemosynthetic bacteria. In the Pacific, communities are

usually dominated by Siboglinidae worms (Vrijenhoek, 2010). At GoM cold seeps, the most abundant are usually bathymodiolin mussels (Cordes *et al*., 2007), while Cayman hydrothermal vents are dominated by *Rimicaris* shrimps (Plouviez *et al*., 2015). MAR hydrothermal vents can host either bathymodiolin mussels or *Rimicaris* shrimps (Vrijenhoek, 2010; van der Heijden *et al*., 2012).

This huge biomass accumulation also attracts parasites, predators and scavengers at the top of the food web. A large abundance of parasites, ranging from bacteria and protists to nematodes, copepods, and polychaetes have been described in the deep sea and it is thought that even larger diversity is still to be discovered at hydrothermal vents and cold seeps (Galkin, 2016). Nevertheless, little is known about these parasites and more research is needed to understand their impact on invertebrate populations.

Figure 1.9 Mussel bed picture taken on a Mid-Atlantic Ridge hydrothermal vent site during the 2013 scientific expedition BIOBAZ. ©IFREMER

Dense populations of highly sedentary invertebrates, such as bivalves and gastropods, present in low diversity and high density, could be vulnerable to parasitic infection or an outburst of pathogenic infection (Moreira, 2003; de Buron and Morand, 2004).

Predators include carnivorous species, particularly Bythograeid crabs and Galatheid squat lobsters found worldwide; Chaetopterid polychaetes found on the periphery of MAR vent sites and cephalopod mollusks (such as octopuses) on the East Pacific Rise (Galkin, 2016).

1.6.2 Bathymodiolin mussels

Bathymodiolin mussels (Figure 1.9) are among the most abundant species found at hydrothermal vents and cold seeps worldwide (Lorion *et al*, 2013). They are particularly dominant at MAR and GoM sites. The first species, *Bathymodiolus* thermophilus, was described in 1985 (Kenk and Wilson, 1985). Since then, 21 other species have been described, of which 12 are present in the Pacific, seven in the Atlantic and only one so far in the Indian Ocean.

1.6.2.1 Host diversity

1.6.2.1.1 Troubled phylogeny

Species identification was initially based on very fine morphological details. However, with the development of species characterization through sequencing techniques, the phylogeny of bathymodiolin mussels has emerged to be poly- and para- phyletic (Jones and Vrijenhoek, 2006; Lorion *et al*., 2013; Thubaut *et al*., 2013). The authors of the morphological description of "*Bathymodiolus*" *childressi* and *Bathymodiolus boomerang* suggested differences between the new species and previous studies but concluded that they were too small to justify a new genus (Cosel and Olu, 1998; Gustafson *et al*., 1998). The rise of sequencing technology later revealed that *"B." childressi*, sometimes referred to as *Gigandidas childressi* (Thubaut *et al*., 2013), belongs to a different genus than the original *Bathymodiolus*

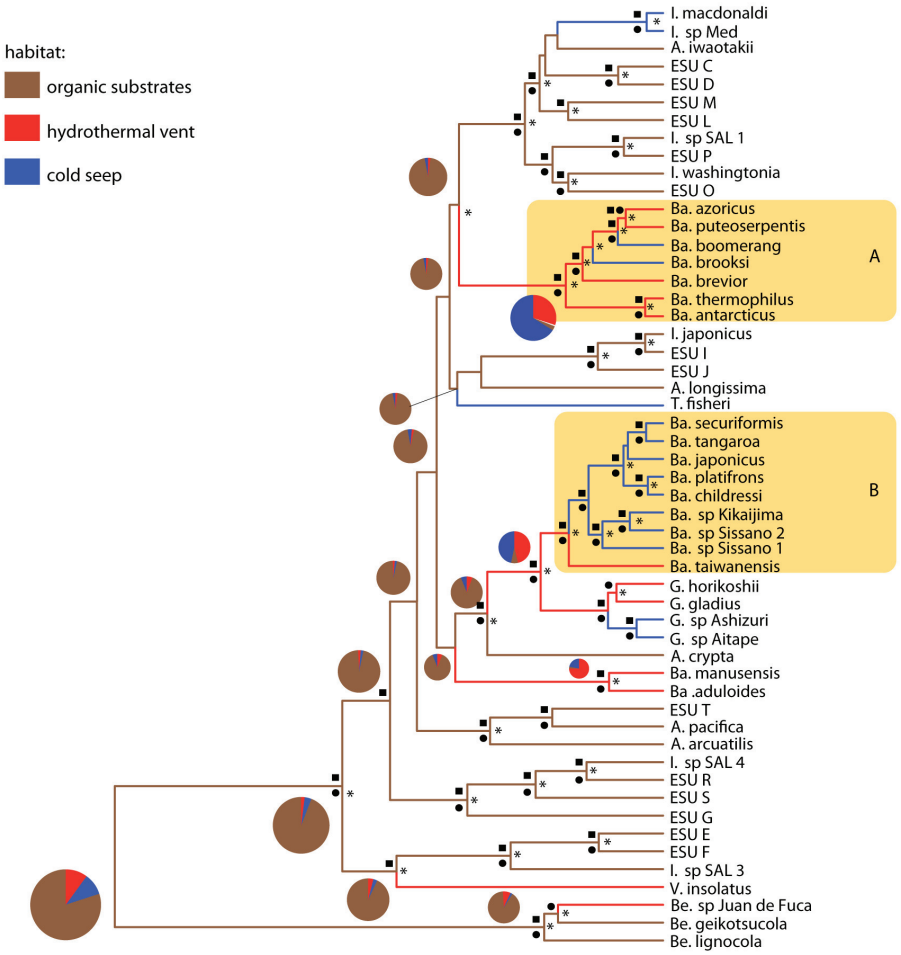


Figure 1.10 Bayesian estimates of Mytilidae phylogenies based on nuclear and mitochondrial genes. Pie charts indicate probabilities of the ancestral habitat. Black squares, circles and asterisks at nodes indicate posterior probabilities greater than or equal to 0.99, bootstrap values greater than or equal to 75% (95% for lineages highlighted in yellow), and nodes inferred in analyses of both nuclear and mitochondrial genes, respectively. **A** "True" *Bathymodiolus* clade and **B** "*Bathymodiolus*" clade. Figure modified from Lorion *et al.*, 2013

genus, which includes *B. boomerang* (Figure 1.10). Despite clear evidence of the presence of two distinct genera, no official consensus has been accepted and the "*Bathymodiolus*" denomination remains.

The phylogeny of bathymodiolin mussels suggests a shallow water origin (Distel *et al.*, 2000; Thubaut *et al.*, 2013). Initially, the stepping stone

hypothesis was proposed as a possible way of colonizing the deep sea. It states that mussels progressively adapted to the deep-sea environment by colonizing chemosynthetic environments such as shallow colds seeps, mud volcanoes, wood and whale falls. More recent phylogeny, however, suggests a more complex evolution, whereby shallow events still appear to be the origin but adaptation to the deep sea evolved independently multiple times (Thubaut *et al*., 2013).

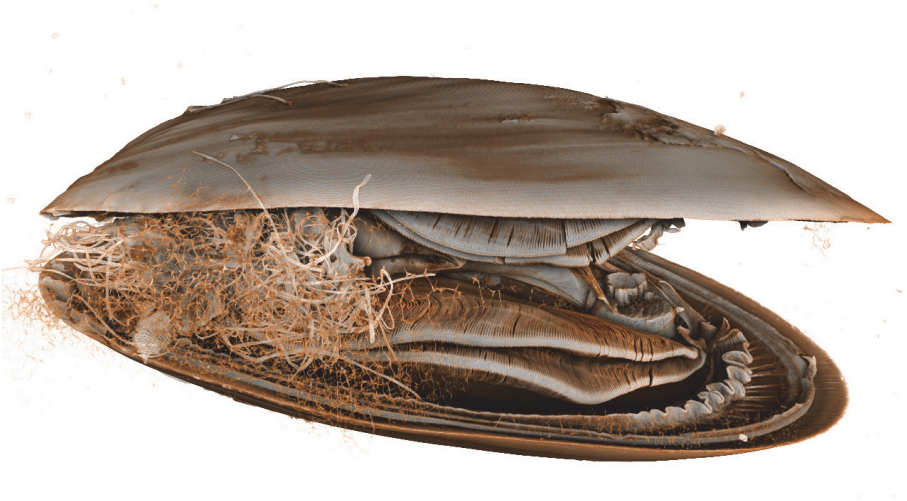


Figure 1.11 Volume rendering of a *Bathymodiolus azoricus* from 3 dimensional data set scans. Courtesy of *B.* Geier

1.6.2.1.2 Host morphological adaptation

The anatomical organization of bathymodiolin mussels is very close to their shallow water relatives from the Mytilidae family (Le Pennec and Hily, 1984; Fiala-Médioni *et al*., 1986). Notable exceptions are a reduced gastrointestinal tract and overgrown gills.

Bathymodiolin mussels host chemoautotrophic bacterial symbionts within their gill epithelia (Figure 1.12) allowing their evolutionary success in colonizing hydrothermal vents and colds seeps (Nelson and Hagen, 1995; Nelson *et al*., 1995). The symbionts are localized in specialized cells, called bacteriocytes, which are exposed to water flowing through the gills. The presence of the bacteria in the respiratory organ of animals, which is exposed to reduced compounds, suggests that the symbionts are fed by the constant water flow created by the host and, in exchange, the symbiont contributes to the host's nutrition (Le Pennec and Hily, 1984; Fiala-Médioni *et al*., 1986; Auffret and Le Pennec, 1992). The exact feeding method is still unknown and hypotheses include speculation as to whether the host digests the symbionts directly or whether the nutrients are transferred from the bacteria to the host (Fiala-Médioni *et al*., 1994; Bettencourt *et al*., 2008; Kádár *et al*., 2008; Detree *et al*., 2016).

1.6.2.1.3 Host distribution

The different species of bathymodiolin mussels have specific distributions around the world. Mussels belonging to the true *Bathymodiolus* genus are endemic to hydrothermal vents, with the exception of *B. heckerae* and *B.* boomerang, which also colonize cold seeps. Different species of *Bathymodiolus* inhabit different oceans; *B. azoricus* and *B.* puteoserpentis are found associated with MAR hydrothermal vent sites, while *B.* thermophilus colonizes the East Pacific Rise (Miyazaki *et al*., 2010).

On the other hand, mussels belonging to the "*Bathymodiolus*" genus are mostly endemic to cold seeps, with the exception of "*Bathymodiolus*" tawainensis, which is found at hydrothermal vents. Similarly, there is a specific species distribution within the "*Bathymodiolus*" genus. *"B." childressi* colonizes colds seeps in the GoM, while *B.* japonicus and *B.* platifrons are found in the Pacific Ocean (Thubaut *et al*., 2013).

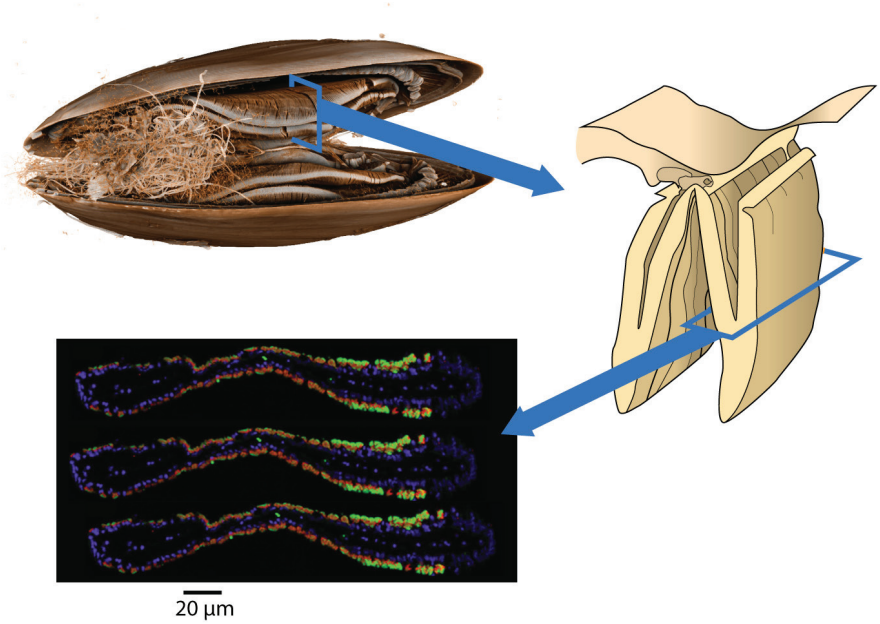


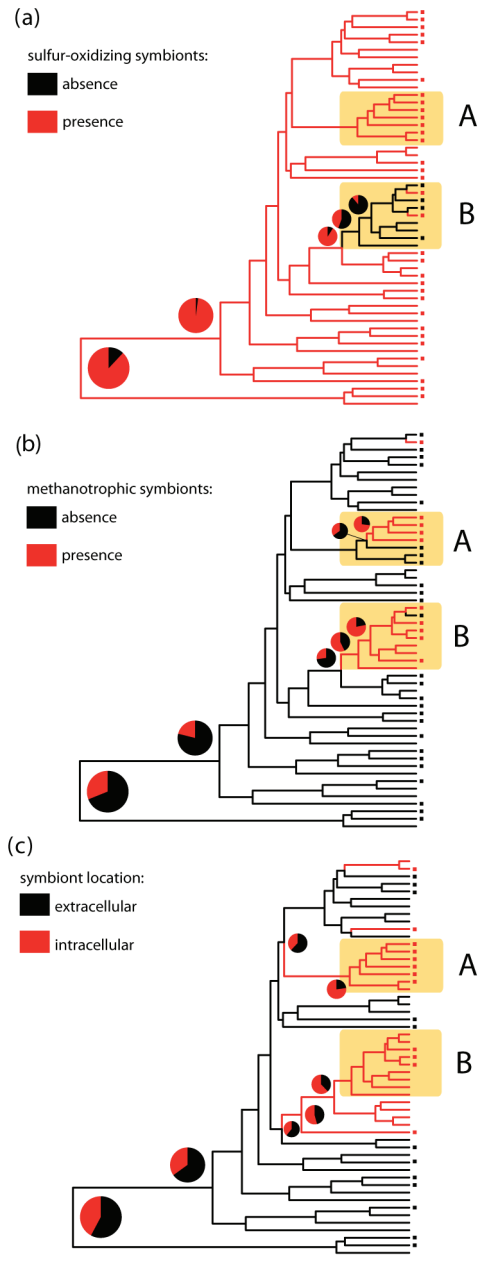
Figure 1.12 Schematic of the gill anatomy and structure of a bathymodiolin mussel. **A.** Volume rendering of micro computed tomography data set. *B.* Schematic of a bathymodiolin gill. **C.** Fluorescence in situ hybridization picture of a cross section of *Bathymodiolus azoricus* gills. In blue is DAPI targeting DNA, red are signal targeting the methane oxidizing endosymbiont and green the sulfur oxidizing one.

1.6.2.2 Symbiont diversity

Bathymodiolin mussels host two main types of endosymbionts. Most of the bathymodiolin species are associated with sulfur oxidizing bacteria, a small group of species are only associated with methane oxidizing bacteria, while the last group of species have both endosymbionts at the same time (Figure 1.13 Duperron *et al*., 2007).

1.6.2.2.1 Thiotrophic symbioses

Many bathymodiolin mussels are associated with thiotrophic bacteria (Nelson *et al*., 1995; Duperron *et al*., 2007), although most of the species in the "*Bathymodiolus*" clade appear to have lost this association, indicating that



the thiotrophic bacteria were lost early in the genus differentiation (Lorion *et al.*, 2013). The sulfur oxidizing endosymbionts all belong to Gammaproteobacteria, more specifically to a deep sea clade within the SUPO5 cluster. They are interspersed by free-living individuals, such as cultivated Candidatus Thioglobus singularis (Sayavedra *et al.*, 2015).

The phylogeny of the mussel host and the symbiotic sulfur oxidizing bacteria are not congruent, which may be explained by a horizontal transmission of the symbiont to the host (Won *et al.*, 2008). This mode of symbiont transmission, in which the host needs to be recolonized every generation by environmental bacteria, may have led to symbiont replacement during the evolutionary history of bathymodiolin mussels.

Figure 1.13 Representation of the (*a*) Evolution of the presence/absence of sulfur-oxidizing symbionts. (*b*) Evolution of the presence/absence of methanotrophic symbionts. (*c*) Evolution of the location of symbionts in the gill epithelium. Squares at tips of chronograms *b*, *c* and *d* indicate available data. Based on the phylogenetic tree of Figure 1.10. **A** "True" *Bathymodiolus* clase and **B** "*Bathymodiolus*" clade. Figure modified from Lorion *et al.*, 2013.

1.6.2.2.2 Methanotrophic symbioses

Methanotrophic bacteria have established a symbiotic association with bathymodiolin mussels at least twice in evolutionary history (DeChaine and Cavanaugh, 2006). Most mussels from the "*Bathymodiolus*" genus are associated with methanotrophic bacteria, rather than thiotrophic bacteria (Lorion *et al*., 2013). The bacteria all belong to an independent clade branching within the gammaproteobacterial family of Methylococcaceae, between the Methylomicrobium and Methylobacter genera (Distel and Cavanaugh, 1994). The closest relative is the recently cultivated Methyloprofundus sedimenti, isolated from hydrothermal vent sediments (Tavormina *et al*., 2015). Similar to thiotrophic endosymbionts, no obvious congruence has been observed between the host and endosymbiont phylogenies (Petersen and Dubilier, 2009).

1.6.2.2.3 Dual symbioses

A branch of the true *Bathymodiolus* clade includes mussels that host both methane and sulfur oxidizing bacteria (Figure 1.14). The symbionts belong to the same groups as in the single symbiont system. The association of the two chemosynthetic endosymbionts confers a metabolic versatility to the host and enables it to adapt to different environmental variations (Duperron *et al*., 2006, 2009; Lorion *et al*., 2013).

1.6.2.2.4 Other associated Bacteria

Some bathymodiolin species have been found to harbor additional bacterial phylotypes. *B. heckerae* mussels harbor two phylogenetically distinct sulfur oxidizing bacterial phylotypes, methane oxidizers and methyl oxidizing bacteria (Duperron *et al*., 2007). The latest probably feed on methanol leaking from the methane oxidation metabolism (Cavanaugh *et al*., 1992). Additionally, Cycloclasticus bacteria have been found associated with *B. heckerae* at asphalt seeps and were initially thought to be hydrocarbon degraders (Raggi *et al*., 2013).

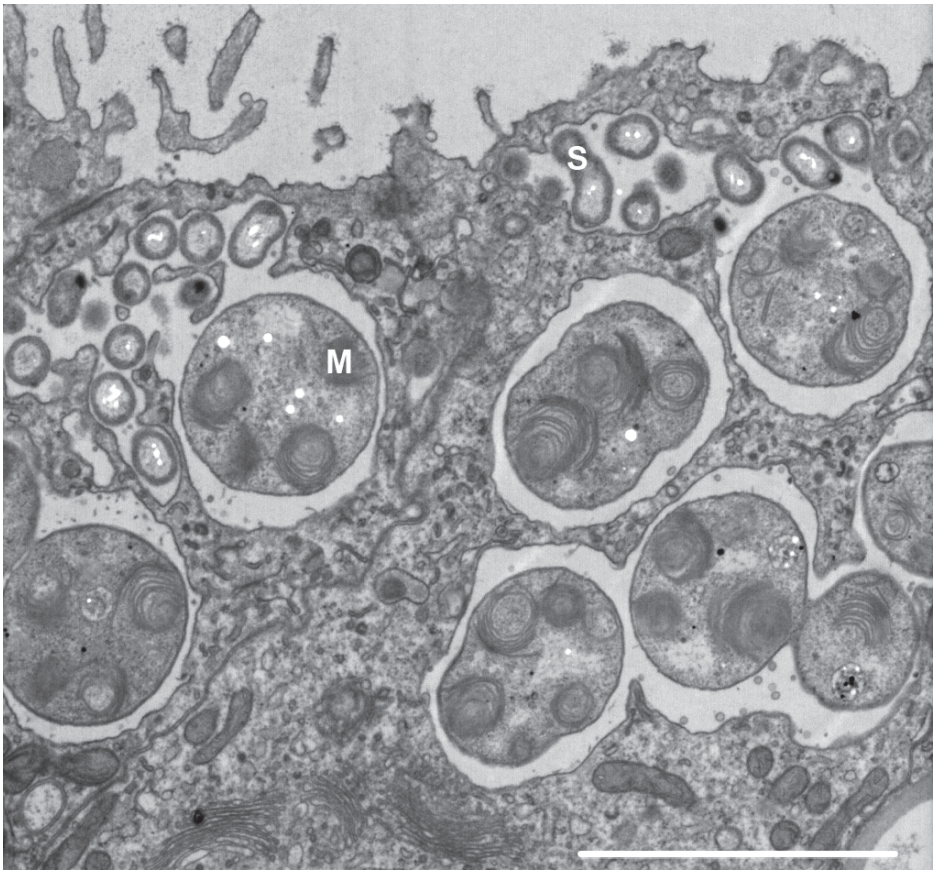


Figure 1.14 Transmission electron microscopy of a *Bathymodiolus* gill filament cross section. **S:** Sulfur oxidizing endosymbiont, **M** Methanotrophic endosymbiont. Scale: 3 µm. Courtesy of N. Leisch

Finally, the intranuclear Bacteria Candidatus Endonuclear bathymodiolii, thought to be pathogens, have been found associated with seven different bathymodiolin species. Preliminary studies have shown that these pathogens are only capable of infecting cells free of symbionts, suggesting a protective role of the endosymbiont (Zielinski *et al*., 2009).

1.6.3 Rimicaris shrimps

Rimicaris shrimp belong to the Alvinocarididae family and to date three species have been described *Rimicaris exoculata* (Figure 1.15) found at hydrothermal vents on the MAR, *Rimicaris hybisae* found on the MCSC and *Rimicaris* kairei found in the Indian Ocean (Nye *et al*., 2011). These shrimps occur in large swarms, which can reach up to several thousand individuals per square meter (Van Dover *et al*., 1988).

To date, three species of *Rimicaris* shrimp have been described: *Rimicaris* excolulata, *Rimicaris kairei* and *Rimicaris hybisae*. Each species is endemic to a different oceanic region. *R. exoculata* colonize vents on the MAR, *R.* kairei are found at the Kairei hydrothermal vent in the Indian Ocean and *R. hybisae* have only recently been discovered on MCSC vents (Watabe and Hashimoto, 2002; Nye *et al*., 2011; Teixeira *et al*., 2012).

Figure 1.15 3 dimensional reconstruction of a *Rimicaris exoculata* from µCT scans. Courtesy of *B.* Geier

1.6.3.1 Host morphological adaptation

One of the most remarkable adaptations of *Rimicaris* shrimps is the presence of an overgrown gill chamber with two mouthpart appendages, the scaphognathite of the second maxillae and the exopodite of the first maxilliped, that are greatly expanded in comparison to other shrimp species (Komai and Segonzac, 2008). These appendages are flat and covered with thick setae and the entire gill chamber, including the appendages, is covered by a thick population of filamentous bacteria. Previous studies have indicated that the shrimps move these appendages in the gill chamber and these movements have been suggested to enable the shrimp to scrub off bacteria and feed on them, as well as create a water current toward the gills and mouth located posteriorly (Van Dover *et al*., 1988).

1.6.3.2 Ectosymbiotic association

R. *exoculata* was the first described species and the presence of a dense population of bacteria was quickly established (Van Dover *et al*., 1988). Molecular analyses showed that Epsilon- and Gammaproteobacteria were the dominant symbionts, with different phylotypes geographically distributed along the various vents on the MAR (Petersen *et al*., 2010). Recent genomic studies have also suggested a low abundance of Zetaproteobacteria (Jan *et al*., 2014). Further studies have also shown that a switch occurs in the dominant taxa during the shrimp's life cycle: juveniles are dominated by Gammaproteobacteria, whereas adults are dominated by Epsilonproteobacteria (Zbinden *et al*., 2012). Additionally, clone libraries showed a stable association between the ectosymbionts and the shrimps, even when taking into account that shrimps molt multiple times during their life cycle.

The bacterial community present in the gill chamber is thought to contribute to the nutrition of its host (Van Dover *et al*., 1988; Gebruk *et al*., 2000). Initial observations of *Rimicaris* shrimp appendage movements have suggested that the shrimps feed on their own bacterial population as well as

particulate material from the environment. A recent incubation experiment, however, challenged this hypothesis by showing carbon transfer from the bacteria to the host (Ponsard *et al*., 2013). More in depth work is necessary fully to understand the source of the *Rimicaris* shrimp nutrition.

1.7 Challenges of deep-sea research

Hydrothermal vents and colds seeps are usually remote and difficult to access. These are some of the main challenges deep-sea research must overcome. Accessing such sites usually requires multi-disciplinary scientific campaigns to be organized and the mobilization of scientific vessels, as well as specific equipment. The first deep-sea fauna samples were collected randomly using trawls, bait traps and TV grabs. Recently, this has been done with remotely (ROV), automated (AUV) and, less frequently, human (HOV) operated vehicles.

With an average depth of 3800 m, the deep ocean has a local pressure of around 390 atmospheric units. Sampling hydrothermal vent or cold seep animals and maintaining them alive under atmospheric pressure conditions is still a challenge. This is not the case, however, for individuals sampled above a 1000 m depth, for example at continental slope sites. The maintenance of *"B." childressi* individuals in an aquarium has recently been documented and there has even been a successful induction of a spawning event (Arellano and Young, 2009; Bettencourt *et al*., 2010). However, the cultivation of host and symbionts remains inaccessible for many symbiotic systems. This demands the fixation of samples as soon as possible after sampling, with the drawback of mainly providing snapshots in time of biological systems. Nevertheless, a lot has been achieved with these samples, from morphological descriptions and quantification of symbionts, to the identification and localization of specific symbionts within the host's tissues. In depth biological mechanisms requiring a temporal component to be considered, such as molecular interaction mechanisms or in depth analysis of the influence of external factors, have yet to be conducted routinely for deep-sea organisms.

1.8 Sequencing technologies

In the past 30 years, sequencing technologies have quickly evolved from the initial two-dimensional chromatography of the early 1970's, whereby sequencing was a time consuming method with very short sequence sizes, to the advent of the Sanger chain termination method, which has allowed faster and systematic sequencing of DNA fragments. Finally, the rise of next generation sequencing (NGS) in the past decade has transformed sequencing into an affordable high throughput method of choice for the analysis of biological systems (reviewed in Loman and Pallen, 2015).

1.8.1 Sanger sequencing

In 1977, the same year hydrothermal vents were discovered on the Galapagos rift, F. Sanger designed the chain termination sequencing method, which became one of the most used techniques until the mid-2000's (Sanger *et al*., 1977). This method was relatively easy to perform and produced reliable sequencing of sequences around 1000 bp length. However, this was still not enough to sequence large chromosome sequences. Robert Staden later proposed the first computer program automatically to read sequencing gels and assemble/align the first sequencing results (Staden, 1979). To overcome the sequence length limitation, shotgun sequencing was developed; it includes a step in which long DNA chains are sheared into smaller fragments, allowing reliable sequencing (Fleischmann *et al*., 1995). It ultimately led to the development of large projects, such as the first Human Genome sequencing project in 2001 (Lander *et al*., 2001).

1.8.2 Next generation sequencing

In the second part of the 2000's multiple NGS platforms became widely accessible (reviewed in Metzker, 2010). Roche 454 pyrosequencing and the Illumina (Solexa) platform were the beginning of the NGS era. The pyrosequencing method is a non-electrophoretic method based on the measurement of bioluminescent nucleotides as they are incorporated into a

new DNA strand. The order and intensity of the light peaks are recorded, which reveals the underlying DNA sequence (Ronaghi *et al*., 1996; Ronaghi, 1998). Although the drawback of such a method was that it generated small length sequences, the method compensated this bias with a fast turnover and high throughput. One of the first applications to marine microbial diversity used sequence tags from hyper-variable regions of the 16S ribosomal RNA (rRNA) gene to discern the phylogenetic diversity of deep water masses in the North Atlantic. It discovered a "rare biosphere" previously unknown (Sogin *et al*., 2006). Soon the field of bioinformatics generated new methods of processing and assembling small length sequences, allowing the combination of shotgun sequencing and NGS. The NGS technology has evolved at fast pace since taking over the sequencing market, with new and more efficient platforms. These include the Ion Torrent Personal Genome Machine, Applied Biosystems SOLiD system and Illumina MiSeq or HiSeq (Reviewed in Metzker, 2010). The last named is the most widely used sequencing technology, due to offering low cost sequencing, short turnover and increased read length as well as read number (Schirmer *et al*., 2016).

In comparison to Sanger sequencing, in which a complete human genome cost 100 million dollars in 2001 and took years to generate and assemble, this can now be sequenced within a week for only 1000 $ (Lander *et al*., 2001; Check Hayden, 2014).

Current high-throughput technologies have led to a shift in the traditional way of handling genomic data; from closed genome from pure cultures, studies shifted to the analysis of draft genomes and comparisons of specific regions of interest. Additionally, affordable NGS sequencing projects and recent bioinformatics developments now allow the sequencing of environmental samples of unknown composition. Such approaches allow the recovery of genomic information from multiple organisms present in a single sample and investigation into the genomic information of yet uncultivated organisms, and understanding their role in the environment.

1.9 Deep sea sequencing

Sequencing approaches quickly became the method of choice for the analysis of deep-sea organisms. Early studies using clone library approaches investigated specific genes of interest, for example, phylogenetic markers used to investigate symbiotic systems (DeChaine and Cavanaugh, 2006; Jones *et al*., 2006; Duperron *et al*., 2007). Sanger sequencing enabled the study of bathymodiolin-associated bacterial communities and showed the tight association between certain bacterial phylotypes and the host (Distel *et al*., 1995). It also allowed the resolution of host phylogenies by sequencing phylogenetic marker genes (e.g. mitochondrial cytochrome oxidase I – COI), with which ambiguous anatomical descriptions were not enough to separate bathymodiolin mussels into two distinct genera (Jones *et al*., 2006).

Next generation sequencing also allows the in-depth analysis of whole systems and investigation of the molecular mechanisms involved in the interaction between host and symbiont. For example, a recent study showed the widespread presence of toxin-like related proteins within the sulfur oxidizing symbionts of bathymodiolin mussels, which could be involved in beneficial interactions with their host, or protection against other pathogens (Sayavedra *et al*., 2015).

1.10 Aim of this thesis

The aim of this thesis was to use widely available sequencing technologies to investigate deep sea symbioses. In particular, the ectosymbiotic associations within *Bathymodiolus* mussels and *Rimicaris* shrimps were investigated.

1.10.1 The epibionts of bathymodiolin mussels.

Preliminary studies, using 16S rRNA clone libraries, suggested the presence of low abundance Epsilonproteobacteria in *“B.” childressi*. Previous studies had already shown that certain species of bathymodiolin mussel can be associated with secondary bacterial phylotypes. The aim of this project

was thus to investigate the presence of Epsilonproteobacterial sequences in these clone libraries and assess whether these bacteria have a more widespread association with bathymodiolin mussels. Additionally, the location in the gill tissue and the potential role of these bacteria in the host were investigated. Finally, we used Illumina sequencing of metagenomic and metatranscriptomic libraries to investigate the potential influence, mutualistic or pathogenic, of bacteria associated with their mussel host.

1.10.2 Ectosymbiont populations of *Rimicaris exoculata*

The discovery of two hydrothermal vents fields on the MCSC led to the description of *Rimicaris hybisae*, a new species of deep-sea shrimp. Early analysis showed that, like their close relative *R. exoculata*, this new species also hosted a dense bacterial population within their gill chamber. The Cayman hydrothermal vent fields occur on different geological bases and are located 25 km apart, with a 2500 m depth difference. This unique setting offers the opportunity to investigate the effect of different physical and chemical environmental conditions on the ectosymbiotic population. We sequenced amplicon libraries of the V3-V4 region of the 16S rRNA sequences, using Illumina technologies to compare ectosymbiotic populations of *Rimicaris* individual, to environmental bacterial populations. Finally, we investigated the potential presence of community composition patterns of *Rimicaris* ectosymbionts and whether these are influenced by the environment or specific to the host species.

1.11 References

Abbott, L.K. and Lumley,S. (2014) Mycorrhizal Fungi: Use in Sustainable Agriculture and Land Restoration Solaiman,Z.M., Abbott,L.K., and Varma,A. (eds) Springer Berlin Heidelberg, Berlin, Heidelberg.

Ahmed, M.Z., De Barro, P.J., Ren, S.-X., Greeff, J.M., and Qiu, B.-L. (2013) Evidence for Horizontal Transmission of Secondary Endosymbionts in the *Bemisia tabaci* Cryptic Species Complex. *PLoS One* **8**: e53084.

Aoshima, M. (2007) Novel enzyme reactions related to the tricarboxylic acid cycle: phylogenetic/functional implications and biotechnological applications. *Appl. Microbiol. Biotechnol.* **75**: 249–255.

Arellano, S.M. and Young, C.M. (2009) Spawning, development, and the duration of larval life in a deep-sea cold-seep mussel. *Biol. Bull.* **216**: 149–62.

Auffret, M. and Le Pennec, M. (1992) Microscopical observations on the excretory organs of the hydrothermal mussel *Bathymodiolus thermophilus* (Mollusca: Bivalvia). *J. Mar. Biol. Assoc. United Kingdom. Plymouth [J. Mar. Biol. Assoc. U.K.]* **72**: 503–506.

De Bary, A. (1879) Die erscheinung der symbiose Strassburg, Germany: Verlag von Karl J. Trubner.

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

Berg, I.A., Kockelkorn, D., Buckel, W., and Fuchs, G. (2007) A 3-Hydroxypropionate/4-Hydroxybutyrate Autotrophic Carbon Dioxide Assimilation Pathway in Archaea. *Science (80-.).* **318**: 1782–1786.

Berg, I.A., Kockelkorn, D., Ramos-Vera, W.H., Say, R.F., Zarzycki, J., Hügler, M., *et al*. (2010) Autotrophic carbon fixation in archaea. *Nat. Rev. Microbiol.* **8**: 447–460.

Berg, I.A., Ramos-Vera, W.H., Petri, A., Huber, H., and Fuchs, G. (2010) Study of the distribution of autotrophic CO_2 fixation cycles in Crenarchaeota. *Microbiology* **156**: 256–269.

Bettencourt, R., Costa, V., Laranjo, M., Rosa, D., Pires, L., Colaco, A., *et al*. (2011) Out of the deep sea into a land-based aquarium environment: investigating physiological adaptations in the hydrothermal vent mussel *Bathymodiolus azoricus*. *ICES J. Mar. Sci.* **68**: 357–364.

Bettencourt, R., Dando, P., Rosa, D., Riou, V., Colaço, A., Sarrazin, J., *et al*. (2008) Changes of gill and hemocyte-related bio-indicators during long

term maintenance of the vent mussel *Bathymodiolus azoricus* held in aquaria at atmospheric pressure. *Comp. Biochem. Physiol. - A Mol. Integr. Physiol.* **150**: 1–7.

Bright, M. and Bulgheresi,S. (2010) A complex journey: transmission of microbial symbionts. *Nat. Rev. Microbiol.* **8**: 218–230.

Brooks, J.M., Wiesenburg, D.A., Camel, R.S., Macdon, I.R., Fisher, C.R., Guinasso, N.L., and Bright, J. (1990) Salt, Seeps and Symbiosis in the Gulf of Mexico. **71**: 1986–1988.

Buron, I. de and Morand,S. (2004) Deep-sea hydrothermal vent parasites: why do we not find more? *Parasitology* **128**: S0031182003004347.

Byrne, N., Strous, M., Crépeau, V., Kartal, B., Birrien, J.-L., Schmid, M., *et al*. (2009) Presence and activity of anaerobic ammonium-oxidizing bacteria at deep-sea hydrothermal vents. *ISME J.* **3**: 117–23.

Campbell, B.J. and Cary, CS. (2004) Abundance of Reverse Tricarboxylic Acid Cycle Genes in Free-LivingMicroorganisms at Deep-Sea Hydrothermal Vents. *Appl Env. Microbiol* **70**: 6282–6289.

Cani, P.D. and Delzenne, N.M. (2011) The gut microbiome as therapeutic target. *Pharmacol. Ther.* **130**: 202–212.

Cavanaugh, C.M., Gardiner, S.L., Jones, M.L., Jannasch, H.W., and Waterbury, J.B. (1981) Prokaryotic Cells in the Hydrothermal Vent Tube Worm *Riftia pachyptila* Jones: Possible Chemoautotrophic Symbionts. *Science (17).* **213**: 340–342.

Cavanaugh, C.M., Wirsen, C.O., and Jannasch, H.W. (1992) Evidence for methylotrophic symbionts in a hydrothermal vent mussel (bivalvia: mytilidae) from the Mid-Atlantic Ridge. *Appl. Environ. Microbiol.* **58**: 3799–803.

Charlou, J.., Donval, J.., Douville, E., Jean-Baptiste, P., Radford-Knoery, J., Fouquet, Y., *et al*. (2000) Compared geochemical signatures and the evolution of Menez Gwen (37°50′N) and Lucky Strike (37°17′N) hydrothermal fluids, south of the Azores Triple Junction on the Mid-Atlantic Ridge. *Chem. Geol.* **171**: 49–75.

Check Hayden, E. (2014) Technology: The $1,000 genome. *Nature* **507**: 294–295.

Chow, J., Lee, S.M., Shen, Y., Khosravi, A., and Mazmanian, S.K. (2010) Host–Bacterial Symbiosis in Health and Disease. *Adv Immunol.* **107**: 243–274.

Cordes, E.E., Becker, E.L., Hourdez, S., and Fisher, C.R. (2010) Influence of

foundation species, depth, and location on diversity and community composition at Gulf of Mexico lower-slope cold seeps. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **57**: 1870–1881.

Cordes, E.E., Bergquist, D.C., and Fisher, C.R. (2009) Macro-Ecology of Gulf of Mexico Cold Seeps. *Ann. Rev. Mar. Sci.* **1**: 143–168.

Cordes, E.E., Carney, S.L., Hourdez, S., Carney, R.S., Brooks, J.M., and Fisher, C.R. (2007) Cold seeps of the deep Gulf of Mexico: Community structure and biogeographic comparisons to Atlantic equatorial belt seep communities. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **54**: 637–653.

von Cosel, *R.* and Olu, K. (1998) Gigantism in Mytilidae. A new *Bathymodiolus* from cold seep areas on the Barbados accretionary Prism. *Comptes Rendus l'Académie des Sci. - Ser. III - Sci. la Vie* **321**: 655–663.

DeChaine, E.G. and Cavanaugh, C.M. (2006) Symbioses of methanotrophs and deep-sea mussels (Mytilidae: Bathymodiolinae). In, Overmann,J. (ed), *Progress in molecular and subcellular biology*. Springer-Verlag Berlin, pp. 227–249.

Detree, C., Chabenat, A., Lallier, F.H., Satoh, N., Shoguchi, E., Tanguy, A., and Mary, J. (2016) Multiple I-Type Lysozymes in the Hydrothermal Vent Mussel *Bathymodiolus azoricus* and Their Role in Symbiotic Plasticity. *PLoS One* **11**: e0148988.

Distel, D.L., Baco, a R., Chuang, E., Morrill, W., Cavanaugh, C.M., and Smith, C.R. (2000) Do mussels take wooden steps to deep-sea vents? *Nature* **403**: 725–6.

Distel, D.L. and Cavanaugh, C.M. (1994) Independent phylogenetic origins of methanotrophic and chemoautotrophic bacterial endosymbioses in marine bivalves. *J. Bacteriol.* **176**: 1932–1938.

Distel, D.L., Lee, H.K., and Cavanaugh, C.M. (1995) Intracellular coexistence of methano- and thioautotrophic bacteria in a hydrothermal vent mussel. *Proc. Natl. Acad. Sci. U.S. A.* **92**: 9598–9602.

Douglas, A.E. (1998) Nutritional Interactions in Insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria Buchnera. *Annu. Rev. Entomol.* **43**: 17–37.

Douglas, A.E. (2014) Symbiosis as a General Principle in Eukaryotic Evolution. *Cold Spring Harb. Perspect. Biol.* **6**: a016113–a016113.

Van Dover, C.L. (2000) The Ecology of Deep-sea Hydrothermal Vents Princeton University Press.

Van Dover, C.L., Fry, B., Grassle, J.F., Humphris, S., and Rona, P.A. (1988) Feeding biology of the shrimp *Rimicaris exoculata* at hydrothermal vents on the Mid-Atlantic Ridge. *Mar. Biol.* **98**: 209–216.

Dubilier, N., Amann, R., Erséus, C., Muyzer, G., Park, S., Giere, O., and Cavanaugh, C. (1999) Phylogenetic diversity of bacterial endosymbionts in the gutless marine oligochete *Olavius loisae* (Annelida). *Mar. Ecol. Prog. Ser.* **178**: 271–280.

Duperron, S., Bergin, C., Zielinski, F., Blazejak, A., Pernthaler, A., McKiness, Z.P., *et al*. (2006) A dual symbiosis shared by two mussel species, *Bathymodiolus azoricus* and *Bathymodiolus puteoserpentis* (Bivalvia: Mytilidae), from hydrothermal vents along the northern Mid-Atlantic Ridge. *Environ. Microbiol.* **8**: 1441–1447.

Duperron, S., Lorion, J., Samadi, S., Gros, O., and Gaill, F. (2009) Symbioses between deep-sea mussels (Mytilidae: Bathymodiolinae) and chemosynthetic bacteria: diversity, function and evolution. *C. R. Biol.* **332**: 298–310.

Duperron, S., Sibuet, M., MacGregor, B.J., Kuypers, M.M.M., Fisher, C.R., and Dubilier, N. (2007) Diversity, relative abundance and metabolic potential of bacterial endosymbionts in three *Bathymodiolus* mussel species from cold seeps in the Gulf of Mexico. *Environ. Microbiol.* **9**: 1423–1438.

Durden, J.M., Bett, B.J., Jones, D.O.B., Huvenne, V.A.I., and Ruhl, H.A. (2015) Abyssal hills – hidden source of increased habitat heterogeneity, benthic megafaunal biomass and diversity in the deep sea. *Prog. Oceanogr.* **137**: 209–218.

Eloe-Fadrosh, E.A. and Rasko, D.A. (2013) The Human Microbiome: From Symbiosis to Pathogenesis. *Annu. Rev. Med.* **64**: 145–163.

Eppley, R.W. and Peterson, B.J. (1979) Particulate organic matter flux and planktonic new production in the deep ocean. *Nature* **282**: 677–680.

Felbeck, H. (1981) Chemoautotrophic Potential of the Hydrothermal Vent Tube Worm, *Riftia pachyptila* Jones (Vestimentifera). *Science (80-.).* **213**: 336–338.

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

Fiala-Médioni, A., Michalski, J., Jollés, J., Alonso, C., and Montreuil, J. (1994) Lysosomic and lysozyme activities in the gill of bivalves from deep hydrothermal vents. *Comptes rendus l'Académie des Sci. Série 3, Sci. la*

vie **3**: 239–244.

Fleischmann, R., Adams, M., White, O., Clayton, R., Kirkness, E., Kerlavage, A., *et al*. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science (80-.).* **269**: 496–512.

Franck, A.B. (1877) Uber die biologischen verhaltnisse des thallus eineger krustenflechten. *Beitrage zur Biol. der Pflanzen.* **2**: 123–200.

Fruh-Green, G.L. (2003) 30,000 Years of Hydrothermal Activity at the Lost City Vent Field. *Science (80-.).* **301**: 495–498.

Galkin,S. V. (2016) Structure of Hydrothermal Vent Communities. In, Galkin,S. V. and Demina,L.L. (eds), *Trace Metal Biogeochemistry and Ecology of Deep-Sea Hydrothermal Vent Systems*. Springer International Publishing Switzerland, pp. 41–53.

Gebruk, A.V., Southward, E.C., Kennedy, H., and Southward, A.J. (2000) Food sources, behaviour, and distribution of hydrothermal vent shrimps at the Mid-Atlantic Ridge. *J. Mar. Biol. Assoc. UK* **80**: S0025315400002186.

German, C.R., Bowen, A., Coleman, M.L., Honig, D.L., Huber, J. a, Jakuba, M. V, *et al*. (2010) Diverse styles of submarine venting on the ultraslow spreading Mid-Cayman Rise. *Proc. Natl. Acad. Sci.* **107**: 14020–14025.

Goffredi, S.K. (2010) Indigenous ectosymbiotic bacteria associated with diverse hydrothermal vent invertebrates. *Environ. Microbiol. Rep.* **2**: 479–488.

Gruber-Vodicka, H.R., Dirks, U., Leisch, N., Baranyi, C., Stoecker, K., Bulgheresi, S., *et al*. (2011) Paracatenula, an ancient symbiosis between thiotrophic Alphaproteobacteria and catenulid flatworms. *Proc. Natl. Acad. Sci. U.*S. *A.* **108**: 12078–83.

Guri, M., Durand, L., Cueff-Gauchard, V., Zbinden, M., Crassous, P., Shillito, B., and Cambon-Bonavita, M.-A. (2012) Acquisition of epibiotic bacteria along the life cycle of the hydrothermal shrimp *Rimicaris exoculata*. *ISME J.* **6**: 597–609.

Gustafson, R.G., Turner, R.D., Lutz, *R.* a, and Vrijenhoek, R.C. (1998) A new genus and five new species of mussels (Bivalvia, Mytilidae) from deep-sea sulfide/hydrocarbon seeps in the Gulf of Mexico. *Malacologia* **40'**: 63–112.

Hannington, M.D., De Ronde, C.E.J., and Petersen,S. (2005) Sea-Floor Tectonics and Submarine Hydrothermal Systems. *Econ. Geol.* **100th Anni**: 111–141.

He, X., McLean, J.S., Edlund, A., Yooseph, S., Hall, A.P., Liu, S.-Y., *et al*. (2015) Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. *Proc. Natl. Acad. Sci.* **112**: 244–249.

van der Heijden, K., Petersen, J.M., Dubilier, N., and Borowski, C. (2012) Genetic connectivity between north and south Mid-Atlantic Ridge chemosynthetic bivalves and their symbionts. *PLoS One* **7**: e39994.

Herter, S., Fuchs, G., Bacher, A., and Eisenreich, W. (2002) A bicyclic autotrophic CO_2 fixation pathway in *Chloroflexus aurantiacus*. *J. Biol. Chem.* **277**: 20277–20283.

Hessler, R.R. and Smithey, W.M. (1983) The Distribution and Community Structure of Megafauna at the Galapagos Rift Hydrothermal Vents. In, *Hydrothermal Processes at Seafloor Spreading Centers*. Springer US, Boston, MA, pp. 735–770.

Hodgkinson, M.R.S., Webber, A.P., Roberts, S., Mills, R.A., Connelly, D.P., and Murton, B.J. (2015) Talc-dominated seafloor deposits reveal a new class of hydrothermal system. *Nat. Commun.* **6**: 10150.

Huber, H., Gallenberger, M., Jahn, U., Eylert, E., Berg, I. a, Kockelkorn, D., *et al*. (2008) A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum *Ignicoccus hospitalis*. *Proc. Natl. Acad. Sci. U.*S. *A.* **105**: 7851–7856.

Hügler, M. and Sievert, S.M. (2011) Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean. *Ann. Rev. Mar. Sci.* **3**: 261–289.

Imlay, J.A. (2006) Iron- sulfur clusters and the problem with oxygen. *Mol. Microbiol.* **59**: 1073–1082.

Jan, C., Petersen, J.M., Werner, J., Teeling, H., Huang, S., Glöckner, F.O., *et al*. (2014) The gill chamber epibiosis of deep-sea shrimp *Rimicarisexoculata* : an in-depth metagenomic investigation and discovery of Zetaproteobacteria. *Environ. Microbiol.* **16**: 2723–2738.

Jannasch, H.W. (1995) Microbial interactions with hydrothermal fluids., pp. 273–296.

Jones, W.J.J. and Vrijenhoek, R.C. (2006) Evolutionary relationships within the "*Bathymodiolus*" childressi group. *Cah. Biol. Mar.* **47**: 403–407.

Jousselin, E., Desdevises, Y., and Coeur d'acier, A. (2009) Fine-scale cospeciation between Brachycaudus and Buchnera aphidicola: bacterial genome helps define species and evolutionary relationships

in aphids. *Proc. R. Soc. B Biol. Sci.* **276**: 187–196.

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

Keeling, P.J. (2010) The endosymbiotic origin, diversification and fate of plastids. *Philos. Trans. R. Soc. B Biol. Sci.* **365**: 729–748.

Kelley, D.S., Baross, J.A., and Delaney, J.R. (2002) Volcanoes, Fluids, and Life at Mid-Oceanic Ridge Spreading Centers. *Annu. Rev. Earth Planet. Sci.* **30**: 385–491.

Kenk, V.C. and Wilson, B.R. (1985) A new mussel (Bivalvia, Mytilidae) from hydrothermal vents in the Galapagos Rift zone. *Malacologia* **26**: 253–271.

Kinsey, J.C. and German, C.R. (2013) Sustained volcanically-hosted venting at ultraslow ridges: Piccard Hydrothermal Field, Mid-Cayman Rise. *Earth Planet. Sci. Lett.* **380**: 162–168.

Kojima, S., Hashimoto, J., and Ohta,S. (1995) The distribution and the phylogenies of the species of genus *Calyptogena* and those of Vestimentiferans around Japan. *JAMSTEC J. Deep Sea Res.* **11**: 213–218.

Komai, T. and Segonzac, M. (2008) Taxonomic Review of the Hydrothermal Vent Shrimp Genera *Rimicaris* Williams & Rona and *Chorocaris* Martin & Hessler (Crustacea: Decapoda: Caridea: Alvinocarididae). *J. Shellfish Res.* **27**: 21–41.

Korem, T., Zeevi, D., Suez, J., Weinberger, A., Avnit-Sagi, T., Pompan-Lotan, M., *et al*. (2015) Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science (80-.).* **349**: 1101–1106.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., *et al*. (2001) Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.

Leung, T.L.F. and Poulin, *R.* (2008) Parasitism, commensalism, and mutualism: Exploring the many shades of symbioses. *Vie Milieu* **58**: 107–115.

Li, M., Wang, B., Zhang, M., Rantalainen, M., Wang, S., Zhou, H., *et al*. (2008) Symbiotic gut microbes modulate human metabolic phenotypes. *Proc. Natl. Acad. Sci.* **105**: 2117–2122.

Lilley, M.D., Butterfield, D. a, Lupton, J.E., and Olson, E.J. (2003) Magmatic events can produce rapid changes in hydrothermal vent chemistry.

Nature **422**: 878–881.

Lin, J., Purdy, G.M., Schouten, H., Sempere, J.-C., and Zervas, C. (1990) Evidence from gravity data for focused magmatic accretion along the Mid-Atlantic Ridge. *Nature* **344**: 627–632.

Ljungdahl, L.G. (2009) A Life with Acetogens, Thermophiles, and Cellulolytic Anaerobes. *Annu. Rev. Microbiol.* **63**: 1–25.

Loman, N.J. and Pallen, M.J. (2015) Twenty years of bacterial genome sequencing. *Nat. Rev. Microbiol.* **13**: 787–794.

Lonsdale, P. (1977) Clustering of suspension-feeding macrobenthos near abyssal hydrothermal vents at oceanic spreading centers. *Deep Sea Res.* **24**: 857–863.

Lorion, J., Kiel, S., Faure, B., Kawato, M., Ho, S.Y.W., Marshall, B., *et al*. (2013) Adaptive radiation of chemosymbiotic deep-sea mussels. *Proc. R. Soc. B Biol. Sci.* **280**: 20131243.

Maffei, G., Miozzi, L., Fiorilli, V., Novero, M., Lanfranco, L., and Accotto, G.P. (2014) The arbuscular mycorrhizal symbiosis attenuates symptom severity and reduces virus concentration in tomato infected by Tomato yellow leaf curl Sardinia virus (TYLCSV). *Mycorrhiza* **24**: 179–186.

Manoharan, P.T., Shanmugaiah, V., Balasubramanian, N., Gomathinayagam, S., Sharma, M.P., and Muthuchelian, K. (2010) Influence of AM fungi on the growth and physiological status of *Erythrina variegata* Linn. grown under different water stress conditions. *Eur. J. Soil Biol.* **46**: 151–156.

Margulis, L. (1993) Symbiosis in cell evolution: microbial communities in the Archean and Proterozoic eons New York: W. H. Freeman.

Margulis, L. (1975) Symbiotic theory of the origin of eukaryotic organelles; criteria for proof. *Symp Soc Exp Biol* 21–38.

Markert, S., Arndt, C., Felbeck, H., Becher, D., Sievert, S.M., Hügler, M., *et al*. (2007) Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila*. *Science (80-.).* **315**: 247–250.

Marshall, K.T. and Morris, R.M. (2015) Genome Sequence of " *Candidatus* Thioglobus singularis" Strain PS1, a Mixotroph from the SUP05 Clade of Marine Gammaproteobacteria. *Genome Announc.* **3**: e01155–15.

Martin, B.D. and Schwab, E. (2013) Current Usage of Symbiosis and Associated Terminology. *Int. J. Biol.* **5**: 32–45.

Martin, B.D. and Schwab, E. (2012) Symbiosis: "Living Together" in Chaos.

Stud. Hsitory Biol. **4**: 8–25.

McCutcheon, J.P. and Moran, N. a. (2011) Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**: 13–26.

McDermott, J.M., Seewald, J.S., German, C.R., and Sylva, S.P. (2015) Pathways for abiotic organic synthesis at submarine hydrothermal fields. *Proc. Natl. Acad. Sci.* **112**: 7668–7672.

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

Metzker, M.L. (2010) Sequencing technologies - the next generation. *Nat. Rev. Genet.* **11**: 31–46.

Miyazaki, J.-I., Martins, L. de O., Fujita, Y., Matsumoto, H., and Fujiwara, Y. (2010) Evolutionary Process of Deep-Sea *Bathymodiolus* Mussels. *PLoS One* **5**: e10363.

Moran, N. a (2006) Symbiosis. *Curr. Biol.* **16**: R866–R871.

Moreira, D. (2003) Are hydrothermal vents oases for parasitic protists? *Trends Parasitol.* **19**: 556–558.

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

Nakabachi, A., Yamashita, A., Toh, H., Ishikawa, H., Dunbar, H.E., Moran, N.A., and Hattori, M. (2006) The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science (80-.).* **314**: 267.

Nakagawa,S. and Takai, K. (2008) Deep-sea vent chemoautotrophs: Diversity, biochemistry and ecological significance. *FEMS Microbiol. Ecol.* **65**: 1–14.

Nelson, D.C. and hagen, K.D. (1995) Physiology and Biochemistry of Symbiotic and Free-Living Chemoautotrophic Sulfur Bacteria. *Am. Zool.* **35**: 91–101.

Nelson, D.C., Hagen, K.D., and Edwards, D.B. (1995) The gill symbiont of the hydrothermal vent mussel *Bathymodiolus thermophilus* is a psychrophilic, chemoautotrophic, sulfur bacterium. *Mar. Biol.* **121**: 487–495.

Nicolas, A. (1995) The Mid-Oceanic Ridges Springer Berlin Heidelberg, Berlin, Heidelberg.

Nybakken, J.W. and Bertness, M.D. (2005) Marine biology: an ecological

approach.

Nye, V., Copley, J., and Plouviez,S. (2011) A new species of *Rimicaris* (Crustacea: Decapoda: Caridea: Alvinocarididae) from hydrothermal vent fields on the Mid-Cayman Spreading Centre, Caribbean. *J. Mar. Biol. Assoc. United Kingdom* **92**: 1057–1072.

Nyholm,S. V and McFall-Ngai, M. (2004) The winnowing: establishing the squid–vibrio symbiosis. *Nat. Rev. Microbiol.* **2**: 632–642.

Oliver, K.M., Moran, N. a, and Hunter, M.S. (2006) Costs and benefits of a superinfection of facultative symbionts in aphids. *Proc. R. Soc. B Biol. Sci.* **273**: 1273–1280.

Oliver, K.M., Moran, N. a, and Hunter, M.S. (2005) Variation in resistance to parasitism in aphids is due to symbionts not host genotype. *Proc. Natl. Acad. Sci.* **102**: 12795–12800.

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

Perner, M., Hansen, M., Seifert, R., Strauss, H., Koschinsky, A., and Petersen,S. (2013) Linking geology, fluid chemistry, and microbial activity of basalt- and ultramafic-hosted deep-sea hydrothermal vent environments. *Geobiology* **11**: 340–355.

Perner, M., Hentscher, M., Rychlik, N., Seifert, R., Strauss, H., and Bach, W. (2011) Driving forces behind the biotope structures in two low-temperature hydrothermal venting sites on the southern Mid-Atlantic Ridge. *Environ. Microbiol. Rep.* **3**: 727–737.

Petersen, J.M. and Dubilier, N. (2009) Methanotrophic symbioses in marine invertebrates. *Environ. Microbiol. Rep.* **1**: 319–335.

Petersen, J.M., Ramette, A., Lott, C., Cambon-Bonavita, M.-A., Zbinden, M., and Dubilier, N. (2010) Dual symbiosis of the vent shrimp *Rimicaris exoculata* with filamentous gamma- and Epsilonproteobacteria at four Mid-Atlantic Ridge hydrothermal vent fields. *Environ. Microbiol.* **12**: 2204–2218.

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

Petersen, S., Kuhn, K., Kuhn, T., Augustin, N., Hékinian, R., Franz, L., and Borowski, C. (2009) The geological setting of the ultramafic-hosted

Logatchev hydrothermal field (14°45′N, Mid-Atlantic Ridge) and its influence on massive sulfide formation. *Lithos* **112**: 40–56.

Pindell, J. L. *(*1985*),* Alleghenian reconstruction and subsequent evolution of the Gulf of Mexico, Bahamas, and Proto-Caribbean*, Tectonics,* **4***(*1*),* 1–39*.*

Plouviez, S., Jacobson, A., Wu, M., and Van Dover, C.L. (2015) Characterization of vent fauna at the Mid-Cayman Spreading Center. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **97**: 124–133.

Ponsard, J., Cambon-Bonavita, M.-A., Zbinden, M., Lepoint, G., Joassin, A., Corbari, L., *et al*. (2013) Inorganic carbon fixation by chemosynthetic ectosymbionts and nutritional transfers to the hydrothermal vent host-shrimp *Rimicaris exoculata*. *ISME J.* **7**: 96–109.

Raggi, L., Schubotz, F., Hinrichs, K.-U., Dubilier, N., and Petersen, J.M. (2013) Bacterial symbionts of *Bathymodiolus* mussels and *Escarpia* tubeworms from Chapopote, an asphalt seep in the southern Gulf of Mexico. *Environ. Microbiol.* **15**: 1969–1987.

Ragsdale, S.W. (2003) Pyruvate ferredoxin oxidoreductase and its radical intermediate. *Chem. Rev.* **103**: 2333–2346.

Ramirez-Lloodra, E., Shank, T.M., and German, C.R. (2007) Biodiversity and Biogeography of hydrothermal Vent Species. *Oceanography* **20**: 30–41.

Rau, G.H. (1981) Hydrothermal Vent Clam and Tube Worm 13C/12C: Further Evidence of Nonphotosynthetic Food Sources. *Science (80-.).* **213**: 338–340.

Reeves, E.P., Yoshinaga, M.Y., Pjevac, P., Goldenstein, N.I., Peplies, J., Meyerdierks, A., *et al*. (2014) Microbial lipids reveal carbon assimilation patterns on hydrothermal sulfide chimneys. *Environ. Microbiol.* **16**: 3515–3532.

Reveillaud, J., Reddington, E., McDermott, J., Algar, C., Meyer, J.L., Sylva, S., *et al*. (2015) Subseafloor microbial communities in hydrogen-rich vent fluids from hydrothermal systems along the Mid-Cayman Rise. *Environ. Microbiol.*

Rex, M.A. and Etter, R.J. (2010) Deep-sea Biodiversity: Pattern and Scale Harvard University Press.

Rillig, M.C. and Mummey, D.L. (2006) Mycorrhizas and soil structure. *New Phytol.* **171**: 41–53.

Robidart, J.C., Roque, A., Song, P., and Girguis, P.R. (2011) Linking

Hydrothermal Geochemistry to Organismal Physiology: Physiological Versatility in *Riftia pachyptila* from Sedimented and Basalt-hosted Vents. *PLoS One* **6**: e21692.

Ronaghi, M. (1998) DNA SEQUENCING:A Sequencing Method Based on Real-Time Pyrophosphate. *Science (80-.).* **281**: 363–365.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., and Nyrén, P. (1996) Real-Time DNA Sequencing Using Detection of Pyrophosphate Release. *Anal. Biochem.* **242**: 84–89.

Rosencrantz, E., Ross, M.I., and Sclater, J.G. (1988) Age and spreading history of the Cayman Trough as determined from depth, heat flow, and magnetic anomalies. *J. Geophys. Res.* **93**: 2141.

Sandwell, D.T. and Smith, W.H.F. (2009) Global marine gravity from retracked Geosat and ERS-1 altimetry: Ridge segmentation versus spreading rate. *J. Geophys. Res.* **114**: B01411.

Sanger, F., Nicklen, S., and Coulson, a *R.* (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.*S. *A.* **74**: 5463–7.

Sayavedra, L., Kleiner, M., Ponnudurai, R., Wetzel, S., Pelletier, E., Barbe, V., et al. (2015) Abundant toxin-related genes in the genomes of beneficial symbionts from deep-sea hydrothermal vent mussels. *eLife* **4**: e07966.

Schirmer, M., D'Amore, R., Ijaz, U.Z., Hall, N., and Quince, C. (2016) Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics* **17**: 125.

Schmidt, K., Koschinsky, A., Garbe-Schönberg, D., de Carvalho, L.M., and Seifert, *R.* (2007) Geochemistry of hydrothermal fluids from the ultramafic-hosted Logatchev hydrothermal field, 15°N on the Mid-Atlantic Ridge: Temporal and spatial investigation. *Chem. Geol.* **242**: 1–21.

Sibuet, J.-C., Deffontaines, B., Hsu, S.-K., Thareau, N., Le Formal, J.-P., and Liu, C.-S. (1998) Okinawa trough backarc basin: Early tectonic and magmatic evolution. *J. Geophys. Res. Solid Earth* **103**: 30245–30267.

Sievert, S.M., Hügler, M., Taylor, C.D., and Wirsen, C.O. (2008) Sulfur oxidation at deep-sea hydrothermal vents. *Microb. Sulfur Metab.* 238–258.

Singh, Y., Ahmad, J., Musarrat, J., Ehtesham, N.Z., and Hasnain, S.E. (2013) Emerging importance of holobionts in evolution and in probiotics. *Gut Pathog.* **5**: 12.

Smith, D.C. (1979) From extracellular to intracellular: the establishment of a

symbiosis. *Proc. R. Soc. Lond. B. Biol. Sci.* **204**: 115–30.

Smith, S.E. and Smith, F.A. (2012) Fresh perspectives on the roles of arbuscular mycorrhizal fungi in plant nutrition and growth. *Mycologia* **104**: 1–13.

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., *et al*. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere." *Proc. Natl. Acad. Sci.* **103**: 12115–12120.

Staden, *R.* (1979) A strategy of DNA sequencing employing computer programs Nucleic. *Nucleic Acid Res.* **6**: 2601–2610.

Stern, R.J., Lin, P.-N., Morris, J.D., Jackson, M.C., Fryer, P., Bloomer, S.H., and Ito, E. (1990) Enriched back-arc basin basalts from the northern Mariana Trough: implications for the magmatic evolution of back-arc basins. *Earth Planet. Sci. Lett.* **100**: 210–225.

Streit, K., Bennett, S.A., Van Dover, C.L., and Coleman, M. (2015) Sources of organic carbon for *Rimicaris hybisae*: Tracing individual fatty acids at two hydrothermal vent fields in the Mid-Cayman rise. *Deep. Res. Part I Oceanogr. Res. Pap.* **100**: 13–20.

Sutton, T.T., Porteiro, F.M., Heino, M., Byrkjedal, I., Langhelle, G., Anderson, C.I.H., *et al*. (2008) Vertical structure, biomass and topographic association of deep-pelagic fishes in relation to a mid-ocean ridge system. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **55**: 161–184.

Szathmáry, E. and Smith, J.M. (1995) The major evolutionary transitions. *Nature* **374**: 227–232.

Tabita, F.R., Hanson, T.E., Li, H., Satagopan, S., Singh, J., and Chan,S. (2007) Function, Structure, and Evolution of the RuBisCO-Like Proteins and Their RuBisCO Homologs. **71**: 576–599.

Tavormina, P.L., Hatzenpichler, R., McGlynn, S., Chadwick, G., Dawson, K.S., Connon, S.A., and Orphan, V.J. (2015) *Methyloprofundus sedimenti* gen. nov., sp. nov., an obligate methanotroph from ocean sediment belonging to the "deep sea-1" clade of marine methanotrophs. *Int. J. Syst. Evol. Microbiol.* **65**: 251–259.

Teixeira, S., Serrão, E.A., and Arnaud-Haond,S. (2012) Panmixia in a Fragmented and Unstable Environment: The Hydrothermal Shrimp *Rimicaris exoculata* Disperses Extensively along the Mid-Atlantic Ridge. *PLoS One* **7**: e38521.

Thubaut, J., Puillandre, N., Faure, B., Cruaud, C., and Samadi,S. (2013) The contrasted evolutionary fates of deep-sea chemosynthetic mussels

(Bivalvia, Bathymodiolinae). *Ecol. Evol.* **3**: 4748–4766.

Tivey, M. (2007) Generation of Seafloor Hydrothermal Vent Fluids and Associated Mineral Deposits. *Oceanography* **20**: 50–65.

Toft, C. and Andersson, S.G.E. (2010) Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat. Rev. Genet.* **11**: 465–475.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., and Gordon, J.I. (2007) The Human Microbiome Project. *Nature* **449**: 804–810.

Vrijenhoek, R.C. (2010) Genetic diversity and connectivity of deep-sea hydrothermal vent metapopulations. *Mol. Ecol.* **19**: 4391–4411.

Wallace, B.D. and Redinbo, M.R. (2013) The human microbiome is a source of therapeutic drug targets. *Curr. Opin. Chem. Biol.* **17**: 379–384.

Wang, F.P., Zhou, H.Y., Meng, J., Peng, X.T., Jiang, L.J., Sun, P., *et al*. (2009) GeoChip-based analysis of metabolic diversity of microbial communities at the Juan de Fuca Ridge hydrothermal vent. *Proc. Natl. Acad. Sci. U.*S. *A.* **106**: 4840–4845.

Wang, S., Xiao, X., Jiang, L., Peng, X., Zhou, H., Meng, J., and Wang, F. (2009) Diversity and Abundance of Ammonia-Oxidizing Archaea in Hydrothermal Vent Chimneys of the Juan de Fuca Ridge. *Appl. Environ. Microbiol.* **75**: 4216–4220.

Watabe, H. and Hashimoto, J. (2002 A new species of the genus *Rimicaris* (Alvinocarididae: Caridea: Decapoda) from the active hydrothermal vent field, “Kairei Field,” on the Central Indian Ridge, the Indian Ocean. *Zoolog. Sci.* **19**: 1167–1174.

Wolff, T. (2005) Composition and endemism of the deep-sea hydrothermal vent fauna. *Cah. Biol. Mar.* **46**: 97–104.

Won, Y., Jones, W.J., and Vrijenhoek, R.C. (2008) Absence of Cospeciation Between Deep-Sea Mytilids and Their Thiotrophic Endosymbionts. *J. Shellfish Res.* **27**: 129–138.

Youle, M., Knowlton, N., Rohwer, F., Gordon, J., and Relman, D. a (2013) Superorganisms and Holobionts. *Microbe Mag.* **8**: 152–153.

Zarzycki, J., Brecht, V., Müller, M., and Fuchs, G. (2009) Identifying the missing steps of the autotrophic 3-hydroxypropionate CO_2 fixation cycle in *Chloroflexus aurantiacus*. *Proc. Natl. Acad. Sci. U.*S. *A.* **106**: 21317–22.

Zielinski, F.U., Pernthaler, A., Duperron, S., Raggi, L., Giere, O., Borowski,

C., and Dubilier, N. (2009) Widespread occurrence of an intranuclear bacterial parasite in vent and seep bathymodiolin mussels. *Environ. Microbiol.* **11**: 1150–1167.

Zilber-Rosenberg, I. and Rosenberg, E. (2008) Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol. Rev.* **32**: 723–735.

Zimmermann, J., Wentrup, C., Sadowski, M., Blazejak, A., Gruber-Vodicka, H., Kleiner, M., *et al*. (2016) Closely coupled evolutionary history of ecto- and endosymbionts from two distantly-related animal phyla. *Mol. Ecol.* n/a–n/a.

Zug, *R.* and Hammerstein, P. (2015) Bad guys turned nice? A critical assessment of *Wolbachia* mutualisms in arthropod hosts. *Biol. Rev.* **90**: 89–111

Chapter 2 **A specific and widespread association between deep-sea Bathymodiolus mussels and a novel family of Epsilonproteobacteria**

Authors: Adrien Assié[1], Christian Borowski[1], Karina van der Heijden[1], Luciana Raggi[2], Benedikt Geier[1], Nikolaus Leisch[1], Mario P. Schimak[1], Nicole Dubilier[1,4,5], Jillian M Petersen[1,3]

Published in "Environmental Microbiology Reports"

A. Assié, C. Borowski, K. van der Heijden, L. Raggi, N. Leisch, N. Dubilier, J. M. Petersen. *Environmental Microbiology Reports* (2016) A specific and widespread association between deep-sea *Bathymodiolus* mussels and a novel family of Epsilonproteobacteria. doi:10.1111/1758-2229.12442

2.1 Manuscript

environmental microbiology reports

Environmental Microbiology Reports (2016) **8**(5), 805–813 doi:10.1111/1758-2229.12442

A specific and widespread association between deep-sea *Bathymodiolus* mussels and a novel family of Epsilonproteobacteria

Adrien Assié,[1] Christian Borowski,[1] Karina van der Heijden,[1] Luciana Raggi,[2] Benedikt Geier,[1] Nikolaus Leisch,[1] Mario P Schimak,[1] Nicole Dubilier[1,3] and Jillian M Petersen[1,4]*

[1]*Max Planck Institute for Marine Microbiology, Celsiusstr. 1, Bremen 28359, Germany.*
[2]*CIGoM, Instituto de Biotecnologia, UNAM, Av. Universidad 2001, C.P.62210, Cuernavaca, Morelos, Mexico.*
[3]*MARUM, University of Bremen, Germany.*
[4]*Division of Microbial Ecology, Department of Microbiology and Ecosystem Science, University of Vienna, Althanstr. 14, Vienna 1090, Austria*

Summary

***Bathymodiolus* mussels dominate animal communities at many hydrothermal vents and cold seeps. Essential to the mussels' ecological and evolutionary success is their association with symbiotic methane- and sulfur-oxidizing gammaproteobacteria, which provide them with nutrition. In addition to these well-known gammaproteobacterial endosymbionts, we found epsilonproteobacterial sequences in metatranscriptomes, metagenomes and 16S rRNA clone libraries as well as by polymerase chain reaction screening of *Bathymodiolus* species sampled from vents and seeps around the world. These epsilonproteobacterial sequences were closely related, indicating that the association is highly specific. The *Bathymodiolus*-associated epsilonproteobacterial 16S rRNA sequences were at most 87.6% identical to the closest cultured relative, and 91.2% identical to the closest sequences in public databases. This clade therefore represents a novel family within the Epsilonproteobacteria. Fluorescence *in situ* hybridization and transmission electron microscopy showed that the bacteria are filamentous epibionts associated with the gill epithelia in two *Bathymodiolus* species. In animals that host highly specific symbioses with one or a few types of endosymbionts, other less-abundant members of the microbiota can be easily overlooked. Our work highlights how widespread and specific associations with less-abundant microbes can be. Possibly, these microbes play an important role in the survival and health of their animal hosts.**

Received 16 February, 2016; accepted 13 July, 2016. *For correspondence. E-mail jmpeters@mpi-bremen.de or petersen@microbial-ecology.net; Tel. +43 1 427776606.
[Correction added on 16 August 2016, after first online publication: Affiliation footnotes for Nicole Dubilier and Jillian M Petersen were corrected].

Introduction

Deep-sea hydrothermal vents and cold seeps sustain thriving ecosystems through the release of fluids rich in hydrogen sulfide, hydrogen and methane, which fuel chemosynthetic primary production. Several invertebrate species have adapted to these habitats by establishing relationships with ecto- or endosymbiotic chemosynthetic bacteria that can use the energy sources provided by the vents and seeps (reviewed by Dubilier *et al.*, 2008). Prominent examples include mussels of the mytilid subfamily Bathymodiolinae that are endemic to reduced environments such as hydrothermal vents, cold seeps, and wood and whale falls (Duperron *et al.*, 2008; 2009). The symbiotic chemosynthetic bacteria are associated as ecto- or endosymbionts with the epithelia of the mussel's gills and provide nutrition to their hosts (DeChaine and Cavanaugh, 2006; Nelson *et al.*, 1995; Petersen and Dubilier, 2009; Petersen *et al.*, 2011).

The Bathymodiolinae subfamily is polyphyletic and consists of multiple poorly resolved clades. Among these, two main groups are recognized: (i) the *Bathymodiolus thermophilus* clade that consists of true *Bathymodiolus* species, and (ii) the *'Bathymodiolus' childressi* clade that differs from *Bathymodiolus* sensu stricto and is sometimes referred to as the genus *Gigantidas* (Jones *et al.*, 2006; Thubaut *et al.*, 2013). Species from these two groups typically host one or two gammaproteobacterial endosymbiotic phylotypes. For example, *'Bathymodiolus' childressi*, which colonizes cold seeps in the Gulf of Mexico, hosts methane-oxidizing endosymbionts (Duperron *et al.*, 2008), whereas *B. azoricus*, which colonizes hydrothermal

vents along the north Mid-Atlantic Ridge, hosts sulfur-oxidizing and methane-oxidizing endosymbionts (Duperron *et al*., 2006). Most research has focused on these highly abundant gammaproteobacterial endosymbionts (e.g., Dubilier *et al*., 2008; Duperron *et al*., 2009; Rodrigues *et al*., 2013). However, additional bacterial phylotypes can also associate with bathymodiolin mussels. For example, *Bathymodiolus heckerae* can harbor up to five endosymbiotic phylotypes and numerous bathymodiolin species are infected with intranuclear bacteria (Duperron *et al*., 2007; Raggi *et al*., 2013; Zielinski *et al*., 2009).

Epsilonproteobacteria are abundant at hydrothermal vents and they often associate as ecto- or endosymbionts with diverse invertebrate animals including gastropods, shrimp, crabs and polychaetes, but they have never been found in association with bathymodiolin mussels (Campbell *et al*., 2006; Dubilier *et al*., 2008; Goffredi, 2010; Petersen *et al*., 2010). This study was inspired by our recent discovery of two closely related epsilonproteobacterial 16S rRNA sequences in clone libraries from two distantly related bathymodiolin species: '*B.*' *childressi* from the Gulf of Mexico, and *Bathymodiolus* sp. from the Pakistan Margin in the Arabian Sea, referred to here as '*B.*' sp (Makran). The occurrence of closely related epsilonproteobacteria in two distinct host species from two different geographic regions was surprising. Our aim was therefore to explore the distribution and specificity of the association between bathymodiolin mussels and epsilonproteobacteria by polymerase chain reaction (PCR) screening of a large collection of samples and screening of metagenomes and metatranscriptomes. Furthermore, we used fluorescence in situ hybridization (FISH) and electron microscopy to identify and localize the epsilonproteobacteria in mussel gill sections.

Results and discussion

The association between epsilonproteobacteria and bathymodiolin mussels is widespread and highly specific

Clone libraries of the bacterial 16S rRNA genes amplified from *'B.' childressi* and *'B.'* sp (Makran) gill DNA contained epsilonproteobacterial sequences in addition to the known gammaproteobacterial endosymbionts. In *'B.' childressi*, 12 out 46 clones screened were related to epsilonproteobacteria, and in *'B.'* sp (Makran), 3 out 86 clones were related to epsilonproteobacteria. All 15 epsilonproteobacterial clone sequences (1201 base pairs) shared 98.3% identity. The closest relative in the NCBI database was an unpublished sequence named 'Uncultured *Bathymodiolus platifrons* gill symbiont' from a mussel that was collected from seeps in the Okinawa Trough, East China Sea (99.1% identity, accession number AB250697). The next closest relative in the database was an uncultured and unclassified epsilonproteobacterium associated with the bivalve *Thyasira flexuosa* (91.2% sequence identity – FN600361). The closest cultured relative was *Sulfurovum lithotrophicum*, a mesophilic chemolithoautotroph isolated from hydrothermal vent sediments (87.6% sequence identity; Inagaki *et al*., 2004).

The closely related bathymodiolin epsilonproteobacterial sequences were found in two distantly related host species sampled from sites thousands of kilometers apart, suggesting that these hosts might regularly associate with this specific group of epsilonproteobacteria. To test this hypothesis, we screened DNA extracted from seven different bathymodiolin species from vent and seep sites around the world (Table S1) using specific forward and reverse 16S rRNA PCR primers. We also screened metagenomes and metatranscriptomes for epsilonproteobacterial 16S rRNA sequences using the PhyloFlash script (https://github.com/HRGV/phyloFlash). In four of these species, epsilonproteobacteria were found in all individuals screened: *'B.' childressi* ($n = 32$) collected from six sites across the Gulf of Mexico; *B. azoricus* ($n = 11$) from three sampling sites along the North Mid-Atlantic Ridge (NMAR); *'B.' manusensis* ($n = 4$) from two sampling sites in Lau Basin; *'B.' mauritanicus* ($n = 3$) from the Barbados Accretionary Prism (Tables 1 and S1). A detailed analysis of the metagenomes and metatranscriptomes will be published elsewhere.

The epsilonproteobacterial 16S rRNA sequences recovered by PCR and next-generation sequencing were remarkably similar: sequences from different host species ranged from 97.2% (26 substitutions) to almost identical (99.9%; 1 substitution). In addition, all of the sequences recovered from a single species were identical or deviated by a single substitution. If present, substitutions were always located at conserved positions in the 16S rRNA gene, suggesting that they were likely the result of PCR or sequencing errors. Associations between bathymodiolin mussels and this group of epsilonproteobacteria are therefore host species-specific.

We did not detect epsilonproteobacteria in all species screened. PCRs were negative in *B. brooksi* ($n = 1$) from the Gulf of Mexico, *B. septemdierum* ($n = 3$) from Lau Basin (previously described as *B. brevior*, taxonomy revised in Breusing *et al*., 2015) and *B. thermophilus* ($n = 4$) from the East Pacific Rise (Tables 1 and S1). This could indicate that these host species do not harbor epsilonproteobacteria or that the primers we used did not match because they were designed based on epsilonproteobacterial 16S rRNA sequences obtained from two other host species. However, screening of metagenome libraries from *B. brooksi* ($n = 2$) and *B.*

Table 1. Summary of samples screened for the presence of epsilonproteobacteria.

Host species	Sampling location	Epsilon-proteobacteria	Number of host individuals screened			Number of sampling sites	Imaging	Associated gammaproteobacterial endosymbionts
			PCRs	Metagenome	Transcriptome			
'B.' childressi	Gulf of Mexico	Yes	20	6	3	6	FISH, SEM and TEM	MOX
B. azoricus	NMAR	Yes	8	2	1	3	FISH	SOX and MOX
'B.' mauritanicus	Barbados Accretionary Prism	Yes	3	–	–	1	–	MOX
'B.' manusensis	Manus Basin	Yes	4	–	–	2	–	SOX
***'B.'* sp**	Makran seeps	Yes	2	–	–	1	–	MOX
***B.* sp**	SMAR	Yes	–	6	–	1	–	SOX and MOX
'B.' platifrons	Okinawa Trough	Yes	1 sequence present in public databases			–	–	SOX
B. brooksi	Gulf of Mexico	N. D.	1	2	–	3	–	SOX and MOX
B. puteoserpentis	NMAR	N. D	–	3	–	2	–	SOX and MOX
B.septemdierum	Lau Basin	N. D	4	–	–	2	–	SOX
B. thermophilus	EPR	N. D	4	–	–	2	–	SOX

For more information see Table S1. NMAR, North Mid-Atlantic Ridge; SMAR, South Mid-Atlantic Ridge; EPR, East Pacific Rise; N.D., Not detected; SOX, Sulfur-oxidizing endosymbionts; MOX, Methane-oxidizing endosymbionts.

thermophilus ($n = 3$) for 16S rRNA sequences from the epsilonproteobacteria were also negative, mirroring the negative PCR screening results for two species. Metagenome libraries from an additional Mid-Atlantic Ridge species, *B. puteoserpentis* ($n = 4$), were also negative. It is therefore possible that not all bathymodiolin species host epsilonproteobacteria.

There was no clear factor explaining the presence of epsilonproteobacteria in some host species but not in others. The host species in which epsilonproteobacteria were detected did not group together based on their cytochrome oxidase I (COI) phylogeny (Fig. S1A). Neither were they all sampled from close geographic locations, or from similar habitats (Fig. S1B). These associations therefore do not appear to be restricted to a particular clade of mussels, a particular geographic region, or even to mussels occurring only at vents or only at seeps, but are widely distributed across diverse host species in disparate habitats. This distribution pattern could indicate that associations within this clade of epsilonproteobacteria are an ancient feature found in the ancestor of all bathymodiolins, but were subsequently lost by a number of different species. Alternatively, distantly related bathymodiolin mussels may have formed associations with this clade of epsilonproteobacteria multiple times in multiple locations independently. If this is the case, then these epsilonproteobacteria must have been uniquely 'primed' for associating with marine invertebrates, in particular with bathymodiolin mussels.

Novel epsilonproteobacterial family

Phylogenetic analyses placed all of our epsilonproteobacterial 16S rRNA sequences in a well-supported monophyletic clade that only contained *Bathymodiolus*-associated sequences. This clade is a sister group to uncultured and unclassified sequences from deep-sea hydrothermal vent invertebrates such as bivalves (FN600361), gastropods (FM994656) and corals (DQ917867 and GU117971) (Fig. 1). The unpublished sequence from *'B.' platifrons* also fell into the *Bathymodiolus*-associated epsilonproteobacterial group, indicating that this host species may also associate with epsilonproteobacteria. The full 16S rRNA sequence recovered from a '*B.' childressi* metagenome was 87.6% similar to the closest cultured relative *Sulfurovum lithotrophicum* (NR_024802), which belongs to the Thiovulgaceae family. According to the guidelines of Yarza *et al*. (2014) the *Bathymodiolus*-associated clade represents a new family within the Epsilonproteobacteria, which possibly also includes unclassified sequences from other host-associated epsilonproteobacteria (Fig. 1).

Epsilonproteobacteria are epibionts on mussel gills

We did FISH on gill filament sections to identify and localize the epsilonproteobacteria in two host species, *B. azoricus* and '*B.' childressi*. We designed two specific probes: BCE141 and BCE1422 to target each end of the epsilonproteobacterial 16S rRNA sequences

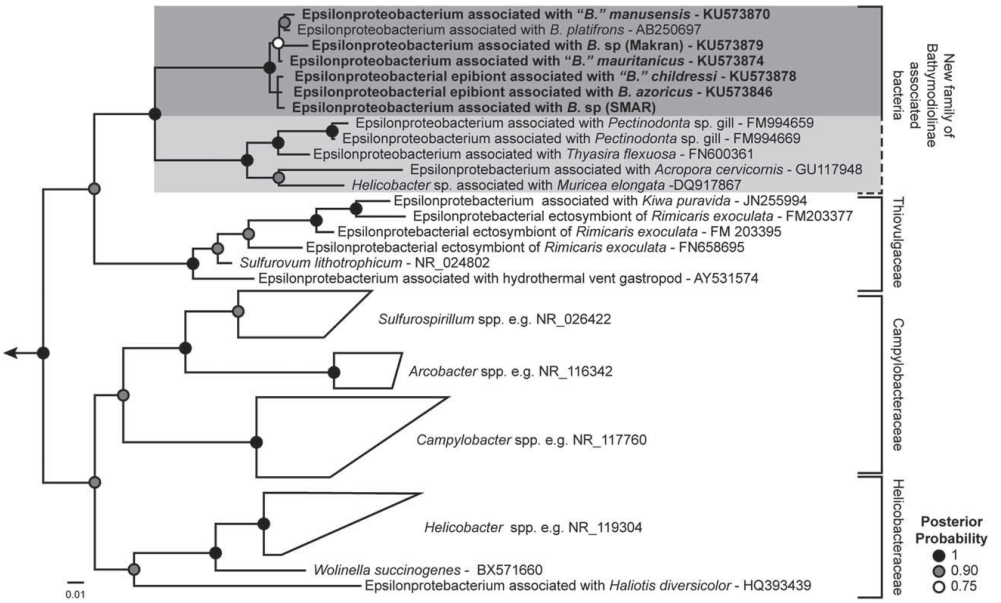


Fig. 1. Bayesian inference tree of 16S rRNA gene sequences under a General Time Reversible model with Gamma-distributed rates of evolution and a proportion of invariant sites. Analyses were performed with 10 million generations using four parallel Monte Carlo Markov chains. Sample trees were taken every 5000 generations. *Thiomicrospira halophila* (NR_043780) was used as root for the tree. Sequence names in bold are from this study. Square black brackets show described epsilonproteobacterial families. The colored boxes show sequences that form a new family of epsilonproteobacteria: the dark gray box shows the bathymodiolin sequences from this study, the light gray box shows published sequences which form a monophyletic group with the bathymodiolin sequences.

from bathymodiolins. Both probes showed at least three strong mismatches with all available sequences in the SILVA database (release 123, Quast *et al.*, 2013). FISH on '*B.' childressi* showed signals close to the gill epithelium (BCE141 signals displayed in Fig. 2, data not shown for BCE1422). These long, thin filamentous epibionts were concentrated toward the ciliated edges of the gill filaments. In contrast, on *B. azoricus*, the probes hybridized with thick filaments on the gill epithelium that were distributed evenly across the gill surface (Fig. 2 A and B). 3D rendering of whole mount FISH analyses of '*B.' childressi* gill filaments also identified thick epsilonproteobacterial filaments in patches located on the epithelial cells (Fig. 2C). These results show that at least on the two species we investigated with FISH, the epsilonproteobacteria are filamentous epibionts.

Filamentous bacteria could also be seen by scanning and transmission electron microscopy (respectively SEM and TEM) of *'B.' childressi* gill sections. SEM revealed a dense coating of filaments occurring on the epithelial cells (Fig. 4). This correlated with the FISH volume reconstruction, however, it was difficult to distinguish between bacterial filaments and host cilia in SEM. TEM clearly showed filamentous bacteria interspersed with host cilia and mucus (Fig. 3). Previous studies showed that the gill epithelium of bathymodiolin mussels is coated with a mucus layer (Bettencourt *et al.*, 2011; Fiala-Medioni *et al.*, 2002). Settling in mucus is a common strategy for some host-associated epsilonproteobacteria. For example, members of the Helicobacteraceae specifically colonize the mucus layer lining the mammalian gastrointestinal tract (for reviews see Fox and Lee, 1997; Oxley and McKay, 2004; Solnick and Schauer, 2001).

Potential function of the epibionts

The novel *Bathymodiolus*-associated epsilonproteobacterial family is a sister clade to the Thiovulgaceae family, which contains sulfur oxidizing epibionts of various invertebrate species such as the deep-sea shrimp *Rimicaris excoculata* (Jan *et al.*, 2014; Petersen *et al.*, 2010), the crab *Kiwa hirsuta* (Goffredi *et al.*, 2008) and polycheates such as *Alvinella pompejana* (Campbell *et al.*, 2001). These bacteria are suggested to play a role in the nutrition of their hosts (Desbruyères *et al.*, 1998; Ponsard

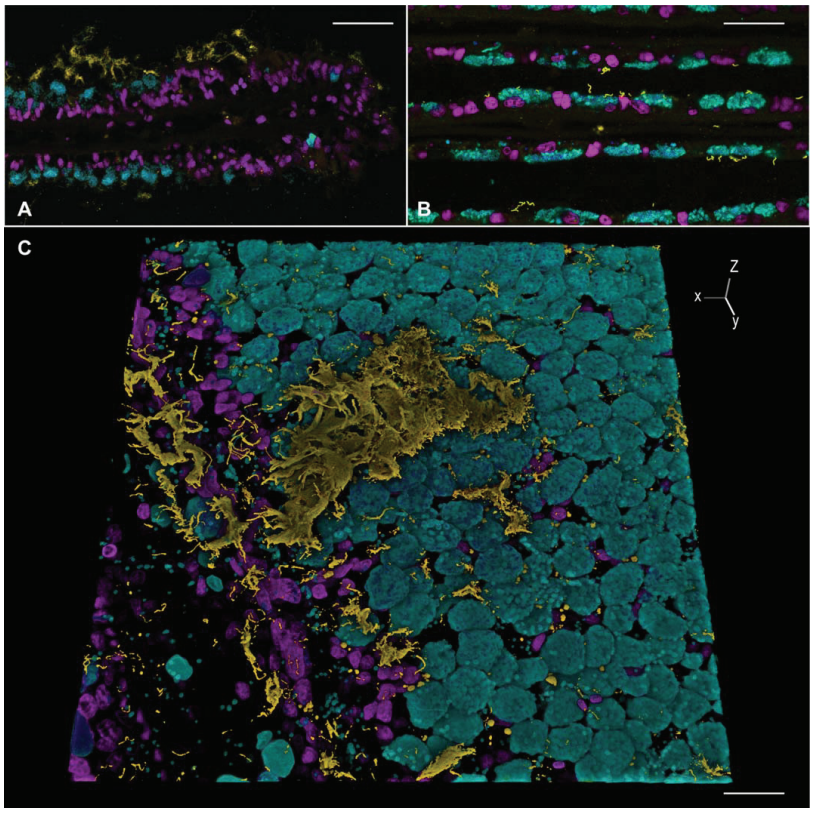


Fig. 2. Fluorescence in situ hybridization of epsilonproteobacterial epibionts in bathymodiolin gill tissues (yellow – specific probe for epsilonproteobacterial epibionts BCE141, Cy3) and gammaproteobacterial endosymbionts (teal – general bacterial probe EUB338, 6FAM) and DNA (purple – DAPI). A. Section of a '*B.*' *childressi* gill filament, scale 50 μm. B. Section of *B. azoricus* gill filaments, scale 30 μm. These hosts have far fewer epsilonproteobacteria than '*B.*' *childressi* mussels. C. 3D rendering of a z-stack series of a '*B.*' *childressi* gill filament, scale 20 μm.

et al., 2013; Thurber *et al.*, 2011). Considering that their closest cultivated relatives are sulfuroxidizers of the Thiovulgaceae family, the novel epsilonproteobacterial epibionts associated with bathymodiolin mussels may also be sulfur oxidizers. If this is the case, their epibiotic location on bathymodiolin gills would be advantageous for accessing reduced sulfur compounds from fluids that are pumped across the gills of the mussels. The role these epibionts play in the symbiosis is currently unclear. Epibionts have been proposed to transfer nutrients to their host by diffusion of small molecules, or by endocytosis and intracellular digestion of the bacteria (e.g. Ponsard *et al.*, 2013 and Zbinden *et al.*, 2015). It is also possible that the epibionts provide protection against biofouling of the gills by pathogens.

Our initial observations suggest that epsilonproteobacteria may be more abundant on bathymodiolin mussels that host only methane-oxidizing symbionts. Bathymodiolins hosting both sulfur- and methane-oxidizing endosymbionts seem to host fewer or no epsilonproteobacterial epibionts, and we have not yet found epsilonproteobacteria on mussels that only have sulfur-oxidizing symbionts. One possible explanation for the observed differences in epibiont abundance could be active prevention of their growth by the sulfur-oxidizing endosymbionts. Sayavedra *et al.* (2015) recently showed that the sulfur-oxidizing endosymbionts encode a vast array of toxin-like genes in their genomes, some of which may be involved in antagonistic interactions between bacteria.

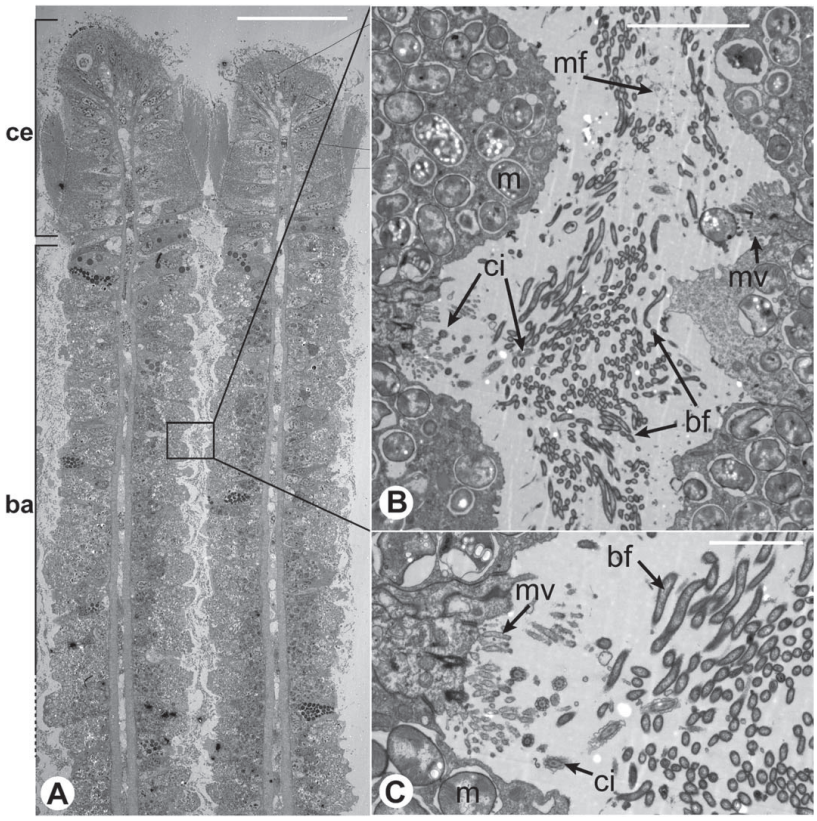


Fig. 3. A. Transmission electron micrograph of '*B.*' *childressi* gill filaments: (ba) indicates the region harboring bacteriocytes and (ce) the ciliated edge free of gammaproteobacterial endosymbionts (Scale: 50μm). B and C. Close-up of the space between gill filaments. Filamentous bacteria (bf) are located among mucus filaments (mf) between the cells containing methanotrophic gammaproteobacterial endosymbionts (m). Intercalary cells (symbiont-free cells) can be identified by the presence of microvilli (mv) and cilia (ci) (Scale B: 5 μm; C: 2 μm).

Another intriguing hypothesis would be that a sulfur-oxidizing epibiont could replace the sulfur-oxidizing gammaproteobacterial endosymbiont, which was most likely lost by some bathymodiolins during their evolutionary history (Lorion *et al.*, 2013). The loss of the sulfur-oxidizing endosymbiont may open up a new ecological niche, which sulfur-oxidizing epsilonproteobacteria could take advantage of. These two explanations, the replacement of an ecological niche and the inhibition of growth are not mutually exclusive, and both might play a role.

Evolutionary importance of low-abundance symbionts

Despite their relatively low abundance on some host species, these epsilonproteobacteria could have a significant evolutionary importance. First, the epsilonproteobacterial epibionts regularly associate with bathymodiolin mussels, and given the right conditions, they might be able to displace the dominant gammaproteobacterial symbionts. It has been shown in corals, and more extensively in insects, that during stress or following a shift in the host's source of nutrition, rare opportunistic bacteria can replace the primary symbiont population because they are presumably better adapted to the new environmental conditions (Moran and Yun, 2015; Pettay *et al.*, 2015). Second, this association could be a nascent endosymbiosis. An epibiotic life stage has been suggested to be an intermediate stage of endosymbiosis evolution (Lorion *et al.*, 2013; Miyazaki *et al.*, 2010; Smith, 1979). In fact, a few bathymodiolin mussels host their sulfur-oxidizing gammaproteobacterial symbionts outside of host cells (Fujiwara *et al.*, 2010; Lorion *et al.*, 2009). Similar associations

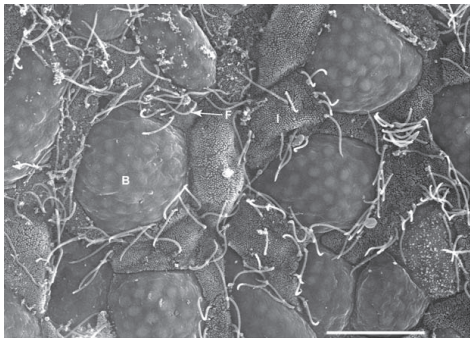


Fig. 4. Scanning electron micrograph of a '*B.*' *childressi* gill filament. The bacteriocytes (B) are smooth and the intercalary cells (I) are covered with microvilli. Numerous filamentous (F) structures can be seen on the surface of the cells (Scale: 10 μm).

could be evolving with the epsilonproteobacterial epibionts, especially in species such as *'B.' childressi* where gammaproteobacterial sulfur-oxidizing endosymbionts have been lost.

Acknowledgements

We thank Silke Wetzel for excellent technical assistance, Lizbeth Sayavedra, Juliane Wippler and Harald Gruber-Vodicka for advice on bioinformatics and Sébastien Duperron for scientific input and ideas. We also thank the two anonymous reviewers who helped us to improve the manuscript. This project would not have been possible without the dedicated captains and crews of the research vessels involved in the sampling. For the sampling effort we also thank Silke Wetzel, Lizbeth Sayavedra, Dennis Fink, Heiko Sahling, Stéphane Hourdez, Chuck Fisher, Stephanie Markert, Ruby Ponnudurai, Matthew Saxton, Mandy Joye, Bernie Ball, Cindy Van Dover and the Connectivity In Western Atlantic Seep Populations: Oceanographic and Life-History Processes Underlying Genetic Structure (SeepC) project funded by NSF Ocean Sciences (NSF OCE). This work was funded by the Max Planck Society, the DFG Cluster of Excellence 'The Ocean in the Earth System' at MARUM (University of Bremen), a European Research Council Advanced Grant (BathyBiome, Grant 340535) and a Gordon and Betty Moore Foundation Marine Microbiology Initiative Investigator Award through Grant GBMF3811 to ND, and the European Union (EU) Marie Curie Actions Initial Training Network (ITN) SYMBIOMICS (contract number 264774). JMP was supported by the Vienna Science and Technology Fund (WWTF) through project VRG14-021.

References

Bettencourt, R., Costa, V., Laranjo, M., Rosa, D., Pires, L., Colaco, A., *et al.* (2011) Out of the deep sea into a land-based aquarium environment: investigating physiological adaptations in the hydrothermal vent mussel *Bathymodiolus azoricus*. *ICES J Mar Sci* **68**: 357–364.

Breusing, C., Johnson, S.B., Tunnicliffe, V., and Vrijenhoek, R.C. (2015) Population structure and connectivity in Indo-Pacific deep-sea mussels of the *Bathymodiolus septemdierum* complex. *Conserv Genet* **16**: 1415–1430.

Campbell, B.J., Jeanthon, C., Kostka, J.E., Luther, G.W., and Cary, S.C. (2001) Growth and phylogenetic properties of novel bacteria belonging to the epsilon subdivision of the proteobacteria enriched from *Alvinella pompejana* and deep-sea hydrothermal vents. *Appl Environ Microbiol* **67**: 4566–4572.

Campbell, B.J., Engel, A.S., Porter, M.L., and Takai, K. (2006) The versatile epsilon-proteobacteria: key players in sulphidic habitats. *Nat Rev Microbiol* **4**: 458–468.

DeChaine, E.G., and Cavanaugh, C.M. (2006) Symbioses of methanotrophs and deep-sea mussels (Mytilidae: Bathymodiolinae). *Prog Mol Subcell Biol* **41**: 227–249.

Desbruyères, D., Chevaldonné, P., Alayse, A.M., Jollivet, D., Lallier, F.H., Jouin-Toulmond, C., *et al.* (1998) Biology and ecology of the 'Pompeii worm' (*Alvinella pompejana* Desbruyères and Laubier), a normal dweller of an extreme deep-sea environment: a synthesis of current knowledge and recent developments. *Deep Sea Res Part II Top Stud Oceanogr* **45**: 383–422.

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

Duperron, S., Bergin, C., Zielinski, F., Blazejak, A., Pernthaler, A., McKiness, Z.P., *et al.* (2006) A dual symbiosis shared by two mussel species, *Bathymodiolus azoricus* and *Bathymodiolus puteoserpentis* (Bivalvia: Mytilidae), from hydrothermal vents along the northern Mid-Atlantic Ridge. *Environ Microbiol* **8**: 1441–1447.

Duperron, S., Sibuet, M., MacGregor, B.J., Kuypers, M.M.M., Fisher, C.R., and Dubilier, N. (2007) Diversity, relative abundance and metabolic potential of bacterial endosymbionts in three *Bathymodiolus* mussel species from cold seeps in the Gulf of Mexico. *Environ Microbiol* **9**: 1423–1438.

Duperron, S., Laurent, M.C.Z., Gaill, F., and Gros, O. (2008) Sulphur-oxidizing extracellular bacteria in the gills of Mytilidae associated with wood falls. *FEMS Microbiol Ecol* **63**: 338–349.

Duperron, S., Lorion, J., Samadi, S., Gros, O., and Gaill, F. (2009) Symbioses between deep-sea mussels (Mytilidae: Bathymodiolinae) and chemosynthetic bacteria: diversity, function and evolution. *C R Biol* **332**: 298–310.

Fiala-Medioni, A., McKiness, Z.P., Dando, P.R., Boulegue J., Mariotti, A.A.A.D., *et al.* (2002) Ultrastructural, biochemical, and immunological characterization of two populations of the mytilid mussel *Bathymodiolus azoricus* from the Mid-Atlantic Ridge: evidence for a dual symbiosis. *Mar Biol* **141**: 1035–1043.

Fox, J.G., and Lee, A. (1997) The role of *Helicobacter* species in newly recognized gastrointestinal tract diseases of animals. *Lab Anim Sci* **47**: 222–255.

Fujiwara, Y., Kawato, M., Noda, C., Kinoshita, G., Yamanaka, T., Fujita, Y., *et al.* (2010) Extracellular and mixotrophic symbiosis in the whale-fall mussel *Adipicola*

pacifica: a trend in evolution from extra- to intracellular symbiosis. *PLoS One* **5**: e11808.
Goffredi, S.K. (2010) Indigenous ectosymbiotic bacteria associated with diverse hydrothermal vent invertebrates. *Environ Microbiol Rep* **2**: 479–488.
Goffredi, S.K., Jones, W.J., Erhlich, H., Springer, A., and Vrijenhoek, R.C. (2008) Epibiotic bacteria associated with the recently discovered Yeti crab, *Kiwa hirsuta*. *Environ Microbiol* **10**: 2623–2634.
Inagaki, F., Takai, K., Nealson, K.H., and Horikoshi, K. (2004) *Sulfurovum lithotrophicum* gen. nov., sp. nov., a novel sulfur-oxidizing chemolithoautotroph within the E-Proteobacteria isolated from Okinawa Trough hydrothermal sediments. *Int J Syst Evol Microbiol* **54**: 1477–1482.
Jan, C., Petersen, J.M., Werner, J., Teeling, H., Huang, S., Glöckner, F.O., *et al*. (2014) The gill chamber epibiosis of deep-sea shrimp *Rimicaris exoculata*: an in-depth metagenomic investigation and discovery of *Zetaproteobacteria*. *Environ Microbiol* **16**: 2723–2738.
Jones, W.J., Won, Y.J., Maas, P.A.Y., Smith, P.J., Lutz, R.A., Vrijenhoek, R.C., *et al*. (2006) Evolution of habitat use by deep-sea mussels. *Mar Biol* **148**: 841–851.
Lorion, J., Duperron, S., Gros, O., Cruaud, C., and Samadi, S. (2009) Several deep-sea mussels and their associated symbionts are able to live both on wood and on whale falls. *Proc R Soc Lond B Biol Sci* **276**: 177–185.
Lorion, J., Kiel, S., Faure, B., Kawato, M., Ho, S.Y.W., Marshall, B., *et al*. (2013) Adaptive radiation of chemosymbiotic deep-sea mussels. *Proc R Soc Lond B Biol Sci* **280**: 20131243.
Miyazaki, J.I., Martins, L., de, O., Fujita, Y., Matsumoto, H., and Fujiwara, Y. (2010) Evolutionary process of deep-sea *Bathymodiolus* mussels. *PLoS One* **5**: e10363.
Moran, N.A., and Yun, Y. (2015) Experimental replacement of an obligate insect symbiont. *Proc Natl Acad Sci U S A* **112**: 2093–2096.
Nelson, D.C., Hagen, K.D., and Edwards, D.B. (1995) The gill symbiont of the hydrothermal vent mussel *Bathymodiolus thermophilus* is a psychrophilic, chemoautotrophic, sulfur bacterium. *Mar Biol* **121**: 487–495.
Oxley, A.P.A., and McKay, D.B. (2004) Fecal shedding of *Helicobacter* spp. by co-housed Australian sea lions (*Neophoca cinerea*) and Australian fur seals (*Arctocephalus pusillus doriferus*). *Vet Microbiol* **101**: 235–243.
Petersen, J.M., and Dubilier, N. (2009) Methanotrophic symbioses in marine invertebrates. *Environ Microbiol Rep* **1**: 319–335.
Petersen, J.M., Ramette, A., Lott, C., Cambon-Bonavita, M.A., Zbinden, M., and Dubilier, N. (2010) Dual symbiosis of the vent shrimp *Rimicaris exoculata* with filamentous gamma- and epsilonproteobacteria at four Mid-Atlantic Ridge hydrothermal vent fields. *Environ Microbiol* **12**: 2204–2218.
Petersen, J.M., Zielinski, F.U., Pape, T., Seifert, R., Moraru, C., Amann, R., *et al*. (2011) Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* **476**: 176–180.
Pettay, D.T., Wham, D.C., Smith, R.T., Iglesias-Prieto, R., and LaJeunesse, T.C. (2015) Microbial invasion of the Caribbean by an Indo-Pacific coral *Zooxanthella*. *Proc Natl Acad Sci U S A* **112**: 7513–7518.
Ponsard, J., Cambon-Bonavita, M.A., Zbinden, M., Lepoint, G., Joassin, A., Corbari, L., *et al*. (2013) Inorganic carbon fixation by chemosynthetic ectosymbionts and nutritional transfers to the hydrothermal vent host-shrimp *Rimicaris Exoculata*. *ISME J* **7**: 96–109.
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., *et al*. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590–D596.
Raggi, L., Schubotz, F., Hinrichs, K.U., Dubilier, N., and Petersen, J.M. (2013) Bacterial symbionts of *Bathymodiolus* mussels and *Escarpia* tubeworms from Chapopote, an asphalt seep in the southern Gulf of Mexico. *Environ Microbiol* **15**: 1969–1987.
Rodrigues, C.F., Cunha, M.R., Génio, L., and Duperron, S. (2013) A complex picture of associations between two host mussels and symbiotic bacteria in the Northeast Atlantic. *Naturwissenschaften* **100**: 21–31.
Sayavedra, L., Kleiner, M., Ponnudurai, R., Wetzel, S., Pelletier, E., Barbe, V., *et al*. (2015) Abundant toxin-related genes in the genomes of beneficial symbionts from deep-sea hydrothermal vent mussels. *eLife* **4**: e07966.
Smith, D.C. (1979) From extracellular to intracellular: the establishment of a symbiosis. *Proc R Soc Lond B Biol Sci* **204**: 115–130.
Solnick, J.V., and Schauer, D.B. (2001) Emergence of diverse *Helicobacter* species in the pathogenesis of gastric and enterohepatic diseases. *Clin Microbiol Rev* **14**: 59–97.
Thubaut, J., Puillandre, N., Faure, B., Cruaud, C., and Samadi, S. (2013) The contrasted evolutionary fates of deep-sea chemosynthetic mussels (Bivalvia, Bathymodiolinae). *Ecol Evol* **3**: 4748–4766.
Thurber, A.R., Jones, W.J., and Schnabel, K. (2011) Dancing for food in the deep sea: bacterial farming by a new species of Yeti crab. *PLoS One* **6**: 1–12.
Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.H., *et al*. (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* **12**: 635–645.
Zbinden, M., Marqué, L., Gaudron, S.M., Ravaux, J., Léger, N., and Duperron, S. (2015) Epsilonproteobacteria as gill epibionts of the hydrothermal vent gastropod *Cyathermia naticoides* (North East-Pacific Rise). *Mar Biol* **162**: 435–448.
Zielinski, F.U., Pernthaler, A., Duperron, S., Raggi, L., Giere, O., Borowski, C., and Dubilier, N. (2009) Widespread occurrence of an intranuclear bacterial parasite in vent and seep bathymodiolin mussels. *Environ Microbiol* **11**: 1150–1167.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. A. Bayesian inference tree of 16S rRNA gene sequences of epsilonproteobacterial epibionts (right) and

COI gene sequences of bathymodiolin mussels (left). B. Geographical distribution of the collected samples. There are no evident correlations between the epsilonproteobacterial phylogeny and the host phylogeny, the habitat, or the geographic location of sampling.

Fig. S2. Bayesian inference tree of all the epsilonproteobacterial 16S rRNA (1049 bp) sequences generated in this study. Analyses were performed with 10 million generations using four parallel Monte Carlo Markov chains. Sample trees were taken every 5000 generations. *Thiomicrospira halophila* (NR_043780) was used as root for the tree. This tree show the high level of identity of sequences associated to same hosts as well as different ones.

Table S1. Meta-data table of the samples used in this study.

2.2 Experimental procedures and supplementary figures and table

Sample collection

Ten bathymodiolin species were collected during numerous cruises: *Bathymodiolus azoricus, "B." childressi, "B." manuensis, "B." mauritanicus, B. brooski, B. puteoserpentis, B. thermophilus, B. brevior* and two currently unnamed species. Sampling and fixation details can be found in Digital Supplementary Table 2.1.

DNA extraction

Genomic DNA was extracted from gill tissues of the mussels according to Zhou et al., (1996) with the following modifications: An initial incubation step was performed at 37°C in 360 µl of extraction buffer (100 mM Tris-HCl [pH 8.0], 100 mM sodium EDTA [pH 8.0], 100 mM sodium phosphate [pH 8.0], 1.5 M NaCl, 1% CTAB) and 40 µl of proteinase K (10 mg/ml) overnight. The quality of the DNA was assessed with a Qubit® 2.0 Fluorometer (Invitrogen, Eugen, USA).

RNA extraction

RNA *was extracted from* the gill tissues of *three B. childressi individuals*. A fragment of the gill was dissected and incubated overnight at 37°C in 360 µl of extraction buffer (see DNA extraction) and 40 µl of proteinase K (10 mg/ml). RNA was extracted with the AllPrep DNA/RNA Micro Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The quality of the RNA was assessed with a Qubit® 2.0 Fluorometer (Invitrogen)

Host identification

The cytochrome oxidase I (COI) genes of the different species were sequenced with the PCR primers: COI1F (ATY GGN GGN TTY GGN AAY TG) and COIR (ATN GCR AAN ACN GCN CCY AT) (Matsumoto and Hayami, 2000) and a high-quality Taq DNA polymerase (error rate

2.7×10^{-5}; Eppendorf, Hamburg, Germany). The PCR was performed with the following conditionss: an initial denaturation of five minute at 5°C, 35 cycles of denaturation of one minute at 95°C, one minute annealing step at 58°C for the COIF/R primer pair and a elongation step of two minute at 72°C, and a final elongation step of 10 min at 72°C. Sequencing was performed using the BigDye Terminator v3.1 Cycle Sequencing Kit along with the Genetic Analyzer Abiprism 3130 (Applied Biosystems, Foster City, USA). Sequences have been deposited in Genbank the accession numbers are displayed in Digital Supplementary table 2.1

Clone libraries

Initial PCR was performed using the universal bacterial primers GM3F and GM4R (Muyzer et al., 1995) and a high-quality Taq DNA polymerase (Eppendorf) with the same conditions as above except the annealing temperature, which was 45°C. Amplification products were purified and ligated into the pCR™4-TOPO® Vector (Invitrogen). One Shot TOP10 competent *E. coli* cells (Invitrogen) were subsequently transformed. A total of 96 positive transformants from one *"B."* sp (Makran) individuals and 46 positive transformants from a *"B." childressi* individual were picked by blue/white screening and grown overnight in V96 MicroWell Plates (Nunc, Wiesbaden, Germany) containing 200 ml Luria–Bertani/ampicillin (100µg/ml) broth per well. The clones from both individuals were screened by PCR using the M13F/M13R primer pair (Yanisch-Perron et al., 1985). 83 clones from *"B."* sp (Makran) and 45 from *"B." childressi* had an insert with the expected size of approximately 1200 bp. The clones with these inserts were partially sequenced. PCR products were purified in MultiScreen-HV plates (Millipore, Darmstadt, Germany) using Sephadex G50 Superfine resin (Amersham Biosciences, Uppsala, Sweden) and sequenced using the BigDye Terminator v2.0 Cycle Sequencing Kit along with the Genetic Analyzer Abiprism 3100 (Applied Biosystems). The GM3F oligonucleotide (Muyzer et al., 1995) was used as sequencing primer. Sequences were analyzed using the

Geneious software version (V6) (http://www.geneious.com, Kearse et al., 2012). Sequence ends with more than 5% chance of error per base were trimmed. The trimmed sequences were compared with the NCBI nucleotide database using nucleotide BLAST (Madden, 2002) and the Silva rRNA database (Release 123) using the SINA web-aligner (Pruesse et al., 2012; Quast et al., 2013). 3 out 43 *"B."* sp (Makran) partial sequences and 12 out 45 *"B." childressi* partial sequences were related to Epsilonproteobacteria. One *"B." childressi* and two *"B."* sp (Makran) Epsilonproteobacteria-related sequences were chosen to be fully sequenced using the same protocol as above and the PCR primer GM4R.

Next-generation sequencing

Genomic libraries of *B. puteoserpentis* M64/2-2-244-9 were prepared by Genoscope (Centre National de Séquençage, Evry, France) and sequenced on an Illumina HiSeq 2000 platform. Genomic libraries from *B. azoricus* and *B. puteoserpentis* (M2-5, M3-12, M4-16) were prepared by the Max Planck Genome Center (Cologne, Germany) and sequenced on an Illumina HiSeq 2500 platform. All libraries were generated with the Illumina TruSeq DNA Sample Preparation Kit according to manufacturer recommendations and the sequencing details are summarized in Digital Supplementary Table 2.1.

Transcriptomic libraries of *"B." childressi* were generated with the Illumina TruSeq RNA Sample Preparation Kit according to manufacturer recommendations and the sequencing details are summarized in Digital Supplementary Table 2.1. All sequencing runs were performed on an Illumina HiSeq 2500 platform at the Max Planck Genome Center. Transcriptome libraries of *B.* sp (SMAR) were prepared and sequenced as described in Sayavedra *et al.*, 2015.

Screening of next-generation sequence libraries

The PhyloFlash 2.0 (https://github.com/HRGV/phyloFlash)

script was used to perform RNA small sub unit (SSU) screening and reconstruction from metagenomic and metatranscriptomic libraries. In libraries where Epsilonproteobacteria were detected but not assembled we performed read mapping using BBmap (Bushnell B. - sourceforge.net/projects/bbmap/) against the 16S rRNA of the *"B." childressi* associated Epsilonproteobacteria or the cytochrome oxidase I (COI) sequence of the mussels. The mapped reads were then reassembled with SPAdes (Bankevich et al., 2012).

Sequencing the Epsilonproteobacteria from other bathymodiolin species

Primers specifically targeting the 16S rRNA gene of the epsilonproteobacterial sequence were designed using the reverse and complementary sequences of the specific epsilonproteobacterial FISH probes developed in this study (see below): probe BCE141 as a forward primer (BCE141F: TCGGCGCTTATCCCCTGCT) and probe BCE1422 as a reverse primer (BCE1422R: CCGACTTCAGGTGAATTC), resulting in amplification products of 1281 bp. The primer pair was used to screen individuals from ten bathymodiolin species (details in Digital Supplementary Table 2.1). PCR was performed using the same conditions as described for the amplification of the host COI gene, using an annealing temperature of 58°C for the BCE141/BCE1422 primers. Sequencing was performed using the BigDye Terminator v3.1 Cycle Sequencing Kit with the Genetic Analyzer Abiprism 3130 (Applied Biosystems). Sequences have been deposited in Genbank, the accession numbers are displayed in Digital Supplementary Table 2.1

Phylogenetic reconstruction

Sequences were analyzed using the Geneious software. Sequences were trimmed as above. The trimmed sequences were compared with the NCBI nucleotide database using nucleotide BLAST (Madden, 2002) and the Silva rRNA database using the SINA web-aligner

(Pruesse et al., 2012). Highly similar sequences and cultivable reference sequences from the class Epsilonproteobacteria were included in the analysis and aligned using MUSCLE (v3.6.) (Edgar, 2004)

For the bacterial 16S rRNA sequence the final alignment comprised 1049 positions, for the mussel host COI sequences the final alignment comprised 500 positions. Phylogenetic analyses were performed using Bayesian and maximum likelihood analysis. The Bayesian analysis was performed with MrBayes (v3.2) (Huelsenbeck et al., 2001; Ronquist and Huelsenbeck, 2003) under a General Time Reversible model with Gamma-distributed rates of evolution and a proportion of invariant sites. Analyses were performed for 2 million generations using four parallel Monte Carlo Markov chains. Sample trees were taken every 5000 generations. Posterior probabilities calculated with PHYML using a GTR substitution model with 5000 bootstraps were used as support values for nodes in the tree.

FISH probe design

Based on the almost complete 16S rRNA full sequence (1468 bp) of the Epsilonproteobacteria from *"B." childressi,* two probes were designed using the probe design tool of ARB (Ludwig, 2004): (BCE141: AGCAGGGGATAACGCCGA) and (BCE1422: CCGACTTCAGGTGAATTC). Their specificity was verified with the online Silva TestProbe Tool (Quast et al., 2013). Both probes were confirmed to specifically target the Epsilonproteobacteria from *"B." childressi*. 16S rRNA sequences of closely related Epsilonproteobacteria from other bathymodiolin species obtained from metagenomic libraries showed that the probes did not have a mismatch to *B. B. azoricus* and *B.* sp (SMAR) phylotypes. All probes were fluorescently labeled (biomers. net, Ulm, Germany). Probe specificity was determined by increasing formamide concentrations in the hybridization buffer (Pernthaler et al., 2002). Both probes hybridized equally well with the target organism, strong FISH signals were visible up to 60% formamide, but not at 70%.

We therefore used 60% formamide in the hybridization buffer for all FISH experiments. Probe EUB338 (Amann et al., 1990), which hybridizes with most eubacteria was used as a positive control and the NON338 probe (Wallner et al., 1993) as a control for background autofluorescence.

Fluorescence in situ hybridization

For FISH analyses, gill, of *B. azoricus* and *B. childressi,* and gill tissues of *B. brooksi* were subsampled. The samples were fixed in 1 x phosphate buffered saline (PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4 ,2 mM KH2PO4) containing 4% paraformaldehyde at 4°C for 9–18 hours. Samples were washed three times by placing them in fresh 1 x PBS for 10 min each, and subsequently transferred to a cold PBS/ethanol solution containing 1 x PBS and pure ethanol in equal parts. Samples were kept at 4°C on board the research vessel, transported back to the laboratory at 4°C and finally stored at -20°C. Fixed specimens were embedded in Steedman's Wax (Steedman, 1957) and sectioned with a microtome into 8 µm thick sections. The sections were placed on 3-aminopropyltriethoxysilane AminoSilane-Sticky (ATPS) coated slides, dewaxed in three successive baths of absolute ethanol, then rehydrated in three successive baths of 80%, 70% and 50% ethanol for 15 min each and air dried. 20 µl of hybridization buffer (Pernthaler et al., 2002) containing fluorescently labeled oligonucleotide probes (5 $ng.ml^{-1}$ final concentration) was then applied on the sections and the slides where placed in a humid chamber and hybridized in an oven at 46°C for 3-4 hours. The slides were then rinsed in two baths of washing buffer: one dip in the first, then 15 min at 48°C in the second. Finally, the slides were washed in Milli-Q water and absolute ethanol.

Whole mount

Whole mount FISH was performed on *"B." childressi* gill filaments that were fixed for FISH as described above. Dissected gill filaments were directly transferred to 200 µl hybridization buffer (as above) and placed

in a heat shaker in the dark at 46°C for 3-4 hours. After hybridization, the filaments were dipped twice in fresh washing buffer (as above) then washed in fresh buffer for 15 min at 48°C. Finally, the samples were washed in Milli-Q water. Microscopy slides with four layers of regular adhesive tape placed on each side of the slide were used to analyze the gill filaments.

Confocal microscopy.

Confocal images were taken on a Zeiss 780 inverted confocal microscope with Zen 2010 software (Carl Zeiss SAS, Jena, Germany), with a 40×/1.3 Oil Plan Neofluar lens. Pictures were acquired in three track mode with excitations at 405 nm (DAPI), 561 nm (Cy3) and 633 nm (Cy5). Z-stack data set are composed of 40 to 111 pictures taken every 0,4 µm. Where appropriate, tiling and digital zooming were applied. No spectral unmixing was employed to distinguish the multicolor labels. Color was digitally adjusted with the software.

3D reconstruction.

3D rendering was performed on a workstation with Windows 7 containing a 16-core CPU with 64 GB of main RAM and a 2 GB NVIDIA graphics card. The free 3D imaging software Drishti 2.6.1 (Limaye, 2012 - http://sf.anu.edu.au/Vizlab/drishti/index.shtml) was used for visualization of the volumetric CLSM datasets. Volumes of the three channels were imported simultaneously into Drishti. Transfer functions for false color rendering and transparency were adjusted with the 2D histogram. Shading was performed using the shader widget in Drishti. Clipping planes allowed virtual sectioning of the 3D model.

Electron microscopy

For TEM analyses whole gills were dissected from live mussels and cut in smaller pieces for fixation. Samples were fixed for 12 hours at 4°C using an isosmotic fixative (Leisch, unpublished) containing 2.5% glutaraldehyde, 1.5X PHEM buffer (90 mM PIPES, 37.5 mM HEPES,

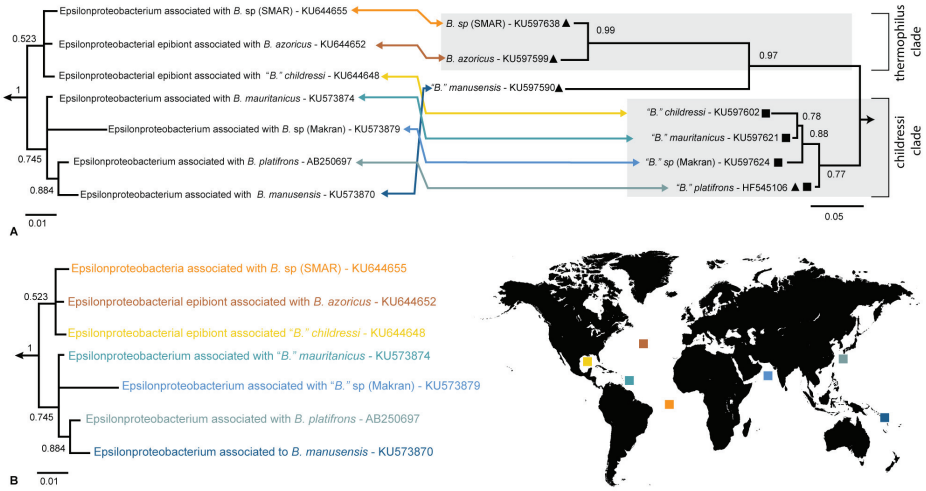
15 mM EGTA and 3 mM MgCl2) and 9% sucrose. For improved mucus retention 0.1% Alcian Blue was added to the fixative immediately before fixation (Erlandsen, Kristich, Dunny, & Wells, 2004). Samples were transported in washing buffer and after rinsing three times, they were post-fixed with 1% osmium tetroxide in ddH_2O for one hour. The samples were then dehydrated in a graded ethanol series (30%, 50%, 70%, 100% twice), transferred into 100% dry acetone, and infiltrated using centrifugation (modified from McDonald (2014)) in 2ml tubes sequentially with 25%, 50%, 75% and 2x 100% LVR resin. During this process, the samples were placed into the tube and centrifuged for 30 seconds with a bench top centrifuge (Heathrow Scientific, USA) at 2000 g for each step. After the second pure resin step, they were transferred into fresh resin in embedding molds and polymerized at 60°C for 24 hours.

Ultra-thin (70 nm) sections were cut with an Ultracut UC7 (Leica Microsystem, Austria), mounted on formvar coated slot grids and contrasted with 0.5% aqueous uranyl acetate (Science Services, Germany) for 20 min and with 2% Reynold's lead citrate for 6 min. Ultrathin sections were imaged at 20 kV on a Quanta FEG 250 scanning electron microscope (FEI Company, USA) equipped with a STEM detector using the xT microscope control software ver. 6.2.6.3123. Images were cropped and contrast values were adjusted using Adobe Photoshop and Illustrator (Adobe Systems, Inc., USA).

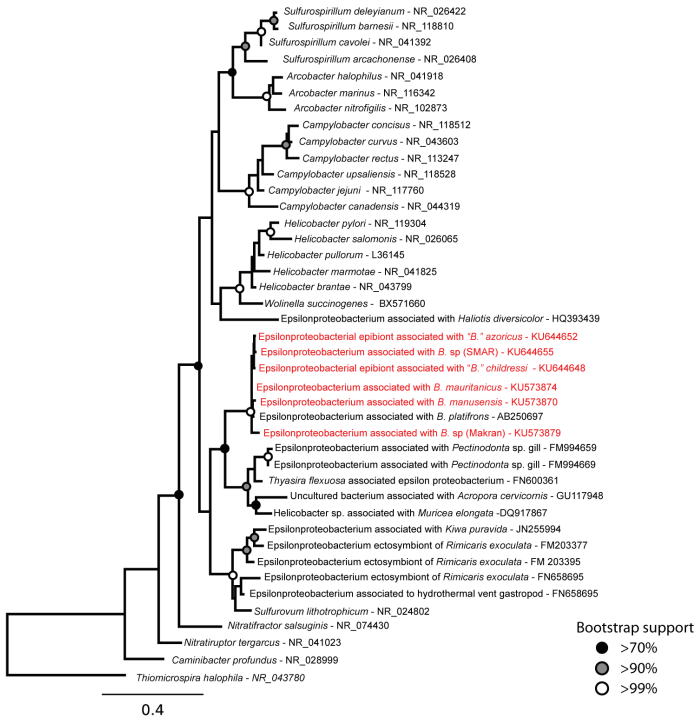
For SEM analyses whole *Bathymodiolus childressi* gill filaments were fixed in Trump's fixative (EMS, Hatfield, USA) for 24 hours and stored in 50% seawater until further processing. Specimens were post-fixed with 1% osmium tetroxide in 0,1 M Cacodylate-buffer (EMS, Hatfield, USA) for one hour and subsequently dehydrated in a graded Acetone series (50%, 60%, 70%, 80%, 90% 100% twice) for 30 minutes at each concentration. All buffers and fixatives contained 10% sucrose. Critical point drying of specimens was done using an EM CPD300 (Leica, Wetzlar, Germany) followed by subsequent sputtering with Gold/Palladium with an Emitech

SC7640 sputter coater (Polaron, New Haven, U.K) for 4 minutes at 10^{-1} mbar and 1 kV.

Scanning electron micrographs were attained with a Quanta FEG 250 (FEI Company, USA) at 20 kV in high vacuum mode using the xT microscope control software ver. 6.2.6.3123.



Supplementary Figure 2.1 A. Bayesian inference tree of 16S rRNA gene sequences of epsilonproteobacterial epibionts (right) and COI gene sequences of bathymodiolin mussels (left). *B.* Geographical distribution of the collected samples. There are no evident correlations between the epsilonproteobacterial phylogeny and the host phylogeny, the habitat, or the geographic location of sampling.



Supplementary Figure 2.2 . Posterior probabilities of the epsilonproteobacterial 16S rRNA (1049 bp) tree calculated with PHYML using a GTR substitution model. 5000 bootstraps were used as support values for nodes in the tree.

2.3 Experimental procedures References

Amann, R.I., Krumholz, L., and Stahl, D. A. (1990) Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *J. Bacteriol.* **172**: 762–770.

Bankevich, A., Nurk, *S.*, Antipov, D., Gurevich, A. a., Dvorkin, M., Kulikov, A.S., et al. (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**: 455–477.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acid Res.* **32**: 1792–1797.

Erlandsen, *S.*L. (2004) High-resolution visualization of the microbial glycocalyx with low-voltage scanning electron microscopy: dependence on cationic dyes. *J. Histochem. Cytochem.* **52**: 1427–1435.

Guindon *S.*, Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**:307-21

Huelsenbeck, J.P., Ronquist, F., Nielsen, R., and Bollback, J.P. (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294**: 2310–2314.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, *S.*, Cheung, M., Sturrock, *S.*, et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–9.

Limaye, A. (2012) Drishti: a volume exploration and presentation tool. *SPIE 8506, Dev. X-Ray Tomogr. VIII, 85060X* **8506**: 85060X.

Ludwig, W. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**: 1363–1371.

Matsumoto, M. (2000) Phylogenetic analysis of the family Pectinidae (Bivalvia) based on mitochondrial cytochrome c oxidase subunit I. *J. Molluscan Stud.* **66**: 477–488.

McDonald, K.L. (2014) Rapid embedding methods into epoxy and lr white resins for morphological and immunological analysis of cryofixed biological specimens. *Microsc. Microanal.* **20**: 152–163.

Muyzer, G., Teske, A., Wirsen, C.O., and Jannasch, H.W. (1995) Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Arch. Microbiol.* **164**: 165–172.

Pernthaler, A., Pernthaler, J., and Amann, *R.* (2002) Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl. Environ. Microbiol.* **68**: 3094–3101.

Pruesse, E., Peplies, J., and Glockner, F.O. (2012) SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823–1829.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590–D596.

Ronquist, F. and Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Sayavedra, L., Kleiner, M., Ponnudurai, R., Wetzel, *S.*, Pelletier, E., Barbe, V., et al. (2015) Abundant toxin-related genes in the genomes of beneficial symbionts from deep-sea hydrothermal vent mussels. *Elife* **4**: e07966.

Wallner, G., Amann, R., and Beisker, W. (1993) Optimizing fluorescent in situ hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. *Cytometry* **14**: 136–43.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. *Gene* **33**: 103–119.

Zhou, J., Bruns, M.A., and Tiedje, J.M. (1996) DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**: 316–322.

Chapter 3 **I wanna be like you: Multiple horizontal gene transfers in an epsilonproteobacterial epibiont of *Bathymodiolus***

Authors: Adrien Assie[1], Harald Gruber-Vodicka[1], Samantha Joye[2], Matthew Saxton[2], Halina Tegetmeyer[1], Nicole Dubilier[1,4,5], Jillian M. Petersen[1,3]

1 Symbiosis Department, Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, D-28359 Bremen, Germany.

2 Department of Marine Sciences, The University of Georgia, Room 159, Marine Sciences Bldg. Athens, GA 30602-3636

3 Division of Microbial Ecology, Department of Microbiology and Ecosystem Science, Research Network Chemistry Meets Microbiology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria.

4 Faculty of Biology / Chemistry, University of Bremen, Bibliothekstrasse 1, D-28359 Bremen, Germany

5 Marum - Zentrum für Marine Umweltwissenschaften der Universität Bremen, Leobener Str. 2, 28359 Bremen

Manuscript in Preparation

3.1 Abstract

Microbial communities at cold seeps and hydrothermal vents are often dominated by autotrophic Epsilon- and Gammaproteobacteria, which harness energy from reduced chemicals to fuel carbon fixation. All currently known epsilonproteobacterial autotrophs use the reductive tricarboxylic acid (rTCA) cycle to fix carbon. In contrast, most known autotrophs from the Gammaproteobacteria rely on a different pathway, the Calvin Benson Bassham (CBB) cycle for carbon fixation. We recently discovered a novel family of epsilonproteobacterial epibionts associated with bathymodiolin mussels from hydrothermal vents and cold seeps worldwide. To understand the nature of the association between epibionts and mussel host we assembled draft genomes of epsilonproteobacterial epibionts from two bathymodiolin species. The genomes revealed that these Epsilonproteobacteria are sulfur oxidizers that have the genetic potential to fix inorganic carbon through the CBB cycle instead of the rTCA cycle. Remarkably our phylogenetic reconstructions of key CBB genes suggested that the genes were acquired through two distinct horizontal gene transfer events. The gene sequences coding for the key enzyme of the CBB cycle, the 1,5-ribulose bisphosphate carboxylase, are similar to gammaproteobacterial sequences closely related to the sulfur-oxidizing endosymbionts of bathymodiolin mussels. The other genes of the cycle appear to have multiple betaproteobacterial origins. Additionally, we show that the key gene of the sulphur oxidizing (SOX) multi-enzyme pathway also has a gammaproteobacterial origin. We hypothesize that these horizontal gene transfer events enabled Epsilonproteobacteria to establish a successful symbiosis with bathymodiolin mussels. Additionally very few pathogens are autotrophic bacteria suggesting these epibionts are in a commensalistic or mutualistic relationship with their host.

3.2 Introduction

In the open ocean, primary production is predominantly driven by photosynthetic inorganic carbon fixation via the Calvin Benson Bassham cycle (CBB) (Raven, 2009). However, little to no light penetrates below 200 m depth. The dark deep-sea life therefore mainly relies on organic matter falling from the surface as nutrient input (Turner, 2002). Exceptions are hydrothermal vents and cold seeps - geological features where local release of fluids enriched in reduced compounds fuels a rich microbial community that thrives through chemoautotrophic processes. The fixation of inorganic carbon by these microorganisms is driven by energy derived from the oxidation of sulfur, hydrogen or methane (Walsh *et al.*, 2009; Orcutt *et al.*, 2011; Swan *et al.*, 2011).

Autotrophs convert inorganic carbon to organic compounds later available to higher trophic levels. Although the CBB cycle is the most widespread CO_2 fixation pathway on Earth, several other pathways exist (Reviewed in Hügler and Sievert, 2011). Multiple studies have shown that the reductive tricarboxylic acid (rTCA) cycle and the CBB cycle are the predominant pathways to fix inorganic carbon at hydrothermal vents and cold seeps (Campbell *et al.*, 2006; Nakagawa and Takai, 2008; Sievert and Vetriani, 2012; Reeves *et al.*, 2014). Additionally, CBB and rTCA cycles appear to be associated with different bacterial classes, for example the CBB cycle is associated with Alpha-, Beta- and Gammaproteobacteria, as well as some Chloroflexi and iron-oxidizing Firmicutes, whereas rTCA is associated with Chlorobiales, Epsilonproteobacteria and Aquificales (Hügler and Sievert, 2011). Many studies of deep-sea hydrothermal vent and cold seep microbial communities have shown the preference of certain bacterial classes for various local environments (Nakagawa and Takai, 2008; Sievert *et al.*, 2008; Pop Ristova *et al.*, 2015). In environments close to hydrothermal vent fluids, characterized by high temperature, low oxygen concentration and high fluid flow rates, Aquificales and Epsilonproteobacteria are often the dominant microbes (Campbell *et al.*, 2006; Huber *et al.*, 2007; Akerman *et al.*, 2013). The

autotrophic members of these bacterial classes have thus far been thought to rely on the rTCA cycle for CO_2 fixation. On the other hand, environments further away from hot fluids, with higher oxygen concentration and lower, more stable temperatures, are often dominated by Gammaproteobacteria, which mainly utilize the CBB cycle (Yamamoto and Takai, 2011; Anderson *et al.*, 2013; Oshkin *et al.*, 2014).

Another unique feature of hydrothermal vents and cold seeps is the presence of dense macrofaunal populations (Desbruyères *et al.*, 2006; Galkin, 2016). The bathymodiolins are a subfamily of deep-sea bivalves occurring around the world in both vent and seep habitats. These mussels have established mutualistic relationships with chemoautotrophic bacteria, allowing them to thrive in these environments. Gammaproteobacterial endosymbionts, capable of oxidizing sulfur and/or methane to produce energy and fix carbon, are found in the mussel's gill epithelial cells (Duperron *et al.*, 2009). In exchange for a stable ecological niche with a constant flux of water rich in reduced compounds, the prokaryotic partners provide nutrition to their eukaryotic host (Reviewed in Dubilier *et al*., 2008).

A major challenge in studying the interaction within a deep-sea symbiotic system, such as the bathymodiolin mussel system, lies in the difficulty of sampling and maintaining laboratory cultures of the host and symbionts. Molecular techniques have been of great help in understanding the diversity and distribution of both partners (Bayer *et al.*, 2009; Rinke *et al.*, 2009; Petersen *et al.*, 2010; Zwirglmaier *et al.*, 2015). The recent rise of next generation sequencing, such as Roche 454 and Illumina technologies, has helped to investigate genomes and gene expressions of symbiotic systems (Kuwahara *et al.*, 2007; Newton *et al.*, 2007; Grzymski *et al.*, 2008; Jan *et al.*, 2014).

We recently described the diversity and distribution of a novel Epsilonproteobacteria family widely associated as gill epibionts with bathymodiolin mussels worldwide (Assié *et al*., 2015). We showed that two different bathymodiolin genera, *B. azoricus* and *"B." childressi* (also referred

to as *Gigandidas childressi* - Jones and Vrijenhoek*, 2006;* Thubaut *et al., 2013)* were colonized by closely related epsilonproteobacterial epibionts. However, little is known of the nature of the relationship between the epibionts and their host. This new Epsilonproteobacteria family, based on a 16S rRNA gene sequence identity, is equally distant to the *Sulfurovum* clade, which groups many free living and associated chemoautotrophic bacteria, and to the *Helicobacter* clade, a known group of gastrointestinal pathogens. We therefore aimed to understand the genetic potential of the epibionts through the use of metagenomic and metatranscriptomic approaches. The current study focuses on the epibionts metabolic pathway to fix inorganic carbon and gain energy. Furthermore, we discuss the origin and the potential for either a mutualistic or pathogenic relationship of the epsilonproteobacterial epibionts with their mussel host.

3.3 Material and Methods

Sample collection

For metagenome analysis, *"B." childressi* individuals were collected at the GC246 site in the Gulf of Mexico during the R/V Atlantis AT26-13 cruise and *B. azoricus* individuals were collected at the Lucky Strike site on the north mid-Atlantic ridge (NMAR) during the Biobaz cruise (summarized in Digital Supplementary Table 3.1). For transcriptome analysis, *"B." childressi* individuals were collected at the brine pool site GC233 in the Gulf of Mexico during the Nautilus NA043 cruise. During all cruises, the mussel's gills were dissected on board and gill fragments were later fixed in RNA (Sigma, Germany) according to the manufacturer's instructions overnight; the fixative was then removed and the samples were stored at −80 ˚C.

Genomic DNA was extracted from the gill tissue of the mussels according to Zhou *et al.* (1996) with the following modifications: an initial incubation step was performed at 37°C in 360 µl of extraction buffer (100 mM Tris-HCl [pH 8.0], 100 mM sodium EDTA [pH 8.0], 100 mM sodium phosphate [pH 8.0], 1.5 M NaCl, 1% CTAB) and 40 µl of proteinase K (10 mg/ml) overnight.

For transcriptome sequencing, RNA was extracted from the gill tissue with a QIAGEN Allprep(R) DNA/RNA micro kit (Hilden, Germany) according to the manufacturer's instructions. The quality of the DNA and RNA was assessed with a Qubit® 2.0 Fluorometer (Invitrogen, Eugen, USA).

Metagenome sequencing and assembly

We sequenced metagenomes of one *"B." childressi* and one *B. azoricus* mussel. The epsilonproteobacterial draft genome associated with *B. azoricus* is referred to as EpsA and that associated with *"B." childressi* as EpsC.

The *"B." childressi* metagenomic sample was sequenced by the Center for Biotechnology at the University of Bielefeld (Bielefeld, Germany). A total of 471,459,598 paired-end reads (150 bp long) and 7,739,150 paired-end reads were generated on Illumina MiSeq and Hiseq machines, respectively. The *B. azoricus* sample was sequenced by the Max Planck Genome Center (Cologne, Germany) and 159,408,731 paired-end reads (250 bp long) were generated on an Illumina HiSeq 2500.

We screened the metagenomic and metatranscriptomic libraries for 16S rRNA sequences to assess the presence or absence of Epsilonproteobacteria. The same methods were used as described in Chapter 2. We used PhyloFlash 2.0 (https://github.com/HRGV/phyloFlash) to perform RNA small sub unit (SSU) screening and reconstructions.

Assembly was performed as follows: first the Raw reads were quality trimmed (Q=2) and Illumina adapters were removed using BBduk v3.5 (Bushnell B. - sourceforge.net/projects/bbmap/). An initial assembly was then performed using Megahit software (Li *et al.*, 2014) with default settings. The outputted assembly file was then analyzed with metawatt V1.7 binning tools (Strous *et al.*, 2012), and genomics groups, called here bins, were established by analyzing contig tetranucleotide frequency, coverage and GC values. Contigs belonging to bins with an epsilonproteobacterial taxonomic signature were then extracted. The quality trimmed metagenomic reads were then mapped against the epsilonproteobacterial bin contigs using

Bbmap filtering reads with a minimum identity of 98%. The mapped reads were then used for a new assembly using the single cell assembler SPAdes 3.4.2 (Bankevich *et al.*, 2012) with default settings.

Bin quality was checked with CheckM software (Parks *et al.*, 2015) and a new iteration of taxonomic binning, mapping assembly was performed until no contamination from other bacterial strains or host remained in the assembly. Contigs smaller than 500 bp were excluded from the subsequent analyses because they were unlikely to hold any relevant genetic information. Genome annotations were conducted with RAST and crosschecked with IMG ER web servers (Meyer *et al.*, 2008; Markowitz *et al.*, 2012). Genes of interest were manually curated against public databases. Genome average nucleotide identity (ANI) and average amino acid identity (AAI) were calculated using the AAI and ANI calculator from the enveomics collection (Rodriguez-R and Konstantinidis, 2016) with the default settings.

On Supplementary Figure 3.1 we plotted each contig from the *B. azoricus* metagenomic assembly according to the GC content and the coverage values. The figure was generated with the genome binning tool R package (Seah and Gruber-Vodicka, 2015) using the Phyla Amphora package (Wang and Wu, 2013) to identify marker genes.

Transcriptome sequencing and processing

Transcriptomes of three *"B." childressi* individuals were sequenced at the Max Planck Genome Center (Cologne, Germany). The libraries are referred to as LibA; LibB and LibC. For each library, 150 bp long paired-end sequencing was done and 53,094,890, 54,169,778 and 39,669,403 reads were generated, respectively.

Reads were quality trimmed (Q=2) and Illumina adapters were removed with BBduk. Transcriptome reads were mapped against the EpsC draft genome with BBmap, with a minimum similarity of 98%. The number of transcriptome reads mapping to each gene was estimated with featureCounts (Liao *et al.*, 2014). To compare the transcriptome counts between individuals,

a normalization factor per library was estimated with a trimmed mean of M-values (TMM) (calcNormFactors, edgeR package) (Oshlack *et al.*, 2010). The counts were then converted to reads per kilobase per million (RPKM) (Rsubread package).

Gene screening

In order to check the presence or absence of the rTCA cycle in the different metagenomic libraries, we created three individual databases containing published amino acid sequences of epsilonproteobacterial rTCA key genes. These genes were ATP citrate lyase, 2-oxoglutarate ferredoxin oxidorecductase and pyruvate ferredoxin kinase. The first metagenomic assembly iterations, as well as the final epsilonproteobacterial bins, were screened using blastx against the respective database to detect the presence of potential rTCA related genes.

Phylogenetic analysis

The IMG ER pipeline detected, based on sequence homologies to their database, the presence of genes with a gammaproteobacterial origin and we extracted and analyzed the sequences with Geneious software. Sequences of interest were compared to the NCBI nucleotide and amino acid databases using nucleotide and amino acid BLAST. We retrieved closely related sequences from the blastx results on the NCBI non-redundant database. Additionally, other reference sequences were included in the analysis and all sequences were aligned using MUSCLE (v3.6.) (Edgar, 2004). In order to detect the best substitution model to use for amino acid sequences phylogenetic reconstruction, we used the ProtTest3 package (Darriba *et al.*, 2011). Phylogenetic analyses were then performed using Bayesian and Maximum Likelihood analysis. Bayesian analysis was performed with MrBayes (v3.2) (Huelsenbeck *et al.*, 2001; Ronquist *et al.*, 2011) under a General Time Reversible model with the best-fitted substitution model. Analyses were performed for two million generations using four parallel Monte Carlo Markov chains. Sample trees were taken every 1000 generations. Maximum

likelihood trees were calculated with PHYML (Guindon *et al.*, 2010) using the best fitted substitution model. We used 1000 bootstraps as support values for nodes in the tree.

Codon usage analysis

The codon usage of every EpsA and EpsC gene was determined with CodonW (Peden, 2011) using default parameters. The outputted PCA analyses results were then plotted with R (version 3.2.3). To check whether unusual taxonomic affiliation was due to a recent horizontal gene transfer (HGT) we performed codon usage analysis. This analysis calculated the codon usage frequency and performed a PCA test on the results (Supplementary Figure 3.2).

3.4 Results

Genome assembly

Metagenomes of *"B." childressi* and *B. azoricus* gill tissues were deeply sequenced in this study. Using similar assembly and binning methods on both metagenomic data sets, we were able to identify and selectively assemble Epsilonproteobacteria-related draft genomes. The *"B." childressi* metagenome initial assembly yielded a single bin with an Epsilonproteobacteria signature. We were able to assemble a draft genome, which included a 16S rRNA sequence identical to previously published epsilonproteobacterial sequences (Chapter 2).

The metagenomic assemblies of the "B. *azoricus*" sample revealed the presence of three distinct epsilonproteobacterial bins (Supplementary Figure 3.1), suggesting the presence of multiple epsilonproteobacterial species. This was supported by the SSU reconstruction, which predicted three 16S rRNA sequences. One related to the *"B." childressi* epsilonproteobacterial epibiont, one related to *Sulfurovum* sp. (AQWF01000016) and the last one to Sulfurimonas sp. (KC682116). Two bins were present in low abundance, below an average coverage of 5x. The third bin was present at around 20x

coverage. We compared the average nucleotide and amino acid identity (respectively ANI and AAI) of the different bins with representative genomes to identify to which epsilonproteobacterial family they belong. We used the epsilonproteobacterial epibiont draft genome from *"B." childressi* and draft genome of*S. lithotrophicum* and*S.* dentirifans as references, since they are the closest cultivable relative to the epsilonproteobacterial sequences predicted by SSU screening. The ANI values were too divergent to be comparable, the AAI values are given in Table 3.1.

Table 3.1 AAI values of the different epsilonproteobacterial bins annotated from the *B. azoricus* metagenomic assembly compared to reference datasets.

	1.	***2.***	***3.***	**4.**	**5.**	**6.**
1. EpsC						
2. S. ***lithotrophcium***	43.86					
3. S. ***denitrificans***	45.56	51.76				
4. BinA	76.26	44.64	45.97			
5. BinB	47.40	51.89	49.35	50.17		
6. BinC	45.14	51.84	66.20	46.37	48.14	

BinA, with the highest AAI value, was closely related to the epibiont genome from *"B". childressi* EpsC. The AAI value of BinC indicated that it was closely related to*S.* denitrificans and BinB seemed to be most similar to*S. lithotrophicum*.

The low abundance Epsilonproteobacteria from BinB and BinC are related to deep-sea free living and surface colonizers. There was no indication from our previous FISH study that other filamentous Epsilonproteobacteria were present on *B. azoricus* gills. From the published information about *S. lithotrophicum* and*S. denitrificans*, plus the lack of physical evidence of the presence of other Epsilonproteobacteria in *Bathymodiolus* gills, we hypothesize that BinB and BinC might be contamination of environmental bacteria. However, we cannot exclude that the Epsilonproteobacteria associated with *B. azoricus* are part of a resident community of ectosymbionts. The present study focused on comparison of the epsilonproteobacterial epibionts. Description of the two additional epsilonproteobacterial bins is

beyond the scope of this paper and will be addressed elsewhere.

The draft genomes associated with the novel epsilonproteobacterial epibiont were analyzed in detail in this study and for clarity we refer to epsilonproteobacterial bin "BinA" associated with *B. azoricus* as EpsA and the epsilonproteobacterial bin assembled in *"B." childressi* as EpsC.

For genome analysis, contigs smaller than 500 bp or without any predicted ORFs were excluded. The EpsA genome was 92% complete with 1950 contigs, while the EpsC genome was 92% complete with 843 contigs. The genomes were 2.8 (EpsA) and 2.1 (EpsC) mb and 16S rRNA sequences were reconstructed for each library. These SSU sequences were identical to previously published sequences in Assié *et al*., (Chapter 2). The draft genomes had an average G+C-content of 35% and median coverage of 262 to 40 fold.

Annotations were predicted by the RAST web server and verified by the IMG web server. The draft genome of EpsC was composed of 2552 protein coding genes and 44 tRNA-encoding genes and the draft genome of EpsA was composed of 2393 protein coding genes and 40 tRNAs (genes are summarized in Digital Supplementary Table 3.2). Annotation details are beyond the scope of this paper and are discussed at length elsewhere (Chapter 4)

Calvin cycle

In both reconstructed epsilonproteobacterial genomes, all essential genes for carbon fixation via the CBB pathway were present. Although most of the genes involved in the CBB cycle are also used in other metabolic pathways, there are two genes that are unique to this cycle: phosophoribulokinase and ribulose 1,5-bisphosphate carboxylase/oxidase (RuBisCO). In both draft genomes, the RuBisCO genes were present in the same order in the cluster: a RuBisCO activation protein *cbbQ* followed by a conserved hypothetical gene and small and large RuBisCO subunit genes (*cbbS* and *cbbL*). EpsC had an additional RuBisCO activation protein *cbbO* coding region on the 5' side; in the EpsA genome the *cbbO* gene was located on a separate small contig (Figure 3.1). Phosphoribulokinase was located

on different contigs in the two genomes, within the same gene cluster. In both cases, the phosphoribulokinase gene was surrounded by the same five genes: fructose-1,6-bisphosphatase, transketolase, phosphoglycolate phosphatase, fructose-bisphosphate aldolase and ribulose-phosphate 3-epimerase (Figure 3.1). All these genes are involved in the CBB cycle.

Figure 3.1 Graphic display of the EpsA and EpsC contig with the CBB related genes. The read coverage is plotted above the contigs and the phylogenetic origin of the surrounding genes is displayed. A. EpsA contigs *B.* EpsC contigs.

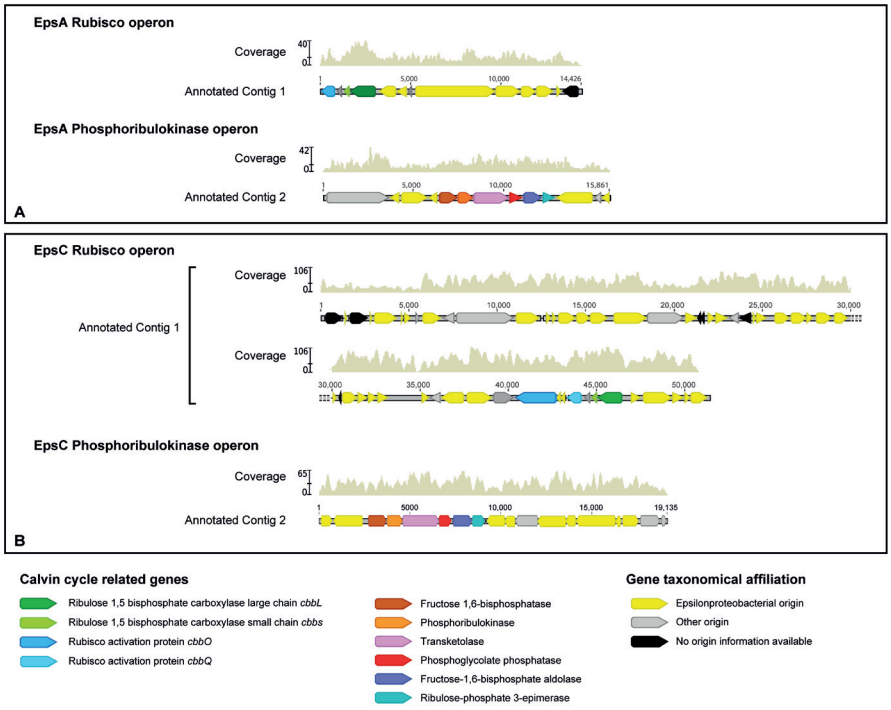
We performed a Blast search of translated amino acid sequences of the large RuBisCO subunit. It showed 96% identity to a large RuBisCO subunit sequence of the free-living sulfur oxidizing Gammaproteobacteria *Candidatus* Thioglobus sp. EF1 (WP_053951896) and 95% identity to the "endosymbiont of *Bathymodiolus* sp." (WP_010646812). The small RuBisCO subunit showed similar results, with 87% similarity to the *Candidatus* Thioglobus sp. EF1 (WP_053951897) sequence and 82.6% similarity to the "thioautotrophic gill symbiont of *B. azoricus*" (CRN09809) sequences.

The phosphoribulokinase sequences had a 64.5% nucleotide sequence identity to the closest published sequence from the gammaproteobacterial endosymbiont of vesicomyid clams, *Ca.* Ruthia magnifica (CP000488). The amino acid sequence was 63.2% identical to the *Ferrovum* sp. JA12 (WP_056930131) phosphoribulokinase sequence. The other genes present in the "phosphoribulokinase cluster" had various betaproteobacterial best blast hits (summarized in Digital Supplementary Table 3.3). The most closely related sequences for all genes were from either the Gammaproteobacteria or Betaproteobacteria classes.

Absence of rTCA cycle in epsilonproteobacterial epibiont genomes

We screened both initial metagenomic assemblies and final genomic bin assemblies against the amino acid database of the epsilonproteobacterial rTCA cycle key genes, which include pyruvate:ferredoxin oxidorecducatase *porABCD,* 2-oxoglutarate oxidoreductase genes *oorABDG,* and ATP citrate lyase genes *aclBA*. The initial metagenomic assemblies of the *B. azoricus*

metagenome revealed hits with a significant e value for every gene involved in the rTCA cycle. This metagenome assembly contained three distinct epsilonproteobacterial bins and include EpsA described in the present study, BinB related to the *Sulfurovum* and BinC to the *Sulfurimonas* genus. The final three epsilonproteobacterial bins all shared four annotations for *porABCD*



genes; only BinB was predicted to have *oorABDG* genes; and finally, all three bins had genes related to *aclA* but only BinB and BinC had gene *aclB.* No significant hits for the ATP citrate lyase subunits were recovered from the initial assemblies or the final epsilonproteobacterial bins of the *"B." childressi* metagenomic libraries. Genes *oorDG* and *porAB* were found in the same gene cluster in the final EpsC bin.

Phylogeny of the genes

We performed phylogenetic reconstruction of the *soxB* and CBB cycle

gene, including references sequences to compare the phylogenetic origins. We performed a substitution model test independently for every set of genes and determined that LG inverse Gamma was the best model for all of them. We then used this setting to performed Bayesian inference and maximum likelihood based phylogenetic reconstruction of the amino acid sequences of the genes involved in the CBB cycle.

The ribulose 1,5-bisphosphate carboxylase large and small subunits fell basal to a group of sequences associated with sulfur-oxidizing bacteria from the gammaproteobacterial group SUP05. Many of the related sequences were from bacteria associated with deep-sea invertebrates (Figure 3.2 A and B). The epsilonproteobacterial epibiont phosphoribulokinase sequences were placed on an unresolved branch basal to both gammaproteobacterial and betaproteobacterial clades (Figure 3.4). For the other genes coding for CBB proteins, fructose 1,6-bisphophatase, 1,6-bisphophate aldolase, transketolase, ribulose phosphate 3-epimerase and phosphoglycolate phosphatase, each of their amino acid sequences grouped consistently with betaproteobacterial sequences. However the sequences did not share the same betaproteobacterial origin.

From the ten genes related to the SOX system, three genes were closely related to gammaproteobacterial sequences: the essential gene soxB and two copies of the accessory gene soxH. Bayesian inference and Maximum likelihood phylogenetic reconstruction of the soxB genes showed that the genes were closely related to sequences from symbiotic Gammaproteobacteria (Figure 3.3) associated with deep-sea invertebrates, whereas most of the recorded Epsilonproteobacteria soxB genes cluster in distinct different clades. Additionally, transcriptome analyses revealed that all the SOX genes were expressed, some of the genes involved in the sulfur oxidation being among the most expressed genes in every transcriptome.

Codon usage

We plotted the PCA result of gene codon usage from both draft genomes.

Once plotted, genes with a similar codon usage clustered closer together, while genes with a different codon usage were located away from the main plot. The plot is displayed in Supplementary Figure 3.2.

Gene expression

Transcriptome analysis of three *"B." childressi* individuals indicated that

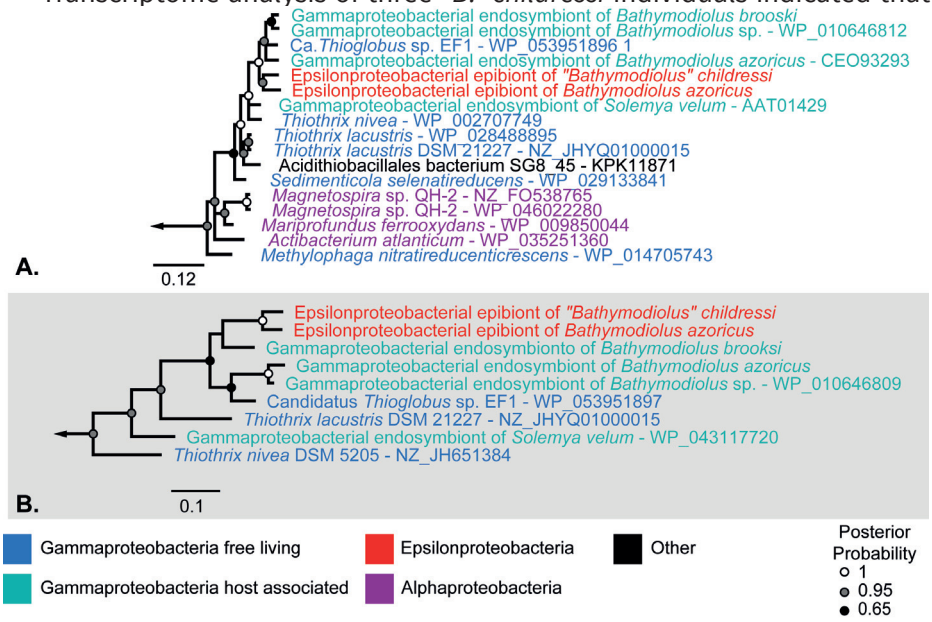


Figure 3.2 Bayesian inference based phylogenetic tree of **A.** Large RuBisCO chain and *B.* Small RuBisCO chain amino acid sequences phylogenies. The phylogenies were calculated using an LG inverse gamma substitution model. Complete trees are displayed as Digital Supplementary Figures 17 and 18. Maximum likelihood trees are available as Digital Supplementary Figure 7 and 8.

the key genes of the CBB cycle were expressed. RuBisCO large and small genes were among the ten most expressed annotated genes (**Table 3.2** Gene expression table of the various genes involved in the CBB cycle. Only genes involved in CBB are shown. The gene expression ranking is based on genes with known annotations. The gene expression is quantified by counting the number of transcriptome reads mapping to the EpsC draft genome, which are then normalized with the TMM method.).

Table 3.2 Gene expression table of the various genes involved in the CBB cycle. Only genes involved in CBB are shown. The gene expression ranking is based on genes with known annotations. The gene expression is quantified by counting the number of transcriptome reads mapping to the EpsC draft genome, which are then normalized with the TMM method.

Rank	Gene ID	Library A	Library B	Library C	Average

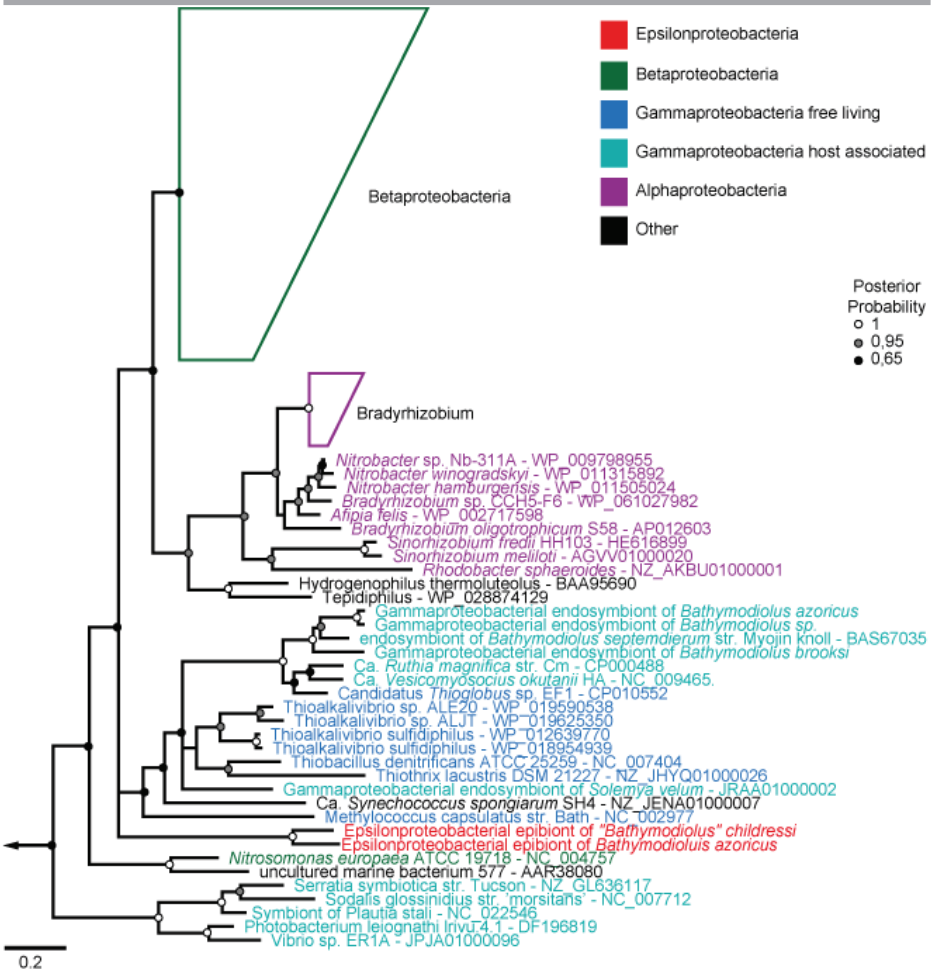


Figure 3.3 Bayesian inference-based phylogenetic tree of the phosphoribulokinase gene amino acid sequences. The phylogenies were calculated using an LG inverse gamma substitution model. Complete trees are displayed as Digital Supplementary Figure 19 and maximum likelihood tree is available as Digital Supplementary Table 9.

Figure 3.4 Bayesian inference-based phylogenetic tree of the ribulose phosphate 3-epimerase. The phylogenies were calculated using an LG inverse gamma substitution model. Complete trees are displayed as Digital Supplementary Figure 14.

5	RuBisCO small subunit	8649.29	2887.96	3972.75	5170.00
6	RuBisCO large subunit	9942.60	2029.42	2748.89	4906.97
72	Sulfate thiol esterase SoxB	578.10	91.74	182.29	284.04
97	Fructose bisphosphate aldolase	413.72	92.83	142.29	216.28
101	Transketolase	382.74	57.87	197.11	212.57

161	D-fructose 16bisphosphatase	279.27	57.69	29.48	122.15
170	Ribulosephosphate 3epimerase	207.69	0	141.28	116.32
213	Phosphoribulokinase	195.26	22.60	23.10	80.32
400	Phosphoglycolate phosphatase	49.22	0	30.91	26.71

3.5 Discussion

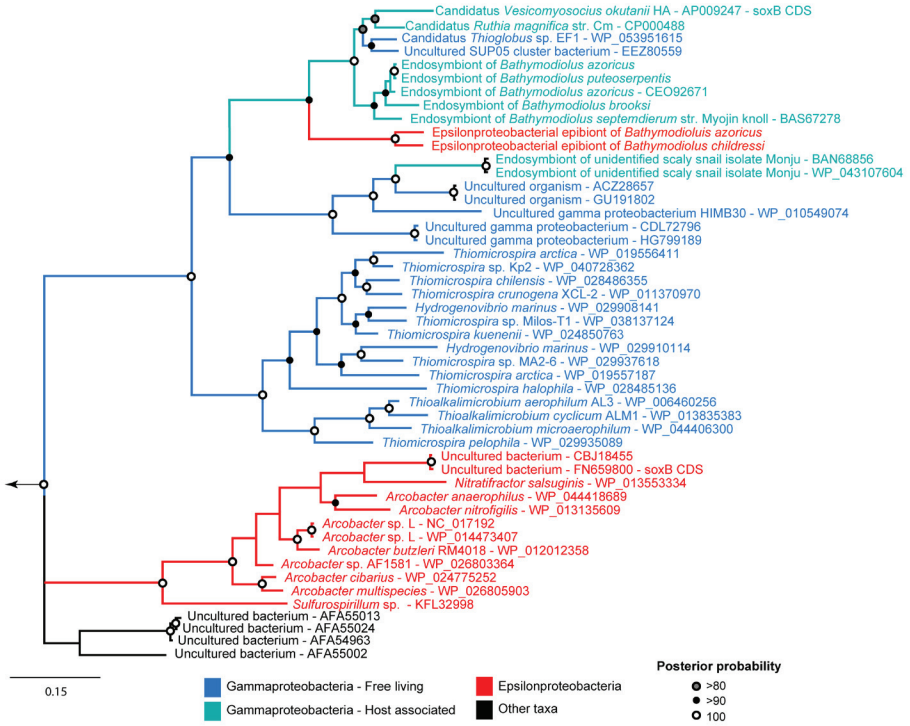


Figure 3.5 Bayesian inference phylogenetic reconstruction of the amino acid sequences of the soxB genes. The Epsilon- and Gamma- proteobacteria sequences are separated into two distinct clades. The epsilonproteobacterial epibiont sequences cluster within the gammaproteobacterial group. The whole tree is available as Digital Supplementary Figure 15 and Maximum likelihood tree is available as Digital Supplementary Figure 20.

rTCA Absence

The rTCA cycle appears to be absent or non-functional in both epsilonproteobacterial draft genomes. EpsC and EpsA both have an incomplete rTCA cycle, but different genes were missing in the two genomes.

Both draft genomes lack the key gene *aclB*. EpsA possesses *aclA* but lacks the whole *oor* operon, while EpsC lacks *aclA* as well as subunits from the *oor* and *por* genes. Due to the very fragmented nature of the genomes, we cannot exclude that these genes have eluded the genome assembly. However, it is unlikely because the *"B." childressi* transcriptome showed that RuBisCO genes were highly expressed, suggesting this is the main inorganic carbon fixation pathway. It would instead suggest a genome erosion process has taken place in the epsilonproteobacterial epibiont and the unused rTCA genes are being progressively lost.

Some of the key genes involved in the rTCA cycle could also be used in accessory pathways. The presence of pyruvate ferredoxin oxidoreductase in both bins could be explained by the use of the enzyme in other metabolic pathways, such as pyruvate fermentation to acetyl-CoA (Ragsdale, 2003; Hug *et al.*, 2010). Additionally, the presence of two additional 2-oxoglutarate:ferredoxin oxidoreductase subunits (*oorDG*) in EpsC could be correlated with previous observations, whereby in *Helicobacter pylori por* and *oor* were hypothesized to be involved in oxidation of NADPH using it as an electron donor source(Hughes *et al.*, 1998). The uneven loss of genes linked to the rTCA cycle could also indicate that the genetic loss of the rTCA cycle happened recently in both epibiont species.

Origin of the Calvin Cycle

Our analysis showed that the genes involved in the CBB cycle were grouped into two clusters. The first included RuBisCO related genes, such as small and large chain, as well as an activator protein. The other cluster included all other accessory genes, such as phosphoribulokinase and fructose 1-6-fructose bisphosphate aldolase. Phylogenetic reconstruction suggested that the genes had two distinct origins. RuBisCO related genes appeared to belong to a gene cluster related to autotrophic Gammaproteobacteria, many of which were related to endosymbionts of deep-sea bivalves. All the other genes present in the second cluster were related to various Betaproteobacteria (Digital Supplementary Table 3.2)

Horizontal gene transfer (HGT) is a ubiquitous genetic event where DNA exchange occurs between two organisms. This process was deemed critical to the evolution of Bacteria, by shaping their genetic background, either by reordering the genome synteny or expanding the genetic potential. These events are the principal source of phenotypic innovation in Bacteria, leading to niche adaptation (Dutta and Munmun, 2015). HGT is often described as a genetic element copied from one organism and transferred to another. Various transfer mechanisms exist, such as integration via a viral vector (e.g. bacteriophages), the acquisition of plasmids or transposable elements and also by non-homologous recombination (Mell and Redfield, 2014). Epsilonproteobacterial DNA uptake mechanisms are poorly documented but some studies have described uptake via a specific Type 2 secretion system homologous to the Type 4 secretion system described as the most common DNA uptake system in Bacteria (Juhas *et al.*, 2008; Gilbreath *et al.*, 2011). A Type 2 secretion system is present in both our draft epsilonproteobacterial genomes, although more functional analyses must be done to confirm its ability to uptake DNA in the epibionts genomes. The genome neighborhoods of the CBB clusters are different in our two draft genomes, suggesting either that the clusters have been acquired independently or, and this is more likely, that the Epsilonproteobacteria underwent genome rearrangement. Additionally, no obvious transposases or mobile element sequences were found close to these clusters. We thus hypothesize that a natural recombination event happened within the epibionts and could be responsible for the DNA uptake. This is supported by the observed high micro-diversity of Epsilonproteobacteria in deep-sea environments, which is thought to be due to a lack of DNA repair genes, a trait common to many Epsilonproteobacteria (Nakagawa *et al.*, 2005; Campbell *et al.*, 2006).

The genes coding for other CBB cycle enzymatic reactions are present in a second cluster, designated here the “phosphoribulokinase cluster”, probably originating from a different organism. This is supported by the presence of a fructose 1,6-bisphosphatase gene in both EpsA and EpsC, which absent in

the gammaproteobacterial group to which the RuBisCO related sequences belong. These bacteria use an alternative CBB cycle in which sedoheptulose 1-7 and fructose 1,6-bisphosphatase are absent (Sayavedra *et al.*, 2015). Additionally the genes present in the "phosphoribulokinase cluster" did not share the same phylogenetic affiliation. The phylogenetic reconstruction of the different proteins suggested a distant and unclear origin of the genes, from ´within the Betaproteobacteria.

We can still hypothesize that the epibionts acquired the CBB cycle by two independent horizontal gene transfers. Our codon usage analyses showed that the codon usage of the CBB genes was similar to the core codon usage. Genes with an unusual codon usage could be either heavily transcribed and less affected by genetic shifts or by recent HGT. Our analysis showed that all the genes related to the CBB cycle were located in the average codon usage of the genome. The adaptation of the gene codon usage to the host's average codon usage suggests that if the genes were acquired horizontally, then the transfer was probably an ancient event and the foreign codon usage of the genes has evolved to match that of the host. This indicates that the HGT was not a recent event but probably happened to a common ancestor of the two epibiotic Epsilonproteobacteria.

In the case of two separate HGT, we could hypothesize multiple scenarios as to how the two clusters came into the epsilonproteobacterial epibiont. The two HGT may have happened randomly in a shorter interval. However, one key factor that might have helped the acquisition of the pathway is that the genes present in the "phosphoribulokinase" cluster are present in additional metabolic pathways. When this cluster was transferred from the unknown Betaproteobacteria, a whole set of genes were carried over and started to be expressed. However, before genome erosion occurred to remove any useless genes, such as phosphoribulokinase in the present case, a second HGT may have happened that introduced the RuBisCO genes and the co-expression of RuBisCO genes with the phosphoribulokinase closed the CBB cycle. This allowed the epibionts to fix inorganic carbon more efficiently than with the

rTCA cycle, which was subsequently lost.

Sulfur oxidation genes

Similar to the RuBisCO genes, the close homology of the soxB amino acid sequence between the epsilonproteobacterial epibionts and deep-sea Gammaproteobacteria suggests a potential HGT event. Multiple studies have shown that the phylogeny of gammaproteobacterial and epsilonproteobacterial related soxB genes cluster separately (Meyer *et al*., 2007; Yamamoto *et al*., 2010; Hügler and Sievert, 2011). The acquisition of a homologous gammaproteobacterial version of the soxB gene and the subsequent loss of the original epsilonproteobacterial one suggest that the presence of this gene has a selective advantage, such as better enzymatic efficiency in the gills or oxygen tolerance, over the epsilonproteobacterial version in an environment such as a mussel's gill. Future studies should investigate the enzymatic differences between the two soxB groups.

Why a Calvin Cycle?

To date, all chemoautotrophic Epsilonproteobacteria have been shown to use the rTCA cycle to fix inorganic carbon (Hugler *et al.*, 2005; Campbell *et al.*, 2006). This and the residual presence of genes related to the rTCA cycle strongly indicate that the CBB cycle was subsequently acquired. The rTCA cycle requires less energy to incorporate carbon than the Calvin cycle (two versus seven molecules of ATP are required for the synthesis of pyruvate - Hügler and Sievert, 2011), so since the CBB cycle is not more energy efficient than the rTCA cycle, the epibionts must have gained another strong evolutionary advantage by acquiring the more expensive CBB cycle.

One possible reason for acquiring the CBB cycle could be the difference between the rTCA and CBB cycles' enzymatic tolerance to oxygen. The rTCA cycle relies on ferredoxin-based enzymes, which are quickly oxidized by the presence of oxygen (Ragsdale, 2003; Imlay, 2006) and, as a result, a majority of the organisms that use the rTCA cycle are anaerobes or microaerobes. Nevertheless, organisms using this cycle in full oxygen saturation have been

found in the deep sea (Berg, 2011; Meier *et al.*, 2016). Epsilonproteobacteria and Aquificales that have been found using the rTCA cycle at hydrothermal vents are usually found in abundance in warm and chemically fluctuating environments such as vent fluids and vent chimneys (Campbell *et al.*, 2006; Reveillaud *et al.*, 2015) 2015), as are *Rimicaris* shrimps (Van Dover *et al.*, 1988) and *Alvinella* worms (Desbruyères *et al.*, 1998), which colonize hydrothermal vent chimneys. On the other hand, CBB cycle enzymes are not affected by the presence of oxygen (Berg, 2011).Gammaproteobacteria with the CBB cycle have been predominantly found in organisms associated with colder and more stable environments, such as cold seeps (Vigneron *et al.*, 2014; Pop Ristova *et al.*, 2015) and areas further away from hot vents. These bacteria are also often endosymbionts of invertebrates located in the same area (Cordes *et al.*, 2007; Beinart *et al.*, 2012; Anderson *et al.*, 2013; Oshkin *et al.*, 2014), such as vesicomyid clams or bathymodiolin mussels.

Our study shows two Epsilonproteobacteria species colonizing niches usually dominated by Gammaproteobacteria and, moreover, possessing and expressing key genes of an unusual metabolic pathway that was probably obtained from Gammaproteobacteria. The genome plasticity of Epsilonproteobacteria allows constant adaptation to very variable environments and the acquisition of the CBB cycle and functional SOX genes probably enabled the niche adaptation of the epibiont to the bathymodiolin gill.

A step towards symbiosis?

From the findings of our study we hypothesize that epsilonproteobacterial epibionts may be in a commensal or mutualistic association with their bathymodiolin host. The phylogenetic position of the new Epsilonproteobacteria family (Assié *et al*., Chapter 2) to which the epibionts belong, indicates that they are equally related to a known clade of deep-sea epsilonproteobacterial mutualistic symbionts associated with various invertebrates (Van Dover *et al.*, 1988; Desbruyères *et al.*, 1998; Goffredi *et al.*, 2008) and to families of known gastrointestinal pathogens in mammals and

birds (De Groote *et al.*, 2000; Solnick and Schauer, 2001; Oxley and McKay, 2004). Additionally, early metagenomic studies of Epsilonproteobacteria showed the presence of evolutionary precursors of virulence factor, such as N-linked glycosylation, which can help pathogenic Epsilonproteobacteria invade their host's tissue (Gilbreath *et al*., 2011). It has thus been suggested that symbiotic epsilonproteobacterial ancestors evolved from pathogenic bacteria (Nakagawa *et al.*, 2007; Nakagawa and Takai, 2008). Because many molecular mechanisms are shared between mutualists and pathogens (Hentschel *et al.*, 2000), the line between being a mutualist and a pathogen is defined by the interaction between host and symbiont. Epsilonproteobacteria may have initially been pathogens that turned into mutualists.

To date, most pathogenic bacteria have been described as chemoorganotrophic or heterotrophic (Buzolyova and Somov, 1999; Dubilier *et al.*, 2008; Köpke *et al.*, 2013). Pathogens usually depend on their host for many nutrients and metabolites (Schaible and Kaufmann, 2005). Chemoautotrophic bacteria, on the other hand, are capable of independently generating energy and organic carbon. If the original state of Epsilonproteobacteria was as pathogens, the acquisition of the ability to oxidize sulfur might have been one of the factors toward evolution to a symbiotic lifestyle. A similar mechanism may have contributed to the origin of the hypothesized mutualistic association between the epibiotic Epsilonproteobacteria and bathymodiolin mussels.

Additionally, our study shows that epsilonproteobacterial epibionts shared similar metabolic pathways with gammaproteobacterial endosymbionts. Some genes had the same phylogenetic origin, strongly suggesting their acquisition via multiple HGT. It has been previously observed that HGT is a key element in the evolution and maintenance of both pathogenesis and mutualism. The acquisition of new genetic material can allow bacteria to go through an evolutionary bottleneck (Hacker and Carniel, 2001; Juhas *et al.*, 2009; Dutta and Munmun, 2015) by acquiring a set of a molecular mechanisms to invade a specific host cell (Gorkiewicz *et al.*, 2010) or genes

for protecting the host (Pinto-Carbó *et al.*, 2016). The HGT events observed in this study may have given the epsilonproteobacterial epibiont tools from a gammaproteobacterial endosymbiont that had already settled in a symbiotic association and could certainly be one of the main factors in the establishment of this symbiosis.

3.6 Conclusion and outlook

This study illustrates the extent of epsilonproteobacterial versatility to colonize diverse environments. Our metagenome analyses showed for the first time that two closely related Epsilonproteobacteria, colonizing different host species, have the genetic ability to fix inorganic carbon using the CBB cycle. Transcriptome screening of three *"B." childressi* individuals confirmed expression of the key genes involved in the CBB cycle, namely the RuBisCO large and small subunits and phosphoribulokinase. Additionally, the genes involved in the CBB cycle appear to have originated from two different HGT events. The RuBisCO genes have a gammaproteobacterial origin and have an amino acid sequence similar to the deep-sea sulfur-oxidizing group of Gammaproteobacteria related to the endosymbionts of bathymodiolin mussels. Additionally, the key gene of the SOX multi enzyme sulfur oxidation pathway, soxB, also appears to have a similar gammaproteobacterial origin. We hypothesize here that these HGT events could have enabled the Epsilonproteobacteria successfully to colonize bathymodiolin gills.

Recent studies of chemoautotrophy used functional genes screening approaches to investigate the presence of the CBB and rTCA cycles in the environment. These genes are then later used to do indirect prediction on the communities compositions (Perner *et al.*, 2007; Hügler *et al.*, 2010; Böhnke and Perner, 2014; Liu *et al.*, 2015). However the investigation of functional genes closely related to gammaproteobacterial organism such as the RuBisCO sequences we present in this study could not predict the cycle acquisition by a different bacterial class. The discovery of RuBisCO genes in Epsilonproteobacteria should be taken into account in such further studies.

More work should be undertaken correctly to evaluate the impact of this new family of genes in the scope of environmental analyses.

3.7 Acknowledgements

We would like to thank Silke Wetzel and Miriam Sadowsky for excellent technical assistance, Lizbeth Sayavedra and Juliane Wippler for advice on bioinformatics analyses and Petra Pjevack for commenting on this manuscript. This project would not have been possible without the dedicated captains and crews of the “Pourquoi pas?” and “Atlantis” research vessels involved in the sampling. This work was funded by the Max Planck Society, the DFG Cluster of Excellence “The Ocean in the Earth System” at MARUM (University of Bremen), an European Research Council Advanced Grant (BathyBiome, Grant 340535) and a Gordon and Betty Moore Foundation Marine Microbiology Initiative Investigator Award through Grant GBMF3811 to ND, and the European Union (EU) Marie Curie Actions Initial Training Network (ITN) SYMBIOMICS (contract number 264774).

3.8 References

Akerman, N.H., Butterfield, D.A., and Huber, J.A. (2013) Phylogenetic diversity and functional gene patterns of sulfur-oxidizing subseafloor Epsilonproteobacteria in diffuse hydrothermal vent fluids. *Front. Microbiol.* **4**: 185.

Anderson, *R.*E., Beltrán, M.T., Hallam, *S.*J., and Baross, J.A. (2013) Microbial community structure across fluid gradients in the Juan de Fuca Ridge hydrothermal system. *FEMS Microbiol. Ecol.* **83**: 324–339.

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: 805–813. doi:10.1111/1758-2229.12442

Bankevich, A., Nurk, *S.*, Antipov, D., Gurevich, A. a., Dvorkin, M., Kulikov, A.S., *et al*. (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **19**: 455–477.

Bayer, C., Heindl, N.R., Rinke, C., Lücker, *S.*, Ott, J.A., and Bulgheresi,S. (2009) Molecular characterization of the symbionts associated with marine nematodes of the genus *Robbea*. *Environ. Microbiol. Rep.* **1**: 136–144.

Beinart, *R.* a, Sanders, J.G., Faure, *B.*, Sylva, *S.*P., Lee, *R.*W., Becker, E.L., *et al*. (2012) Evidence for the role of endosymbionts in regional-scale habitat partitioning by hydrothermal vent symbioses. *Proc. Natl. Acad. Sci.* **109**: E3241–E3250.

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

Böhnke,S. and Perner, M. (2014) A function-based screen for seeking RuBisCO active clones from metagenomes: novel enzymes influencing RuBisCO activity. *ISME J.* **9**: 1–11.

Buzolyova, L.S. and Somov, G.P. (1999) Autotrophic assimilation of CO_2 and C1-compounds by pathogenic bacteria. *Biochem. Biokhimiia* **64**: 1146–9.

Campbell, *B.*J., Engel, A.S., Porter, M.L., and Takai, K. (2006) The versatile epsilon-proteobacteria: key players in sulphidic habitats. *Nat. Rev. Microbiol.* **4**: 458–68.

Cordes, E.E., Carney, *S.*L., Hourdez, *S.*, Carney, *R.*S., Brooks, J.M., and Fisher, C.R. (2007) Cold seeps of the deep Gulf of Mexico: Community structure and biogeographic comparisons to Atlantic equatorial belt seep communities. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **54**: 637–653.

Darriba, D., Taboada, G.L., Doallo, *R.*, and Posada, D. (2011) ProtTest-HPC: Fast Selection of Best-Fit Models of Protein Evolution. In, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*., pp. 177–184.

Desbruyères, D., Chevaldonné, P., Alayse, A.-M., Jollivet, D., Lallier, F.H., Jouin-Toulmond, C., *et al*. (1998) Biology and ecology of the "Pompeii worm" (Alvinella pompejana Desbruyères and Laubier), a normal dweller of an extreme deep-sea environment: A synthesis of current knowledge and recent developments. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **45**: 383–422.

Desbruyères, D., Segonzac, M., and Bright, M. (2006) Handbook of deep-sea hydrothermal vent fauna *Denisia*.

Van Dover, C.L., Fry, *B.*, Grassle, J.F., Humphris, *S.*, and Rona, P.A. (1988) Feeding biology of the shrimp *Rimicaris exoculata* at hydrothermal vents on the Mid-Atlantic Ridge. *Mar. Biol.* **98**: 209–216.

Dubilier, N., Bergin, C., and Lott, C. (2008) Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat. Rev. Microbiol.* **6**: 725–40.

Duperron, *S.*, Lorion, J., Samadi, *S.*, Gros, O., and Gaill, F. (2009) Symbioses between deep-sea mussels (Mytilidae: Bathymodiolinae) and chemosynthetic bacteria: diversity, function and evolution. *C. R. Biol.* **332**: 298–310.

Dutta, C. and Munmun,S. (2015) Horizontal Gene Transfer and Bacterial Diversity. In, Nelson,K.E. (ed), *Genes, Genomes and Metagenomes: Basics, Methods, Databases and Tools*. Springer US, Boston, MA, pp. 1–23.

Edgar, *R.*C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acid Res.* **32**: 1792–1797.

Galkin,S. V. (2016) Structure of Hydrothermal Vent Communities. In, Galkin,S. V. and Demina,L.L. (eds), *Trace Metal Biogeochemistry and Ecology of Deep-Sea Hydrothermal Vent Systems*. Springer International Publishing Switzerland, pp. 41–53.

Gilbreath, J.J., Cody, W.L., Merrell, D.S., and Hendrixson, D.R. (2011) Change is good: variations in common biological mechanisms in the epsilonproteobacterial genera *Campylobacter* and *Helicobacter*. *Microbiol. Mol. Biol. Rev.* **75**: 84–132.

Goffredi, *S.*K., Jones, W.J., Erhlich, H., Springer, A., and Vrijenhoek, *R.*C. (2008) Epibiotic bacteria associated with the recently discovered Yeti crab, *Kiwa hirsuta*. *Environ. Microbiol.* **10**: 2623–34.

Gorkiewicz, G., Kienesberger, *S.*, Schober, C., Scheicher, *S.*R., Gully, C., Zechner, *R.*, and Zechner, E.L. (2010) A Genomic Island Defines Subspecies-Specific Virulence Features of the Host-Adapted Pathogen *Campylobacter fetus* subsp. *venerealis*. *J. Bacteriol.* **192**: 502–517.

De Groote, D., Ducatelle, *R.*, and Haesebrouck, F. (2000) *Helicobacter*'s of possible zoonotic origin: a review. *Acta Gastroenterol. Belg.* **63**: 380–387.

Grzymski, J.J., Murray, A.E., Campbell, *B.*J., Kaplarevic, M., Gao, G.R., Lee, C., *et al*. (2008) Metagenome analysis of an extreme microbial symbiosis reveals eurythermal adaptation and metabolic flexibility. *Proc. Natl. Acad. Sci. U.*S. *A.* **105**: 17516–21.

Guindon, *S.*, Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* **59**: 307–321.

Hacker, J. and Carniel, E. (2001) Ecological fitness, genomic islands and bacterial pathogenicity: A Darwinian view of the evolution of microbes. *EMBO Rep.* **2**: 376–381.

Hentschel, U., Steinert, M., and Hacker, J. (2000) Common molecular mechanisms of symbiosis and pathogenesis. *Trends Microbiol.* **8**: 226–31.

Huber, J.A., Welch, D.M., Morrison, H.G., Huse, *S.*M., Neal, P.R., Butterfield, D.A., and Sogin, M.L. (2007) Microbial population structures in the deep marine biosphere. *Science (80-.).* **318**: 97–100.

Huelsenbeck, J.P., Ronquist, F., Nielsen, *R.*, and Bollback, J.P. (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294**: 2310–2314.

Hug, L.A., Stechmann, A., and Roger, A.J. (2010) Phylogenetic Distributions and Histories of Proteins Involved in Anaerobic Pyruvate Metabolism in Eukaryotes. *Mol. Biol. Evol.* **27**: 311–324.

Hughes, N.J., Clayton, C.L., Chalk, P.A., and Kelly, D.J. (1998) *Helicobacter pylori* porCDAB and oorDABC genes encode distinct pyruvate:flavodoxin and 2-oxoglutarate:acceptor oxidoreductases which mediate electron transport to NADP. *J. Bacteriol.* **180**: 1119–28.

Hügler, M., Gärtner, A., and Imhoff, J.F. (2010) Functional genes as markers for sulfur cycling and CO_2 fixation in microbial communities of hydrothermal vents of the Logatchev field. *FEMS Microbiol. Ecol.* **73**: 526–37.

Hügler, M. and Sievert, *S.*M. (2011) Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean. *Ann. Rev. Mar. Sci.* **3**: 261–289.

Hugler, M., Wirsen, C.O., Fuchs, G., Taylor, C.D., and Sievert, *S.*M. (2005) Evidence for Autotrophic CO_2 Fixation via the Reductive Tricarboxylic Acid Cycle by Members of the Subdivision of Epsilonproteobacteria. *J. Bacteriol.* **187**: 3020–3027.

Imlay, J.A. (2006) Iron- sulfur clusters and the problem with oxygen. *Mol. Microbiol.* **59**: 1073–1082.

Jan, C., Petersen, J.M., Werner, J., Teeling, H., Huang, *S.*, Glöckner, F.O., *et al*. (2014) The gill chamber epibiosis of deep-sea shrimp *Rimicaris exoculata* : an in-depth metagenomic investigation and discovery of Zetaproteobacteria. *Environ. Microbiol.* **16**: 2723–2738.

Jones, W.J.J. and Vrijenhoek, *R.*C. (2006) Evolutionary relationships within the "*Bathymodiolus*" *childressi* group. *Cah. Biol. Mar.* **47**: 403–407.

Juhas, M., Crook, D.W., and Hood, D.W. (2008) Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. *Cell. Microbiol.* **10**: 2377–2386.

Juhas, M., van der Meer, J.R., Gaillard, M., Harding, *R.*M., Hood, D.W., and Crook, D.W. (2009) Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* **33**: 376–393.

Köpke, M., Straub, M., and Dürre, P. (2013) Clostridium difficile Is an Autotrophic Bacterial Pathogen. *PLoS One* **8**: e62157.

Kuwahara, H., Yoshida, T., Takaki, Y., Shimamura, *S.*, Nishi, *S.*, Harada, M., *et al*. (2007) Reduced Genome of the Thioautotrophic Intracellular Symbiont in a Deep-Sea Clam, *Calyptogena okutanii*. *Curr. Biol.* **17**: 881–6.

Li, D., Liu, C.M., Luo, *R.*, Sadakane, K., and Lam, T.W. (2014) MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**: 1674–1676.

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

Liu, J.-F., Sun, X.-B., Yang, G.-C., Mbadinga, *S.*M., Gu, J.-D., and Mu, *B.*-Z.B. (2015) Analysis of Microbial Communities in the Oil Reservoir Subjected to CO_2-Flooding by Using Functional Genes as Molecular Biomarkers for Microbial CO_2 Sequestration. *Front. Microbiol.* **6**: 1–15.

Markowitz, V.M., Chen, I.M.A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., *et al*. (2012) IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* **40**: 115–122.

Meier, D., Bach, W., Girguis, P.R., Gruber-Vodicka, H., Reeves, E.P., Richter, M., *et al*. (2016) Heterotrophic Proteobacteri a in the vicinity of diffuse

hydrothermal venting. *Environ. Microbiol.* n/a–n/a.

Mell, J.C. and Redfield, *R.*J. (2014) Natural Competence and the Evolution of DNA Uptake Specificity. *J. Bacteriol.* **196**: 1471–1483.

Meyer, *B.*, Imhoff, J.F., and Kuever, J. (2007) Molecular analysis of the distribution and phylogeny of the soxB gene among sulfur-oxidizing bacteria - evolution of the Sox sulfur oxidation enzyme system. *Environ. Microbiol.* **9**: 2957–77.

Meyer, F., Paarmann, D., D'Souza, M., Olson, *R.*, Glass, E.M., Kubal, M., *et al*. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.

Nakagawa,S. and Takai, K. (2008) Deep-sea vent chemoautotrophs: Diversity, biochemistry and ecological significance. *FEMS Microbiol. Ecol.* **65**: 1–14.

Nakagawa, *S.*, Takaki, Y., Shimamura, *S.*, Reysenbach, A.-L., Takai, K., and Horikoshi, K. (2007) Deep-sea vent epsilonproteobacterial genomes provide insights into emergence of pathogens. *Proc. Natl. Acad. Sci.* **104**: 12146–12150.

Nakagawa, *S.*, Takai, K., Inagaki, F., Hirayama, H., Nunoura, T., Horikoshi, K., and Sako, Y. (2005) Distribution, phylogenetic diversity and physiological characteristics of E-proteobacteria in a deep-sea hydrothermal field. *Environ. Microbiol.* **7**: 1619–1632.

Newton, I.L.G., Woyke, T., Auchtung, T.A., Dilly, G.F., Dutton, *R.*J., Fisher, M.C., *et al*. (2007) The Calyptogena magnifica Chemoautotrophic Symbiont Genome. *Science (80-.).* **315**: 998–1000.

Orcutt, *B.*N., Sylvan, J.B., Knab, N.J., and Edwards, K.J. (2011) Microbial Ecology of the Dark Ocean above, at, and below the Seafloor. *Microbiol. Mol. Biol. Rev.* **75**: 361–422.

Oshkin, I.Y., Wegner, C., Lüke, C., Glagolev, M. V, Filippov, I. V, Pimenov, N. V, and Liesack, W. (2014) Gammaproteobacterial Methanotrophs Dominate Cold Methane Seeps in Floodplains of West Siberian Rivers. *Appl. Environ. Microbiol.* **80**: 5944–5954.

Oshlack, A., Robinson, M.D., and Young, M.D. (2010) From RNA-seq reads to differential expression results. *Genome Biol* **11**: 220.

Oxley, A.P. a and McKay, D.B. (2004) Fecal shedding of *Helicobacter* spp. by co-housed Australian sea lions (*Neophoca cinerea*) and Australian fur seals (*Arctocephalus pusillus doriferus*). *Vet. Microbiol.* **101**: 235–43.

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 Res.* **25**: 1043–1055.

Peden, J.F. (2011) Analysis of codon usage. *Biosystems.* **106**: 45–50.

Perner, M., Seifert, *R.*, Weber, *S.*, Koschinsky, A., Schmidt, K., Strauss, H., *et al*. (2007) Microbial CO_2 fixation and sulfur cycling associated with low-temperature emissions at the Lilliput hydrothermal field, southern Mid-Atlantic Ridge (9°S). *Environ. Microbiol.* **9**: 1186–1201.

Petersen, J.M., Ramette, A., Lott, C., Cambon-Bonavita, M.-A., Zbinden, M., and Dubilier, N. (2010) Dual symbiosis of the vent shrimp *Rimicaris exoculata* with filamentous gamma- and Epsilonproteobacteria at four Mid-Atlantic Ridge hydrothermal vent fields. *Environ. Microbiol.* **12**: 2204–2218.

Pinto-Carbó, M., Sieber, *S.*, Dessein, *S.*, Wicker, T., Verstraete, *B.*, Gademann, K., *et al*. (2016) Evidence of horizontal gene transfer between obligate leaf nodule symbionts. *ISME J.* **in press**: 1–14.

Pop Ristova, P., Wenzhöfer, F., Ramette, A., Felden, J., and Boetius, A. (2015) Spatial scales of bacterial community diversity at cold seeps (Eastern Mediterranean Sea). *ISME J.* **9**: 1306–1318.

Ragsdale, *S.*W. (2003) Pyruvate ferredoxin oxidoreductase and its radical intermediate. *Chem. Rev.* **103**: 2333–2346.

Raven, J. (2009) Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments. *Aquat. Microb. Ecol.* **56**: 177–192.

Reeves, E.P., Yoshinaga, M.Y., Pjevac, P., Goldenstein, N.I., Peplies, J., Meyerdierks, A., *et al*. (2014) Microbial lipids reveal carbon assimilation patterns on hydrothermal sulfide chimneys. *Environ. Microbiol.* **16**: 3515–3532.

Reveillaud, J., Reddington, E., McDermott, J., Algar, C., Meyer, J.L., Sylva, *S.*, *et al*. (2015) Subseafloor microbial communities in hydrogen-rich vent fluids from hydrothermal systems along the Mid-Cayman Rise. *Environ. Microbiol.*

Rinke, C., Schmitz-esser, *S.*, Loy, A., Horn, M., Wagner, M., and Bright, M. (2009) High genetic similarity between two geographically distinct strains of the sulfur-oxidizing symbiont "*Candidatus* Thiobios zoothamnicoli". *FEMS Microbiol. Ecol.* **67**: 229–41.

Rodriguez-R, L.M. and Konstantinidis, K.T. (2016) The enveomics collection : a toolbox for specialized analyses of microbial genomes and metagenomes. *Peer J Prepr.* **4**: e1900v1.

Rodriguez-R, L.M. and Konstantinidis, K.T. (2014) Bypassing cultivation to identify bacterial species. *Microbe* **9**: 111–118.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D., *et al*., (2012) MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst Biol* 61 (**3**): 539-542

Schaible, U.E. and Kaufmann, *S.*H.E. (2005) A nutritive view on the host–pathogen interplay. *Trends Microbiol.* **13**: 373–380.

Seah, *B.*K.B. and Gruber-Vodicka, H.R. (2015) gbtools: Interactive Visualization of Metagenome Bins in *R. Front. Microbiol.* **6**: 1451.

Sievert,S. and Vetriani, C. (2012) Chemoautotrophy at Deep-Sea Vents: Past, Present, and Future. *Oceanography* **25**: 218–233.

Sievert, *S.*M., Scott, K.M., Klotz, M.G., Chain, P.S.G., Hauser, L.J., Hemp, J., *et al*. (2008) Genome of the epsilonproteobacterial Chemolithoautotroph *Sulfurimonas denitrificans*. *Appl. Environ. Microbiol.* **74**: 1145–1156.

Solnick, J. V and Schauer, D.B. (2001) Emergence of diverse *Helicobacter* species in the pathogenesis of gastric and enterohepatic diseases. *Clin. Microbiol. Rev.* **14**: 59–97.

Strous, M., Kraft, *B.*, Bisdorf, *R.*, and Tegetmeyer, H.E. (2012) The binning of metagenomic contigs for microbial physiology of mixed cultures. *Front. Microbiol.* **3**: 1–11.

Swan, *B.*K., Martinez-Garcia, M., Preston, C.M., Sczyrba, A., Woyke, T., Lamy, D., *et al*. (2011) Potential for Chemolithoautotrophy. *Science (80-.).* **333**: 1296–9.

Thubaut, J., Puillandre, N., Faure, *B.*, Cruaud, C., and Samadi,S. (2013) The contrasted evolutionary fates of deep-sea chemosynthetic mussels (Bivalvia, Bathymodiolinae). *Ecol. Evol.* **3**: 4748–4766.

Turner, J.T. (2002) Zooplankton fecal pellets, marine snow and sinking phytoplankton blooms. *Aquat. Microb. Ecol.* **27**: 57–102.

Vigneron, A., Cruaud, P., Pignet, P., Caprais, J.-C., Gayet, N., Cambon-Bonavita, M.-A., *et al*. (2014) Bacterial communities and syntrophic associations involved in anaerobic oxidation of methane process of the Sonora Margin cold seeps, Guaymas Basin. *Environ. Microbiol.* **16**: 2777–2790.

Walsh, D.A., Zaikova, E., Howes, C.G., Song, Y.C., Wright, J.J., Tringe, *S.*G., *et al*. (2009) Metagenome of a Versatile Chemolithoautotroph from Expanding Oceanic Dead Zones. *Science (80-.).* **326**: 578–582.

Wang, Z. and Wu, M. (2013) A Phylum-Level Bacterial Phylogenetic Marker Database. *Mol. Biol. Evol.* **30**: 1258–1262.

Yamamoto, M., Nakagawa, *S.*, Shimamura, *S.*, Takai, K., and Horikoshi, K. (2010) Molecular characterization of inorganic sulfur-compound metabolism in the deep-sea epsilonproteobacterium *Sulfurovum* sp. NBC37-1. *Environ. Microbiol.* **12**: 1144–1153.

Yamamoto, M. and Takai, K. (2011) Sulfur metabolisms in epsilon- and gamma-proteobacteria in deep-sea hydrothermal fields. *Front. Microbiol.* **2**: 192.

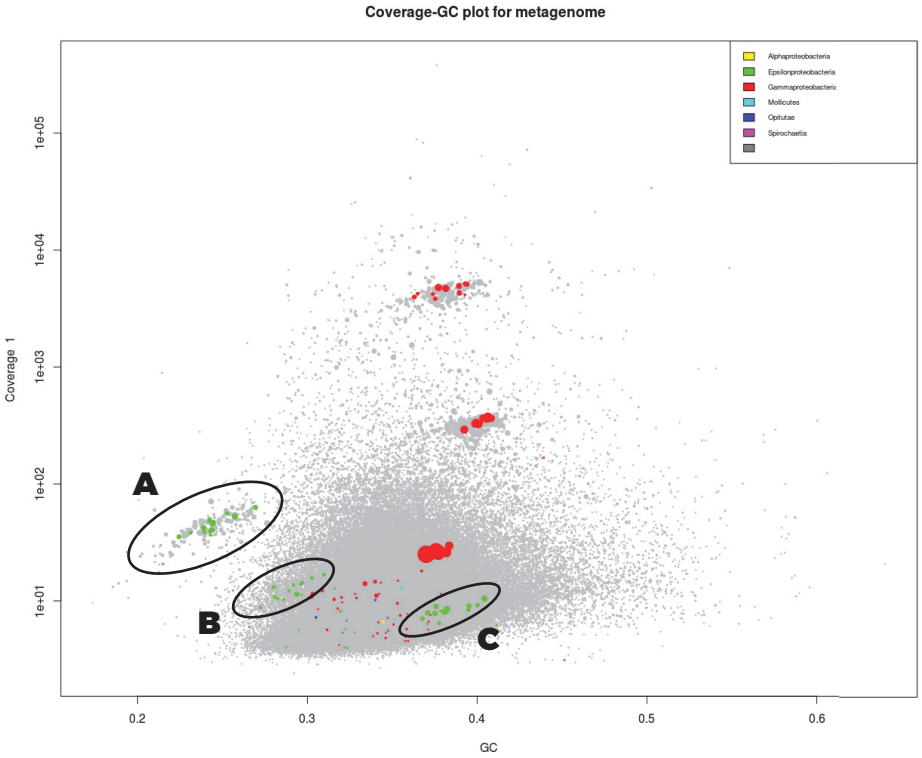
Zhou, J., Bruns, M.A., and Tiedje, J.M. (1996) DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**: 316–322.

Zwirglmaier, K., Reid, W.D.K., Heywood, J., Sweeting, C.J., Wigham, *B.*D., Polunin, N.V.C., *et al*. (2015) Linking regional variation of epibiotic bacterial diversity and trophic ecology in a new species of Kiwaidae (Decapoda, Anomura) from East Scotia Ridge (Antarctica) hydrothermal vents. *Microbiologyopen* **4**: 136–150.

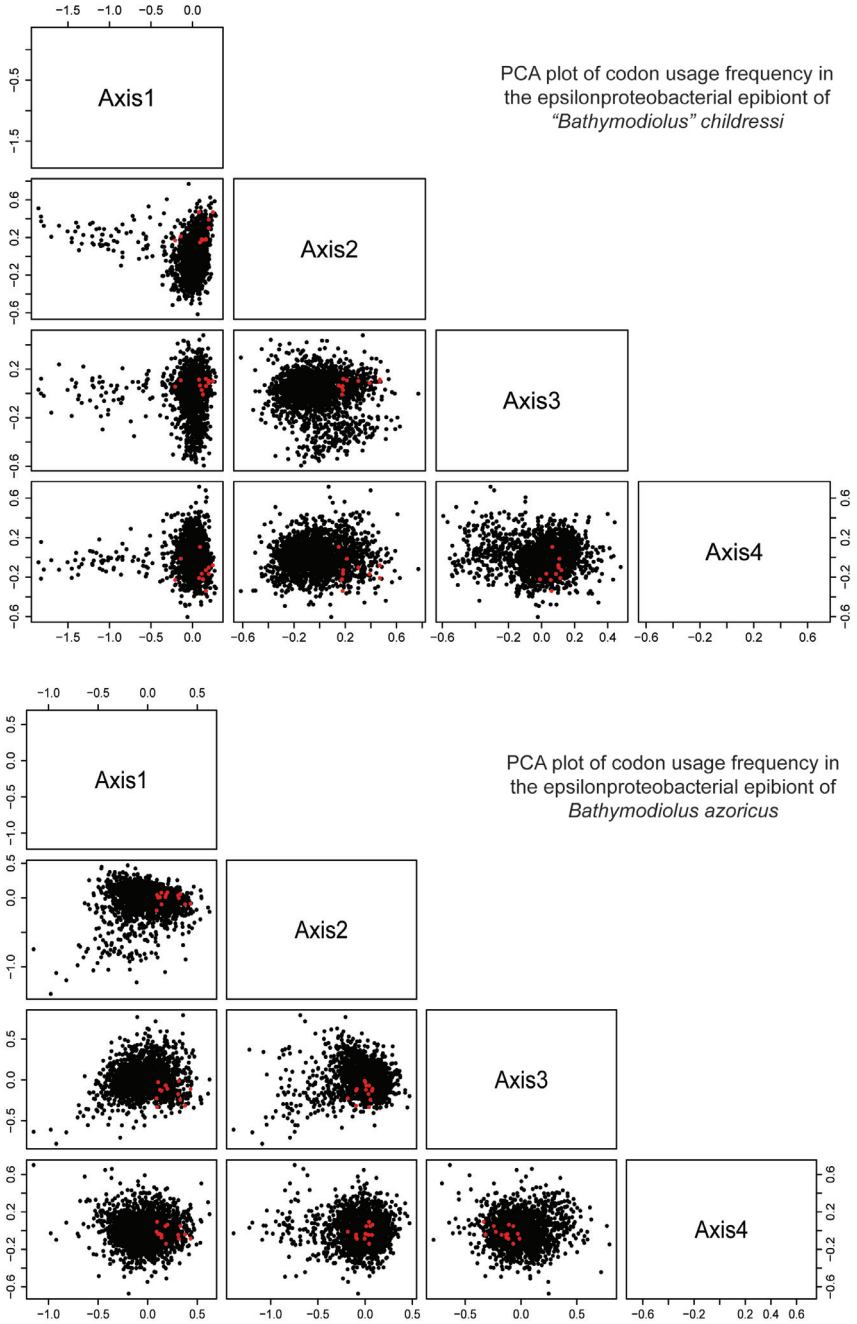
3.9 Supplementary figures and tables

Supplementary Figure 3.4 Bayesian inference phylogenetic reconstruction of the amino acid sequences of transketolase genes. The maximum likelihood tree is available as Digital Supplement 11

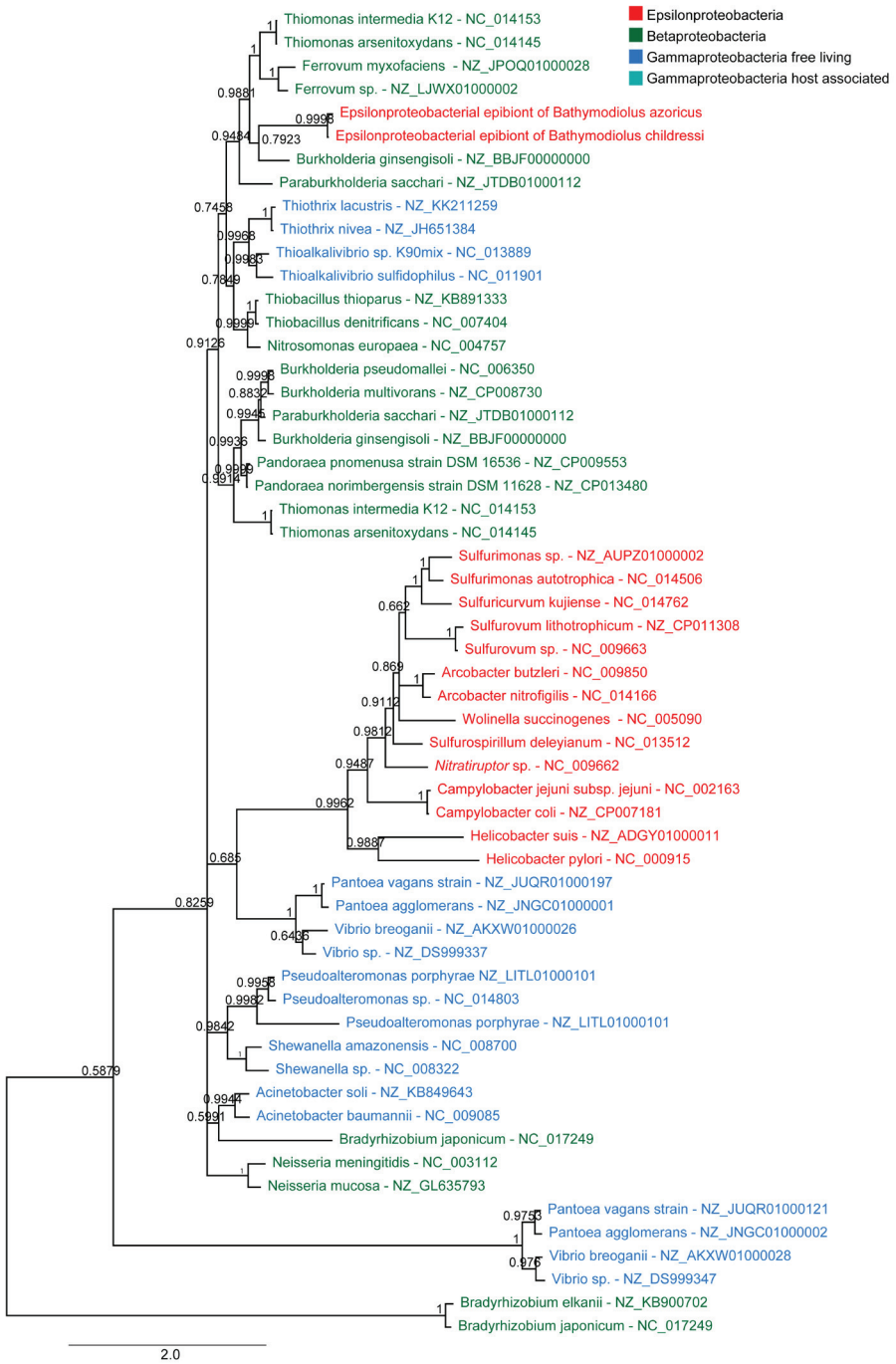
Supplementary Figure 3.5 Bayesian inference phylogenetic reconstruction of the amino acid sequences of phosphoglycolate phosphatase genes. The maximum



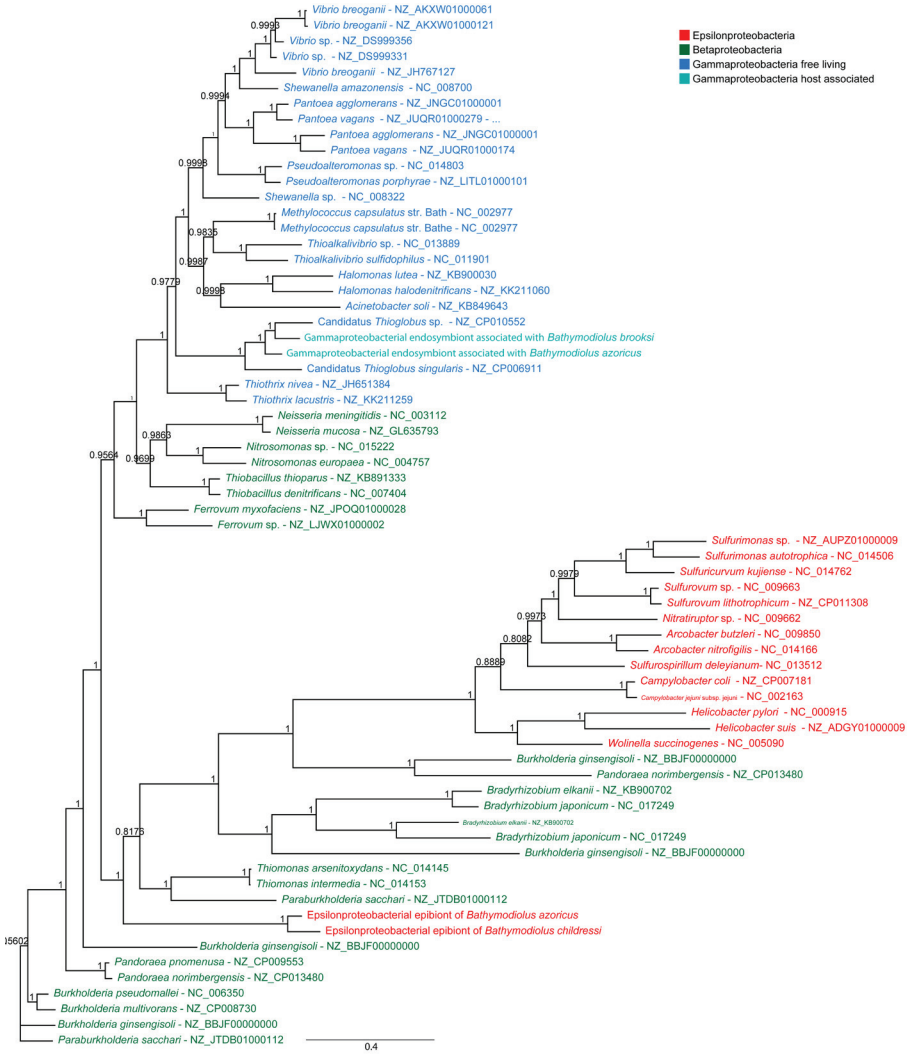
Supplementary Figure 3.1 Plot displaying contig from the initial assembly of the *B. azoricus*e metagenomic libraries. Each dot represent a contig displayed according to his G/C% and Coverage values. In green are the contigs which host a epsilonproteobacterial marker gene, in red gammaproteobacterial. **A.** Initial EpsA bin, *B.* Initial BinB bin, **C.** Initial BinC bin.



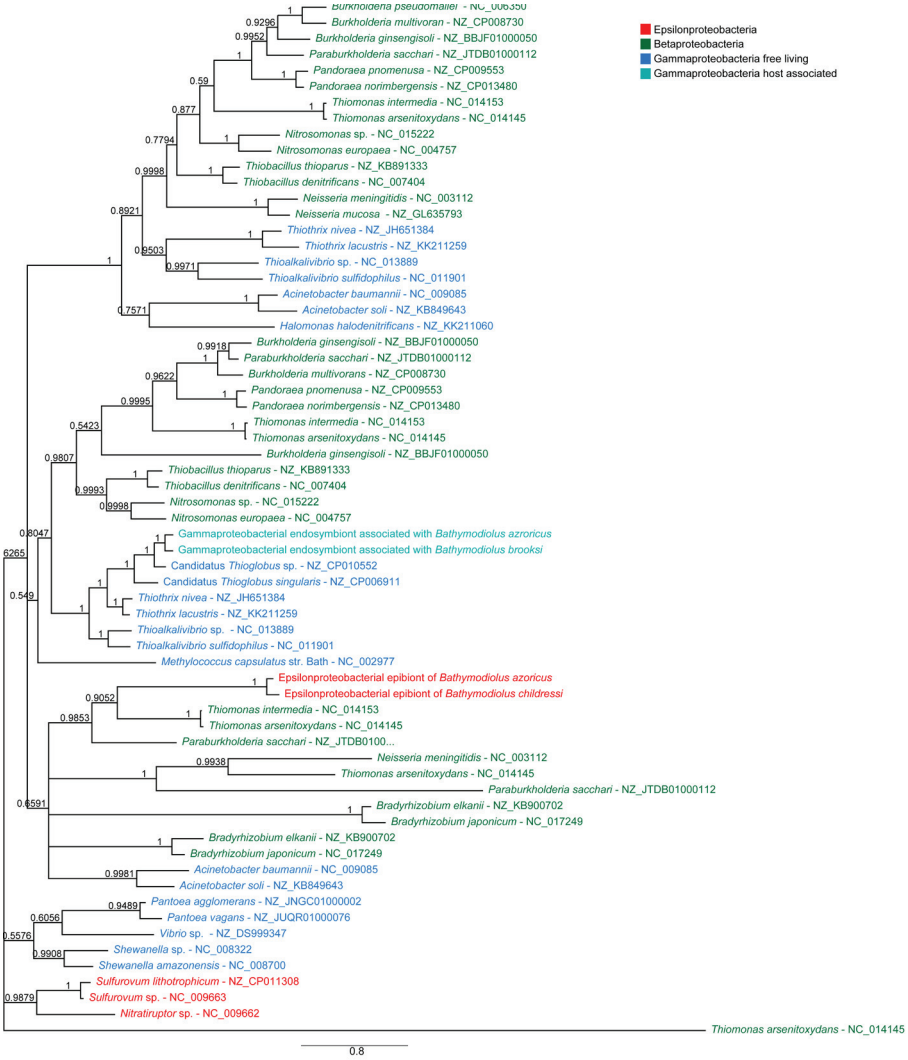
Supplementary Figure 3.2 Plot of the different PCA values calculated for the codon usage of each EpsA and EpsC genes. In red are displayed the CBB and *soxB* genes.



Supplementary Figure 3.3 Bayesian inference phylogenetic reconstruction of the amino acid sequences of the fructose bisphosphatase genes. The maximum likelihood tree is available as Digital Supplement 10



likelihood tree is available as Digital Supplement 12



Supplementary Figure 3.6 Bayesian inference phylogenetic reconstruction of the amino acid sequences of fructose 1,6-bisphosphate aldolase genes. The maximum likelihood tree is available as Digital Supplement 13


Epsilonproteobacteria
Betaproteobacteria
Gammaproteobacteria free living
Gammaproteobacteria host associated
Neisseria meningitidis - NC_003112
Neisseria mucosa - NZ_GL635799
Acinetobacter baumannii - NC_009085
Acinetobacter soli - NZ_KB849643
Shewanella sp. - NC_008322
Shewanella amazonensis - NC_008700
Pseudoalteromonas sp. - NC_014803
Pseudoalteromonas porphyrae - NZ_LITL01000101
Halomonas halodenitrificans - NZ_KK211060
Halomonas lutea - NZ_KB900030
Thiomonas arsenitoxydans - NC_014145
Thiomonas intermedia - NC_014153
Methylococcus capsulatus str. Bath - NC_002977
Methylococcus capsulatus str. Bath - NC_002977
Thiobacillus denitrificans ATCC 25259 - NC_007404
Thiobacillus thioparus - NZ_KB891333
Thiothrix nivea - NZ_JH651384
Thiothrix lacustris - NZ_KK211259
Gammaproteobacterial endosymbiont associated with Bathymodiolus brooksi
Gammaproteobacterial endosymbiont associated with Bathymodiolus septemdierum
Candidatus Thioglobus sp. - NZ_CP010552
Thioalkalivibrio sulfidophilus - NC_011901
Thioalkalivibrio sp. - NC_013889
Burkholderia pseudomallei - NC_006350
Burkholderia multivorans - NZ_CP008730
Burkholderia ginsengisoli- NZ_BBJF01000050
Paraburkholderia sacchari - NZ_JTDB01000112
Pandoraea pnomenusa - NZ_CP009553
Pandoraea norimbergensis - NZ_CP013480
Nitrosomonas europaea - NC_004757
Nitrosomonas sp - NC_015222
Epsilonproteobacterial epibiont of Bathymodiolus azoricus
Epsilonproteobacterial epibiont of Bathymodiolus childressi
Ferrovum myxofaciens - NZ_JPOQ01000028
Ferrovum sp. - NZ_LJWX01000002
Paraburkholderia sacchari - NZ_JTDB01000112
Burkholderia ginsengisoli - NZ_BBJF01000050
Helicobacter pylori - NC_000915
Wolinella succinogenes - NC_005090
Helicobacter suis - NZ_ADGY01000059
Bradyrhizobium japonicum - NC_017249
Bradyrhizobium elkanii - NZ_KB900702
Bradyrhizobium elkanii - NZ_KB900702
Campylobacter jejuni subsp. jejuni - NC_002163
Campylobacter coli RM4661 - NZ_CP007181
Nitratiruptor sp. - NC_009662
Sulfurovum lithotrophicum - NZ_CP011308
Sulfurovum sp. - NC_009663
Sulfurospirillum deleyianum - NC_013512
Sulfurimonas sp. - NZ_AUPZ01000004
Sulfurimonas autotrophica - NC_014506
Sulfuricurvum kujiense - NC_014762
Arcobacter butzleri - NC_009850
Arcobacter nitrofigilis - NC_014166
Pantoea agglomerans - NZ_JNGC01000001
Pantoea vagans - NZ_JUQR01000279
Vibrio sp. - NZ_DS999356
Vibrio breoganii - NZ_AKXW01000061
Pseudoalteromonas porphyrae - NZ_LITL01000101
Burkholderia ginsengisoli - NZ_BBJF01000050
Ferrovum myxofaciens - NZ_JPOQ01000128
Bradyrhizobium japonicum - NC_017249
Bradyrhizobium elkanii - NZ_KB900702
Thiobacillus denitrificans - NC_007404
Thiobacillus thioparus - NZ_KB891333
Thioalkalivibrio sp. - NC_013889
Candidatus Thioglobus singularis PS1 - NZ_CP006911
Paraburkholderia sacchari - NZ_JTDB01000112
Nitrosomonas sp. - NC_015222
Pantoea vagans - NZ_JUQR01000311
Halomonas lutea - NZ_KB900030
Thiomonas arsenitoxydans - NC_014145
Vibrio sp. - NZ_DS999327
1.0

Chapter 4 **Same but different : Genomes of epsilonproteobacterial epibionts associated with bathymodiolin mussels**

Authors: Adrien Assie[1], Harald Gruber-Vodicka[1], Samantha Joye[2], Matthew Saxton[2], Halina Tegetmeyer[1], Nicole Dubilier[1,4,5], Jillian M. Petersen[1,3]

1 Symbiosis Department, Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, D-28359 Bremen, Germany.

2 Department of Marine Sciences, The University of Georgia, Room 159, Marine Sciences Bldg. Athens, GA 30602-3636

3 Division of Microbial Ecology, Department of Microbiology and Ecosystem Science, Research Network Chemistry Meets Microbiology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria.

4 Faculty of Biology / Chemistry, University of Bremen, Bibliothekstrasse 1, D-28359 Bremen, Germany

5 Marum - Zentrum für Marine Umweltwissenschaften der Universität Bremen, Leobener Str. 2, 28359 Bremen

Manuscript in preparation

4.1 Abstract

Bathymodiolin mussels occur worldwide in reducing environments such as cold seeps and hydrothermal vents. They have evolved a mutualistic relationship with intracellular Gammaproteobacteria capable of chemolithoautotrophy. The bacterial symbionts have settled in a unique niche that gives them access to reduced chemicals such as hydrogen, hydrogen sulfide and methane as energy and carbon sources. In exchange, they provide their mussel host with nutrients. Recently, we described a novel family of Epsilonproteobacteria living as epibionts in a widespread association with bathymodiolin mussels. A specific and widespread association between the bacteria and mussels was shown as well as their chemoautotrophic metabolism. Analyzing the metagenome and metatranscriptome of two bathymodiolin species, *“B.” childressi* and *B.azoricus,* the aim of this study was to investigate the complete genetic potential of the epsilonproteobacterial epibiont associated with these hosts. We reconstructed the complete metabolic potential of the epibiont and investigated the potential adhesion and interaction mechanisms used to settle on bathymodiolin gill filaments. Our analysis showed that, contrary as initially thought to be two different strains of the same species based on 16S rRNA sequences phylogeny, the two epsilonproteobacterial genomes were significantly different and the two epibiotic organisms are actually two distinct species. We finally compared the metagenomic data to other available Epsilonproteobacteria to assess the phylogenetic position of the epibiont family within other Epsilonproteobacteria.

4.2 Introduction

Bathymodiolin mussels are among the dominant invertebrate species present in deep sea reduced environments, such as hydrothermal vents and cold seeps around the world (Duperron *et al.*, 2009). The mussels have established associations with chemolitoautotrophic bacteria in their gill epithelial cells. Water flow across the mussel's gill filaments provides the symbiotic bacteria with reduced chemicals such as methane and hydrogen sulfide, which serve as a source of energy and carbon. In exchange, the bacteria support the growth and maintenance of the host's biomass (Nelson *et al.*, 1995; DeChaine and Cavanaugh, 2006; Petersen and Dubilier, 2009). Most of the bathymodiolin mussels have been described as being associated with a methane- or sulfur-oxidizing gammaproteobacterial symbiont, and some species even host both phylotypes in a dual symbiosis (Reviewed in Dubilier *et al*., 2008).

We recently described the widespread association of a novel family of Epsilonproteobacteria living as gill epibionts in bathymodiolin mussels (Assié *et al*., in review – Chapter 2). We additionally investigated the central metabolism of two closely related epibiont strains, each specifically associated with bathymodiolin mussels from different genera: *Bathymodiolus azoricus* and *"Bathymodiolus" childressi*. We showed that the epibionts were chemoautrophic, sulfur oxidizing bacteria and suggested that the epibionts were likely beneficial to the host (Chapter 3). From our initial analyses we observed a noteworthy difference in epsilonproteobacterial epibiont abundance between *B. azoricus* and the *"B." childressi* episymbioses. In *"B." childressi,* the epibionts form a dense coat on the host gill filaments whereas in *B. azoricus,* Epilonproteobacteria colonization is sparser.

Whereas our previous work investigated the foreign origin of key metabolic pathways that could explain the epibionts' ability to colonize bathymodiolin gill tissues, the present study aimed to explore further our metagenomic and metatranscriptomic datasets and compare the epsilonproteobacterial epibionts associated with *B. azoricus* and

"B." childressi. We focused on the phylogenetic position of the new Epsilonproteobacteria family using multi-gene phylogeny. We also reconstructed an overview of the whole bacterial metabolisms then closely examined the differences between the two draft genomes. Finally, we put this information in perspective to other epsilonproteobacterial families.

4.3 Material and Methods

Sample collection, DNA and RNA extraction and metatranscriptome and metagenome sequencing

Analysis of the *"B." childressi* and *B. azoricus* metagenomic samples used in this study was done as described in Chapter 3. The protocols used for DNA sequencing, assembling and annotating these samples have been described in detail in Chapter 3. One library was built from a *"B." childressi* individual from a cold seep in the Gulf of Mexico and is referred to in the text as EpsC. The second library, built from a *B. azoricus* individual, referred to in the text as EpsA, was sampled from the Lucky Strike hydrothermal vent field on the north mid-Atlantic ridge (NMAR). Annotations were performed with the IMG and RAST online annotation platforms (Meyer *et al.*, 2008; Markowitz *et al.*, 2012). Average nucleotide and amino acid identities (ANI and AAI) were calculated using the "Kostas lab" web service: http://enve-omics.ce.gatech.edu/ with default parameters.

Similarly, the treatment of the three *"B." childressi* samples for transcriptome sequencing and the gene expression analyses have also been described in Chapter 3. Briefly, to compare the transcriptomes, we mapped the three transcriptome reads to the EpsC draft genome. We then counted the number of reads mapping to annotated genes for each individual, and a normalization factor per library was estimated with trimmed mean of M-values (TMM) (calcNormFactors, edgeR package) (Oshlack *et al.*, 2010). The counts were then converted to reads per kilobase per million (RPKM) (Rsubread package).

Phylogenomic reconstruction

Phylogenomic trees were calculated using Phylogenomics-tools (Brandon Seah - 2014 Phylogenomics-tools https://github.com/kbseah/phylogenomics-tools). The two epsilonproteobacterial epibiont draft genomes EpsC and EpsA were compared to 79 epsilonproteobacterial draft genome representatives of the phylogenetic class taxonomy. In brief, epsilonproteobacterial marker genes were screened using PHYLAmphora software (Wang and Wu, 2013). Genes present in one copy in every draft genome were selected for the phylogenomic reconstruction. Each gene set was aligned using MUSCLE.

For maximum-likelihood analysis, RAxML (Stamatakis, 2014) was run ten times using the generalized time-reversible model with the best substitution matrix for each gene and then a SH-test was performed for each gene's best tree and constraint tree. This phylogeny was rooted using *Nitratriruptor* sp. as an outgroup.

4.4 Results and Discussion

Phylogenomic tree reconstruction

We calculated phylogenomic trees using the epsilonproteobacterial unique marker genes set of the Phyla Amphora package (Wang and Wu, 2013). Sixty-three genes were used to calculate a RAxML tree with representative genomes of the publicly available Epsilonproteobacteria genomes. A total of 81 genomes were used to calculate the tree (Figure 4.1). The epsilonproteobacterial epibionts were located on a deep long branch clustering with the Arcobacter clade. These results are different from the 16S rRNA phylogeny previously published (Chapter 2), in which the epibionts belong to a deep branching sister group/family of the *Sulfurovum* clade.

However, the two datasets agree on the divergence of the *Bathymodiolus* associated Epsilonproteobacteria from the other families composing the class. The AAI values between EpsC and the closest relatives predicted by this phylogenomic tree (Arcobacter butzleri) and the previously published

16S rRNA (*Sulfurovum lithotrophicum*) are 46.4% and 43.2%, respectively (Supplementary Figure 4.1). According to the guideline laid out in Rodriguez-R and Konstantinidis (2014), the epibionts belong to a different taxonomic division above the genus level. This data supports the observation of our previous work that the epibionts form a novel family of Epsilonproteobacteria.

The incongruence between the 16S rRNA and Phyla Amphora marker genes phylogenies can be explained by the low taxonomic representation of Epsilonproteobacteria in genome databases and large AAI values of divergence between the epibionts and closely related clades, resulting in the long branch formation. The Epsilonproteobacteria phylogeny is highly debated and in need of revision (Campbell *et al*., 2006; Ménard *et al*., 2016). Thus the exact placement of the epsilonproteobacterial epibiont novel family cannot be settled exactly.

Our previous work (Assié *et al.*, 2016) showed that the EpsA and EpsC 16S rRNA sequences had 1 out of 1109 bp difference. Based on the most recent guidelines on 16S rRNA classification (Yarza *et al*., 2014) these two sequences were considered to be the same species. However, our current genome comparison shows that EpsA and EpsC draft genomes have an ANI value of 83.1% (Figure 4.2), indicating that the two genomes belong to two different species (Rodriguez-R and Konstantinidis, 2014). Whole genome sequence comparison clearly showed that, contrary to the 16S rRNA suggestion, EpsA and EpsC were not different strains but are two distinct species belonging to the same genus.

Metabolism

Chemoautotrophy

We previously showed that epsilonproteobacterial epibionts are chemoautotrophic bacteria. We showed (Chapter 3) the unusual presence of a Calvin Benson Bassham (CBB) cycle within Epsilonproteobacteria. We also discussed at length that these epibionts might be commensal or mutualistic in their relationship with their bathymodiolin host. Finally, we

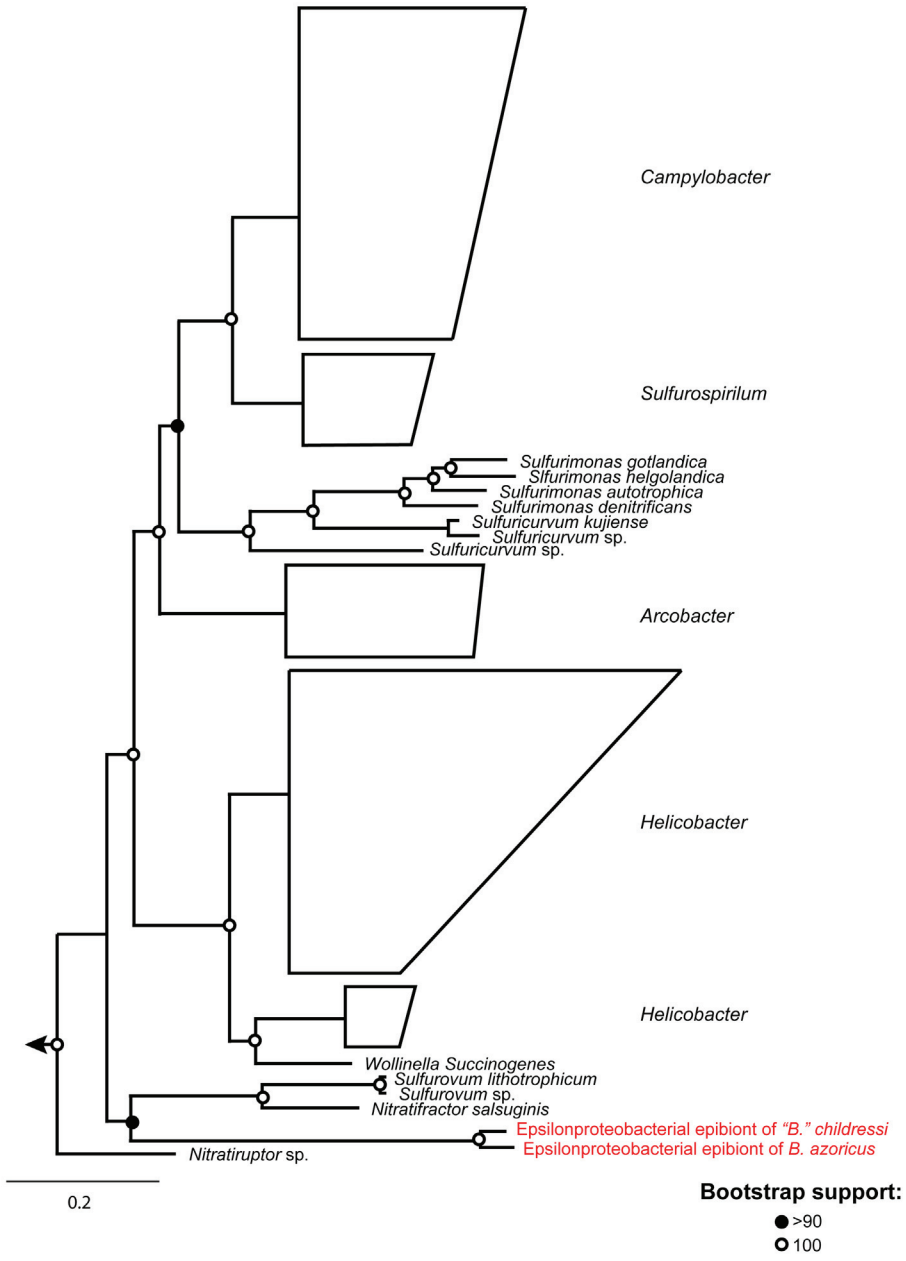


Figure 4.1 Phylogenomic reconstruction of 80 epsilonproteobacterial genomes. 63 genes were aligned then concatenated to calculate a RAxML tree. The tree shows the epsilonproteobacterial epibiont (in red) cluster together on a deep sister branch to the Arcobacter strains.

showed that these bacteria mainly rely on the oxidation of reduced sulfur compounds as a source of energy. Both genomes encode for the soxABCDXYZ genes, but our present analysis shows that EpsC additionally encodes for a flavocytochrome c dehydrogenase (fccAB) and type I sulfide quinone oxidoreductase (sqr) genes. These genes are involved in the oxidation of hydrogen sulfide to elemental sulfur and the transfer of electrons via cytochrome c and menaquinone molecules along the electron chain (Chen *et al*., 1994; Friedrich *et al*., 2001; Brito *et al*., 2009).

The presence of highly expressed fcc gene and less expressed sqr gene in EpsC suggested its ability to use hydrogen sulfide as a source of energy and hydrogen. Hydrogen sulfide is toxic at high concentrations because it inhibits oxygen respiration mechanisms (Vaquer-Sunyer and Duarte, 2008). The ability to oxidize the compound would give an alternative source of energy

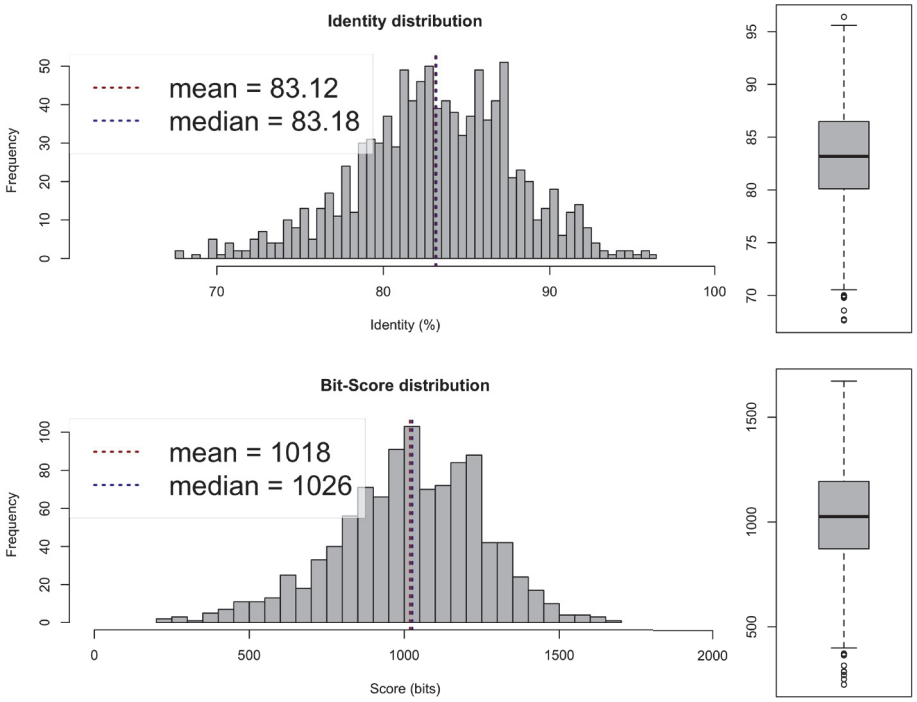


Figure 4.2 Barplot representing the similarities frequencies of contigs average nucleotide identities between the EpsC and EpsA draft genomes.

to the bacteria, while also detoxifying the host's environment. Interestingly, sqr or fcc genes were not found in the EpsA genome. *B. azoricus* hosts sulfur oxidizing Gammaproteobacteria, which possess similar proteins in their genome. This indicates that the EpsA epibionts might not be competing with the endosymbionts for hydrogen sulfide. However, we cannot exclude that those genes are simply missing because the genome is fragmented.

Additional source of electrons

EpsA and EpsC both had a set of genes coding for potential auxiliary electron donors from secondary energy sources other than sulfur oxidation. The two genomes encoded for quinones reductases, which allow direct oxidation of the pool of quinone using various substrates, such as malate:quinone oxidoreductase (Mqo), succinate dehydrogenase, Complex II and III homologs. The presence of a quinol cytochrome c Complex III potentially links the oxidation of quinol to the generation of a proton membrane gradient and the reduction of terminal electron acceptors, such as oxygen.

The EpsA genome also encoded for an uptake hydrogenase. The genome harbored one NiFe hydrogenase and an operon of accessory genes (hyp) required for the maturation of the enzyme (Jacobi *et al*., 1992). Hydrogenases generate a highly electron negative reductant that the epibiont could use to reduce its pool of quinones (Vignais and Colbeau, 2004). The uptake hydrogenase is more similar to the one present inS. *lithotrophicum*. No hydrogenase related genes were predicted in the genome of EpsC. Although we cannot rule out the possibility of the gene being simply missing from the current assembly due to fragmentation, this observation correlates with previous studies showing the presence of hydrogenases in hydrothermal vent symbioses but absence from cold seep associations (Petersen *et al*., 2011). EpsC might have the ability to use formate as an electron source via a formate dehydrogenase complex. The use of formate is usually associated with facultative anaerobes (Enoch and Lester, 1975).

Electron acceptors

EpsA and EpsC genomes both showed the ability of the bacteria to respire oxygen. They contained the entire set of genes for terminal oxygen acceptor Complexs: cytochrome c and cbb3 were present. EpsA was predicted to have the ability to respire nitrate. The genes for assimilatory nitrate reduction napFGBAH and the large and small nitrite reductase subunit genes were present in the EpsA genome. Respiration using nitrate is an alternative pathway commonly used by bacteria to have an alternate electron receptor in the presence of a low oxygen concentration (Moreno-vivián *et al*., 1999).

Oxygen tolerance

Multiple oxygen tolerant systems present in the genomes suggested that the epibionts were adapted to aerobic environments. Firstly, the presence of cytochrome oxidase indicated that oxygen is a possible electron acceptor. Additionally, multiple genes in the epibiont genomes encoded for catalase, peroxidases and alkyl hydroperoxide reductases (ahp), which indicated resistance to oxidative stress. A recent study showed that this protein was the major antioxidant enzyme in the endosymbiont of the deep-sea tubeworm *Riftia pachyptila* (Markert *et al*., 2007). Other metabolic pathways have also been described associated with anaerobic or facultative anaerobic bacteria, such as the oxidation of formate, a byproduct of fermentation, and assimilatory nitrate reduction pathways. Those pathways may highlight the metabolic versatility of the epibionts to adapt to fluctuating oxygen concentrations.

Central carbon metabolism

For both bacterial genomes, glycogenesis, glycolysis, non-oxidative branch of the pentose phosphate pathway and tricarboxylic acid (TCA) were predicted in the genomes. In each case, one gene was missing. The glycolysis pathway was missing 6-phosphofructokinase, which allows the utilization of sucrose or fructose, while the non-oxidative branch of the pentose phosphate pathway was lacking a transaldolase and the TCA cycle was

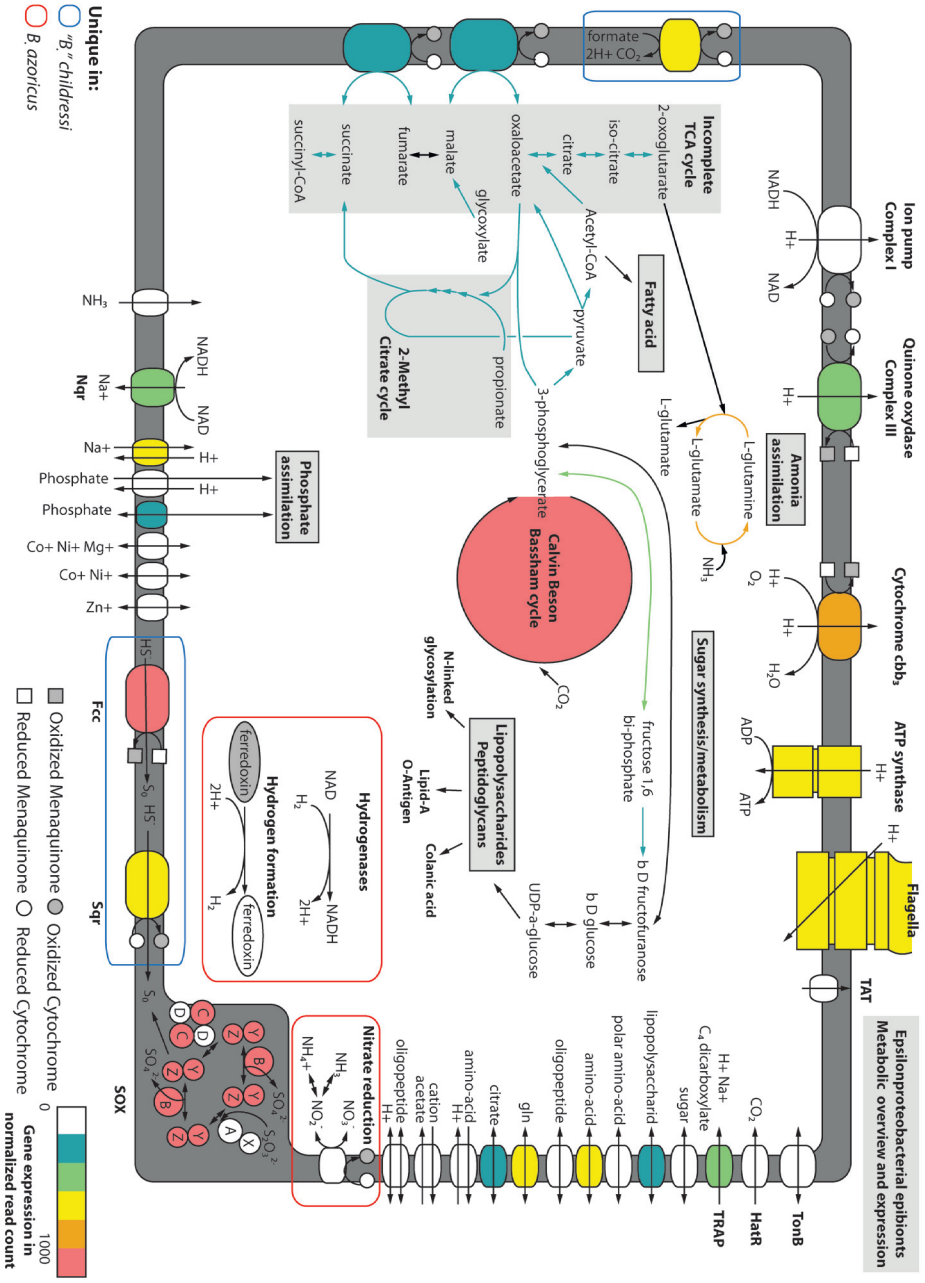


Figure 4.3 Schematic displaying the metabolic summary of the epsilonproteobacterial epibionts.

lacking 2-oxoglutarate dehydrogenase. The lack of the last enzyme has been suggested to indicate obligate autotrophy in other bacteria (Peeters *et al*., 1970; Wood *et al*., 2004). Carbon fixed via the Calvin cycle can enter the TCA cycle through phosphoenolpyruvate and could follow biosynthetic routes, either to fumarate or to 2-oxoglutarate. The presence of a 2-methyl citrate cycle was also predicted in both genomes, granting the ability to incorporate propionate into succinate and pyruvate. Pyruvate could then be fermented into acetyl coA and used as a cofactor in the TCA cycle.

These genomic and transcriptomic predictions support a chemoautotrophic metabolism. The whole metabolic network could be fueled by the CBB cycle, which would generate 3-phosphoglycerate that can be converted into pyruvate. Pyruvate would then be transformed into acetyl CoA and injected into the partial TCA cycle and potentially distribute carbon molecules to the various accessory metabolic pathways connected to it. Additionally, oxaloacetate and propionate via the 2-methyl citrate cycle could recharge the pyruvate and succinate pool.

Ammonium fixation

The epsilonproteobacterial epibionts have both the ability to fix ammonia through the use of glutamine synthetases and high-affinity uptake transporters. This enzyme assimilates ammonia into glutamine with high affinity at very low ammonia concentrations but requires an energy rich environment (Hua *et al*., 2004).

Transporters

We also predicted the presence of transporters in the genomes, such as ABC-like transporters, which can import sugar, lipids and amino acids into the cytoplasm. These transporters could provide a source of substrate to feed various metabolic pathways, for example, by injecting sugars into the glycolysis or glycogenesis pathway and lipids into the fatty acid generation or lipopolysaccharide synthesis pathways (Higgins, 1992).

Multiple tripartite ATP-independent periplasmic transporters (TRAP

transporters) were identified in the two draft genomes. These transporters are a large family of solute transporters found in Bacteria and Archaea, but not in eukaryotes, which appear to be specific for the uptake of organic acids. This transporter family utilizes a substrate binding protein in combination with a secondary transporter. These transporters have been described to import four carbon molecules, such as malate, fumarate or succinate, into the cytoplasm of bacteria (Mulligan *et al*., 2011). They can induce anapleurotical incorporation of fumarate or citrate into the different pathways. This may be a mechanism of supporting an incomplete TCA cycle by directly importing intermediate compounds (Ullmann *et al*., 2000).

Additionally, Na+ translocating NADH:quinone oxidoreductase (Na+-NQR) was present in the genomes. This membrane protein complex couples the oxidation of NADH to generate the export of Na+ into the bacterial periplasm. This efflux maintains a Na+ motive force, which fuels an array of different membrane symporters (Dimroth, 1997).

Amino acid and Vitamin biosynthesis

Both EpsA and EpsC genomes have genes for the synthesis of most amino acids. EpsC did not have the phenylalanine synthesis pathways and EpsA lacked the synthesis pathway for phenylalanine, alanine, proline and selenocysteine. It is not uncommon for bacteria to miss some amino acid synthesis pathways and the genes could still be missing in the assembly because of the fragmented nature of both epsilonproteobacterial epibiont genome drafts. Multiple amino acid transporters have been predicted in the genome; they could directly import those molecules from the environment.

The epibionts also had a partial pathway capable of de novo synthesis of vitamin B6 (pyridoxal 5'-phosphate – PLP) and thiamin, which are both dietary requirements for animals. The key genes of PLP synthesis were present and expressed but some of the precursor metabolic reactions of the pathway were not present. These compounds could, however, be generated by other metabolic pathways or may have been missing in the assembly. All

the genes for the biosynthesis of thiamin were present in EpsC, whereas EpsA lacked the genes for thiazole bionsynthesis, as well as the precursor pathway for the biosynthesis of thiamin. These pathways are widespread in bacteria but not in eukaryotic organisms. The presence of the epsilonproteobacterial epibionts may therefore be beneficial for the host's nutrition. However, more work is needed to investigate whether the host has specific molecular mechanisms in its gills for nutrient uptake.

Response to the environment

Both epsilonproteobacterial epibionts possessed genes that could enable them to adapt to variations in the environment. The genomes contained a wide array of mineral transport systems, including detoxification mechanisms for heavy metals such as mercury, cobalt, cadmium and copper. Additionally, genes used by prokaryotes to sense and respond to environmental signals were also present, such as multiple copies of predicted two-component signal transduction systems and diguanylate cyclase (GGDEF) domains, which are commonly used as molecular messengers triggering specific genetic reactions to environmental stimuli (Stock *et al*., 2000; Ryjenkov *et al*., 2005; Nakagawa *et al*., 2007).

Chemotaxis and motility

A fully functional flagellum apparatus was predicted in both epsilonproteobacterial epibiont genomes. We found that 35 genes involved in the assembly of the flagellum were present: EpsC contained 12 copies of the flagellin gene and EpsA had ten copies. Flagellin has been described in epsilonproteobacterial pathogens, such as *Campylobacter* and *Helicobacter*, to be involved in virulence, by promoting motility or being part of the adhesion process to the cell (McSweegan and Walker, 1986; Gilbreath *et al*., 2011). Multiple gene copies of flagellin are commonly found in bacteria and are often associated with a molecular mechanism called phase variation. This process consists of a random genetic switch between bacterial generations, whereby the expression of the gene is replaced by a homologous one. Bacteria such

as *Campylobacter* use such a mechanism to switch the flagellin molecule and evade the eukaryotic immune system (van der Woude and Baumler, 2004).

Additionally, multiple copies of chemotaxis-related genes were present in the epibiont genomes. Chemotaxis is based on receptor proteins anchored in the bacterial membrane and used for sensing external chemical stimuli (Foynes *et al*., 2000). These signals are transferred to internal effectors modulating bacterial motility toward or away from the environmental stimuli. In *Campylobacter* species, chemotaxis has been described as an important mechanism in the bacterial colonization of eukaryotic cells by inducing bacteria to move toward the eukaryotic cell (van Alphen *et al*., 2008; Gilbreath *et al*., 2011). Methyl chemotaxis protein and transductor are chemotaxis receptors. They were present multiple times in both metagenomes: ten times in EpsC and four times in EpsA. Chemotaxis cytoplasmic effector cheAYVW genes have been described to modulate the excitation/inhibition of the flagellum motor mechanism. Twenty-six copies of che related genes were present in the EpsC and ten in the EpsA genome (Spohn and Scarlato, 2001).

Adhesion and virulence

Genes involved in the N-linked glycosylation were also present in the draft genomes. N-linked glycosylation is a common protein modification mechanism occurring in all domains of life. It has been extensively studied in the Epsilonproteobacteria class, especially in *Campylobacter jejuni*, in which it has been shown to play a role in the evasion of the host immune system and in the adhesion and invasion process of the gut epithelium (Nothaft and Szymanski, 2010). The cluster of genes has also been found in two Epsilonproteobacteria genomes isolated from hydrothermal vents, *Sulfurovum litotrophicum* and *Nitratiruptor* sp. They are thought to be involved in the association with invertebrates (Nakagawa *et al*., 2007).

Lipooligosaccharide (LOS) related genes were present in both genomes. In EpsC, the LOS biosynthesis pathway was predicted to be functional. This includes genes for the synthesis of Lipid A, O antigen and were present in

the genome. Lipid A and O play a role in the adhesion to eukaryotic cells and immune cell evasion. These molecules belong to the endotoxin category and possess a long hydrophobic chain, which allows the bacteria to anchor to the eukaryotic host's cell membrane (Rubin and Trent, 2013).

The synthesis of colanic acid was also predicted to be possible in both draft genomes. Colanic acid (CA) is a highly viscous capsular polysaccharide. It has been described as a key component of biofilms secreted by Enterobacteria and functions to protect cells under conditions of stress, such as exposure to osmotic variation, and it is not involved in pathogenicity (Hanna *et al*., 2003; Hug and Feldman, 2011).

Other genes related to virulence or invasion in Epsilonproteobacteria were also present in both genomes. Multiple copies of fibronectin/fibrinogen binding proteins were annotated in both epibiont genomes. These genes have been associated with flagella to bind fibronectin and fibrinogen molecules, which are present in the extracellular matrix of eukaryotic epithelia (Monteville *et al*., 2003). Additionally, other accessory genes are present, such as invasion antigen CiaB and paralyzed flagella protein PflA. It has been shown in previous studies that when these genes are inactivated in gastrointestinal pathogens such as *Helicobacter* or *Campylobacter*, the virulence of the bacteria decreased. These genes probably play a role in the initial adhesion step of the bacteria to the host gut epithelia (Onozato *et al*., 2009; Gilbreath *et al*., 2011).

Similarity to other Epsilonproteobacteria

We compared the genetic potential of EpsA and EpsC with other publicly available epsilonproteobacterial genomes. The two genomes share common features with other chemolithotrophic Epsilonproteobacteria, the main difference being the presence of a CBB cycle to fix inorganic carbon.

Additionally, multiple genes described in gastrointestinal pathogens were present in both EpsC and EpsA draft genomes. These mechanisms are mainly involved in adhesion to the host cell and potentially avoiding the host innate

immune system. The parallel could be made between mammalian guts and bivalve gills, in the sense that both organs represent an interface between the host and the environment and both are protected with a layer of mucus in which filamentous Epsilonproteobacteria can settle.

Epsilonproteobacteria are traditionally separated into two groups: one associated with the human and animal gastrointestinal tract and other grouping environmental bacteria. Gut associated bacteria usually belong to the Helicobacteraceae and Campylobacteraceae families. Because they have an important economic and social impact, they have been extensively studied in the past. The second group is composed of various less described families, such as Arcobacteraceae and Thiovulgaceae. The latter were described as being one of the major free-living players in deep-sea reduced environments (Campbell *et al*., 2006) and, due to sampling and cultivation difficulties, only a few genomes are currently available. Perceived similarities may therefore be the result of over-representation in annotations of species from medical studies and may be biasing the interpretation of our gene function predictions. Nevertheless, these gene homologies could suggest a very ancient origin of adhesion mechanisms to biological surfaces in Epsilonproteobacteria (Nakagawa *et al*., 2007; Nakagawa and Takaki, 2009; Zhang and Sievert, 2014). Although there are clear similarities to other published adhesion mechanisms in the draft genomes of EpsA and EpsC, there is no indication for pathogenicity. No genes suggesting a pathogenic role, such as immune system modulating or toxin related genes, were found in the studied genomes. However, we cannot exclude pathogenic genes that have not yet been identified and annotated.

4.5 Conclusion

Our metagenomic analyses showed that, although initially thought to be different strains of the same species based on 16S rRNA sequences identity, ANI value comparison suggests that the two epibionts are in fact two distinct species. We observed a number of metabolic differences between EpsA

and EpsC draft genomes (Figure 3). EpsC was predicted to oxidize hydrogen sulfide through fcc and sqr enzyme systems and oxidize formate to generate energy, whereas EpsA did not. On the other hand, the EpsA draft genome showed the ability to oxidize hydrogen, while EpsC did not. These small-scale differences might be due to the different environments colonized by EpsA and EpsC. EpsA has been found associated with mussels living at hydrothermal vent fields, where hydorgenases have been described in many bacteria associated with invertebrates. EpsC, however, is associated with animals endemic to cold seeps and might have acquired a different additional source of energy. In both cases, the general carbon metabolism is consistent with chemoautotrophy. Additionally, the presence of a partial TCA, glycolysis, glucogenesis and glyoxylate cycles showed a global dependence of substrate generation by the CBB cycle.

Our genome annotations reconstruct an overview of the metabolic abilities of the epibionts. This study also shows the genetic versatility of epibionts, with multiple metabolic and genetic systems present and the potential ability to adapt to an array of electron donors, such as hydrogen for EpsA and hydrogen sulfide for EpsC, as well as other potential environmental changes, such as oxygen or heavy metals (such as lead, zinc, mercury or arsenate) concentration variation.

The genome annotations of mechanisms homologous to epsilonproteobacterial gastrointestinal pathogens also gave some insight into how filamentous bacteria could colonize gill epithelia. The presence of chemotaxy and flagellar motility indicated the potential ability of the epsilonproteobacterial filaments to find their host. The flagella could also be involved in adhesion mechanisms, such as fibrinogen and fibronectin binding proteins. Additionally, the synthesis of lipid A/O and colanic acid were predicted in both genomes and could be involved in the settlement of Epsilonproteobacteria on bathymodiolin gill epithelia. We did not find any clear indication of virulence related genes, and the capacity to synthesize essential dietary factors such as vitamin B6 and thiamin supports our previous

hypothesis that the epsilonproteobacterial epibionts are in a commensal or mutualistic relationship with the bathymodiolin mussels.

Future work should investigate in detail the differences between the two draft genomes and what mechanisms are involved in the specific association with their bathymodiolin host. The interaction with the host should also be investigated to understand the nature of the symbiosis. For example, are nutrients transferred from the bacteria to the host or do the epibionts protect the host against harmful concentrations of chemicals, such as heavy metals, or pathogens?

4.6 Acknowledgements

We would like to thank Silke Wetzel and Miriam Sadowsky for excellent technical assistance, Lizbeth Sayavedra and Juliane Wippler for advice on bioinformatics analyses and Pierre Offre for commenting on this manuscript. This project would not have been possible without the dedicated captains and crews of the “Pourquoi pas?” and “Atlantis” research vessels involved in the sampling. This work was funded by the Max Planck Society, the DFG Cluster of Excellence “The Ocean in the Earth System” at MARUM (University of Bremen), an European Research Council Advanced Grant (BathyBiome, Grant 340535) and a Gordon and Betty Moore Foundation Marine Microbiology Initiative Investigator Award through Grant GBMF3811 to ND, and the European Union (EU) Marie Curie Actions Initial Training Network (ITN) SYMBIOMICS (contract number 264774).

4.7 References

van Alphen, L.B., Bleumink-Pluym, N.M.C., Rochat, K.D., van Balkom, *B.*W.M., Wösten, M.M.S.M., and van Putten, J.P.M. (2008) Active migration into the subcellular space precedes *Campylobacter jejuni* invasion of epithelial cells. *Cell. Microbiol.* **10**: 53–66.

Brito, J.A., Sousa, F.L., Stelter, M., Bandeiras, T.M., Vonrhein, C., Teixeira, M., *et al*. (2009) Structural and functional insights into sulfide:quinone oxidoreductase. *Biochemistry* **48**: 5613–5622.

Campbell, *B.*J., Engel, A.S., Porter, M.L., and Takai, K. (2006) The versatile epsilon-proteobacteria: key players in sulphidic habitats. *Nat. Rev. Microbiol.* **4**: 458–68.

Chen, Z.W., Koh, M., Van Driessche, G., Van Beeumen, J.J., Bartsch, *R.*G., Meyer, T.E., *et al*. (1994) The structure of flavocytochrome c sulfide dehydrogenase from a purple phototrophic bacterium. *Science* **266**: 430–432.

DeChaine, E.G. and Cavanaugh, C.M. (2006) Symbioses of methanotrophs and deep-sea mussels (Mytilidae: Bathymodiolinae). In, Overmann,J. (ed), *Progress in molecular and subcellular biology*. Springer-Verlag Berlin, pp. 227–249.

Dimroth, P. (1997) The Na+-translocating NADH : ubiquinone oxidoreductase from. **776**: 770–776.

Dubilier, N., Bergin, C., and Lott, C. (2008) Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat. Rev. Microbiol.* **6**: 725–40.

Duperron, *S.*, Lorion, J., Samadi, *S.*, Gros, O., and Gaill, F. (2009) Symbioses between deep-sea mussels (Mytilidae: Bathymodiolinae) and chemosynthetic bacteria: diversity, function and evolution. *C. R. Biol.* **332**: 298–310.

Enoch, H.G. and Lester, *R.*L. (1975) The purification and properties of formate dehydrogenase and nitrate reductase from Escherichia coli. *J. Biol. Chem.* **250**: 6693–6705.

Foynes, *S.*, Dorrell, N., Ward, *S.*J., Stabler, *R.*A., McColm, A.A., Rycroft, A.N., and Wren, *B.*W. (2000) *Helicobacter pylori* possesses two CheY response regulators and a histidine kinase sensor, CheA, which are essential for chemotaxis and colonization of the gastric mucosa. *Infect. Immun.* **68**: 2016–2023.

Friedrich, C.G., Rother, D., Bardischewsky, F., Quentmeier, A., and Fischer, J. (2001) Oxidation of Reduced Inorganic Sulfur Compounds by Bacteria:

Emergence of a Common Mechanism? *Appl. Environ. Microbiol.* **67**: 2873–2882.

Gilbreath, J.J., Cody, W.L., Merrell, D.S., and Hendrixson, D.R. (2011) Change is good: variations in common biological mechanisms in the epsilonproteobacterial genera *Campylobacter* and *Helicobacter*. *Microbiol. Mol. Biol. Rev.* **75**: 84–132.

Hanna, A., Berg, M., Stout, V., and Razatos, A. (2003) Role of Capsular Colanic Acid in Adhesion of Uropathogenic Escherichia Appl. *Environ. Microbiol*. 69: 4474–4481.

Higgins, C.F. (1992) ABC Transporters: From Microorganisms to Man. *Annu. Rev. Cell Biol.* **8**: 67–113.

Hua, Q., Yang, C., Oshima, T., Mori, H., and Shimizu, K. (2004) Analysis of Gene Expression in *Escherichia coli* in Response to Changes of Growth-Limiting Nutrient in Chemostat Cultures. *Society* **70**: 2354–2366.

Hug, I. and Feldman, M.F. (2011) Analogies and homologies in lipopolysaccharide and glycoprotein biosynthesis in bacteria. *Glycobiology* **21**: 138–151.

Jacobi, A., Rossmann, *R.*, and Böck, A. (1992) The hyp operon gene products are required for the maturation of catalytically active hydrogenase isoenzymes in Escherichia coli. *Arch. Microbiol.* **158**: 444–451.

Markert, *S.*, Arndt, C., Felbeck, H., Becher, D., Sievert, *S.*M., Hugler, M., *et al*. (2007) Physiological Proteomics of the Uncultured Endosymbiont of *Riftia pachyptila*. *Science (80-.).* **315**: 247–250.

Markowitz, V.M., Chen, I.M.A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., *et al*. (2012) IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* **40**: 115–122.

McSweegan, E. and Walker, *R.*I. (1986) Identification and characterization of two *Campylobacter jejuni* adhesins for cellular and mucous substrates. *Infect. Immun.* **53**: 141–8.

Ménard, A., Buissonnière, A., Prouzet-Mauléon, V., Sifré, E., and Mégraud, F. (2016) The GyrA encoded gene: A pertinent marker for the phylogenetic revision of *Helicobacter* genus. *Syst. Appl. Microbiol.* **39**: 77–87.

Meyer, F., Paarmann, D., D'Souza, M., Olson, *R.*, Glass, E.M., Kubal, M., *et al*. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.

Monteville, M.R., Yoon, J.E., and Konkel, M.E. (2003) Maximal adherence and invasion of INT 407 cells by *Campylobacter jejuni* requires the

CadF outermembrane protein and microfilament reorganization. *Microbiology* **149**: 153–165.

Moreno-vivián, C., Cabello, P., Blasco, *R.*, Castillo, F., Cabello, N., Marti, M., and Moreno-vivia, C. (1999) Prokaryotic Nitrate Reduction : Molecular Properties and Functional Distinction among Bacterial Nitrate Reductases. *J. Bacteriol.* **181**: 6573–6584.

Mulligan, C., Fischer, M., and Thomas, G.H. (2011) Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. *FEMS Microbiol. Rev.* **35**: 68–86.

Nakagawa,S. and Takaki, Y. (2009) Nonpathogenic Epsilonproteobacteria. *eLS* 1–11.

Nakagawa, *S.*, Takaki, Y., Shimamura, *S.*, Reysenbach, A.-L., Takai, K., and Horikoshi, K. (2007) Deep-sea vent epsilonproteobacterial genomes provide insights into emergence of pathogens. *Proc. Natl. Acad. Sci.* **104**: 12146–12150.

Nelson, D.C., Hagen, K.D., and Edwards, D.B. (1995) The gill symbiont of the hydrothermal vent mussel *Bathymodiolus* thermophilus is a psychrophilic, chemoautotrophic, sulfur bacterium. *Mar. Biol.* **121**: 487–495.

Nothaft, H. and Szymanski, C.M. (2010) Protein glycosylation in bacteria: sweeter than ever. *Nat. Rev. Microbiol.* **8**: 765–778.

Onozato, J., Kumagai, A., Sekizuka, T., Tazumi, A., Moore, J.E., Millar, *B.*C., and Matsuda, M. (2009) Cloning, sequencing and expression of full-length *Campylobacter* invasion antigen B gene operon from *Campylobacter lari*. *J. Basic Microbiol.* **49**: 342–349.

Oshlack, A., Robinson, M.D., and Young, M.D. (2010) From RNA-seq reads to differential expression results. *Genome Biol* **11**: 220.

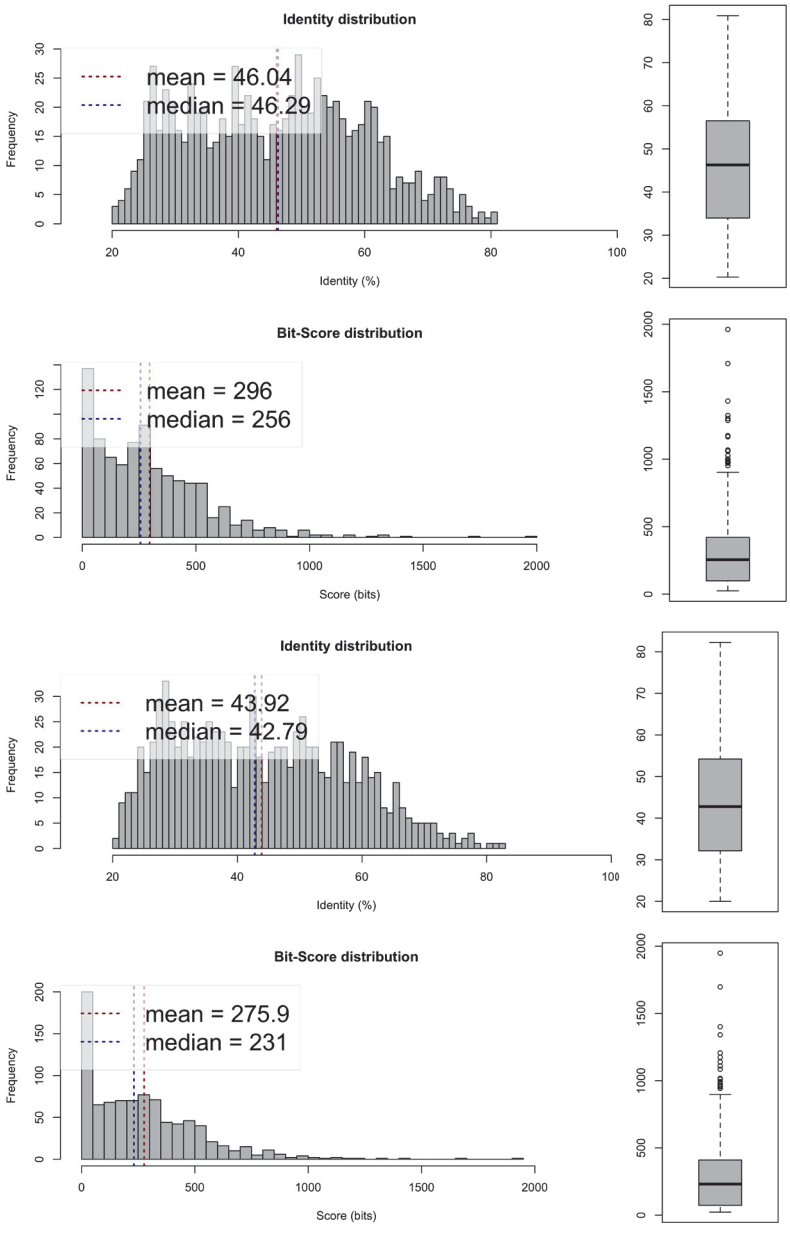
Peeters, T.L., Liu, M.S., and Aleem, M.I.H. (1970) The Tricarboxylic Acid Cycle in Thiobacillus denitrificans and Thiobacillus-A2. *J. Gen. Microbiol.* **64**: 29–35.

Petersen, J.M. and Dubilier, N. (2009) Methanotrophic symbioses in marine invertebrates. *Environ. Microbiol. Rep.* **1**: 319–335.

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

Rodriguez-R, L.M. and Konstantinidis, K.T. (2014) Bypassing cultivation to identify bacterial species. *Microbe* **9**: 111–118.

Rubin, E.J. and Trent, M.S. (2013) Colonize, evade, flourish. *Gut Microbes* **4**: 439–453.



Supplementary Figure 4.1 Average amino acid identity of EpsC against **A.** *Arcobacter butzleri* and *B. Sulfurovum lithotrophicum*.

Ryjenkov, D.A., Tarutina, M., Moskvin, O.V., and Gomelsky, M. (2005) Cyclic Diguanylate Is a Ubiquitous Signaling Molecule in Bacteria: Insights

into Biochemistry of the GGDEF Protein Domain. *J. Bacteriol.* **187**: 1792–1798.

Spohn, G. and Scarlato, V. (2001) Motility, Chemotaxis, and Flagella.- *Helicobacter* pylori: Physiology and Genetics. ASM Press

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

Stock, A.M., Robinson, V.L., and Goudreau, P.N. (2000) Two-Component Signal Transduction. *Annu. Rev. Biochem.* **69**: 183–215.

Ullmann, *R.*, Gross, *R.*, Simon, J., Unden, G., and Kroger, A. (2000) Transport of C4-Dicarboxylates in Wolinella succinogenes. *J. Bacteriol.* **182**: 5757–5764.

Vaquer-Sunyer, *R.* and Duarte, C.M. (2008) Thresholds of hypoxia for marine biodiversity. *Proc. Natl. Acad. Sci. U.*S. *A.* **105**: 15452–15457.

Vignais, P.M. and Colbeau, A. (2004) Molecular biology of microbial hydrogenases. *Curr. Issues Mol. Biol.* **6**: 159–188.

Wang, Z. and Wu, M. (2013) A Phylum-Level Bacterial Phylogenetic Marker Database. *Mol. Biol. Evol.* **30**: 1258–1262.

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 Microbiol. Rev.* **28**: 335–352.

van der Woude, M.W. and Baumler, A.J. (2004) Phase and Antigenic Variation in Bacteria. *Clin. Microbiol. Rev.* **17**: 581–611.

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.-H., *et al*. (2014) Uniting the classification of cultured and uncultured Bacteria and Archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* **12**: 635–645.

Zhang, Y. and Sievert, *S.*M. (2014) Pan-genome analyses identify lineage- and niche-specific markers of evolution and adaptation in Epsilonproteobacteria. *Front. Microbiol.* **5**: 110.

4.8 Supplementary Figures

Chapter 5 **It's all about location: The ectosymbionts of the hydrothermal vent shrimp Rimicaris hybisae are vent dependent.**

Authors: Adrien Assié[1], Julie Huber[2], Julie Reveillaud[2], Cindy van Dover, Christian Borowski[1], Nicole Dubilier[1,4,5], Jillian Petersen[1,3].

1 Symbiosis Department, Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, D-28359 Bremen, Germany.

2 Josephine Bay Paul Center, Marine Biological Laboratory, 7 MBL St., Woods Hole, MA 02543

3 Division of Microbial Ecology, Department of Microbiology and Ecosystem Science, Research Network Chemistry Meets Microbiology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria.

4 Faculty of Biology / Chemistry, University of Bremen, Bibliothekstrasse 1, D-28359 Bremen, Germany

5 Marum - Zentrum für Marine Umweltwissenschaften der Universität Bremen, Leobener Str. 2, 28359 Bremen

Manuscript in preparation

5.1 Abstract

Rimicaris shrimps are endemic to hydrothermal vents, where they occur in very dense swarms. Similar to many other deep sea invertebrate species, they have a symbiotic relationship with chemosynthetic bacteria. Dense ectosymbiotic bacterial communities are maintained in an enlarged gill chamber of the shrimp host. By maintaining a water current rich in reduced compounds, the host provides reduced compounds, which are used as an energy and carbon source by the bacterial communities. The bacteria are hypothesized to provide nutrition to the host. Two hydrothermal vents were recently discovered on the Mid-Cayman Spreading Center (MCSC): Von Damm, located at 2500 m depth, and Piccard, the deepest hydrothermal vent known to date, located at 5000 m depth. These vent fields are heavily colonized by a novel deep sea shrimp species: *Rimicaris hybisae*. The Von Damm and Piccard vent fields are unique systems: located 25 km apart and have a 2500 m depth difference , both vents have distinct chemical signatures due to differences in the underlying rock. In this study, we investigated the microbial communities associated with *R. hybisae* from both vent fields, using next generation sequencing of 16S rRNA genes. We showed that *R. hybisae* hosts a diverse bacterial community dominated by Epsilonproteobacteria. In addition, we compared *Rimicaris*-associated communities with the ectosymbiotic communities associated with *R. exoculata,* a closely related *Rimicaris* species endemic to Mid-Atlantic Ridge vents, and to free living communities sampled at both MCSC sites. These comparisons showed that, although there was some overlap, the symbiotic communities were statistically distinct from the free-living communities. Moreover, the symbiont community structure appears to be driven by geographic location. Our results suggest a local uptake of the ectosymbionts by their crustacean hosts.

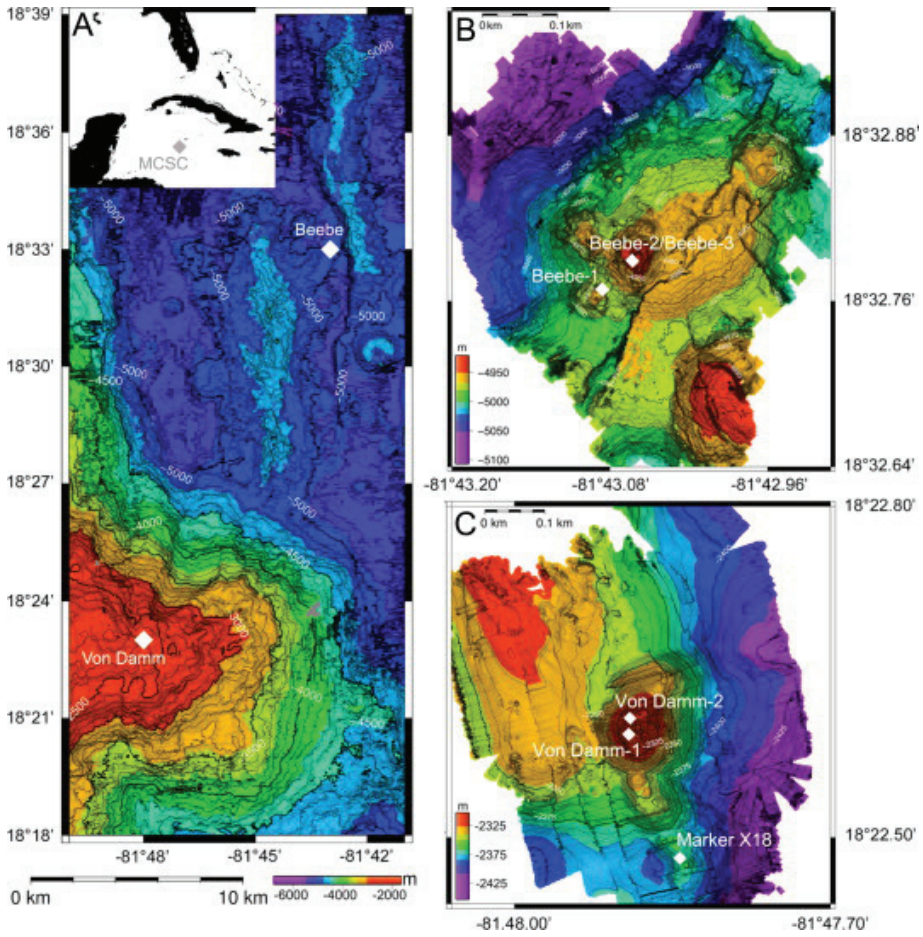


Figure 5.1 Bathymetric map of the Mid-Cayman Spreading Center hydrothermal vent fields. Picture reproduced with permission from Plouviez *et al*., 2015.

5.2 Introduction

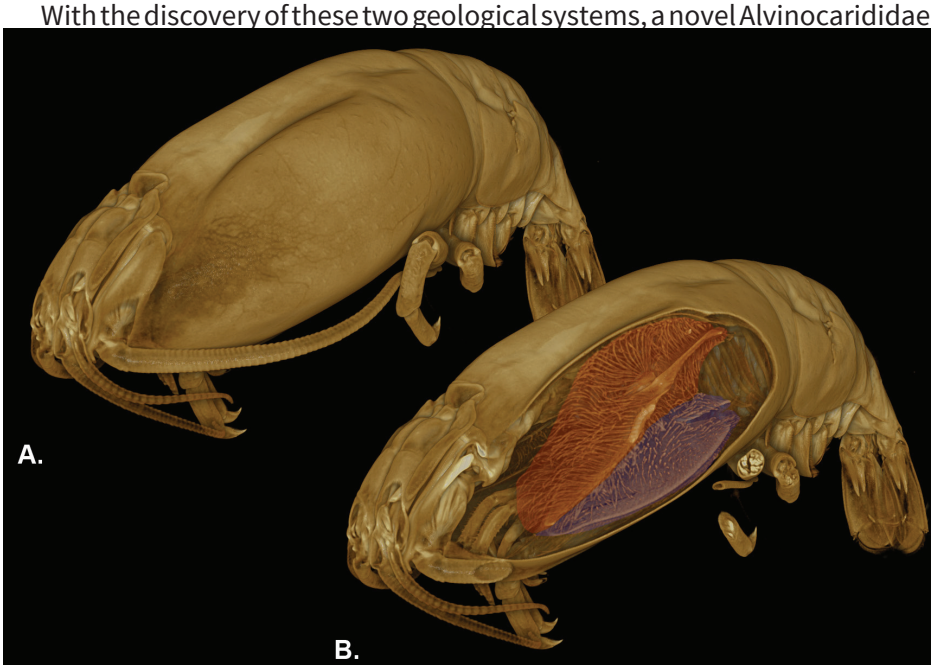
Deep-sea habitats such as hydrothermal vents and cold seeps are unique geological settings in which the local release of fluids with high concentrations of reduced compounds such as methane, hydrogen sulfide, and/or hydrogen sustains rich chemosynthetic communities. In the absence of light, and therefore photosynthesis, many invertebrate species have established symbiotic associations with chemosynthetic bacteria to thrive in such environments (For review see Dubilier *et al*., 2008). The host provides a

unique ecological niche to the bacteria, which in exchange provide a source of nutrients by harnessing reduced compounds present in the vent or seep fluids as energy and carbon sources. Symbioses between Bacteria and marine invertebrates occur in many forms. The types of associations range from endo- to ecto- symbioses. Endosymbiosis, in which intracellular Bacteria live within a specific tissue of the host, have been described in many deep sea species: *Calypotgena* clams and *Bathymodiolus* mussels host sulfur or methane oxidizers in specialized cells within their gill epithelia (Newton *et al.*, 2007; Duperron *et al.*, 2009) and polychaete worms such as *Riftia* host sulfur oxidizers in a special organ called the trophosome. In ectosymbioses, on the other hand, the bacteria colonize various body surfaces of their host, as is the case with the polychaete worm *Alvinella pompejana* hosting filamentous Epsilonproteobacteria on its dorsal setae (Desbruyères *et al.*, 1998) and crustaceans such as *Kiwa* crabs or *Rimicaris* shrimps having specialized body parts densely colonized by Epsilon- and Gammaproteobacteria (Goffredi *et al.*, 2008; Petersen *et al.*, 2010).

In 2010, during Caribbean cruise JC044 to the Mid-Cayman Spreading Center (MCSC), the RRS James Cook discovered two new hydrothermal vent fields: Piccard and Von Damm (Figure 5.1 - German *et al.*, 2010). Both vent fields are hosted in different rock types, which results in stark differences in the vent fluid chemistry between the two sites. Piccard is the deepest hydrothermal vent (5000 m) discovered so far, and is located close to the spreading center, thus the vent is basalt hosted, releasing vent fluids with an acidic pH (3.2), high sulfide (12.3 mM), low methane (0.13 mM) and high hydrogen (20,7 mM) concentrations (Reeves *et al.*, 2014; Mcdermott *et al.*, 2015). The Von Damm vent, on the other hand, is located on a seamount further away from the ridge axis, and the vent fluid passes through mafic rocks. The pH of Von Damm fluids is much higher (5.8) and the fluids contain less sulfide (3.2 mM), more methane (2.84 mM) and similar hydrogen concentrations (19.2 mM) to Piccard (Reeves *et al.*, 2014; Mcdermott *et al.*, 2015). Logatchev vent fluids have been described to have very similar pH,

sulfide (2.5 mM) methane (3.5 mM) and hydrogen (19 mM) concentrations as the Von Damm vent fluids (Schmidt *et al.*, 2007). In addition, they are also located in the same depth range (Von Damm 2500 m and Logatchev 3000 m).

With the discovery of these two geological systems, a novel Alvinocarididae



shrimp, *Rimicaris hybisae,* was described (Nye *et al.*, 2011). *R. hybisae* is closely related to the more intensively studied *Rimicaris exoculata*, which is endemic to hydrothermal vents along the Mid-Atlantic Ridge (MAR). They occur in very dense swarms, which can reach several thousands of individuals per square meter (Van Dover *et al.*, 1988). One of the main features of *R. exoculata* is the dense coating of ectosymbiotic bacteria colonizing the enlarged gill chamber and two overgrown mouth appendages, the scaphognathite and the exopodite ( ). The ectosymbiotic bacteria of *R. exoculata* belong mainly to Epsilon- and Gammaproteobacteria. Additionally, a biogeographic distribution of the symbionts has been suggested across sites on the North Mid-Atlantic Ridge (NMAR) (Petersen *et al.*, 2010). The ectosymbionts have the potential to oxidize sulfur, methane, iron and hydrogen and have been

thought to play a role in chemical detoxification and nutritional support of the host (Van Dover *et al.*, 1988; Ponsard *et al.*, 2013; Jan *et al.*, 2014).

The aim of this study was to investigate the bacterial populations associated with *Rimicaris* shrimps. Early electron microscopy scanning confirmed the presence of Bacteria in the gill chambers of *R. hybisae* but the exact composition of these populations is yet to be determined (Nye *et al.*, 2011). By comparing the bacterial populations from two different host species and from different environmental samples, we wanted to discover whether we could detect any pattern between the different community compositions. In particular, we examined whether the ectosymbiont community compositions are more similar to environmental samples or are specific to the host. The bacterial populations were screened using fluorescence in situ hybridization (FISH) and amplicon libraries targeting the V3-V4 regions of the 16S rRNA genes. Firstly, we analyzed the ectosymbiotic bacterial communities hosted by *R. hybisae* shrimp samples from both MCSC hydrothermal vent fields. We then also compared the *R. hybisae* profiles with water samples taken at each vent field. Finally, we compared the *R. hybisae* associated populations with those associated with *R. exoculata* at the Logatchev site, an ultramafic hydrothermal vent field located on the NMAR.

Figure 5.2 3D reconstruction of a *R. exoculata* shrimp from micro computed tomography data. (A) Entire animal (B) Cut-away view showing inside the gill chamber: the scaphognathite is highlighted in orange and the exopodite in purple.

5.3 Material and Methods

Sample collection

Rimicaris samples

Rimicaris hybisae samples were collected with a sucking device operated by the remotely operated vehicle (ROV) Jason II. Two vent fields were sampled at the Mid-Cayman Spreading Center (MCSC), Caribbean, during the 18th voyage (16th leg) of the RV 'Atlantis' (see Digital Supplementary Table 5.1). Shrimp cephalothoraxes were longitudinally cut in half. One half was

fixed in RNAlater (Sigma Aldrich, Steinheim, Germany), then stored at -80°C, the other half was fixed in 2% formaldehyde in 0.2 µm filtered seawater solution. After fixation, the samples were washed three times at 4°C for 30 min in 0.2 µm filtered seawater, then stored at -20°C in a 50% ethanol, 50% filtered seawater solution.

Rimicaris exoculata samples were collected with a sucking device operated by the Quest ROV on the Logatchev hydrothermal vent field on the North Mid-Atlantic ridge (NMAR) during the Meteor M64/2 cruise (see SUPP table 1). Samples were processed on board immediately whenever possible, or a maximum of 12 hours after retrieval. Entire animals were frozen at -80°C. The individuals analyzed with micro computed tomography (microCT) were fixed at 4°C for 4–10 h in 2% formaldehyde in 0.2 µm filtered seawater solution. After fixation, the samples were washed three times at 4°C for 30 min in 0.2 µm filtered seawater, then stored at -20°C in a 50% ethanol, 50% filtered seawater solution.

Water samples

Water samples from the MCSC Rise were collected according to Reveillaud *et al*., 2015. Briefly; low temperature diffuse hydrothermal fluid samples were collected during the Oases 2012 cruise of the RV 'Atlantis' in January 2012 using the ROV Jason II. Two liters of fluid was filtered through a 0.22 µm SterivexTM filter (Millipore). Upon submersible recovery, filters were flooded with RNAlater, sealed in Male/Female Luer caps and stored in sterile falcon tubes. Filters were stored at 4°C for 18-24 hours, before being stored at -80°C until further processing.

DNA extraction

Genomic DNA was extracted from combined dissected scaphognathite s and exopodites of *Rimicaris*. DNA extraction was performed according to the Zhou *et al*., (1996) protocol, with the following modifications: An initial incubation step was performed at 37°C in 360 µl of extraction buffer (100 mM Tris-HCl [pH 8.0], 100 mM sodium EDTA [pH 8.0], 100 mM sodium phosphate

[pH 8.0], 1.5 M NaCl, 1% CTAB) and 40 µl of proteinase K (10 mg/ml) overnight. Total genomic DNA from the MCSC rise filters was extracted as previously described in (Akerman *et al.*, 2013).

PCR amplification

16S rRNA tagged amplicon libraries were prepared according to (Fadrosh *et al.*, 2014). We used modified 319F and 806R PCR primers to amplify the V3- V4 region of the 16S rRNA gene. A 7bp barcode and a 0 to 8bp heterogeneity spacer were designed using the barcrawl (Frank, 2009) script and added to the primer sequences (for list see Digital Supplementary Table 5.1). The custom primers were synthesized by Biomers (Ulm, Germany). PCRs were run with Phusion polymerase (Thermo Scientific, Ulm, Germany) according to the following protocol: 5 µl of DNA was added to 20 µl Phusion high fidelity buffer, 8 µl dNTPS (10 mM), 5 µl forward and reverse primer, 3 µl 3% DMSO, 1 µl Phusion DNA polymerase and 58 µl water. The PCR mix was split into five technical replicates of 20 µl each. The PCR program consisted of the following steps: an initial denaturation at 98°C for 30 s, then 30 cycles composed of a denaturation step at 98°C for 10 s, an annealing temperature at 58°C for 30 s and extension step at 72°C for 30 s, and a final extension step of 72°C for 10 min.

	Library A	Library B	Library C	Total
Sequenced reads	17155978	9589227	10275770	37020975
After merging / quality trimming	15714554	7394568	4366845	27475967
After Barcode trimming :				2949788

The five technical replicates were pooled together and then amplified DNA was checked on a 15% agarose gel. Bands corresponding to the expected size were cut out and purified using a QIAGEN gel extraction kit (Hilden, Germany). The quality of the DNA was assessed with a Qubit® 2.0 Fluorometer (Invitrogen, Eugen, USA). Three independent pooled amplicon samples were prepared. An equal amount of DNA for each sample was added to a final concentration of 1 µg of DNA. A total of 49 samples, 32 animals and 15 environment samples, as well as two background water column samples,

were amplified and pooled in the libraries.

Sequencing

Amplicon libraries were prepared and sequenced by the Max Planck Genome Center (Cologne, Germany) on an Illumina HISEQ 2500. The libraries were created using a TrueSeq DNA kit according to the manufacturer's recommendations. Three libraries of 250 bp paired end reads yielded 17,155,978 (*Rimicaris* samples), 9,589,227 (environmental samples) and 10,275,770 (mix of *Rimicaris* and environmental samples) reads.

Amplicon data analysis

Amplicon libraries were quality screened and trimmed with the BBduk script. The paired end reads were merged with BBmerge using strict merging settings and no mismatch in the paired reads overlap to minimize false positive (BBmerge and BBduk are part of the BBmap package: Bushnell B. - sourceforge.net/projects/bbmap/). The sequences were then split according to their respective barcodes sequences using Mothur software (Schloss *et al.*, 2009 – split sequence module, using strict settings: no mismatch in the barcode sequences). Primer and barcode sequences were trimmed from the merged reads during this step. The sequence numbers for each sample after quality trimming and barcode splitting are given in Table 5.1 and 2.

Table 5.1 Summary of the amplicon reads through the sequence processing**.**

Sequence composition and distribution were analyzed using minimum entropy decomposition (MED) from the oligotyping pipeline (Eren *et al.*, 2014). Taxonomical units predicted by the oligotyping approach are referred to as "nodes". We set the script to ignore nodes with a minimum substantive abundance (-M) lower than ten sequences and used default settings for other parameters. The taxonomic affiliation of the MED nodes was determined with the online SINA aligner (Pruesse *et al.*, 2012) against the Silva rRNA database (Release 123), query sequences were clustered with similar reference sequences on 97% similarity.

Statistical analyses

The count matrix generated by the oligotyping pipeline was used for further statistical analyses performed in R. To further reduce noise, the data was pruned to analyze only nodes present in more than 10% of the samples. A Mantel test was done to compare original and pruned matrices to check whether the pruning affected the matrix distribution pattern. All statistical analyses were done in R statistical software version 3.2.1 using the community ecology package vegan (Oksanen *et al.*, 2015). Alpha and Beta diversity were estimated and graphically represented using custom R scripts (Personal communication with C. Hassenrueck – https://github.com/chassenr/NGS).

Briefly, Alpha diversity of the community dataset (sequence count data) obtained with MED analysis was explored by first converting the count matrix into a relative abundance matrix. We then calculated species richness and evenness with the Inverse Simpson concentration as implemented by Chao *et al*., 2014. One-way ANOSIM and Permanova tests were used to test for significant differences between sample groups. Posthoc tests were performed to validate significant comparisons using the Bray-Curtis method to generate a distance matrix and the 'fdr' (Benjamini and Hochberg, 1995) method for p-value correction. A Non-metric multidimensional scaling (NMDS) plot based on the communities' relative abundance matrix, calculated using 2 dimensions and 24 iterations, generated a stress value of 0.05.

Fluorescence *in situ* hybridization

Sample preparation was prepared according to Petersen *et al*., (2010). Whole scaphognathite and exopodite tissues were removed from half of the cephalotorax, tissues were embedded in paraffin and then cut into 8 µm thick sections with an RM 2165 microtome (Leica, Germany). The sections were collected on Superfrost Plus slides (Roth, Germany). Wax was removed from paraffin sections by washing three times for 10 min each in Roti-Histol (Roth, Germany). Sections were circled with a wax pen (PAP-pen, Kisker Biotech,

Steinfurt, Germany) and rehydrated in an ethanol series consisting of 1 min in 96% ethanol, 1 min in 80% ethanol, then 1 min in 50% ethanol. Sections were hybridized in a buffer (0.9 M NaCl, 0.02 M Tris/HCl pH 8.0, 0.01% SDS, with the appropriate formamide concentration) containing probes with an end concentration of 8.43ng ml^{-1}. Sections were hybridized for 3 h at 46°C, then washed for 30 min at 48°C with buffer (0.1 M NaCl, 0.02 M Tris/HCl pH 8.0, 0.01% SDS, 5 mm EDTA), then rinsed in distilled water. To stain all DNA, sections were covered with a 1% DAPI solution, left for 3 min, rinsed with distilled water, then dipped in 96% ethanol and air dried. Sections were mounted in a mixture of Citifluor and Vectashield and examined using both a fluorescence microscope (Zeiss Axioskop, Germany) and a confocal laser-scanning microscope (Zeiss CLSM 510, Germany)

Because the amplicon sequences were too short to generate vent-specific probes, we used general FISH probes for the main proteobacterial taxa. General gammaproteobacterial probe GAM42a, with the betaproteobacterial competitor unlabeled probe BET42a was used simultaneously with the general Epsilonproteobacteria probe EPSY914 (Loy, 2003). Additionally, eubacterial probes EUB338 I to III (Amann *et al*., 1990) and non-specific probe NON338 were used as positive and negative controls, respectively. All probes were hybridized using a 35% formamide concentration.

Micro Computed Tomography

Micro Computed Tomography (µCT) was performed on one *R. exoculata* individual. The shrimp's tissue absorption for x-ray radiation was increased with a contrasting solution. The contrasting reagent was based on phosphotungstic acid (EtOH 96%, PTA 1% and DMSO 3%) and contrasting was conducted for three months, following a modified protocol from (Fernández *et al*., 2014). A long incubation was required since the PTA took a long time fully to penetrate all appendages of the crustacean body, i.e. antennae.

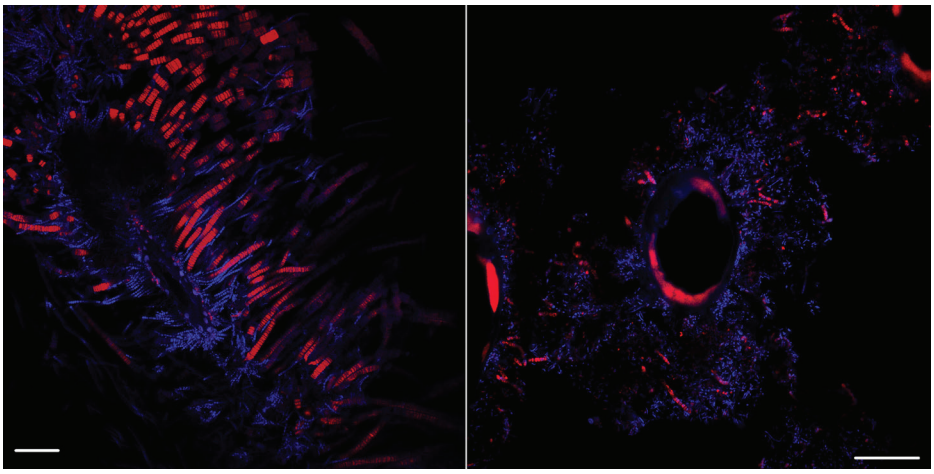
For micro-CT scanning, the specimen was stabilized with agarose chunks while being submerged in contrasting solution. The scanning was carried out

with a Nanotom m (phoenix|x-ray, Wunstorf, Germany) cone beam CT system at a voltage of 110 kV and a current of 70 mA. The 1800 projections were acquired at a resolution of 3072 x 2400 pixels, which resulted in a voxel size of 20 µm after subsequent tomographic reconstruction.

3D rendering was performed on a workstation with Windows 7 containing a 16-core CPU with 64 GB of main RAM and a 2 GB NVIDIA graphics card. The volumetric raw data was imported into the free 3D imaging software Drishti 2.6.1 (Limaye, 2012 - http://sf.anu.edu.au/Vizlab/drishti/index.shtml), which was used for 3D volume rendering. Transfer functions for false color rendering and transparency settings were adjusted with the 2D histogram. Shading was performed using the shader widget in Drishti. Additional clipping planes allowed virtual sectioning of the 3D model.

Table 5.2 Number of sequences recovered per sample after quality filtering and barcode splitting. Sample A-272.6.25 was removed due to a low number of representative sequences

Samples name	Number of reads	Sample name	Number of reads
A-1346	7977	**A-927**	12355
A-1362	23547	**A-934**	35835
A-1370	6081	**A-939**	14614
A-1374	103143	**A-951**	20986
A-1378	16573	**A-955**	57166
A-1382	19084	**A-959**	19218
A-1386	90235	**A-963**	6248
A-272.6.22	25058	**A-967**	3087
A-272.6.23	12423	**E-FS841**	123736
A-272.6.24	11760	**E-FS842**	343830
A-272.6.25	9	**E-FS843**	139420



A-272.6.29	3942	**E-FS844**	111209
A-272.6.3	585503	**E-FS845**	48453
A-272.6.8	10314	**E-FS846**	48234
A-283.7.10	854	**E-FS847**	32912
A-283.7.6	2479	**E-FS848**	33896

A-283.7.8	622	**E-FS849**	27476
A-31	392893	**E-FS850**	37642
A-35	248339	**E-FS852**	38358
A-39	47204	**E-FS853**	42711
A-43	48542	**E-FS854**	12433
A-47	18299	**E-FS855**	10322
A-755	2020	**E-FS856**	8504
A-923	19881	**E-O-CTDBB**	3094
		E-O-CTVD	21267
Total	2949788		

5.4 Results

FISH

We performed FISH analyses on exopodite and schaphognathite sections of *R. hybisae* individuals from both Von Damm and Piccard vent fields. The analyses confirmed the presence of filamentous bacteria attached to host setae. We observed on two different scaphognathite samples from one individual sampled at Von Damm and the other sampled at Piccard specific signal for Gammaproteobacteria but inconsistent signal for the Epsilonproteobacteria. The FISH probes and DAPI signals targeted rectangular cells organized in a string encased in an autofluorescent material (data not shown). These structures appeared to be attached to the host setae (Figure 5.3) at both vent fields. The bacteria cell appeared to have a size different between both vent sites, Gammaproteobacteria from Von Damm had a 2 by 4 µm size against 1 by 2 µm size for the bacteria at Piccard.

Figure 5.3 FISH picture of a *R. hybisae* scaphognathite . General FISH probe against Gammaproteobacteria in red, DNA in blue. Scale 20 µm

Sequencing and oligotyping

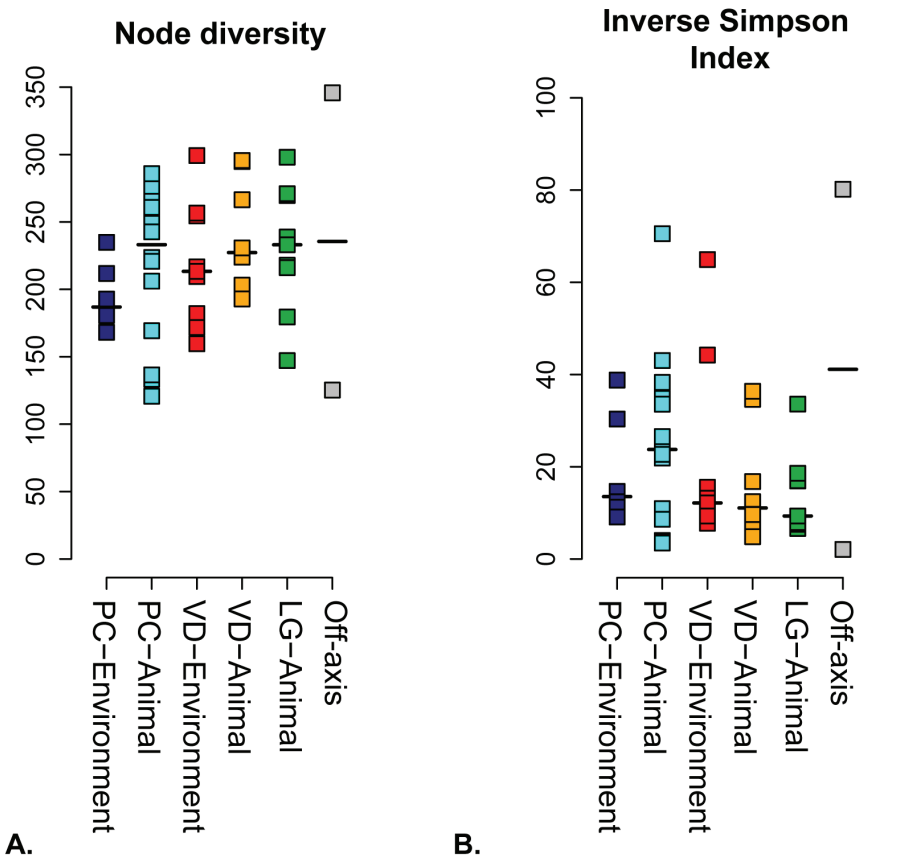


Figure 5.4 Alpha diversity indices. Each square represents diversity (A) or inverse Simpson (B) values. Samples are grouped by sample categories and sampling location, the horizontal lines show the median values.

We sequenced 49 samples in three separate libraries, resulting in a total of 37,020,975 reads. After strict merging, trimming and splitting, a total of 2,949,788 sequences across 48 samples were analyzed with the MED pipeline. The taxonomic unit produced by MED is referenced here as a "node". One *Rimicaris* sample was removed because of a very low number of reads recovered from this sample (9 reads). We removed 437,108 sequences from the analysis because they occurred in nodes composed of fewer than ten sequences. The MED pipeline produced 11,962 unique nodes.

Taxonomic affiliation of theses nodes obtained using the SINA web aligner indicated that the nodes were affiliated to 36 phyla, 90 classes and

R values

	PC-Environment	PC-Animal	VD-Environment	VD-Animal	LG-Animal
PC-Animal	0.2241689				
VD-Environment	0.9317357	0.8887639			
VD-Animal	1	0.6913265	0.3780382		
LG-Animal	1	0.8220222	0.6584362	0.9813368	
Off axis samples	1	1	0.8528529	1	1

P adjusted values

	PC-Environment	PC-Animal	VD-Environment	VD-Animal	LG-Animal
PC-Animal	0.042				
VD-Environment	0.001615385	0.001615385			
VD-Animal	0.003	0.001615385	0.001615385		
LG-Animal	0.001615385	0.001615385	0.001615385	0.001615385	
Off axis sample	0.0378	0.0084	0.019764706	0.0315	0.01976471

412 genera (Supplementary Table 5.2). The nodes were mainly related to Epsilonproteobacteria (74.4%), followed by Gammaproteobacteria (10.7%), Alphaproteobacteria (4.8%) and Flavobacteriia (3.6%).

Statistical analyses

Alpha diversity

Alpha diversity was calculated by estimating richness and evenness of the data set in 100 bootstrapped subsamples (Figure 5.4). Estimation of the Alpha diversity indices suggested that the majority of the samples, regardless of their source (animal or environment) or location (Piccard, Von Damm or Logatchev), were dominated by only a few nodes.

Beta diversity

In order to reduce the noise generated by rare taxa (defined here as being present in fewer than 10% of the samples), filtering of the dataset reduced the data to 2451 nodes distributed across 22 phyla, 49 classes and 141 genera. The pruning did not affect the beta diversity of the dataset (Mantel test with Bray-Curtis: R=0.99, p=0.001 and Jaccard R=0.96, p=0.001). Analyzing the node distribution across the different samples revealed a specific clustering of the sequences based on type and location. Non-parametric multivariate

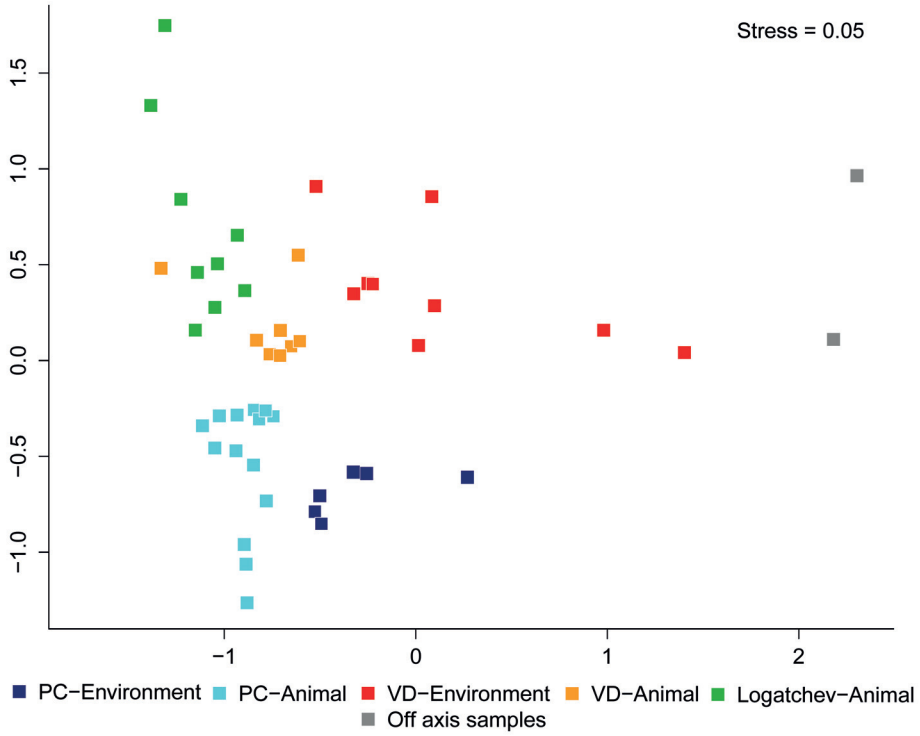


Figure 5.5 NMDS 2D similarity plot comparing the node composition of each sample. Samples are color labeled based on the sample type.

analyses based on dissimilarity matrices were performed on the dataset to test how similar the different community compositions were. ANOSIM, verified with a post hoc test, showed a clustering of the sequences according to the sampling type and location (Table 2). A PERMANOVA test supported this finding with a R^2 value of 0.6 and a p-value of 0.001. An NMDS plot was calculated based on the node relative abundances of the filtered data set (Figure 5.5). The NMDS plot revealed distinct groups consisting of particular sample types (animal-associated or environmental) from particular locations, with the exception of *Rimicaris* samples from Von Damm and Logatchev, in which the two groups overlapped.

Table 5.3 Anomsim R and p adjusted values with Benjamini- Hochberg methods for similarities between sample groups**.** PC: Piccard, VD: Von Damm, LG: Logatchev

ANOSIM and NMDS data indicated that the MCSC samples clustered by location. Samples from Von Damm were significantly different from samples from Piccard. However, environmental and animal samples from the same location were more similar to each other. Animal samples from Logatchev were significantly different from the samples from Piccard and less different from Von Damm samples.

Community composition

The relative abundance of the most abundant nodes for each of the 48 amplicon libraries are shown in Figure 5.6 (detailed population compositions are described in Supplementary Figure 5.3 and Supplementary Table 5.1). Supplementary Table 5.2 gives the identity values of these nodes compared to the reference sequences. Most of the nodes were related to the Epsilonproteobacteria. Other less relatively abundant taxa included Alpha-, Gamma- Delta- and Zetaproteobacteria as well as Flavobacteriia (Supplementary Figure 5.1).

The node distribution among the different samples revealed the dominance of nodes related to Epsilonproteobacteria across all samples (Figure3). A total of 1323 nodes were affiliated with the *Sulfurovum* genus. These sequences appeared to be closely related to *Sulfurovum aggregans (NR_126188 - Mino et al., 2014),* with identities ranging from 97 to 99%. We compared these nodes with the V3-V4 region of the 16S rRNA gene sequences associated with *R. exoculata* from multiple sites along the Mid-Atlantic Ridge. Out of 1323 nodes, 18 nodes were 100% identical and 702 were 99% identical to 19 previously published sequences associated with *R. exoculata* (Supplementary Table 5.2).

Different closely related nodes belonging to the *Sulfurovum* clade could be observed between samples from different vent fields. Node 74707 was relatively most abundant in the samples taken from the Piccard vent field area, whereas nodes 67166 and 76959 were relatively most abundant in

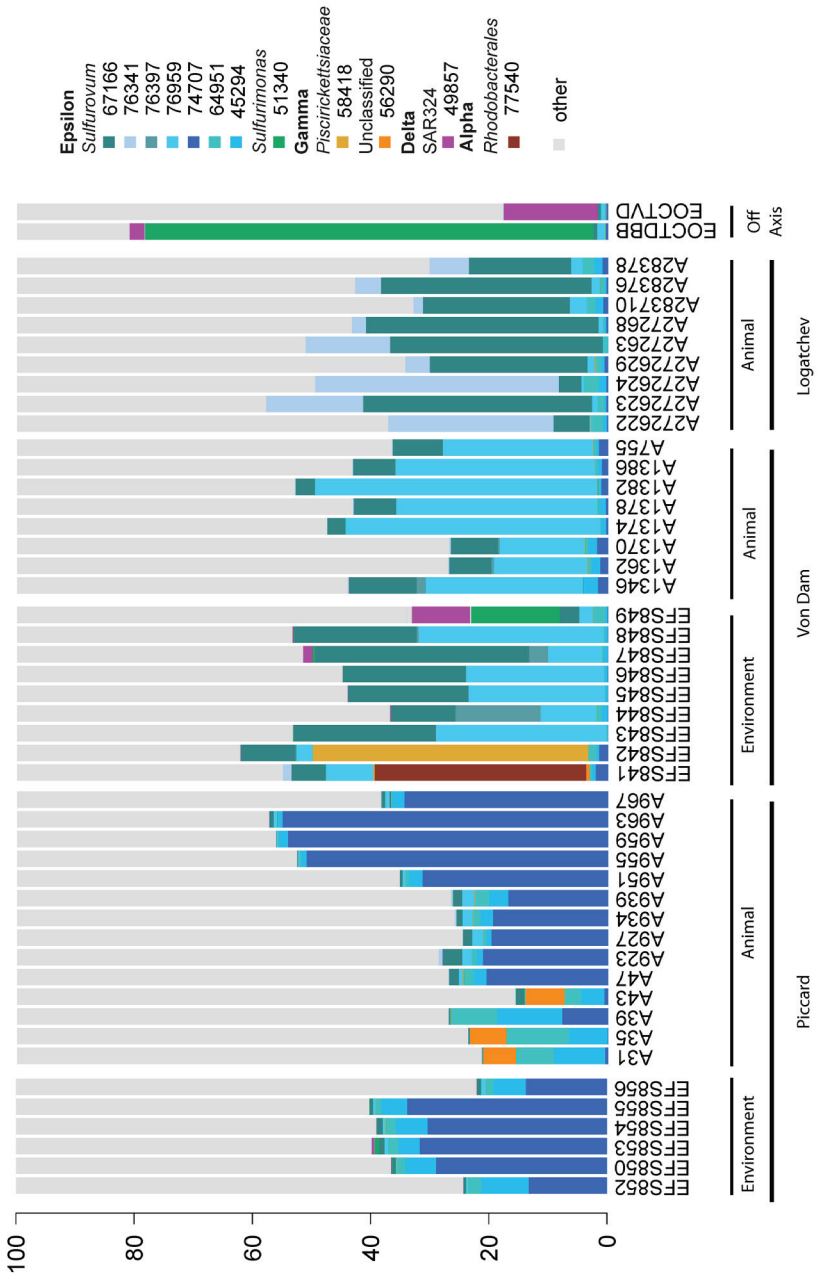


Figure 5.6 Relative abundance of dominant bacterial nodes and their taxonomic assignment for each individual sample. The samples are grouped by geographic location. Are displayed the cumulated first most relatively abundant node per sample to a total of 12 samples, the rest are collapsed into “other”. More detailed bar chart available as Supplemetary figure 5.3. Complete list of nodes is available in Supplementary Table 5.2.

the samples from Von Damm. Nodes 76341 and 67166 were relatively most abundant in the animal samples from Logatchev. This indicates that the abundance of particular nodes, and thus the relative abundance of particular phylotypes, was significantly higher at these sampling locations.

We predicted 111 nodes to belong to the Bacteroidetes phylum. These nodes occurred in all samples, but primarily in the samples from the Piccard vent field, and represented the third relatively most abundant taxa in our dataset. The relatively most abundant nodes belonged to Flavobacteriia and were 80 to 96% identical to *Actibacter sediminis* (NR_044349), a bacterial strain isolated from tidal flats (Kim *et al*., 2008) and 95 to 96% identical to the 16S rRNA sequences from *R. exoculata* sampled at the Rainbow (FN662570) and TAG (FN662637) vents sites on the NMAR.

With 125 and 136 nodes respectively, Gamma- and Alphaproteobacteria were among the relatively most abundant classes present across all samples. We noted that the most common Gammaproteobacteria nodes were mainly related to the Thiotrichales order. The main nodes (57431, 72316 and 61981) were 85% identical to the closest cultivated species *Leucothrix mucor* (NR_044870), an algal epiphyte. Additionally, these sequences were 80 to 95% identical to uncultivated 16S rRNA clone sequences isolated from *R. exoculata* individuals sampled at the NMAR hydrothermal vent fields Logatchev (FR797915) and “Snake Pit” (FN658699). Other Thiotrichales nodes with low relative abundance in our dataset, nodes 76703, 59394 and 71835, were 100% identical to previously published 16S rRNA gene sequences (FR839195 from Rainbow, FM203389 and FM203388 from Logatchev respectively). The relative abundances of the gammaproteobacterial nodes were different between hydrothermal vent fields. Three gammaproteobacterial nodes were relatively most abundant at the Piccard vent site. Four animal samples (A31, A35, A39 and A43) in particular hosted a higher relative abundance of Gammaproteobacteria (node 57431 and 72316), compared to other samples. The samples from Von Damm hosted a lower relative abundance of different Gammaproteobacteria taxa. Nodes 76707 and 61981 were

present in low relative abundance in all individuals except for three, in which they were present in a higher relative abundance. Additionally, one Von Damm environmental sample was dominated by a node (58418) related to Piscirikettsiaca, which was found in very low relative abundance in other samples, or not at all.

Alphaproteobacteria were also present in low relative abundance, the main nodes being related to Rhodobacterales and Rhodospirillaceae. They were similar to sequences associated with *R. exoculata* (AM412521) and were not closely related to any cultivated sequences (83% related to *Pelagibus litoralis*). Both animal and environmental Von Damm samples hosted a relatively more abundant population of Alphaproteobacteria. One environmental sample was dominated by a node (node 77540) related to Rhodobacterales.

There were 106 nodes related to Deltaproteobacteria, the relatively most abundant one occurring across all samples in low relative abundance and being related to the SAR324 clade. These sequences were 97% identical to sequences isolated from *R. exoculata* (FR839088).

Only one node (57386) related to Zetaproteobacteria was present in our data set, in low relative abundances and irregularly found across all samples. The node was affiliated to the Mariprofundus genus and was 97% identical to a sequence isolated from water surrounding *R. exoculata* shrimps (FR839250).

5.5 Discussion

Our work showed that the ectosymbiotic bacterial communities associated with the *R. hybisae* shrimp were statistically different between the two vent sampling sites but were similar to the bacterial community profiles of the surrounding environment. The MCSC hydrothermal vent fields, Von Damm and Piccard, are a unique setting to study ectosymbiotic associations. The two sites are geographically close but are very different in physical, geological and chemical settings. Von Damm vent fluids have a high concentration of methane and low concentration of hydrogen sulfide,

whereas the Piccard vent fluids have higher concentrations of hydrogen sulfide and low concentrations of methane (Reeves *et al*., 2014; Mcdermott *et al*., 2015). This chemical variability was thought to be responsible for the differences in the environmental microbial community compositions (Reveillaud *et al*., 2015). Our study supports such a hypothesis by showing statistically significant differences in the community composition of both the animal-associated and free-living bacteria from Von Damm and Piccard hydrothermal vent fields. Furthermore, we propose the hypothesis that the ectosymbionts are taken up locally from the environment, which would explain the different community profiles from the different sampling sites.

We also compared the *R. hybisae* associated bacterial population profiles from the two MCSC hydrothermal vent fields to *R. exoculata* associated bacterial community profiles from the geographically distant Logatchev hydrothermal vent field on the NMAR. Interestingly, samples from Von Damm and Logatchev were similar to each other, while *Rimicaris*-associated bacteria from Logatchev and Piccard were significantly different. Both Von Damm and Logatchev hydrothermal vent fields are hosted on a ultramafic base and have a similar depth of 2500 m (Schmidt *et al*., 2007); we hypothesize that the geological settings of the hydrothermal vents may have an effect on the animal-associated bacterial community composition. However, we did not have access to environmental samples from the Logatchev hydrothermal vent field, such a sample should be included in a future analysis to confirm the link between geological setting and bacterial communities.

Previous work investigating *R. hybisae* population genetics based on the mitochondrial COI gene and microsatellite markers, showed no population differentiation between the two MCSC shrimp populations (Plouviez *et al*., 2015). This suggested that *R. hybisae*, like other deep-sea shrimps, have a high dispersal potential and are thus not restricted to one vent site but are able to jump between faunal islands. These observations further support our hypothesis that ectosymbiont populations are taken up locally by the host from the environment. Ectosymbionts may have the ability to colonize many

different types of surfaces, biotic or abiotic, and the host may be tolerant to a large diversity and abundance of bacteria colonizing its appendages.

The role of ectosymbionts in the *Rimicaris* symbiosis is still an open question and multiple hypotheses have been discussed. Earlier isotopic studies on *R. exoculata* showed a bacterial origin of the shrimp diet. These observations, coupled with early feeding behavior studies, suggested the shrimp could be "farming" the bacteria on its mouth appendages, and then feeding on them by grooming the scaphognathites and exopodites (Van Dover *et al*, 1988; Gebruk *et al*., 2000). Such "farming" behavior has also been described in other deep-sea crustacean species, such as *Kiwa* crabs (Thurber *et al*., 2011). However, a more recent study showed the absence of scraping or grazing marks on *R. exoculata* gill chambers (Zbinden *et al*., 2004) and direct transintegumental inorganic carbon transfer between bacteria and the host, suggesting a "milking" interaction between the two partners (Ponsard *et al*., 2013). However, "farming" and "milking" are not mutually exclusive. If the host does gain nutrition from its ectosymbionts, then it would profit the most from a bacterial community adapted to local environmental settings. We can hypothesize that the *R. hybisae* association is plastic, hosting a complex ectosymbiotic bacterial community adapted to local physical and chemical conditions. A part of the bacterial community could feed the host through the transfer of nutrients and be grazed on at the same time.

When examined in detail, a taxonomic unit present in large relative abundance often dominated the communities analyzed in our study. These taxonomic units, mainly Epsilonproteobacteria, were phylogenetically closely related. For example, the most abundant node present in the Piccard samples, epsilonproteobacterial node 74707, is 98.5% identical to the most abundant epsilonproteobacterial node 76959 in the Von Damm samples. Because our study only analyzed partial 16S rRNA gene sequences (V3-V4 region), the full-length sequences could prove to be more divergent and these two sequences could be associated with Epsilonproteobacteria from different genera. Such differences could then also be found in metabolic

pathways. However, as mentioned previously, other studies (Reveillaud *et al*., 2015) have examined the community composition differences between free-living populations of the two MCSC vent sites and also investigated the differences in metabolic potential through metagenome analysis. Despite the taxonomic differences, metabolic marker gene analysis showed a remarkable overlap in the functional diversity of the two vent sites. This was unexpected, considering the differences in hydrothermal fluid chemistry at the two sites. Such results suggest metabolic convergence at the two vent sites. The lack of metabolic divergence between hydrothermal vent sites at the MCSC could also support the lack of speciation between bacteria and the *Rimicaris* host, since no particular metabolic capacity is required. However, we cannot exclude that communities behave differently, e.g different gene expression profiles, or are affected by other parameters, such as pressure.

Our very preliminary FISH results from the scaphognathite from the two hydrothermal vent fields showed size difference between the bacteria associated with individuals sampled at Von Damm and Piccard. The Piccard bacteria behind roughly half smaller than the Von Damm one. Many hypotheses could explain this variation, the two different shrimps could have been at two different stage of molting or this could be individual variation. But the pressure could still be a player affecting the morphology of the cells. Nevertheless our low number of replicate cannot answer this question at the moment, but further FISH analyses would statistically support our statement or not.

R. *hybisae* animals from the Von Damm site shared the same dense and rich ectosymbiotic communities as *R. exoculata* individuals from the Logatchev site. From the community taxonomic compositions we could directly hypothesize about the genetic potential of the ectosymbiont populations. These communities are likely to rely on a complex network of metabolic interactions, which could range from supporting the host by fixing inorganic carbon and transferring it to the host, to potential detoxification processes.

The *R. hybisae* samples from both hydrothermal vents were dominated by Epsilonproteobacteria, making them the dominant partner in the ectosymbiosis. We hypothesize that the most abundant bacteria in the community will have a larger impact on the interactions with the host. Previous work on *R. exoculata* has shown that Gamma- and Epsilonproteobacteria remain associated with the shrimp throughout its entire life cycle Even after multiple carapace molts, during which the shrimp is re-colonized by the ectosymbionts, the association appears to stay stable (Zbinden *et al*., 2004, 2012). Moreover, the Gamma- and Epsilonproteobacteria sequences most closely related to the ones predicted in our animal communities have the genetic potential to oxidize sulfur (Ponsard *et al*., 2013; Jan *et al*., 2014). All of this has been previously described to be associated with *R. exoculata* (Petersen *et al*., 2010; Zbinden *et al*., 2012; Jan *et al*., 2014).

The presence of two sulfur oxidizers could be conflicting, because they would compete for the same source of energy. However, the presence of the Gammaproteobacteria sulfur oxidizer could play an important role in the case of an environmental condition switch. Deep-sea Epsilonproteobacteria have been described to be anaerobic to microaerobic (Campbell *et al*., 2006). If the local concentration of oxygen were to increase, therefore, Gammaproteobacteria, which are aerobic, could take over as the main ectosymbiont (Yamamoto and Takai, 2011; Beinart *et al*., 2012) thus adding to the plasticity of the communities to adapt to the local environment.

Other Proteobacteria were present in low relative abundance across all animal samples. The low relative abundance taxa could still play a key role in the symbiosis by maintaining very specific pathways. Zetaproteobacteria have been shown to be associated with *R. exoculata* and genomic investigation indicated the metabolic potential to oxidize iron (Jan *et al*., 2014). These bacteria could add an additional source of energy to the metabolic network present in the host gill chamber or protect it against toxic concentrations of iron.

Finally, to add another player to the network of interactions, Bacteroidetes were consistently present in our amplicon dataset. Two main nodes were predominantly present in the Piccard samples, particularly associated with *R. hybisae* samples. Bacteroidetes, and especially Flavobacteriia, are generally heterotrophic and preferentially consume polymers rather than monomers (Cottrell and Kirchman, 2000). In the oceans, the main lifestyle of Bacteroidetes is particle-attached and polymer-degrading, with the ability to degrade biopolymers such as cellulose and chitin (Kirchman, 2002). As such, the association with crustaceans is thought to be opportunistic, since the Bacteroidetes would have access to a rich source of chitin. A recent study showed that chemolithotrophic Epsilonproteobacteria and chemoorganoheterotrophic Bacteroidetes were tightly associated in hydrothermal vent biofilms (Stokke *et al*., 2015). The study showed that organothrophic bacteroidetes utilize organic polymers and sugar produced by the Epsilonproteobacteria, which in exchange recycle by-production of acetate from the Bacteroidetes into cell material. Our study showed that, at least on a location basis, complex but similar communities with many identical nodes occurred in every sample. Since *R. exoculata* and *R. hybisae* gill chambers have been described as densely covered by bacteria (Komai and Segonzac, 2008; Petersen *et al*., 2010; Nye *et al*., 2011), we hypothesize that a network of beneficial associations within the *Rimicaris* ectosymbioses must be present.

5.6 Conclusion

Our study investigated bacterial populations associated with the deep-sea shrimp *R. hybisae*. Firstly, our work showed that the ectosymbiotic communities associated with *R. hybisae* shrimps are vent specific; these populations are similar to the local environmental free living ones. This finding suggests that the ectosymbiotic association is unspecific and *Rimicaris* shrimps take up bacteria adapted to the local conditions. Finally,

we hypothesize that vent chemistry cannot be the main environmental factor shaping bacterial communities at hydrothermal vents and additional parameters should be investigated. The effect of depth should be next examined to answer this question.

5.7 Acknowledgements

We would like to thank Silke Wetzel and Miriam Sadowski for excellent technical assistance, as well as Rahel Yemanaberhan for her assistance with FISH preparation. We would like to thank Christiane Hassenrueck for her assistance in the statistical analyses, Harald Gruber-Vodicka with advice on amplicon libraries preparations and Pelin Yilmaz for reading and commenting the manuscript. We would like to thank the captains and crews of the Meteor and RV Atlantis ships involved in the gathering of samples for the study.

5.8 References

Akerman, N.H., Butterfield, D.A., and Huber, J.A. (2013) Phylogenetic diversity and functional gene patterns of sulfur-oxidizing subseafloor Epsilonproteobacteria in diffuse hydrothermal vent fluids. *Front. Microbiol.* **4**: 185.

Amann, *R.*I., Krumholz, L., and Stahl, D. a. (1990) Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *J. Bacteriol.* **172**: 762–770.

Beinart, *R.* a., Nyholm,S. V., Dubilier, N., and Girguis, P.R. (2014) Intracellular Oceanospirillales inhabit the gills of the hydrothermal vent snail *Alviniconcha* with chemosynthetic, E-Proteobacterial symbionts. *Environ. Microbiol. Rep.* **6**: n/a–n/a.

Beinart, *R.* a, Sanders, J.G., Faure, *B.*, Sylva, *S.*P., Lee, *R.*W., Becker, E.L., *et al*. (2012) Evidence for the role of endosymbionts in regional-scale habitat partitioning by hydrothermal vent symbioses. *Proc. Natl. Acad. Sci.* **109**: E3241–E3250.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**: 289–300.

Bland, J.A. and Brock, T.D. (1973) The marine bacterium Leucothrix mucor as an algal epiphyte. *Mar. Biol.* **23**: 283–292.

Campbell, *B.*J., Engel, A.S., Porter, M.L., and Takai, K. (2006) The versatile epsilon-proteobacteria: key players in sulphidic habitats. *Nat. Rev. Microbiol.* **4**: 458–68.

Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, *R.*K., and Ellison, A.M. (2014) Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* **84**: 45–67.

Desbruyères, D., Chevaldonné, P., Alayse, A.-M., Jollivet, D., Lallier, F.H., Jouin-Toulmond, C., *et al*. (1998) Biology and ecology of the “Pompeii worm” (*Alvinella pompejana* Desbruyères and Laubier), a normal dweller of an extreme deep-sea environment: A synthesis of current knowledge and recent developments. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **45**: 383–422.

Van Dover, C.L., Fry, *B.*, Grassle, J.F., Humphris, *S.*E., and Rona, P.A. (1988) Feeding biology of the shrimp *Rimicaris exoculata* at hydrothermal vents on the Mid-Atlantic Ridge. *Mar. Biol.* **98**: 209–216.

Dubilier, N., Bergin, C., and Lott, C. (2008) Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat. Rev. Microbiol.* **6**: 725–40.

Duperron, *S.*, Lorion, J., Samadi, *S.*, Gros, O., and Gaill, F. (2009) Symbioses between deep-sea mussels (Mytilidae: Bathymodiolinae) and chemosynthetic bacteria: diversity, function and evolution. *C. R. Biol.* **332**: 298–310.

Eren, A.M., Morrison, H.G., Lescault, P.J., Reveillaud, J., Vineis, J.H., and Sogin, M.L. (2014) Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J.* **9**: 968–979.

Fadrosh, D.W., Ma, *B.*, Gajer, P., Sengamalay, N., Ott, *S.*, Brotman, *R.*M., and Ravel, J. (2014) An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* **2**: 6.

Fernández, *R.*, Kvist, *S.*, Lenihan, J., Giribet, G., and Ziegler, A. (2014) Sine Systemate Chaos? A Versatile Tool for Earthworm Taxonomy: Non-Destructive Imaging of Freshly Fixed and Museum Specimens Using Micro-Computed Tomography. *PLoS One* **9**: e96617.

Frank, D.N. (2009) BARCRAWL and BARTAB: software tools for the design and implementation of barcoded primers for highly multiplexed DNA sequencing. *BMC Bioinformatics* **10**: 362.

Gebruk, A.V., Southward, E.C., Kennedy, H., and Southward, A.J. (2000) Food sources, behaviour, and distribution of hydrothermal vent shrimps at the Mid-Atlantic Ridge. *J. Mar. Biol. Assoc. UK* **80**: 485–499.

German, C.R., Bowen, A., Coleman, M.L., Honig, D.L., Huber, J. a, Jakuba, M. V, *et al*. (2010) Diverse styles of submarine venting on the ultraslow spreading Mid-Cayman Rise. *Proc. Natl. Acad. Sci.* **107**: 14020–14025.

Goffredi, *S.*K. (2010) Indigenous ectosymbiotic bacteria associated with diverse hydrothermal vent invertebrates. *Environ. Microbiol. Rep.* **2**: 479–488.

Goffredi, *S.*K., Jones, W.J., Erhlich, H., Springer, A., and Vrijenhoek, *R.*C. (2008) Epibiotic bacteria associated with the recently discovered Yeti crab, *Kiwa hirsuta*. *Environ. Microbiol.* **10**: 2623–34.

Hodgkinson, M.R.S., Webber, A.P., Roberts, *S.*, Mills, *R.*A., Connelly, D.P., and Murton, *B.*J. (2015) Talc-dominated seafloor deposits reveal a new class of hydrothermal system. *Nat. Commun.* **6**: 10150.

Huber, J.A., Welch, D.M., Morrison, H.G., Huse, *S.*M., Neal, P.R., Butterfield, D.A., and Sogin, M.L. (2007) Microbial population structures in the deep marine biosphere. *Science (80-.).* **318**: 97–100.

Hügler, M. and Sievert, *S.*M. (2011) Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean. *Ann. Rev. Mar. Sci.* **3**: 261–289.

Jan, C., Petersen, J.M., Werner, J., Teeling, H., Huang, *S.*, Glöckner, F.O., *et al*. (2014) The gill chamber epibiosis of deep-sea shrimp *Rimicaris exoculata* : an in-depth metagenomic investigation and discovery of Zetaproteobacteria. *Environ. Microbiol.* **16**: 2723–2738.

Kearse, M., Moir, *R.*, Wilson, A., Stones-Havas, *S.*, Cheung, M., Sturrock, *S.*, *et al*. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–9.

Kim, J.-H., Kim, K.-Y., Hahm, Y.-T., Kim, *B.*-S., Chun, J., and Cha, C.-J. (2008) *Actibacter sediminis* gen. nov., sp. nov., a marine bacterium of the family Flavobacteriaceae isolated from tidal flat sediment. *Int. J. Syst. Evol. Microbiol.* **58**: 139–143.

Kirchman, D.L. (2002) The ecology of Cytophaga-Flavobacteria in aquatic environments. *FEMS Microbiol. Ecol.* **39**: 91–100.

Komai, T. and Segonzac, M. (2008) Taxonomic Review of the Hydrothermal Vent Shrimp Genera *Rimicaris* Williams & Rona and Chorocaris Martin & Hessler (Crustacea: Decapoda: Caridea: Alvinocarididae). *J. Shellfish Res.* **27**: 21–41.

Limaye, A. (2012) Drishti: a volume exploration and presentation tool. *SPIE 8506, Dev. X-Ray Tomogr. VIII, 85060X* **8506**: 85060X.

Loy, A. (2003) DNA Microarray Technology for Biodiversity Inventories of Sulfate-Reducing Prokaryotes. *Ph.D. thesis* - Fakultät Wissenschaftszentrum Weihenstephan

Mcdermott, J.M., Seewald, J.S., German, C.R., and Sylva, *S.*P. (2015) Pathways for abiotic organic synthesis at submarine hydrothermal fields. *Proc. Natl. Acad. Sci. U.*S. *A.* 7668–72.

Mino, *S.*, Kudo, H., Arai, T., Sawabe, T., Takai, K., and Nakagawa,S. (2014) *Sulfurovum aggregans* sp. nov., a hydrogen-oxidizing, thiosulfate-reducing chemolithoautotroph within the Epsilonproteobacteria isolated from a deep-sea hydrothermal vent chimney, and an emended description of the genus *Sulfurovum*. *Int. J. Syst. Evol. Microbiol.* **64**: 3195–3201.

Newton, I.L.G., Woyke, T., Auchtung, T.A., Dilly, G.F., Dutton, *R.*J., Fisher, M.C., *et al*. (2007) The *Calyptogena magnifica* Chemoautotrophic Symbiont Genome. *Science (80-.).* **315**: 998–1000.

Nye, V., Copley, J., and Plouviez,S. (2011) A new species of *Rimicaris* (Crustacea: Decapoda: Caridea: Alvinocarididae) from hydrothermal vent fields on the Mid-Cayman Spreading Centre, Caribbean. *J. Mar. Biol. Assoc. United Kingdom* **92**: 1057–1072.

Oksanen, J., Blanchet, F.G., Kindt, *R.*, Legendre, P., Minchin, P.R., O'Hara, *R.*B., *et al*. (2015) vegan: Community Ecology Package.

Ott, J., Bright, M., and Bulgheresi,S. (2004) Marine microbial thiotrophic ectosymbiosis. *Oceanogr. Mar. Biol. An Annu. Rev.* **42**: 95–118.

Petersen, J.M., Ramette, A., Lott, C., Cambon-Bonavita, M.-A., Zbinden, M., and Dubilier, N. (2010) Dual symbiosis of the vent shrimp *Rimicaris exoculata* with filamentous gamma- and Epsilonproteobacteria at four Mid-Atlantic Ridge hydrothermal vent fields. *Environ. Microbiol.* **12**: 2204–2218.

Plouviez, *S.*, Jacobson, A., Wu, M., and Van Dover, C.L. (2015) Characterization of vent fauna at the Mid-Cayman Spreading Center. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **97**: 124–133.

Ponsard, J., Cambon-Bonavita, M.-A., Zbinden, M., Lepoint, G., Joassin, A., Corbari, L., *et al*. (2013) Inorganic carbon fixation by chemosynthetic ectosymbionts and nutritional transfers to the hydrothermal vent host-shrimp *Rimicaris exoculata*. *ISME J.* **7**: 96–109.

Pruesse, E., Peplies, J., and Glockner, F.O. (2012) SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823–1829.

Reeves, E.P., McDermott, J.M., and Seewald, J.S. (2014) The origin of methanethiol in midocean ridge hydrothermal fluids. *Proc. Natl. Acad. Sci.* **111**: 5474–5479.

Reveillaud, J., Reddington, E., McDermott, J., Algar, C., Meyer, J.L., Sylva, *S.*, *et al*. (2015) Subseafloor microbial communities in hydrogen-rich vent fluids from hydrothermal systems along the Mid-Cayman Rise. *Environ. Microbiol.*

Schloss, P.D., Westcott, *S.*L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al*. (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl. Environ. Microbiol.* **75**: 7537–7541.

Schmidt, K., Koschinsky, A., Garbe-Schönberg, D., de Carvalho, L.M., and Seifert, *R.* (2007) Geochemistry of hydrothermal fluids from the ultramafic-hosted Logatchev hydrothermal field, 15°N on the Mid-Atlantic Ridge: Temporal and spatial investigation. *Chem. Geol.* **242**: 1–21.

Stokke, *R.*, Dahle, H., Roalkvam, I., Wissuwa, J., Daae, F.L., Tooming-Klunderud, A., *et al*. (2015) Functional interactions among filamentous Epsilonproteobacteria and Bacteroidetes in a deep-sea hydrothermal vent biofilm. *Environ. Microbiol.* **17**: 4063–4077.

Thurber, A.R., Jones, W.J., and Schnabel, K. (2011) Dancing for food in the deep sea: bacterial farming by a new species of Yeti crab. *PLoS One* **6**: 1–12.

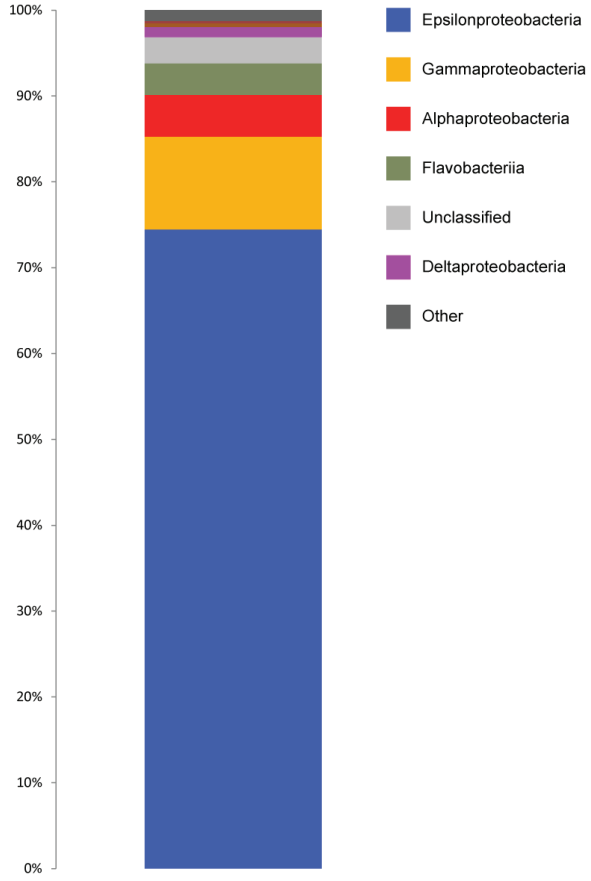
Zbinden, M., Crassous, P., and Shillito, *B.* (2012) Acquisition of epibiotic bacteria along the life cycle of the hydrothermal shrimp *Rimicaris exoculata*. 597–609.

Yamamoto, M. and Takai, K. (2011) Sulfur metabolisms in epsilon- and gamma-proteobacteria in deep-sea hydrothermal fields. *Front. Microbiol.* **2**: 192.

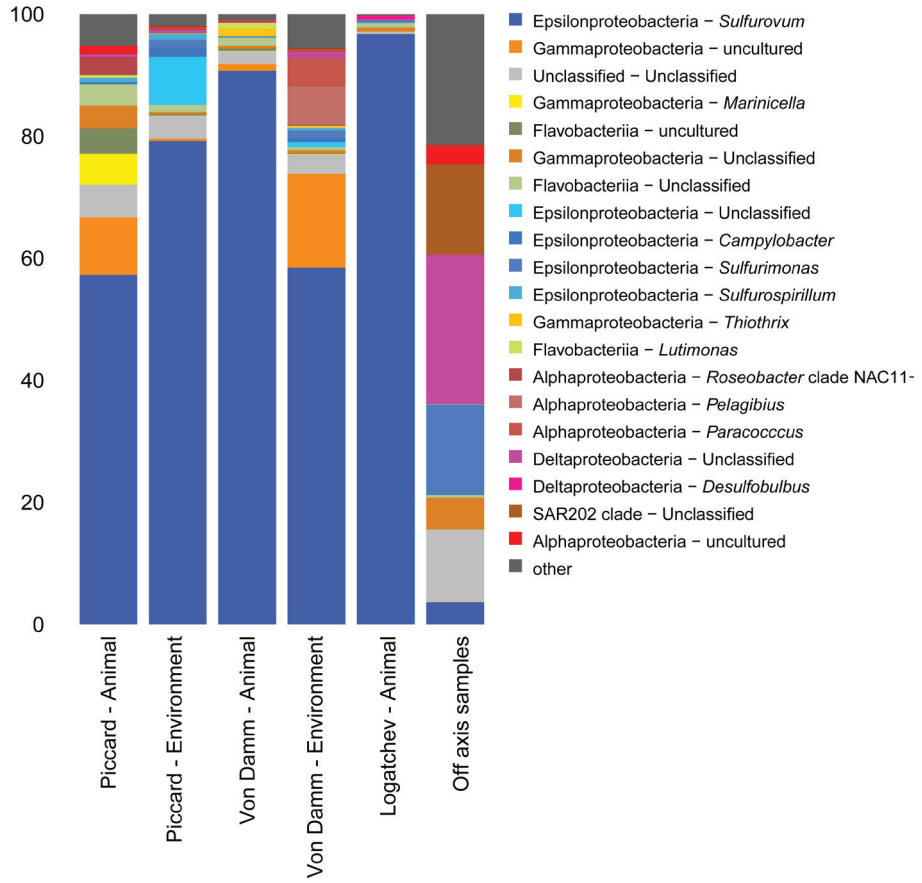
Zhou, J., Bruns, M.A., and Tiedje, J.M. (1996) DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**: 316–322.

Zbinden, M., Crassous, P., and Shillito, *B.* (2012) Acquisition of epibiotic bacteria along the life cycle of the hydrothermal shrimp *Rimicaris exoculata*. 597–609.

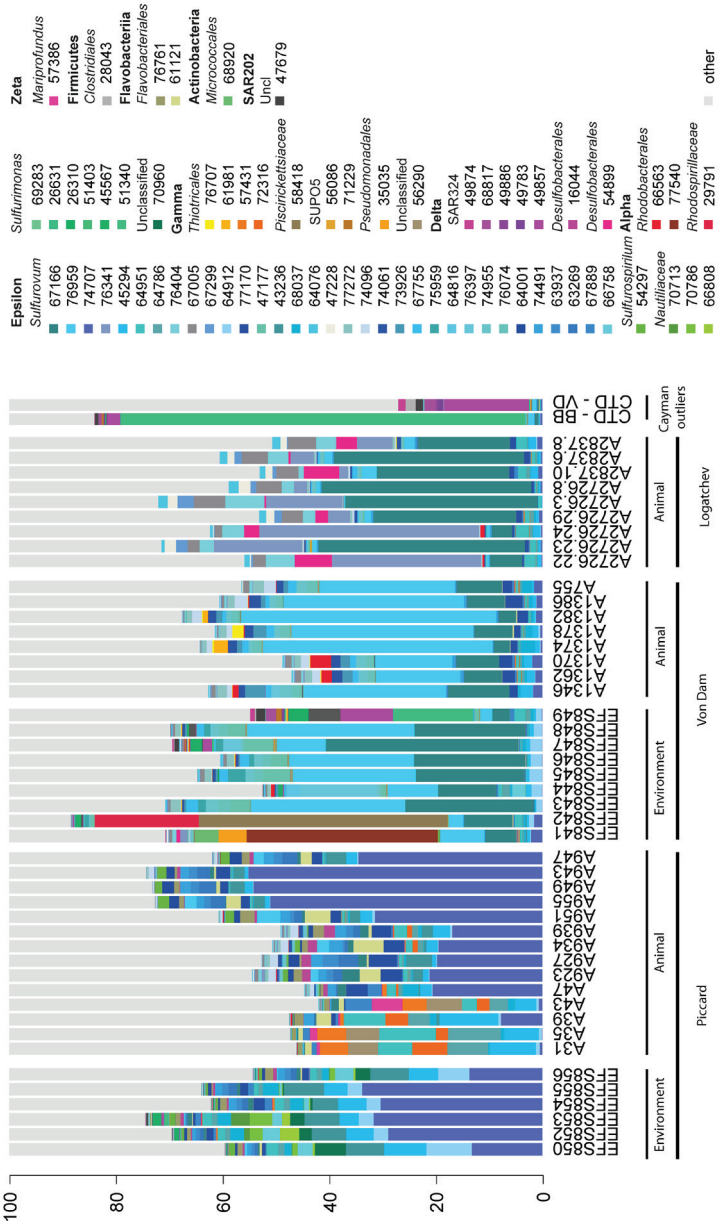
5.9 Supplementary figures and tables



Supplementary Figure 5.1 Relative abundance of dominant bacterial nodes of the entire data set. The eight relatively most abundant bacterial classes are displayed, while the others are collapsed into “other”



Supplementary Figure 5.2 Relative abundance of dominant bacterial nodes of each sample group. The six relatively most abundant bacterial classes are displayed, while others are collapsed into “other”



Supplementary Figure 5.3 Relative abundance of dominant bacterial Nodes and their taxonomic assignment for each individual sample. The samples are grouped by geographic location. The cumulative five most relatively abundant nodes per sample to a total of 77 samples are displayed, the remainder are collapsed into other.

Chapter 6 **General discussion**

6.1 From clone libraries to genome sequencing

For the past two decades, the study of bathymodiolin symbioses has been centered around the gammaproteobacterial endosymbionts present in the gill epithelia. This has been mainly because they are the most abundant players in many different species of bathymodiolin mussels (Fiala-Médioni *et al.*, 1986; Nelson *et al.*, 1995; Miyazaki *et al.*, 2010) and their presence within the host tissue suggests a key role (Duperron *et al.*, 2009) in the symbiotic system. However, multiple other bacteria have been found associated with these systems in low abundance. The work I did during my Ph.D. thesis focused in depth on a second potentially beneficial association between bathymodiolin mussels and a novel family of Epsilonproteobacteria. Firstly, the microscopy analyses reported in Chapter 2 showed that these Epsilonproteobacteria are filamentous epibionts colonizing gill filaments. More importantly, phylogenetic analysis of the 16S rRNA gene also showed that the epibionts formed a new epsilonproteobacterial family that was widespread among bathymodiolin mussels worldwide. These findings are a significant contribution to the study of the diversity and distribution of associations between bacteria and bathymodiolin mussels at hydrothermal vents and cold seeps worldwide. These descriptive data are further complemented by metagenomic and metatranscriptomic approaches presented in Chapters 3 and 4. There I discuss the genetic potential of the epibionts as sulfur oxidizing bacteria capable of fixing carbon via the Calvin Benson Bassham cycle, which is unusual for Epsilonproteobacteria. The datasets presented in Chapters 2, 3 and 4 together suggest a potentially mutualistic association of the epibionts.

The initial phylogenetic reconstruction of the different epsilonproteobacterial 16S rRNA gene sequences discussed in Chapter 2 revealed that the epibionts actually belong to a new family of Epsilonproteobacteria. These sequences cluster together with other

sequences from bacteria associated with bivalves, in a sister family to *Thiovulgaceae (Campbell et al.,* 2006*)*. This family regroups various epsilonproteobacterial genera found free living or associated with invertebrates (e.g. *Sulfurovum*, *Sulfuricurvum*, *Sulfurimonas*, etc.). The genetic information of the two epibiotic species further supported the view that the epsilonproteobacterial epibionts belong to a new family within the Epsilonproteobacteria class. Comparisons of the average amino acid identity (AAI) values with closely related genera, such as *Arcobacter* or *Sulfurovum,* showed a very low sequence similarity, of around 60%, supporting the placement in a different taxonomic unit. However, the phylogenomic reconstruction using multiple Epsilonproteobacteria marker genes was unable significantly to resolve the correct position of the epibionts in the wider Epsilonproteobacteria phylogeny. This was mainly due to a low amount of published Epsilonproteobacteria genomes that would correctly represent the diversity of this bacterial class.

Additionally, the initial 16S rRNA analyses also showed that the two epibionts associated with *"B." childressi* and a *B. azoricus* were probably different strains of the same species, with 1 bp difference between the two 16S rRNA gene sequences (Yarza *et al.*, 2014). However, when I later compared the average nucleotide identity (ANI) of the two genomes, it indicated that the two draft genomes were more different than usually suspected, with an ANI of 82.72%. Based on published studies, this indicates that the two organisms are in fact two clearly defined species (Rodriguez-R and Konstantinidis, 2014), each one specifically associated with its own host.

My genome and transcriptome reconstructions clearly showed that the epibionts were also sulfur oxidizers, capable of oxidizing reduced forms of sulfur such as thiosulfate. So in a symbiosis system such as *B. azoricus,* in which a sulfur oxidizing endosymbiont is present (Duperron *et al.*, 2009), the co-association of two bacteria competing for the same source of energy might be conflicting. My FISH analyses of the gill filaments correlated this and I was able to show that there were different densities of epibionts settled

on the epithelia of the two host species. *"B." childressi,* in which no sulfur oxidizing endosymbionts are present, hosted a dense mat of epibionts on top of its gill epithelia, whereas *B. azoricus* was associated with a sparser population. We hypothesize that the sulfur oxidizing endosymbionts may be protecting their access to reduced sulfur compounds and Sayavedra *et al* (2015) showed that the gammaproteobacterial endosymbiont expressed toxin-like proteins. These toxins, among other mechanisms, may be a way for the gammaproteobacterial endosymbionts to protect their ecological niche exclusivity. Unfortunately, due to a limited amount of fixed material, we could only confirm the epibiotic presence of these bacteria in two species, and comparing the density of associated epsilonproteobacterial epibionts with other host species could be a way to support this hypothesis.

Further analyses of the genome annotations allowed me to assess the genetic potential of the epibionts. One of the most striking features of epsilonproteobacterial epibionts is their ability to fix inorganic carbon through the CBB cycle. The presence of this cycle was highly unusual, because all chemosynthetic Epsilonproteobacteria have been described to fix carbon using the reductive tricarboxylic (rTCA) cycle (Hugler *et al.*, 2005). My analyses showed that the key genes of the rTCA cycle are absent in both draft genomes. Other genes of the rTCA cycle are still present, indicating an ongoing genome erosion process.

Furthermore, I showed that all the genes involved in the CBB cycle probably originated from different horizontal gene transfer (HGT) events. The genes involved in this cycle are present in the two draft genomes in two similar clusters: one grouping the genes coding for the 1,5-bisphosphate ribulose carboxylase (RuBisCO) enzyme and the second grouping all the other genes. The genes coding for the RuBisCO enzyme all share the same gammaproteobacterial origin. On the amino acid level, the genes are all closely related to the gammaproteobacterial clade, which includes the sulfur oxidizing endosymbionts of bathymodiolin mussels. All the other genes involved in the cycle have a betaproteobacterial origin but their exact origin

is unclear since the amino acid phylogenies of the different genes are not consistent. These genes are likely to have originated from another horizontal gene transfer from an organism yet to be discovered.

Another epsilonproteobacterial epibiont gene had a gammaproteobacterial origin, soxB, the key gene of the sulfur oxidation SOX multi enzyme pathway (Wodara *et al.*, 1994). Chemoautotrophic Epsilonproteobacteria usually possess this gene but these genes were significantly different from the gammaproteobacterial one. The gene switch in the epsilonproteobacterial epibionts suggests that having the gammaproteobacterial version of this gene gives them an evolutionary advantage to succeed in colonizing bathymodiolin gills.

Additional genome screening showed that the Epsilonproteobacteria have several genes that could be involved in the adhesion of the bacteria to the mussel's gills. However, no clear indication of contact between the bacteria and host could be made with electron microscopy. Several genes involved in fibronectin adhesion, colanic acid synthesis are present in the genomes of the epibionts and in other Epsilonproteobacteria, such as gastrointestinal tract pathogens *Helicobacter* or *Campylobacter*. These have been shown to play a role in adhesion (Gilbreath *et al.*, 2011).

The genomes annotations also showed small metabolic differences between the epibiont species associated with *"B." childressi* and those of *B. azoricus*. The species associated with *"B." childressi* possess formate dehydrogenase, enabling them to use formate as an additional source of electrons. Additionally, flavocytochrome c dehydrogenase (*fccAB*) and type I sulfide quinone oxidoreductase (*sqr*), both membrane bound enzymes, allow the epibionts additionally to use elemental sulfur as an energy source (Meyer *et al.*, 2000; Brito *et al.*, 2009). The species associated with *B. azoricus,* on the other hand, possess hydrogenases, which allow them to use hydrogen as an energy source, and nitrate reduction to use nitrate as an electron acceptor (Petersen *et al.*, 2011). These examples illustrate the metabolic versatility of epibionts and their ability to adapt to different environmental conditions.

I hypothesized that the epsilonproteobacterial epibionts are mutualists or at least commensalists in their association with the bathymodiolin mussels. My genomic and phylogenetic analyses generated a lot of additional information, which still needs to be carefully processed. First the phylogenetic and phylogenomic analyses of the epsilonproteobacterial epibiont family showed that these bacteria were equally different from predicted mutualistic (Goffredi, 2010; Petersen *et al.*, 2010) and known pathogenic (De Groote *et al.*, 2000) bacteria. The presence of an autotrophic metabolism is a strong indicator against pathogenicity, since very few autotrophic pathogens have been described so far. Despite the fact that I found molecular mechanisms that have been described as playing a role in the adhesion processes of gastrointestinal pathogens (Alemka *et al.*, 2013), these mechanisms could be shared between pathogens and mutualists as a common way of interacting with eukaryotic cells (Hentschel *et al.*, 2000; Nakagawa and Takaki, 2009). The association could be neutral to the host, with little to no advantage for the host and no detriment, but beneficial for the Epsilonproteobacteria. The bacteria could also be beneficial for their host by settling in the mucus secreted by the mussels, where they could protect the gills from environmental infections or transfer nutrients such as vitamins or other metabolites.

Ectosymbiosis has been hypothesized to be the first step toward endosymbiosis (Smith, 1979). Studies have shown that, through the lineage of chemosynthetic Mythilidae, multiple genera of mussels are associated with Gammaproteobacteria at different stages. At the base of the phylogeny, *Benthomodiolus* and *Adipicola* mussels are associated with extracellular symbionts, and further down the phylogeny endosymbiotic associations occur (Miyazaki *et al.*, 2010). We could argue that a similar process is taking place between Epsilonproteobacteria and bathymodiolin mussels. My observation of the difference in density of the epibiont between *"B." childressi* and *B. azoricus* could be explained by the Epsilonproteobacteria taking the ecological niche left vacant by the missing sulfur oxidizing endosymbiont in

"B." childressi. On the other hand, competition could be occurring between two sulfur oxidizing bacteria (i.e. the gammaproteobacterial endosymbiont and the epsilonproteobacterial epibiont) that are using the same energy source in the *B. azoricus* association.

Reasons why these epibionts have not been found in previous studies may be due to the evolution of sample fixation and preparation protocols for both electron microscopy and FISH. Previous harsh treatments may have chemically or mechanically removed the epibionts from the host and recent developments in fixation protocols (Erlandsen, 2004; Montanaro *et al.*, 2016) may have enhanced the retention of mucus and, as a consequence, the retention of epsilonproteobacterial epibionts. Low abundance sequences present in clone libraries have for long been considered to be environmental contaminations, because Epsilonproteobacteria often dominate reduced deep-sea environments. In this study, however, both metagenomic and microscopy approaches confirmed the presence of a novel epibiont associated with bathymodiolin mussels. These bacteria may play a more important role in the bathymodiolin symbiosis than has been previously thought.

6.2 Linking symbiosis and environment

In Chapter 5, I presented the work done investigating the *Rimicaris hybisae* ectosymbiotic bacteria. *R. hybisae* shrimps are a unique system for understanding the relationship between ectosymbiotic bacteria and their host. This species of deep-sea shrimp was found colonizing two hydrothermal vent fields located at the mid-Cayman spreading center (MCSC), which had very distinct chemical and physical characteristics. Von Damm is an ultramafic hydrothermal vent field located at 2500 m, releasing fluids rich in methane but poor in hydrogen sulfide. In contrast, the Piccard vent field is located at a depth of 5000 m has a basaltic base with fluids rich in hydrogen sulfide and poor in methane (Reeves *et al.*, 2014; Hodgkinson *et al.*, 2015).

I analyzed different ectosymbiotic and free-living communities of

bacteria from the two vents using 16S rRNA amplicon libraries. My analysis showed that the Von Damm and Piccard ectosymbiotic communities were significantly different. The global taxonomic distribution appeared to be the same between the two sites when observed on a class or genus level. Epsilonproteobacterial sequences dominated all samples and low relative abundances of Gamma- and Zetaproteobacteria as well as Bacteroidetes were also present. However, when these populations were analyzed on a finer taxonomical level using MED, they appeared to be distinct. Although some overlap still occurred, certain sequences were clearly relatively more abundant at one vent site than at the other. Based on these comparisons, I hypothesized that the ectosymbiotic population of bacteria associated with *R. hybisae* were unique to the environment in which the host is found.

Finally, I compared the *R. hybisae* ectosymbiont population to *R. exoculata* ectosymbionts sampled at the Logatchev vent fields on the Mid-Atlantic Ridge (MAR). Comparing the ectosymbiotic community profiles showed that the *R. exoculata* samples were similar to the *R. hybisae* sampled at Von Damm. Von Damm and Logatchev vent fields share similar environmental settings: they are both located at the same depth range (2500 m and 3000 m, respectively), both are ultramafic vents and share the same chemical profiles (Schmidt *et al.*, 2007). Unfortunately, we lack an environmental free-living bacteria sample from Logatchev properly to complete our data set. Nevertheless, the animal sample sets still support the hypothesis that the ectosymbiont population is influenced by the hydrothermal vent geology and chemistry.

A recent study looking at the metabolism of free living bacterial communities present at the Von Damm and Piccard vent sites based on metagenomic data (Reveillaud *et al.*, 2015) showed that despite a different taxonomic distribution, no obvious metabolic differences could be predicted between the two sites. This would suggest that other parameters than vent chemistry are driving the community composition, such as pressure. Piccard being the deepest vent colonized with megafauna discovered so far would suggest that pressure (500 atm) might play an active role in selecting the

different communities.

The differences in ectosymbiont community composition are similar to those observed when comparing free living community samples from Von Damm and Piccard vent fields. Interestingly, previous studies looking at the population genetic distribution of the host, *R. hybisae*, showed that active genetic flow occurs between the two vents, suggesting that the host is not restricted to one location (Nye *et al.*, 2013). This strongly indicates that the *Rimicaris* shrimp might not disperse with their associated bacteria but instead reacquire them directly from the environment. I hypothesize that this could be beneficial for the host, because they are directly picking up a bacterial population adapted to the local environment.

The interactions between the apparently complex ectosymbiotic communities and the *Rimicaris* hosts are still debated. Many early observations suggested that the shrimps were "farming" bacteria, using their overgrown mouth parts as a base for surface colonizers, which are later rubbed against the gill chamber to collect biofilm excess (Van Dover *et al.*, 1988; Gebruk *et al.*, 2000). Early isotopic studies showed that *R. exoculata* shrimps were likely to have a diet with a bacterial origin. More recent studies have suggested that carbon transfer occurs between the bacteria and the host through the gill chamber, indicating some level of association in which the bacteria could be used for indirect nutritional benefit (Ponsard *et al.*, 2013), whereby the host collects leaking molecules that contribute to its nutrition (i.e. "milking" the bacteria). Nevertheless, "farming" and "milking" are not mutually exclusive and both processes could be happening simultaneously.

6.3 Outlook

6.3.1 Epibionts associated with bathymodiolin mussels

6.3.1.1 Chasing the carbon

The analyses performed during my Ph.D. study revealed the presence of an overlooked Epsilonproteobacteria associated with bathymodiolin

mussels. By using next generation sequencing techniques, we were able to assess the genetic potential of these epibionts. Using metagenome and metatranscriptome analyses, we showed that the epibionts were using an unusual inorganic carbon fixation pathway that is not typically associated with chemoautotrophic Epsilonproteobacteria.

Future research should go back to traditional molecular biology methods to verify and analyze my *in silico* predictions. We have set up an incubation experiment, in collaboration with Nikolaus Leisch, in which *"B." childressi* mussels are incubated with ^{13}C labeled CO_2 to monitor whether labeled compounds are actually incorporated by the epibionts (Appendix A). Preliminary results of bulk isotopic measurements of mussel tissues showed a significant uptake in comparison to non-incubated samples.

Additionally, further analyses could look at the natural abundance of ^{13}C in the epibionts. The different inorganic carbon fixation pathways incorporate different amounts of CO_2 molecules and this has a direct impact on the natural delta ^{13}C values of the produced biomass. This can be used to determine the contribution of individual CO_2 fixation pathways to total primary production and elucidate carbon flow in microbial communities (Biddle *et al*., 2006, Schubotz *et al*., 2011; Olins *et al*., 2013). The CBB and rTCA cycles have distinct isotopic signatures (Figure 1), so using sensitive methods (e.g. NANOSIMS scans), we could directly screen fixed epsilonproteobacterial filaments to measure their natural delta ^{13}C composition. To confirm the use of the CBB cycle by the epsilonproteobacterial epibionts, we could then compare the measured values to other known deep-sea chemoautotrophs, such as the recently cultivated Gammaproteobacteria *Ca*. *Thioglobus singularis*, also relying on the CBB cycle to fix CO_2 (Marshall and Morris, 2015) and *Sulfurovum lithotrophicum,* an Epsilonproteobacteria using the rTCA cycle (Campbell *et al.*, 2006). However, current isotopic knowledge has been derived from laboratory-based experiments and the variation within environmental samples could be harder to interpret because there is often more than one factor influencing the delta ^{13}C signature.

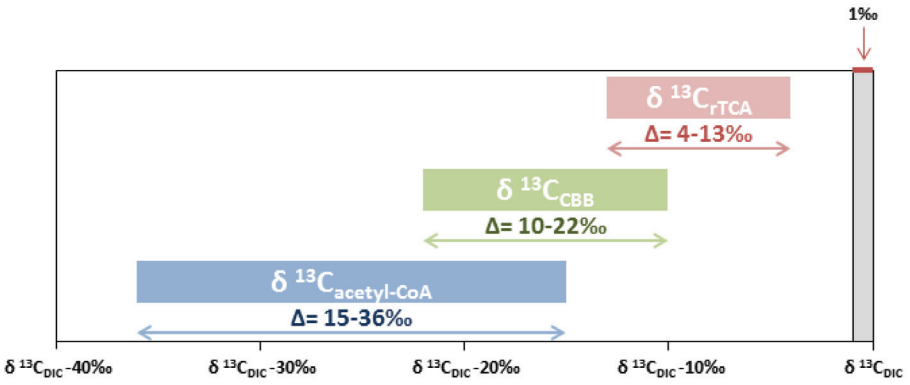


Figure 6.1 Overall carbon isotope fractionation associated with different pathways of autotrophic carbon fixation. Based on data presented in Hayes (2001).

6.3.1.2 There can be only one

Another intriguing feature worth investigation is the general absence of ectosymbiotic-associated bacteria within bathymodiolin mussel species in which no Epsilonproteobacteria have been found. In contrast to the *Rimicaris* ectosymbiotic system, in which a high diversity of ectosymbionts is found, there is only one ectosymbiont species present in *"B." childressi*. On the other hand, metagenomic analyses of *B. azoricus* samples showed the presence of two other likely Epsilonproteobacteria, which have yet to be identified with microscopy. Moreover, 16S rRNA screening of metagenomic libraries of bathymodiolin species such as *B. puteoserpentis,* in which no Epsilonproteobacteria have been detected by PCR, also showed the absence of additional bacteria other than the known gammaproteobacterial endosymbionts.

Bacteria are ubiquitous and will colonize any environment if nothing prevents them, especially biological tissue. Active prevention or protection must be taking place in the mussel gills to prevent any opportunistic and pathogenic bacteria from the environment to harm the host. Previous work with various mussel groups, including *Bathymodiolus azoricus*, has suggested that mucus secreted by the gill tissue acts as a protective layer (Bettencourt *et al.*, 2008). The question raised here is how epsilonproteobacterial

epibionts bypass this protection. The epibionts must have a molecular mechanism enabling them to escape the host's innate immune defenses. Epsilonproteobacteria colonizing mucosal surfaces are found within the *Helicobacter* and *Campylobacter* families, common pathogens of mammalian gastrointestinal tract. Parallels have already been made in the genome description of the epibionts (Chapter 4), discussing that similar mechanisms, such as the N-glycosylation pathway, are found in Epsilonproteobacteria colonizing gills and the pathogens colonizing the gut. This is a complex pathway found in *Helicobacter* and *Campylobacter* used to coat molecules of sugar onto surface proteins in order to avoid detection by the innate immune system of the host (Alemka *et al.*, 2013).

One way to investigate further which mechanism is involved in the epsilonproteobacterial epibiont colonization of the gills would be to analyze the association in a laboratory setting. In recent years, progress in the cultivation of bathymodiolin mussels, with *"B." childressi* and *B. azoricus* in particular (Arellano and Young, 2009; Bettencourt *et al.*, 2011), has allowed the maintenance of mussels in the aquarium for up to a year. Such laboratory setups could be used to investigate various symbiosis mechanisms, by employing both classical molecular microbiology techniques and next generation sequencing. Analysis of the distortion of a symbiotic system by antibiotics, for example, and monitoring its reaction with omics methods such as metatranscriptomics or metabolomics could pin-point the mechanisms involved in the association. What molecules are not expressed when epsilonproteobacterial epibionts or gammaproteobacterial endosymbionts are suddenly absent? What molecules are over-expressed when either of the partners is recolonizing the system? These questions could be answered with a properly monitored experiment.

6.3.2 In depth analysis of *R. hybisae* ectosymbionts

6.3.2.1 Who are the main players?

We showed, using 16S rRNA amplicon libraries, that two populations of *R. hybisae* were colonized by different bacterial communities. This work could be supplemented with microscopy based methods, such as FISH analyses using specific probes. Unfortunately, our 16S rRNA amplicon sequences were too short to design FISH probes specific enough to discriminate the main Epsilonproteobacteria or Gammaproteobacteria players colonizing the two different shrimp species. Further studies involving metagenomics or metatranscriptomics could overcome these limitations by generating full-length small subunit ribosomal sequences. At the same time, these would provide more genetic information to compare and investigate the different genetic potentials associated with the two *R. hybisae* populations. This could highlight genetic patterns playing a role in community shaping, such as adhesion, reconnaissance, chaperone proteins, which could be involved in environmental adaptations.

6.3.2.2 What are the interactions within the ectosymbiotic population?

My Ph.D. thesis paved the way to understanding fully the complexity of *Rimicaris* ectosymbiotic communities. Although mainly dominated by Epsilonproteobacteria, our study showed the presence of multiple additional taxa in low abundance. These taxa could be playing a unique role in a dense metabolic network. Previous work has shown that Zetaproteobacteria, present in all our *Rimicaris* samples, are iron oxidizers and use iron as an energy source (Jan *et al.*, 2014). Bacteroidetes have also been shown to have a syntrophic association with Epsilonproteobacteria, each organism complementing part of a metabolic pathway (Stokke *et al.*, 2015). This diversity present within the *Rimicaris* gill chamber could, taken as a whole, be adapted to a versatile environment and always have bacterial taxa able to dominate the population when the environmental conditions are not suitable for the current one. Additionally, these bacteria could also participate in a

detoxification process, reducing the local impact of free radical or heavy metals, which could be beneficial for the host and the community.

6.3.3 Deep Se(a)quencing

As challenging as deep sea research is, next generation sequencing techniques and new bioinformatics protocols have been extremely useful and are opening new research paths to understanding symbiotic interactions. Data generated by one metagenome can be used to answer multiple questions about the various bathymodiolin symbioses. For example, one could investigate the composition of the associated microbial communities, the metabolic capacity of the whole system, as well as provide genomic information on the eukaryotic host and the prokaryotic symbionts. As more metagenomes of bathymodiolin mussels are sequenced, they will provide information to look at symbioses on the population level and what are the genetic or evolution relationships among samples coming from different sampling sites and different years. Population genomics will allow in depth reliable analyses of population dynamics and will help answer questions such as migration patterns or species radiation across different sites.

Further investigation could be focused on understanding interactions within the holobiont. What gene compositions are unique to one symbiotic system? What are the common gene families or domains when comparing multiple bathymodiolin species? These questions can and will be answered with progress in bioinformatics. Biology is entering an era of “Big Data”, in which affordable next generation sequencing is producing very large amount of data (Baker, 2010; Marx, 2013; Dolinski and Troyanskaya, 2015) and needs new methods to be analyzed, such as Network analyses and Machine learning algorithms. These new giants bring new questions, understanding systems not only one gene at the time but all of them with a tremendous number of replications. Software such as ‘SpeakEasy‘ has already been used to try to understand the relationship between samples in large and complex datasets (Gaiteri *et al.*, 2015). Such network analyses could be performed to compare

the co-occurrence of known and unknown eukaryotic and prokaryotic genes and identify genes involved in specific interaction processes between host and symbiont. Including free-living communities of closely related partners in such complex analyses could also help to identify sets of genes or gene families necessary to define one bacterial species as a symbiont.

References

Alemka, A., Nothaft, H., Zheng, J., and Szymanski, C.M. (2013) N-glycosylation of *Campylobacter* jejuni surface proteins promotes bacterial fitness. *Infect. Immun.* **81**: 1674–1682.

Arellano, *S.*M. and Young, C.M. (2009) Spawning, development, and the duration of larval life in a deep-sea cold-seep mussel. *Biol. Bull.* **216**: 149–62.

Baker, M. (2010) Next-generation sequencing: adjusting to data overload. *Nat. Methods* **7**: 495–499.

Bettencourt, *R.*, Costa, V., Laranjo, M., Rosa, D., Pires, L., Colaco, A., *et al*. (2011) Out of the deep sea into a land-based aquarium environment: investigating physiological adaptations in the hydrothermal vent mussel *Bathymodiolus azoricus*. *ICES J. Mar. Sci.* **68**: 357–364.

Bettencourt, *R.*, Dando, P., Rosa, D., Riou, V., Colaço, A., Sarrazin, J., *et al*. (2008) Changes of gill and hemocyte-related bio-indicators during long term maintenance of the vent mussel *Bathymodiolus azoricus* held in aquaria at atmospheric pressure. *Comp. Biochem. Physiol. - A Mol. Integr. Physiol.* **150**: 1–7.

Brito, J.A., Sousa, F.L., Stelter, M., Bandeiras, T.M., Vonrhein, C., Teixeira, M., *et al*. (2009) Structural and functional insights into sulfide:quinone oxidoreductase. *Biochemistry* **48**: 5613–5622.

Campbell, *B.*J., Engel, A.S., Porter, M.L., and Takai, K. (2006) The versatile epsilon-proteobacteria: key players in sulphidic habitats. *Nat. Rev. Microbiol.* **4**: 458–68.

Dolinski, K. and Troyanskaya, O.G. (2015) Implications of Big Data for cell biology. *Mol. Biol. Cell* **26**: 2575–2578.

Van Dover, C.L., Fry, *B.*, Grassle, J.F., Humphris, *S.*, and Rona, P.A. (1988) Feeding biology of the shrimp *Rimicaris exoculata* at hydrothermal vents on the Mid-Atlantic Ridge. *Mar. Biol.* **98**: 209–216.

Duperron, *S.*, Lorion, J., Samadi, *S.*, Gros, O., and Gaill, F. (2009) Symbioses between deep-sea mussels (Mytilidae: Bathymodiolinae) and chemosynthetic bacteria: diversity, function and evolution. *C. R. Biol.* **332**: 298–310.

Erlandsen, *S.*L. (2004) High-resolution Visualization of the Microbial Glycocalyx

with Low-voltage Scanning Electron Microscopy: Dependence on Cationic Dyes. *J. Histochem. Cytochem.* **52**: 1427–1435.

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

Gaiteri, C., Chen, M., Szymanski, *B.*, Kuzmin, K., Xie, J., Lee, C., *et al*. (2015) Identifying robust communities and multi-community nodes by combining top-down and bottom-up approaches to clustering. *Sci. Rep.* **5**: 16361.

Gebruk, A.V., Southward, E.C., Kennedy, H., and Southward, A.J. (2000) Food sources, behaviour, and distribution of hydrothermal vent shrimps at the Mid-Atlantic Ridge. *J. Mar. Biol. Assoc. UK* **80**: S0025315400002186.

Gilbreath, J.J., Cody, W.L., Merrell, D.S., and Hendrixson, D.R. (2011) Change is good: variations in common biological mechanisms in the epsilonproteobacterial genera *Campylobacter* and *Helicobacter*. *Microbiol. Mol. Biol. Rev.* **75**: 84–132.

Goffredi, *S.*K. (2010) Indigenous ectosymbiotic bacteria associated with diverse hydrothermal vent invertebrates. *Environ. Microbiol. Rep.* **2**: 479–488.

De Groote, D., Ducatelle, *R.*, and Haesebrouck, F. (2000) *Helicobacter*s of possible zoonotic origin: A review. In, *Acta Gastro-Enterologica Belgica*., pp. 380–387.

Hentschel, U., Steinert, M., and Hacker, J. (2000) Common molecular mechanisms of symbiosis and pathogenesis. *Trends Microbiol.* **8**: 226–31.

Hodgkinson, M.R.S., Webber, A.P., Roberts, *S.*, Mills, *R.*A., Connelly, D.P., and Murton, *B.*J. (2015) Talc-dominated seafloor deposits reveal a new class of hydrothermal system. *Nat. Commun.* **6**: 10150.

Hugler, M., Wirsen, C.O., Fuchs, G., Taylor, C.D., and Sievert, *S.*M. (2005) Evidence for Autotrophic CO2 Fixation via the Reductive Tricarboxylic Acid Cycle by Members of the Subdivision of Epsilon Proteobacteria. *J. Bacteriol.* **187**: 3020–3027.

Jan, C., Petersen, J.M., Werner, J., Teeling, H., Huang, *S.*, Glöckner, F.O., *et al*. (2014) The gill chamber epibiosis of deep-sea shrimp R imicaris

exoculata : an in-depth metagenomic investigation and discovery of Z etaproteobacteria. *Environ. Microbiol.* **16**: 2723–2738.

Marshall, K.T. and Morris, *R.*M. (2015) Genome Sequence of " *Candidatus* Thioglobus singularis" Strain PS1, a Mixotroph from the SUP05 Clade of Marine Gammaproteobacteria. *Genome Announc.* **3**: e01155–15.

Marx, V. (2013) Biology: The big challenges of big data. *Nature* **498**: 255–260.

Meyer, T.E., Cusanovich, M.A., Kostanjevecki, V., Brige, A.N.N., Guisez, Y., and Beeumen, J.V.A.N. (2000) A Membrane-Bound Flavocytochrome FCC from E.vacuolata. **182**: 3097–3103.

Miyazaki, J.-I., Martins, L. de O., Fujita, Y., Matsumoto, H., and Fujiwara, Y. (2010) Evolutionary Process of Deep-Sea *Bathymodiolus* Mussels. *PLoS One* **5**: e10363.

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

Nakagawa,S. and Takaki, Y. (2009) Nonpathogenic Epsilonproteobacteria. *eLS* 1–11.

Nelson, D.C., Hagen, K.D., and Edwards, D.B. (1995) The gill symbiont of the hydrothermal vent mussel *Bathymodiolus* thermophilus is a psychrophilic, chemoautotrophic, sulfur bacterium. *Mar. Biol.* **121**: 487–495.

Nye, V., Copley, J.T., and Tyler, P.A. (2013) Spatial Variation in the Population Structure and Reproductive Biology of *Rimicaris hybisae* (Caridea: Alvinocarididae) at Hydrothermal Vents on the Mid-Cayman Spreading Centre. *PLoS One* **8**: e60319.

Petersen, J.M., Ramette, A., Lott, C., Cambon-Bonavita, M.-A., Zbinden, M., and Dubilier, N. (2010) Dual symbiosis of the vent shrimp *Rimicaris exoculata* with filamentous gamma- and epsilonproteobacteria at four Mid-Atlantic Ridge hydrothermal vent fields. *Environ. Microbiol.* **12**: 2204–2218.

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

Ponsard, J., Cambon-Bonavita, M.-A., Zbinden, M., Lepoint, G., Joassin, A., Corbari, L., *et al*. (2013) Inorganic carbon fixation by chemosynthetic

ectosymbionts and nutritional transfers to the hydrothermal vent host-shrimp *Rimicaris exoculata*. *ISME J.* **7**: 96–109.

Reeves, E.P., McDermott, J.M., and Seewald, J.S. (2014) The origin of methanethiol in midocean ridge hydrothermal fluids. *Proc. Natl. Acad. Sci.* **111**: 5474–5479.

Reveillaud, J., Reddington, E., McDermott, J., Algar, C., Meyer, J.L., Sylva, *S.*, *et al*. (2015) Subseafloor microbial communities in hydrogen-rich vent fluids from hydrothermal systems along the Mid-Cayman Rise. *Environ. Microbiol.*

Schmidt, K., Koschinsky, A., Garbe-Schönberg, D., de Carvalho, L.M., and Seifert, *R.* (2007) Geochemistry of hydrothermal fluids from the ultramafic-hosted Logatchev hydrothermal field, 15°N on the Mid-Atlantic Ridge: Temporal and spatial investigation. *Chem. Geol.* **242**: 1–21.

Smith, D.C. (1979) From extracellular to intracellular: the establishment of a symbiosis. *Proc. R. Soc. Lond. B. Biol. Sci.* **204**: 115–30.

Stokke, *R.*, Dahle, H., Roalkvam, I., Wissuwa, J., Daae, F.L., Tooming-Klunderud, A., *et al*. (2015) Functional interactions among filamentous Epsilonproteobacteria and Bacteroidetes in a deep-sea hydrothermal vent biofilm. *Environ. Microbiol.* **17**: 4063–4077.

Wodara, C., Kostka, *S.*, Egert, M., Kelly, D.P., and Friedrich, C.G. (1994) Identification and sequence analysis of the soxB gene essential for sulfur oxidation of Paracoccus denitrificans GB17. *J. Bacteriol.* **176**: 6188–6191.

List of publications and manuscripts with author's contribution

Chapter 2 - A specific and widespread association between deep-sea *Bathymodiolus* mussels and a novel family of Epsilonproteobacteria

Authors: Adrien Assié, Christian Borowski, Karina van der Heijden, Luciana Raggi, Benedikt Geier, Nikolaus Leisch, Mario P. Schimak, Nicole Dubilier, Jillian M Petersen

Published in "Environmental Microbiology Reports"

Author contributions: 'Generated sequence data that was essential for developing the project: K.vdH and L.R, developed the concept A.A., C.B, J.P and N.D., PCR and NGS libraries screening: A.A.; Phylogenetic analyses: A.A.; FISH analyses A.A, 3D reconstruction: *B.*G, electron microscopy: N.L. and M.S., figures and tables A.A., writing of manuscript A.A. and J.P., all authors read and commented on the manuscript.

Chapter 3 - I wanna be like you: Multiple horizontal gene transfers in an epsilonproteobacterial epibiont of bathymodiolin mussels.

Authors: Adrien Assie, Harald Gruber-Vodicka, Samantha Joye, Matthew Saxton, Halina Tegetmeyer, Nicole Dubilier, Jillian M. Petersen

Manuscript in preparation

Author contributions: developed the concept A.A., J.P and N.D., Bioinformatic analyses: A.A and H.G.V, figures and tables A.A., sequencing: H.T., sample collection and processing in the field: *S.*J and M.S., writing of manuscript A.A. and J.P., not all co-authors have read and commented on this draft version.

Chapter 4 - Same but different: Genomes of epsilonproteobacterial epibionts associated with bathymodiolin mussels.

Authors: Adrien Assie, Harald Gruber-Vodicka, Samantha Joye, Matthew Saxton, Halina Tegetmeyer, Nicole Dubilier, Jillian M. Petersen

Manuscript in preparation

Author contributions: concept A.A., J.P and N.D., Bioinformatic analyses: A.A and H.G.V, figures and tables A.A., sequencing: H.T., sample collection and processing in the field: *S.*J and M.S., writing of manuscript A.A. and J.P., not all co-authors have read and commented on this draft version.

Chapter 5- It's all about location: The deep sea shrimp *Rimicaris hybisae* is associated with an ectosymbiotic population of bacteria adapted to the hydrothermal vent field settings.

Authors: Adrien Assié, Julie Huber, Julie Reveillaud, Cindy van Dover, Benedikt Geier, Christian Borowski, Nicole Dubilier, Jillian Petersen.

Manuscript in preparation

Author contributions: concept A.A., J.P, C.B. and N.D., Bioinformatic analyses: A.A, FISH: A.A, figures and tables A.A., µC.T: *B.*G, sequencing: H.T., sample collection and processing in the field: J.H, J. *R.* and C.D., writing of manuscript: A.A. and J.P., not all co-authors have read and commented on this draft version.

Acknowledgment

So many people, so many thanks:

Prof. Dr. Nicole Dubilier for giving me the opportunity to do my Ph.D. thesis in the Symbiosis department. Thank you for teaching me the scientific writing, to communicate my ideas the clear way, the valuable scientific and career advice.

Dr. Jillian Petersen, I could never thank you enough for the mentoring along the past years, for your infinite patience, your scientific advice and all the help you provided for me to grow as a scientist!

Dr. Sebastien Duperron for accepting to review this thesis. And also for all the different scientific advices and discussions we had across Europe and the Atlantic.

Prof. Dr. Friedrich for kindly accepting to be the chair of the thesis examination board, I truly appreciate.

Clara Martinez Petez and **Cedric Han** for having accepted to be part of my thesis examination board.

Dr. Christian Borowski For the supervision, the support, the introduction to the German way and for that time where you took me to the middle of the Atlantic and we tried to sample half of it.

Liz, Juliane and **Harald** for your friendship and your endless patience and support when I started to ask questions about bioinformatic!

Silke, Miriam and **Martina** for the beautiful lab expertise everything would have been so different without your support!

Mario, compagnon de galère, for being there from the start and through the ups and downs, the whisky, the drum and the bass!

Niko for the support, the scientific discussion and the friendship. For being around for speed bowling idea sessions, the Friday beers and other feasts!

Benedikt for the magnificent 3D expertise, the animal scans and other pretty pictures!

Pelin, Petra and **Pierre** for the reviews and the invaluable suggestions on the different chapters of this thesis.

The Symbionts and the Molies for being such a “Wunderbar” team and friends beyond colleagues !

Bernd Stickfort, for always being able to find all those publications, including obscure French ones.

The friends and flatmates, Manu, Judith, Rodrigo, Dimitri, Clara, Saar, Jime, Gerd, Georgio, Niels, Cameron, Oliver, Greta, Christianne, the Marina(s) and Union 60 rugby for being part of this crazy roller-coaster across the long day and the endless night that the past four years have been in Bremen.

A bunch of European Raccoons including Mariana, Julia, Flavia, Alejandro, Dani, Kamil and all the others for being amazing friend through the Marie Curie ITN7 program !

The other french Ambroise, Magalie, Bastien, Antonin, Lucie, Camille, Mario et tous les autres pour toujours avoir été la quand je rentre au pays!

My Family for the unconditional love and support since the very beginning! *“Tchembé rèd pa moli...”*

Sara for being Crazy with me, and endorsing my ever growing insanity. For being always there to correct my gibberish and broken English and especially always up to go on the next stupid adventure! May they be plenty!

And you! you whom I have forgot, because I always forgot someone, you whom have open my thesis even if you only read the acknowledgments!

Appendix A. Incubation experiment to check the incorporation of CO_2 by the Epsilonproteobacteria.

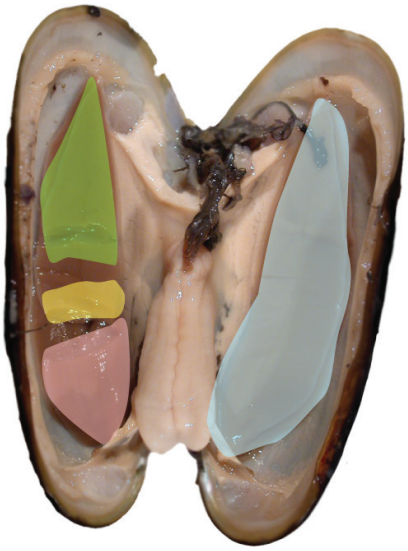


Figure Appendix 1 Photo of a an open *"B." childressi* with dissection plan.

Appendix A.1 Introduction

The first part of this Ph.D. work explained the in silico prediction of a Calvin Benson Bassham (CBB) cycle in chemoautotrophic Epsilonproteobacteria associated with bathymodiolin mussels. One suggested outlook was to correlate the metagenomic and metatranscriptomic prediction with molecular biology methods. To do so we planned an incubation experiment to trace the incorporation of ^{13}C labeled bicarbonate by the epsilonproteobacterial epibiont associated with *B. childressi*. This incubation experiment took place on the Atlantis cruise (NA 058) in April 2015 and was performed by Nikolaus Leisch.

"B." childressi offered multiple advantages for the incubation. Firstly, our analyses have shown that this mussel species hosts a large community of epsilonproteobacterial epibionts. Secondly, the gammaproteobacterial

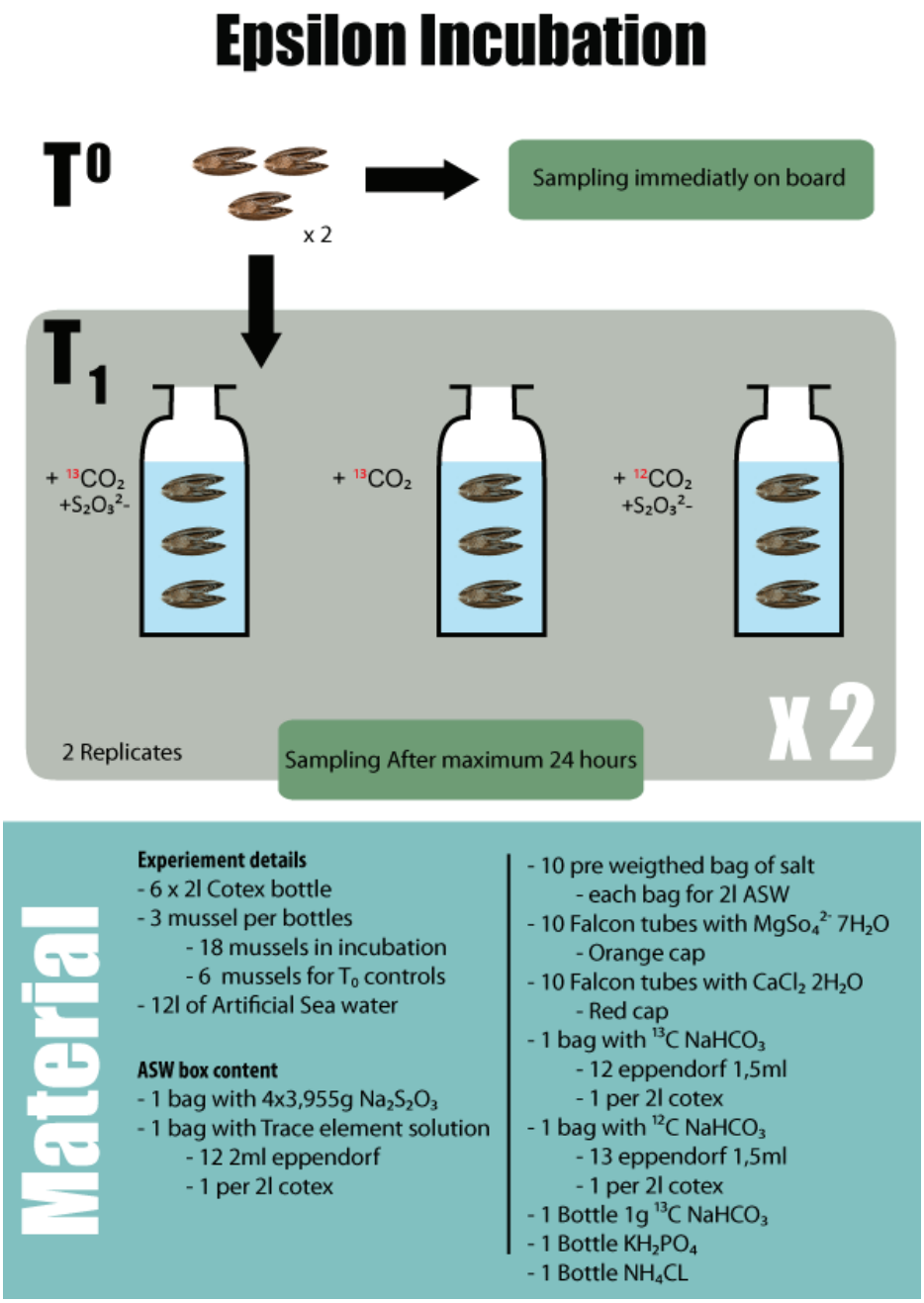


Figure Appendix 2 Schematic summarizing the incubation experiment.

DNA analysis would be used to show the presence of the

Epsilonproteobacteria in the sample.

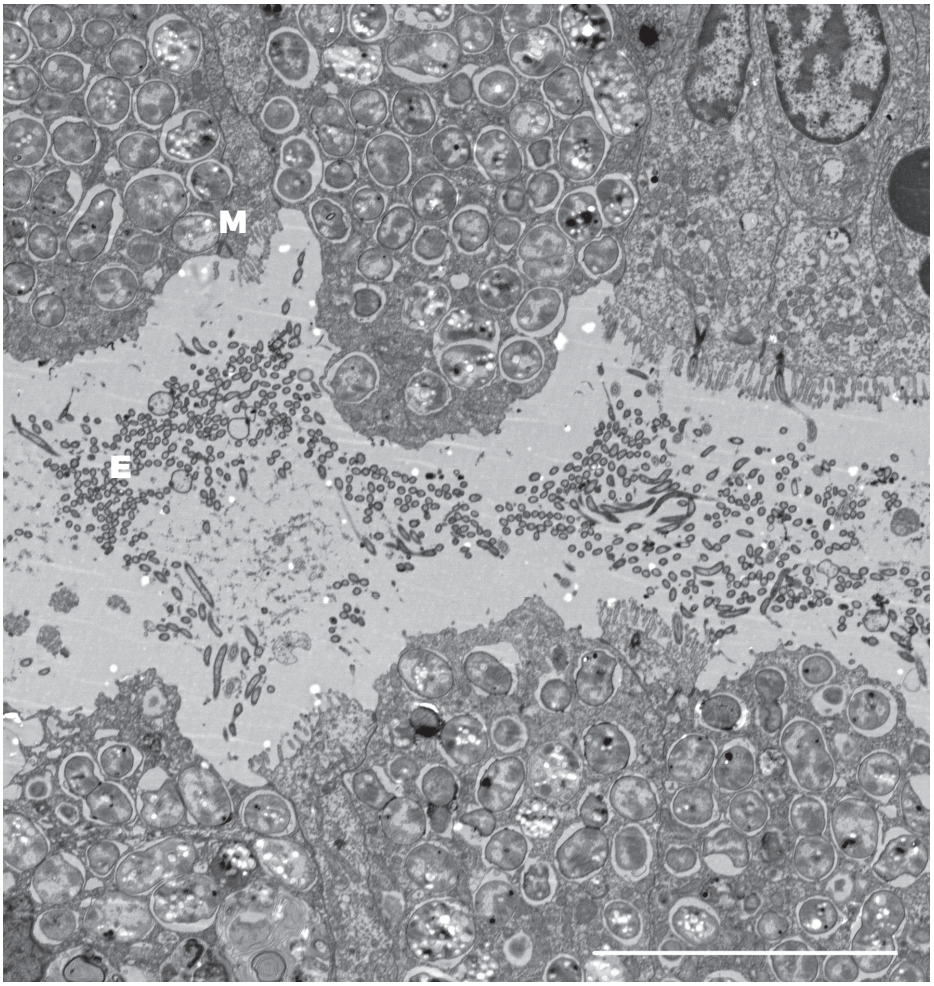


Figure Appendix 3 Transmission electron microscopy picture of a gill cross section from an incubated "B." childressi sample. **M.** Metanotroph, **E.** Epsilonproteobacterial epibiont

Chemical concentration:

The water to incubate the mussels in should have been artificial see water prepared without $^{13/12}C$ source. Depending on the size of the mussel they would be placed in a 1 or 2 liter cotex bottle with an aquarium pump. Then we would add Thiosulfate and Bicarbonate to the final concentration of 1 mM of $S_2O_3^{2-}$ and 1 mM $^{13}CO_2$ to the medium.

However due to unforeseeable issues during transit, we were not able to use pre weighted chemicals for the incubation thus we were not able to prepare artificial sea water or control the end concentration of $^{13}CO_2$. Filtered deep-seawater was used instead and an estimated weight of $^{13}CO_2$

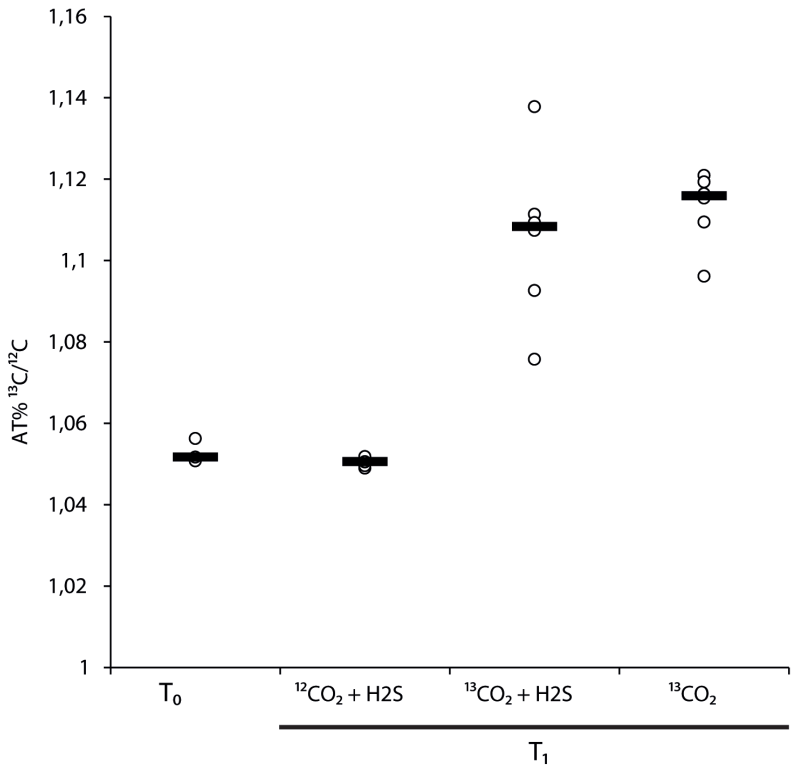


Figure Appendix 4 Plot displaying the 13C/12C ratio value of the different samples from the incubation experiment. Each circle represent a sample, the line the median values.

was used for the incubation.

Appendix A.3 Preliminary methods and results

DNA extraction and Epsilonproteobacteria screening

At the end of the incubation we had a total of 24 samples to process. We performed PCR with specific primers for the epsilonproteobacterial

epibionts according the method described in Chapter 1 and confirmed the epibiont presence in all samples.

Transmission electron microscopy

We did transmission electron microscopy to investigate the abundance of the epibionts in different samples and if the samples were fit for further analyses such as NanoSim scans. Analyses revealed an abundant presence of bacterial filaments between the mussel's gill filaments. Figure 2 below is an example of the filament abundance between the gills.

Mass spectrometry analyses

To check whether or not the incubated animals took up $^{13}CO_2$ we did bulk isotopic measurements using an EA-IRMS. We analyzed all 24 samples using the same protocol, as follows. We sampled 10 mg of gill tissues, rinsed them in one time PBS solution and then dried the sample in a glass vial overnight at 60°C. The dry samples were homogenized and around 200 µg of tissue was taken for mass spectrometry measurement. Samples were measured by Clara Martinez Perez.

Preliminary $^{13}C/^{12}C$ ratios of the different samples are displayed in Figure 4 below.

These preliminary analyses showed high $^{13}C/^{12}C$ ratio in samples incubated with 13C. Although our experiment would have greatly benefited from an additional negative control such as dead mussel incubated with $^{13}CO_2$ or mussel incubated without the addition of CO_2 this incubation, the different values of T0 and T1 with $^{13}CO_2$ still suggest the active uptake of $^{13}CO_2$.

Appendix A.4 Outlook

Our preliminary results confirmed the abundant presence of epsilonproteobacterial epibionts in the different samples we used for the incubation and the preliminary bulk isotopic ratio measurement indicated that inorganic carbon was taken up during the experiment. From these results we now need to confirm the specific uptake of compounds by the

epsilonproteobacterial epibionts. To do so we plan to do NanoSIM scans of gill section and measure the local ratio of $^{13}C/^{12}C$, hopefully confirming the specific uptake of inorganic carbon by the Epsilonproteobacteria.

Appendix B. Digital supplements

Digital supplements are supportive information of the different chapter which are not fitting within the printed format of the thesis. They were originally present on a CD with the physical copy of the thesis.

These documents are available on request from Prof. Dr. Nicole Dubilier: ndubilie@mpi-bremen.de or from the Max Planck Institute for Marine Microbiology: contact@mpi-bremen.de.