Dynamics of Symbiont Abundance in Bathymodiolin Deep-sea Symbioses

Dissertation
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
- Dr. rer. nat. -

dem Fachbereich Biologie/Chemie der
Universität Bremen
vorgelegt von

Dennis Fink

Bremen
September 2011

1. Gutachterin: Dr. Nicole Dubilier

2. Gutachter: Prof. Dr. Ulrich Fischer

Tag der Verteidigung: 19. Oktober 2011

Die vorliegende Arbeit wurde in der Zeit von April 2008 bis September 2011 in der Symbiose-Gruppe am Max-Planck-Institut für marine Mikrobiologie in Bremen angefertigt.

Don't you remember when we were young
and we wanted to set the world on fire?
Cause I still am and I still do.

Rise Against

Abstract

Deep-sea hydrothermal vents and cold seeps are widespread throughout the world's oceans and represent 'oases of life' in a dark and hostile environment. These chemosynthetic habitats are often dominated by mussels of the genus *Bathymodiolus*. The bivalves harbor bacterial symbionts in their gills that use methane and reduced inorganic compounds such as sulfide and hydrogen as energy sources for chemosynthetic primary production. It is well known that the spatial and temporal gradients of these energy sources can be extremely steep at vents and seeps, and some studies have shown that differences in energy source availability affect symbiont abundance in *Bathymodiolus* mussels. However, in-depth analyses of physico-chemical gradients and their effect on symbiont abundance are lacking. One of the basic requirements for these analyses is a reliable quantification method for the symbionts. The goal of my thesis was therefore to develop an accurate and efficient protocol for determining symbiont abundance in bathymodiolin mussels. I used both a stereological and a molecular method for quantifying symbiont abundance. For the former, I employed three-dimensional fluorescence *in situ* hybridization (3D-FISH) to quantify the biovolume of the symbionts inside the mussel gills. For the latter, quantitative polymerase chain reaction (qPCR) was used to assess the relative concentrations of a gene from the host (18S rRNA) versus two symbiont genes (16S rRNA and *recA*). The extensive optimization of these methods was the major part of this thesis work and included improvements in DNA extraction, primer design and automated data analysis. The comparison of these two methods revealed, that qPCR is more reliable and accurate than 3D-FISH in the quantification of symbiont abundances in bathymodiolin symbioses.

After optimizing these methods, I used them to analyze symbiont abundance in *Bathymodiolus* mussels from hydrothermal vents and cold seeps. In an experimental *in situ* approach, mussels from hydrothermal vents on the Mid-Atlantic Ridge were removed from active venting for up to 14 days to simulate a radical temporal decrease in their energy sources. The thiotrophic and methanotrophic populations of *B. puteoserpentis* were almost completely depleted after 10 days of starvation, while both symbiont species in the *B. azoricus* mussels were left with about 30% of their original abundance after 14 days of starvation. When displaced mussels were returned to the hydrothermal fluids that carry their energy sources, methanotrophic symbiont populations of mussels increased, while the thiotrophic symbiont population did not recover. This could indicate symbiont-specific differences in the reaction to changing energy source availability.

The effect of spatial gradients in energy sources on symbiont abundance was analyzed on *Bathymodiolus* mussels from two cold seeps in the Gulf of Mexico. *In situ* measurements of sulfide and methane at the two seep sites revealed significant differences in the concentrations of these two energy sources. These differences were reflected in corresponding differences in symbiont abundance, indicating a link between local energy source availability and symbiont abundance.

My results emphasize the flexibility of the bathymodiolin symbiosis in regard to temporal and spatial variations in energy source concentrations and help to answer the question, why these bivalves dominate hydrothermal vents and cold seeps in the dark ocean.

Zusammenfassung

Hydrothermalquellen und Cold-Seeps sind weit verbreitet in der Tiefsee der Weltmeere und repräsentieren 'Oasen des Lebens' in einem dunklen und lebensfeindlichen Ökosystem. Diese chemosynthetischen Habitate werden von Muscheln der Gattung *Bathymodiolus* dominiert. Diese Muscheln beherbergen bakterielle Symbionten in ihren Kiemen, welche Methan und reduzierte anorganische Stoffe wie Schwefelwasserstoff und Wasserstoff als Energiequelle für chemosynthetische Primärproduktion benutzen. Es ist bereits bekannt, dass die räumlichen und zeitlichen Gradienten dieser Energiequellen sehr steil sein können. Einige Studien zeigten, dass Unterschiede in der Verfügbarkeit der Energiequellen die Menge der Symbionten innerhalb der *Bathymodiolus* Muscheln beeinflusst. Allerdings fehlen tiefergehende Studien zu den physikalisch-chemischen Gradienten und deren Effekt auf die Population der Symbionten in den Muscheln. Eine der grundlegensten Voraussetzungen für diese Analysen ist eine verlässliche Quantifizierung der *Bathymodiolus* Symbionten. Das Ziel dieser Arbeit war es daher, ein präzises und effizientes Protokoll für die Bestimmung dieser Symbionten zu entwickeln. Ich habe sowohl eine stereologische, als auch eine molekulare Methode für diese Quantifizierung der Symbionten benutzt. 3-dimensionale Fluoreszenz *in situ* Hybridisierung (3D-FISH) wurde benutzt, um das Biovolumen der Symbionten innerhalb der Muschelkiemen zu messen. Zusätzlich habe ich quantitative Polymerase Kettenreaktion (qPCR) verwendet, um die relative Häufigkeit eines Gens des Wirtes (18S rRNA) sowie von zwei Genen der Symbionten (16S rRNA und *recA*) zu bestimmen. Die ausführliche Optimierung dieser Methoden hat den Hauptteil dieser Arbeit ausgemacht und beinhaltete Verbesserungen in der Extraktion genomischer DNA, dem Design von Primern und der automatischen Datenanalyse. Der Vergleich dieser Methoden hat ergeben, dass die qPCR verlässlicher und genauer in der Quantifizierung der *Bathymodiolus* Symbionten

ist, als 3D-FISH.

Nach der Optimierung der Methoden wurden diese eingesetzt, um die Menge von Symbionten in *Bathymodiolus* Muscheln zu messen, welche an Hydrothermalquellen und Cold-Seeps vorkommen. In einem *in situ* Experiment wurden Muscheln von Hydrothermalquellen am Mittelatlantischen Rücken von aktiven Quellen für bis zu 14 Tage wegtransportiert. Dies sollte eine spürbare zeitliche Absenkung ihrer Energiequellen simulieren. Die thiotrophen und methanotrophen Symbiontenopulationen der *B. puteoserpentis* Muschel waren nach 10 Tagen des Aushungerns fast komplett verschwunden, während die beiden Symbiontenspezies der *B. azoricus* Muschel nach 14 Tagen des Aushungerns auf 30% der ursprünglichen Menge gesunken waren. Nachdem diese ausgehungerten Muscheln wieder zurück zu den hydrothermalen Fluiden transportiert worden waren (welche ihre Energiequelle enthalten), ist die Menge an methanotrophen Symbionten wieder angestiegen. Die Population der thiotrophen Symbionten hat sich dagegen nicht erholt. Dies zeigt eine für jede Symbiontenspezies spezifische Reaktion auf eine unterschiedliche Verfügbarkeit ihrer Energiequelle.

Der Effekt räumlicher Gradienten von Energiequellen wurde anhand von *Bathymodiolus* Symbiosen an zwei Cold-Seeps im Golf von Mexiko untersucht. *In situ* Messungen von Schwefelwasserstoff und Methan an diesen beiden Habitaten ergaben signifikante Unterschiede in den Konzentrationen der beiden Energiequellen. Diese Unterschiede spiegelten sich ebenfalls in Veränderungen in der betreffenden Menge der Symbionten wieder, was auf eine Verbindung zwischen der räumlichen Verfügbarkeit der Energiequellen und der Menge an Symbionten hindeuten könnte.

Meine Ergebnisse betonen die Flexibilität der *Bathymodiolus* Symbiose im Bezug auf räumliche und zeitliche Variationen in der Verfügbarkeit an Energiequellen und tragen dazu bei, die Frage zu beantworten, warum diese Muschelgattung die Hydrothermalquellen und Cold-Seeps in der dunklen Tiefsee dominieren kann.

Contents

List of Figures

LIST OF FIGURES

List of Tables

GLOSSARY

ATP - Adenosine triphosphate

AOM - Anaerobic oxidation of methane

DAPI - 4,6-diamidino-2-phenylindole

DNA - Deoxyribonucleic acid

EPR - East Pacific Rise

FISH - Fluorescence *in situ* hybridization

ISMS - *In situ* mass spectrometer

ITS - Internal transcribed spacer

MAR - Mid-Atlantic Ridge

PFA - Para-formaldehyde

pMMO - Particulate methane monooxygenase

qPCR - Quantitative polymerase chain reaction

RGE - Reductive genome evolution

ROV - Remotely-operated vehicle

rRNA - Ribosomal ribonucleic acid

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

sMMO - Soluble methane monooxygenase

SSU - Small subunit

Part I

Introduction

1

Chemosynthetic symbioses in the deep-sea

1.1 Definition of symbiosis

The term 'symbiosis' (from the Greek *sýn* = 'with' and *bíōsis* = 'living') was coined by Anton de Bary and Albert Bernhard Frank in the late 1800s (Trappe, 2004) and describes an often close and long lasting association between two different organisms.

Symbiotic relationships are as old as eukaryotic life on Earth. The symbiotic theory of the origin of life suggests a cooperation of different single celled organisms, leading to such a strong dependence over time that their genomes became incorporated into one multicellular organism (Margulis, 1999). The type of symbiosis of two or more different organisms with each other is defined by the impact of their relationship on the organisms fitness. Scientists used to classify symbioses as mutualistic, commensal or parasitic. In a mutualistic relationship, both organisms increase their fitness through the symbiosis (e.g. flowering plants and pollinators such as bees and flies). In contrast to this, parasitism defines a relationship where one organism decreases the fitness of another one, for example the flea as a parasite of the human host. The word commensal (from the latin words *com* and *mensa*, meaning 'sharing a table') defines an association of organisms that is beneficial for one and does not significantly affect the fitness of

the other one (e.g. spiders, building their webs in between plants).

These examples for different symbiotic associations are easy to differentiate, but in many cases, the role of at least one symbiotic partner is not clear or can not be set to only one kind of interaction. In this thesis work, the term symbiosis is used in the broader definition coined by Anton de Bary and Albert Bernhard Frank, declaring all persistent biological interactions (mutualism, parasitism and commensalism) as symbiosis.

1.2 Definition of chemosynthesis

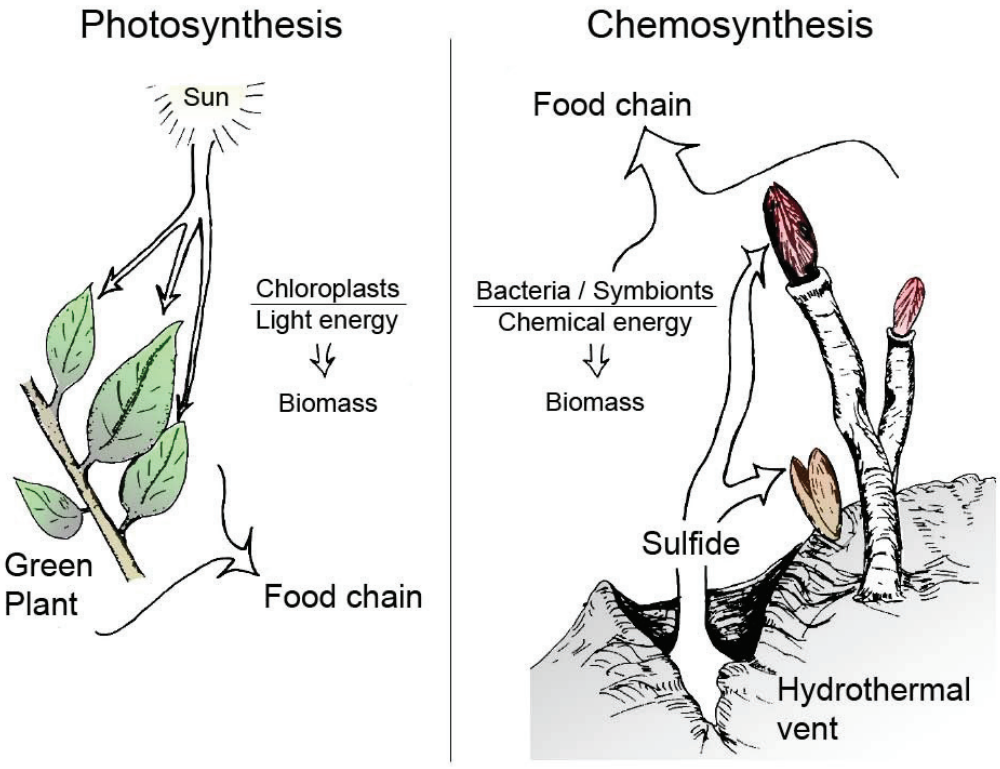

Figure 1.1: Photosynthesis vs. chemosynthesis - Comparison of basic principles of photosynthetic and chemosynthetic primary production. Modified from Somero (1984).

The process known as photosynthesis uses the energy of the sun and CO_2 to produce organic compounds (such as sugars) and O_2. Organisms that use this

process (such as plants, algae and many bacteria) are called photoautotrophs and occur at any place with sufficient sun light (on land, in the water and even under ice). Chemosynthesis, on the other hand, describes the synthesis of biomass from carbon molecules (e.g. CO_2 or CH_4) by the oxidation of a chemical compound. Both processes use carbon molecules to produce carbohydrates and an energy source to fuel the reactions. And both photosynthesis and chemosynthesis result in energy-rich compounds for the organisms, which in turn becomes food for other organisms in the food chain (Figure 1.1). The difference is the energy source, which in case of photosynthesis is sunlight and in case of chemosynthesis is an inorganic or organic molecule.

Chemosynthesis (like photosynthesis) is based on redox reactions, which means energy-rich electrons are transferred between reaction partners. Electrons flow from one partner (the electron donor) to another one (the electron acceptor). These redox reactions can be coupled to the production of biomass, which involves the fixation of carbon. Dependent on which molecule is used, the organisms are classified as chemolithotrophs (using inorganic compounds as electron donors) or chemoorganotrophs (using organic compounds as electron donors). Depending on their carbon source, these organisms are either autotrophic (using CO_2 as carbon source), or heterotrophic (using organic carbon for biomass production).

1.3 Discovery of deep-sea symbioses

It took mankind until the late 1970s to discover the first chemosynthetic ecosystems in the deep-sea (Corliss & Ballard, 1977), located at the Galapagos rift at the East Pacific Rise. Without sunlight, there is no photosynthesis-driven biomass production, and only a minor fraction of the sunlight-derived and energy-rich compounds from the sea-surface make their way down to the ocean floor. At first, the scientists thought that the deep-sea animals themselves would be able to use the inorganic compounds from the hydrothermal vent fluids and turn them into energy for biomass production. This deep-sea animal was the tubeworm *Riftia pachyptila* and the evidence were key enzymes for sulfur oxidation that were found in its body (Felbeck, 1981). It took the scientists not too long, to discover that the true source of these enzymes were sulfur-oxidizing bacteria, harbored in

a specialized organ of the tubeworm, the trophosome (Cavanaugh *et al.*, 1981).

1.4 Marine hosts of chemosynthetic symbioses

Since the discovery of chemosynthetic symbioses at hydrothermal vents, many more habitats and animal hosts have been found (reviewed by Dubilier *et al.* (2008a)). Until today, seven different phyla of marine symbiotic hosts have been discovered, including ciliates, shrimps, mussels, tubeworms and crustaceans (Stewart *et al.*, 2005). A selection of these symbioses is shown in Figure 1.2, differentiating the symbionts in endo- and ectosymbionts. Endosymbionts (or endobionts) are located inside the host tissue and ectosymbionts (or epibionts) are attached to external epithelia of the host. Depending on the evolutionary state of the symbiosis, the association of symbiont and host can lead to morphological adaptations on the host side. With endosymbionts, these adaptations can lead to the development of special organs for harboring the symbionts (e.g. the trophosome of siboglinid tube worms (Cavanaugh *et al.*, 1981) or the enlarged esophageal gland of the scaly foot snail (Goffredi *et al.*, 2004).

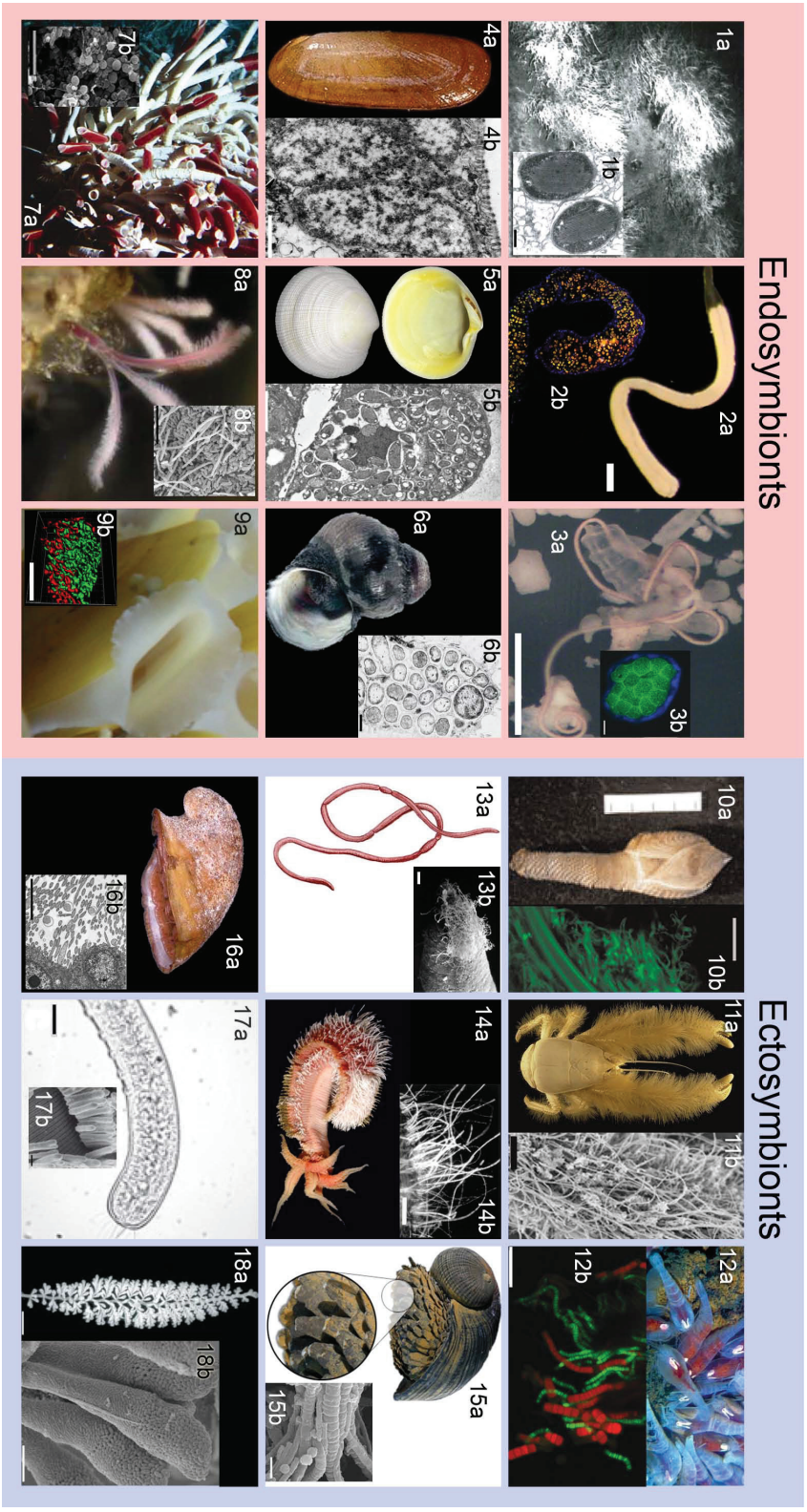

Figure 1.2: Marine hosts with chemosynthetic symbionts - Examples for marine hosts of chemosynthetic symbionts, divided in endosymbioses (left) and ectosymbioses (right). Legend continues on next page.

Figure legend to 1.2:

Endosymbionts

1: Cladorhizid sponge at Barbados trench (1a) with methanotrophic endosymbionts (2b), scale bar = 0.25 μm, from Vacelet *et al.* (1996).

2: The flatworm *Paracatenula galateia* (2a) with sulfur oxidizers (2b, fluorescent signals inside the trophosome, scale bar = 250 μm), from Gruber-Vodicka *et al.* (2011).

3: *Astomonema sp.* worm from Bahamas coral reef sediments (3a, scale bar = 0.5 mm) with sulfur oxidizers (3b, scale bar = 5μm), from Musat *et al.* (2007).

4: The mollusc *Solemya occidentalis* (4a, from femorale.com) with sulfur-oxidizing endosymbionts (4b, scale bar = 1 μm, from Krueger *et al.* (1996)).

5: The lucinid *Codakia orbicularis* (5a, from idscaro.net) with sulfur-oxidizing endosymbionts (5b, scale bar = 5 μm, from Caro *et al.* (2007)).

6: The snail *Ifremeria nautilei* (6a, from chess.lifedesks.org) with sulfur oxidizers (6b, scale bar = 0.5 μm, from Windoffer & Giere (1997)).

7: The tubeworm *Riftia pachyptila* (7a) with sulfur-oxidizing bacteria inside the trophosome (7b, scale bar = 5 μm) from Stewart & Cavanaugh (2006).

8: The bone-eating worm *Osedax mucofloris* (8a) with heterotrophic endosymbionts (8b, scale bar = 5 μm) from Verna (2010).

9: The mollusc *Bathymodiolus puteoserpentis* (9a, courtesy of MARUM Bremen) with 3D reconstruction of methane- (red) and sulfur- (green) oxidizing endosymbionts (scale bar = 5 μm).

Ectosymbionts

10: The barnacle *Vulcanolepas osheai* (10a, scale bar = 1 cm per bold unit) with associated γ- and ε-proteobacteria (10b, FISH-image, scale bar = 20 μm) from Suzuki *et al.* (2009).

11: The yeti crab *Kiwa hirsuta* (11a, from National Geographic) with sulfur-oxidizing episymbionts on the setae (11b, scale bar = 200 μm), from Goffredi *et al.* (2008).

12: Deep-sea shrimp *Rimicaris exoculata* (12a from Rutgers Deep Sea Center with FISH-image of γ- (green) and ε- (red) proteobacteria hosted in the gill chamber (12b, scale bar = 20 μm, from Petersen *et al.* (2010).

13: The sludge worm *Tubificoides benedii* (13a, from biopix.dk) with filamentous sulfur oxidizers attached to it's tail end (13b, scale bar = 20 μm, from Dubilier (1986)).

14: The polychaete *Alvinella pompejana* (14a, from sciencemag.org) with sulfur-oxidizing bacteria on the dorsal integument (14b, scale bar = 100 μm, from Cary *et al.* (1997)).

15: The snail *Crysomallon squamiferum* (15a, image from Anders Warn) with scales colonized

by ε- and δ-proteobacteria (15b, scale bar = 1 μm, from Goffredi *et al.* (2004)).

16: The limpid *Lepetodrilus fucensis* (16a, from Femorale) with episymbionts on its gill lamellae (16b, scale bar = 10 μm, from Bates (2007)).

17: The marine nematode *Robbea sp.* (17a, scale bar = 25 μm) with sulfur oxidizers on the body surface (17b, scale bar = 0.6 μm) from Bayer *et al.* (2009).

18: Colony of the ciliate *Zoothamnium niveum* (18a, scale bar = 0.5 mm) associated with sulfur-oxidizing epibionts (18b, scale bar = 10 μm) from Rinke *et al.* (2007).

1.5 Chemosynthetic primary production at vents and seeps

Hydrothermal vents and cold seeps in the deep-sea provide a variety of electron donors that can be used as energy sources by chemosynthetic bacteria. These electron donors are for example organic matter, hydrogen, methane, reduced sulfur compounds, ammonium, reduced iron and manganese. Electron acceptors available at vents and seeps include (from highest to lowest electron accepting potential) for example oxygen, nitrate, nitrite, manganese and iron oxides, oxidized sulfur compounds and carbon dioxide. The energy yielded during chemosynthetic primary production ($\Delta G^{0'}$) depends on the compounds metabolized and differs between the different processes (see Table 1.1).

Table 1.1: Selection of aerobic metabolisms from chemosynthetic microorganisms. Adapted from Petersen (2009).

Metabolism	$\mathbf{e_d^{-}}$ [a]	$\mathbf{e_{acc}^{-}}$ [b]	**Redox reaction**	$\mathbf{\Delta G^{0'}/e^{-}}$ [c]	**Symbiotic bacteria**[d]
Sulfur oxidation	HS^-	O_2	$HS^- + 2O_2 \rightarrow {SO_4}^{2-} + H^+$	−94	✓
Methane oxidation	CH_4	O_2	$CH_4 + 2O_2 \rightarrow {HCO_3}^- + H^+ + H_2O$	−94	✓
Hydrogen oxidation	H_2	O_2	$H_2 + 0.5O_2 \rightarrow H_2O$	−115	✓
Iron oxidation	Fe^{2+}	O_2	$Fe^{2+} + 0.5O_2 + H^+ \rightarrow Fe^{3+} + 0.5H_2O$	−65	–
Manganese oxidation	Mn^{2+}	O_2	$Mn^{2+} + 0.5O_2 + H_2O \rightarrow MnO_2 + 2H^+$	−25	–
Nitrification	${NH_4}^+$	O_2	${NH_4}^+ + 1.5O_2 \rightarrow {NO_2}^- + 2H^+ + H_2O$	−27.5	–

[a] electron donor
[b] electron acceptor
[c] energy available ($\Delta G^{0'}$) in kJ per mole per electron
[d] ✓ indicates that this metabolism has been identified in symbiotic bacteria

Despite this variety of potential energy sources at deep-sea vents and seeps (and their use by free-living bacteria), chemosynthetic symbionts were believed to exclusively perform sulfur- and methane-oxidation. This was until recently, as scientists found out that hydrogen is a third energy source for symbiotic primary production at hydrothermal vents (Petersen *et al.*, 2011). Hydrogen occurs together with methane and sulfide at hydrothermal vents and the energy yielded by hydrogen oxidation exceeds the energy yielded with any other chemolithoautotrophic pathway (Amend & Shock, 2001).

1.5.1 Sulfur-oxidizing symbionts

Reduced sulfur compounds can be used by free-living and symbiotic bacteria as an energy source, for example with oxygen as an electron acceptor. The sulfur-oxidizing (also called thiotrophic) symbionts at vents and seeps are chemolithoautotroph, which means they use inorganic compounds like H_2S as electron donors and CO_2 as carbon source. The oxygen comes from photosynthesis in the photic zone and is distributed via ocean circulation (so oxygen-dependent chemosynthesis in the deep-sea is not completely independent of photosynthesis).

The key enzymes involved in chemosynthetic sulfur oxidation are shown in Figure 1.3. The enzyme APR (adenosine 5'-phosphosulfate reductase), catalyzing the transformation between sulfite and adenosine 5'-phosphosulfate (APS), has become a diagnostic marker for characterizing sulfur-oxidizing symbionts. The gene *aprA*, encoding for the alpha subunit of the APR, has become one of the marker genes for the presence of sulfur-oxidizing bacteria (Harada *et al.*, 2009).

Studying the sulfur oxidation of symbionts is hindered by the inability to culture these bacteria. However, recent advances in pyrosequencing allowed for metagenomic and metatranscriptomic analyses that made it possible to identify for example some cellular proteins that take part in the sulfur metabolism of unculturable symbionts (Stewart *et al.*, 2011). For the vesicomyid clam *Calyptogena*, gene expression analyses of the sulfur-oxidizing symbionts indicated, that both sulfide and thiosulfate can be used by the symbionts to gain energy for carbon fixation (Harada *et al.*, 2009). The genetic potential or transcribed RNA content of symbionts are only hints in the direction of the pathways used. Environmental

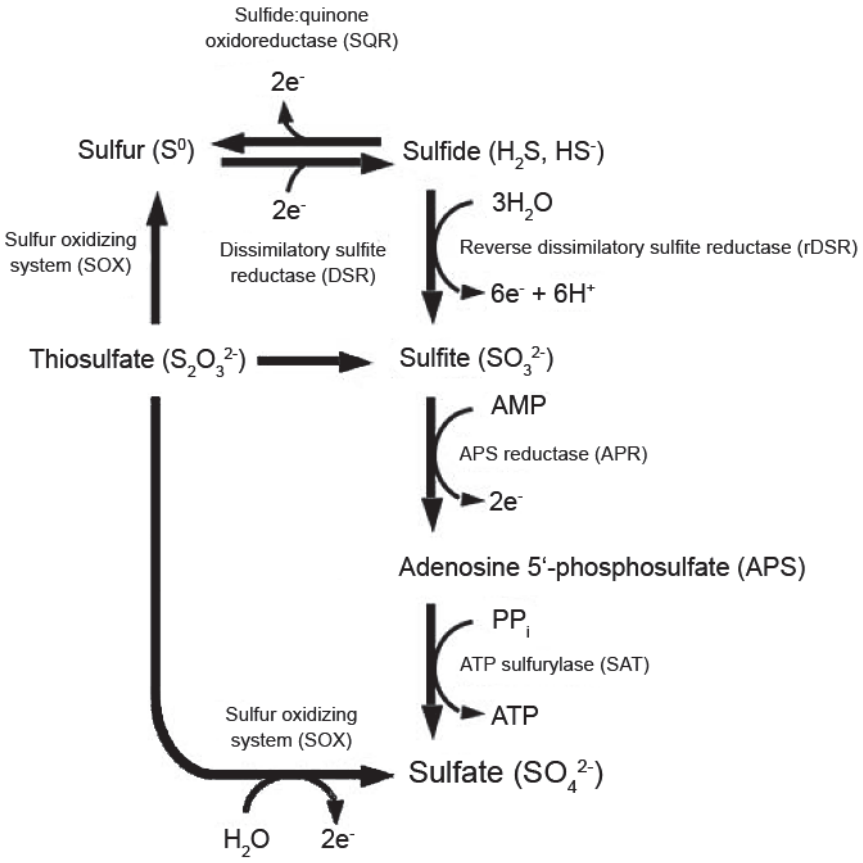

Figure 1.3: Sulfur oxidation pathways in sulfur-oxidizing symbionts - Arrows indicate pathways of oxidation or reduction of sulfur compounds that are used by the sulfur-oxidizing symbionts. Modified from Harada *et al.* (2009).

factors such as that thiosulfate is more stable than sulfide under aerobic conditions might also influence the pathway chosen to oxidize the reduced sulfur compounds.

1.5.2 Autotrophic CO_2 fixation in sulfur-oxidizing symbionts

The redox energy produced by sulfur-oxidizing microorganisms can be funneled into different pathways to fix inorganic carbon. So far, 6 fixation pathways for CO_2 are known (reviewed by Berg (2011)):

- Calvin-Benson-Bassham (CBB) cycle
- Reductive tricarboxylic acid (rTCA) cycle
- Reductive acetyl-CoA pathway
- 3-hydroxypropionate (3-HP) cycle
- 3-hydroxypropionate/4-hydroxybutyrate (3-HP/4-HB) pathway
- Dicarboxylate/4-hydroxybutyrate pathway

Most CO_2-fixing organisms on earth use the Calvin-Benson-Bassham (CBB) cycle. Among them are plants, most photoautotrophic prokaryotes, and many chemotrophic bacteria. Also chemotrophic symbionts at deep-sea vents and seeps use the CBB cycle to fix carbon, like for example the endosymbionts of the mussel *Bathymodiolus azoricus* (Spiridonova *et al.*, 2006) and the tubeworm *Riftia pachyptila* (Robidart *et al.*, 2008). The tubeworm symbionts can additionally use the rTCA cycle for carbon fixation (Markert *et al.*, 2007). The CBB cycle is a series of redox reactions, involving the central enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) that catalyzes the carboxylation of ribulose-1,5-bisphosphate (RuBP) by carbon dioxide into two molecules of 3-phosphoglycerate (3PGA). This is the key step that fixes carbon and makes it available for subsequent metabolic pathways.

1.5.3 Methane-oxidizing symbionts

Methanotrophic symbionts are chemoorganoheterotroph, using methane both as electron donor and carbon source. They are always aerobic bacteria and have so far been found mainly in bathymodiolin mussels, but also in different host phyla like a pogonophoran tubeworm and carnivorous sponges (DeChaine & Cavanaugh, 2006). As endosymbionts, they can co-occur with sulfur-oxidizing symbionts, even in the same host cell (Duperron *et al.*, 2006).

The key enzyme of the aerobic methane oxidation is methane monooxygenase, either membrane-bound (pMMO) or in soluble form (sMMO). This first enzyme of the pathway catalyzes the oxidation of methane to methanol (see Figure 1.4)

and has been found in all methanotrophs except *Methylocella* (Theisen *et al.*, 2005). The gene encoding the active subunit (*pmoA*) is the marker gene used to identify aerobic methanotrophs.

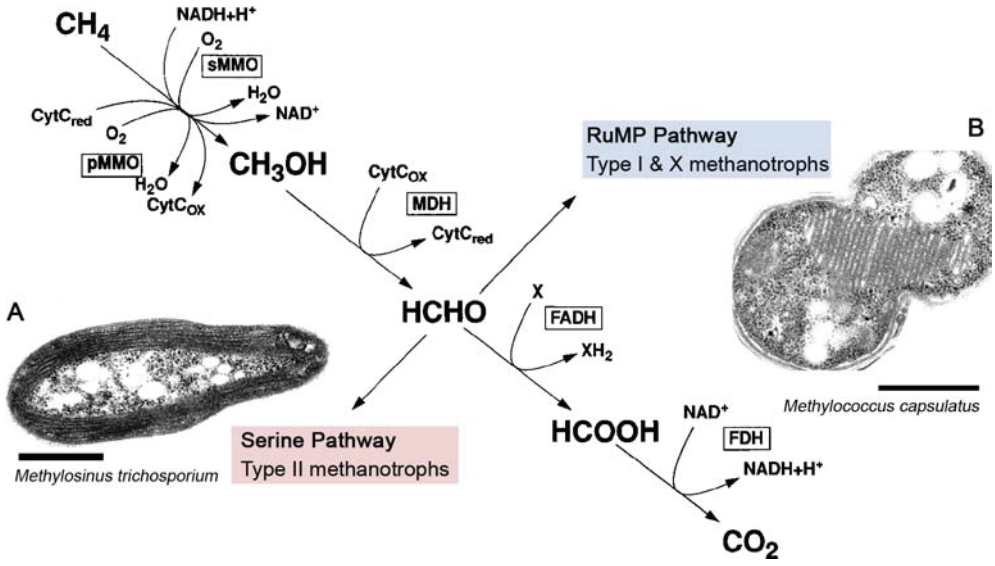

Figure 1.4: Oxidation of methane in methanotrophic bacteria - Pathway for the oxidation of methane and assimilation of formaldehyde. The examples for a type II methanotroph is *Methylosinus trichosporium* (A) (from (Dalton, 2005)) and *Methylococcus capsulatus* for a type X methanotroph (B) (from Smith *et al.* (1970)), scale bars are 0.5 μm. Modified from Hanson & Hanson (1996).

Methanotrophic bacteria are classified according to their morphology and physiology in type I (γ-proteobacteria, e.g. *Methylomonas* and *Methylobacter*), type II (α-proteobacteria e.g. *Methylosinus* and *Methylocystis*) and type X methanotrophs (e.g. *Methylococcus capsulatus*) (Hanson & Hanson, 1996). Type I and type X methanotrophs have intracellular cytoplasmic membranes (ICM) arranged as bundles throughout the cell. Type II methanotrophs have ICM arranged at the periphery of the cell. All methanotrophs oxidize methane to carbon dioxide via the intermediates methanol, formaldehyde, and formate. The formaldehyde can be used in anabolic sub-pathways to incorporate the carbon of methane into biomass. Type I and type X methanotrophs use the ribulose monophosphate (RuMP) pathway and the type II methanotrophs use the serine pathway (Figure 1.4). Type X methanotrophs are distinguished from type I by their ability to grow at higher temperatures and they possess the RubisCO

enzyme from the CBB cycle (Bowman, 2006). Concerning methane-oxidizing symbionts of marine invertebrates, comparative 16S rRNA gene analyses and the presence of ICM bundles inside the cells indicate that these bacteria belong to type I methane oxidizers Petersen & Dubilier (2009).

2

Deep-sea habitats for symbioses

2.1 Hydrothermal vents

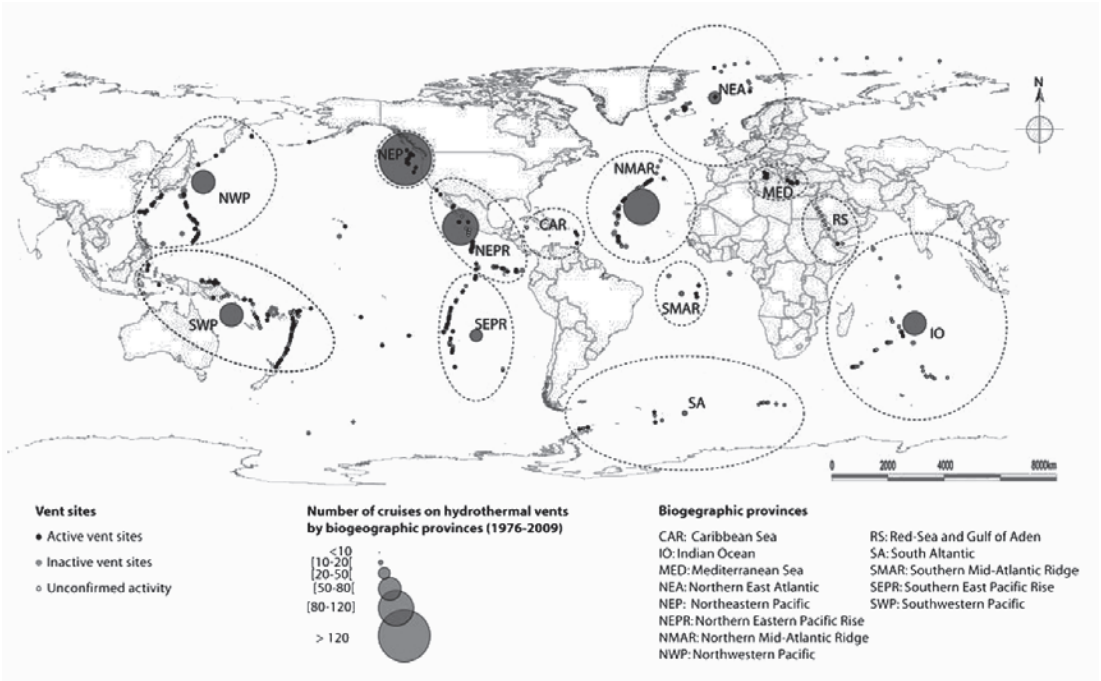

Figure 2.1: Global distribution of hydrothermal vents visited by scientists - The number and distribution of scientific cruises to hydrothermal vents until 2009 is linked to the biogeographic provinces of the vent systems (Godet *et al.* (2011)).

Since the discovery of hydrothermal vents in the Pacific Ocean more than 30 years ago, scientists have been exploring hydrothermal ecosystems all over the world. Due to the remoteness and the logistic challenge of studying deep-sea hydrothermal vents, the eastern Pacific and the northern Atlantic ridge systems were the first habitats to be visited by researchers (because these regions were most easily accessible compared to other vent sites). Over the years, the number of scientific cruises visiting and sampling hydrothermal systems increased and so did the discoveries arising from their research, leading to more and more vent systems to be found (Figure 2.1). The biogeographic province that has been inspected most, is the Northeastern Pacific (NEP), with more than 120 scientific visits until 2009 (see Figure 2.1). Due to this intense research, the amount of scientific publications of hydrothermal vents is today comparable with that of the most well-studied marine ecosystems, coral reefs and sea grass beds (Godet *et al.*, 2011).

2.1.1 The geological setting of hydrothermal vents

Spreading ridges are expressions of the global ocean crust formation and account for the majority of the world's volcanism in the deep-sea (more than 75%). These deep-sea mountain ranges are located at the margins of the tectonic plates and are more than 75.000 km in length (Van Dover & Trask, 2000).

These tectonic plates get pulled apart by forces from the Earth's interior and hot rock material surfaces to fill the gap. The erupted crust moves away from the ridge axis in lateral direction, which means the age of the crust increases with the distance to the origin. Because the seafloor is not spreading everywhere with the same speed, plate separations are classified in slow (10-50 mm yr^{-1}), medium (50-90 mm yr^{-1}) and fast spreading ($\geq$90 mm yr^{-1}) ridges (Lonsdale, 1977). By visualizing the age of the ocean floor (Figure 2.2), it is easy to distinguish between these different ridge systems, for example the slow spreading Mid-Atlantic Ridge (MAR) or the fast spreading East Pacific Rise (EPR). The differences in these ridge dynamics affect their morphology, forming steep central rifts and wide rift valleys in slow-spreading ridges and rather shallow central rifts and more gentle fall-offs in fast-spreading ridges.

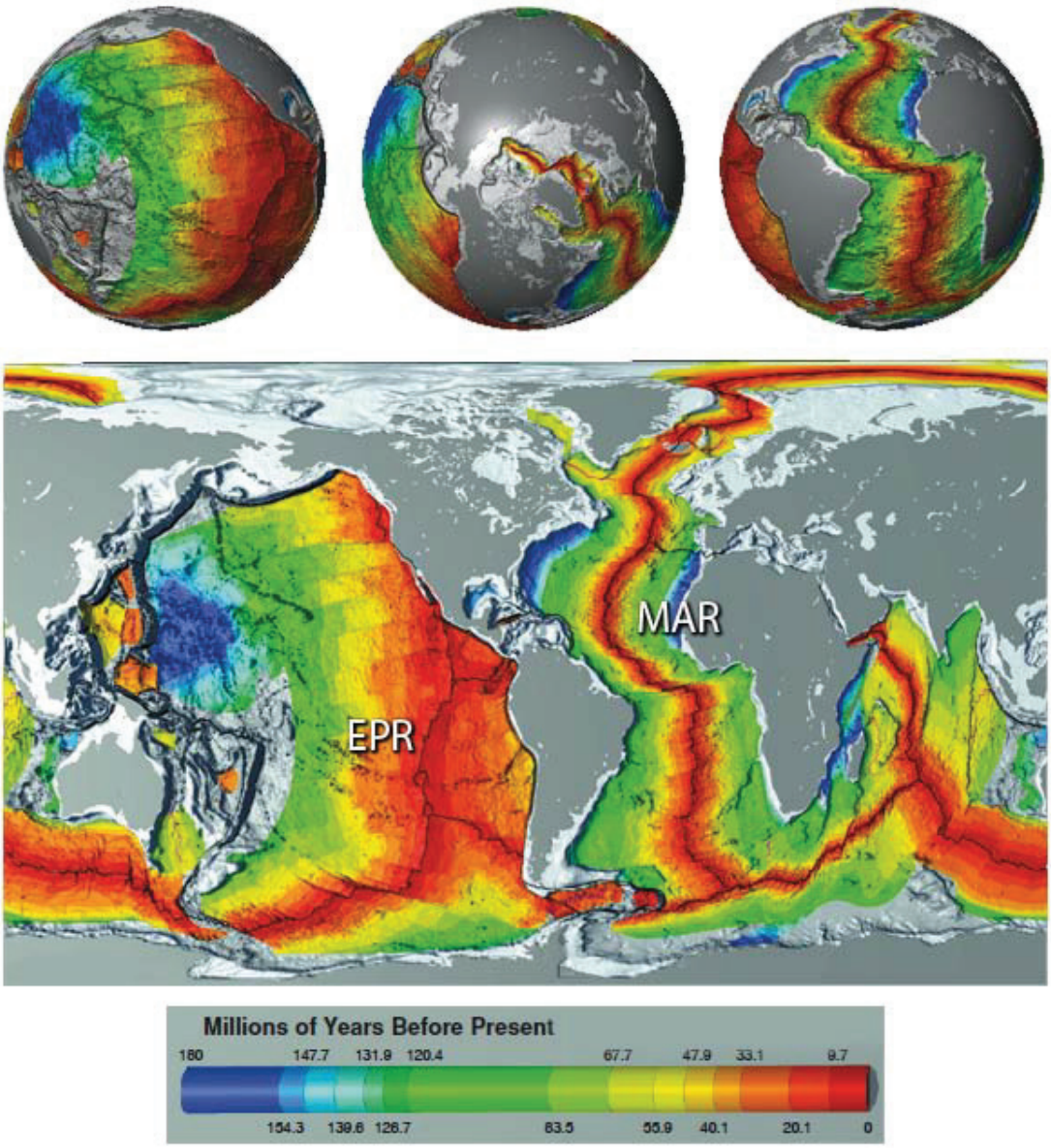

Figure 2.2: Mid-ocean ridges and sea floor spreading - Overview of the world's mid-ocean ridges, modified from the National Geophysical Data Center (NOAA). The age difference between slow spreading ridges (e.g. the Mid-Atlantic Ridge, MAR) and fast spreading ridges (e.g. the East Pacific Rise, EPR) is depicted by the color-coded age of the sea floor.

2.1.1.1 The effect of seafloor spreading rates on vent species diversity

The spreading rate of these ridges is believed to be directly linked to the species diversity at the related vent habitats. First of all, the spreading rate determines the spatial occurrence of hydrothermal vents. Fast-spreading ridge systems harbor more hydrothermal vent fields than slow-spreading ridges, because they can feed more hydrothermal systems by a greater magma budget (Baker, 1996). This abundance of the hydrothermal habitats influences the ability of organisms to disperse and spread over long distances. Therefore, the lower habitat frequency leads to lower species diversity on slow-spreading ridges (Juniper & Tunnicliffe, 1997).

2.1.1.2 Hydrothermal vent fluid chemistry

Deep-sea hydrothermal vents occur along all active mid-ocean ridges and back-arc spreading centers. Their most impressive features are the black smokers with their chimneys and buoyant plumes in the water column. Columnar chimneys can form very early in the evolutionary stage of a hydrothermal vent and within days or weeks the hot vent fluids are already funneled through sulfide chimneys (Haymon *et al.*, 1993). The formation and growth of a black smoker is shown in Figure 2.3.

Hydrothermal venting needs seawater to penetrate through ocean floor (e.g. through cracks or fissures) towards a magma source. As it travels deeper in the direction of the heat source, the seawater becomes depleted in minerals and looses hydroxyl ions, which leads to an acidification of the water and drop of pH. Further down at the magma chamber which is creating this circulation, temperatures above 300°C lead to the precipitation of anhydrite ($CaSO_4$) and the acidic fluids leach minerals from the nearby rocks into the seawater. Additionally, the phase separation happening close to the hydrothermal cell contributes to the composition of the fluids by vapor (rich in gases) and brine (rich in metals). When the heated vent fluid rises back to the surface and gets in contact with the cold, oxygenated and alkaline seawater, the metal sulfides precipitate and resemble black smoke that comes out of a chimney (Figure 2.3 B).

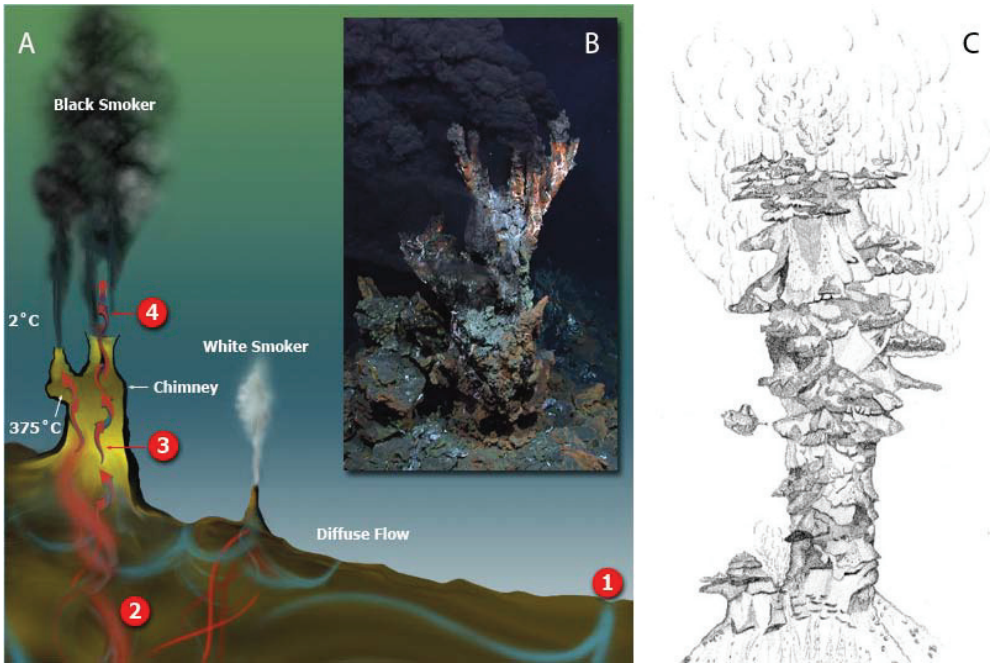

Figure 2.3: The formation of hydrothermal vents - This illustration (A) from the Woods Hole Oceanographic Institution (WHOI) shows the major stages in the formation of a hydrothermal vent. Cold seawater seeps down into the ocean floor (**1**). The water is heated up (350-400°C) and reacts with the rocks in the crust (**2**). As a result, chemical reactions change the water (**3**). Hydrothermal vent fluids emit and mix with cold, oxygenated and alkaline seawater, precipitating as black particles (**4**). Image B shows a black smoker ('candelabra') in 3,000 meters depth at the Logatchev hydrothermal field at the Mid-Atlantic Ridge (courtesy of MARUM University of Bremen). The drawing (C) is an example of one of the largest black smokers ever discovered, called Godzilla. The submersible ALVIN is shown for scale (Robigou *et al.*, 1993).

White smokers in contrast to black ones feature a lower temperature of 100-300°C and they have their name from the precipitation of silica, anhydrite and barite ($BaSO_4$) in form of white particles. Hydrothermal fluids can also leave the seafloor without chimneys but by seeping out of the ground. These diffusive flow fluids already mix with the ambient seawater below the seafloor and have a lower temperature (less than 100°C) and flow slower than hot, focused vent fluids. Diffuse lower temperature fluids are depleted in minerals but can still contain considerably high amounts of sulfide and dissolved gases (Johnson *et al.*, 1986) and therefore fuel large populations of microorganisms (Sogin *et al.*, 2006),

invertebrates (Van Dover *et al.*, 2002) and viruses (Williamson *et al.*, 2008).

2.1.2 Geochemistry of Mid-Atlantic Ridge hydrothermal vents

The thickness and composition of oceanic crust at spreading ridges is relatively homogeneous compared to the continental crust. On average, oceanic crust is 6-7 km thick and basaltic in composition as compared to the continental crust which on average is 35-40 km thick and has a roughly andesitic (intermediate rock type between basalt and dacite) composition (Turekian, 2010). The majority of Mid-Atlantic Ridge (MAR) mantle rocks consist of basalt, but some areas contain peridotite in the upper oceanic crust, a dense and coarse-grained lava rock. Serpentinization, a process of oxidizing and hydrolyzing ultramafic rocks, converts this peridotite to serpentinite and leads to elevated concentrations of hydrogen and methane in the endmember fluids of ultramafic-hosted hydrothermal systems (Schulte *et al.*, 2006).

Table 2.1: Endmember fluid composition of Mid-Atlantic Ridge hydrothermal vent fields[a]

		MG	LS	RB	BS	TAG	SP	LG
T	°C	265-284	152-333	360-365	356-364	270-363	335-356	≥353
pH	(at 25°C)	4.2-4.8	3.5-4.9	2.8-3.1		2.5-3.4	3.7-3.9	≤ 3.3
Si	mmol kg $^{-1}$	8.2-11.2	9.1-17.5	6.9-8.0		18-22	18-20	7.8-2
Cl	mmol kg $^{-1}$	360-400	410-540	≥ 750	469	633-675	550-563	515-522
CO_2	mmol kg $^{-1}$	17-20	8.9-28	≤ 16		2.9-4.1	10.63	
H_2S	mmol kg $^{-1}$	1.5-2	1.4-3.3	1-2.5	9.3	2.5-6.7	2.7-6.1	0.8
CH_4	mmol kg $^{-1}$	1.35-2.63	0.5-0.97	2.2-2.5	0.065	0.14-0.62	0.046-0.062	2.1
Fe	mmol kg $^{-1}$	0.0002-0.018	0.13-0.86	24	1.68-2.16	1.64-5.45	1.8-2.56	2.5
Mn	mmol kg $^{-1}$	0.068	0.45	2.5	0.26	1	0.49	0.33

Mid-Atlantic Ridge hydrothermal vent fields: **MG** = Menez Gwen, **LS** = Lucky Strike, **RB** = Rainbow, **BS** = Broken Spur, **TAG** = Trans Atlantic Geotraverse, **SP** = Snake Pit, **LG** = Logatchev

[a] Table adapted from (Desbruyères *et al.*, 2000, Douville *et al.*, 2002)

The MAR features some of these ultramafic-hosted vent fields, for example Rainbow (36°14'N), the low-temperature vent field Lost City (30°N) and the Logatchev vent field (14°45'N) (Schmidt *et al.*, 2007, Petersen *et al.*, 2009). Two hydrothermal vent fields at the MAR were visited for this study, the Logatchev and Menez Gwen hydrothermal vent fields (see Figure 2.4). The Logatchev field contains elevated amounts of methane, low amounts of sulfide (Table 2.1) and the

highest amounts of hydrogen measured in hydrothermal systems so far (19 mM) (Schmidt *et al.*, 2007). Menez Gwen on the other hand is a basalt-hosted and shallower (871-847 m deep) vent field at 37°50'N. The Menez Gwen hydrothermal fluids contain moderate amounts of hydrogen sulfide and methane (Table 2.1) and low amounts of hydrogen gas (38 μM) compared to other MAR sites (Charlou *et al.*, 2000).

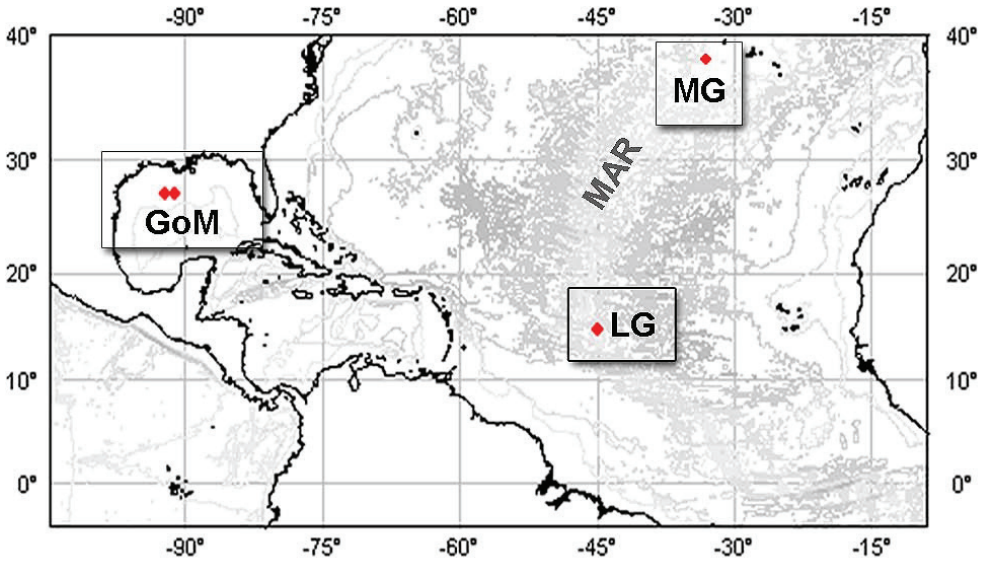

Figure 2.4: Sample sites at Mid-Atlantic Ridge vents and Gulf of Mexico cold seeps - Location of hydrothermal vent fields and cold seeps at which experiments for this study have been conducted (highlighted areas). The red dots show the locations of two cold seeps at the Gulf of Mexico (GoM) and hydrothermal vent fields at Logatchev (LG) and Menez Gwen (MG).

2.2 Cold seeps

Seven years after the discovery of hydrothermal vents in 1977, researchers found another type of seafloor expression that is associated with geologic processes, the cold seeps. They can resemble lakes on the ocean floor (Figure 2.5) and are sites where substances like hydrocarbons and gases (e.g. methane and hydrogen sulfide) flow out of the seafloor. These outflows can occur as pockmarks (seafloor depressions), gas chimneys, brine pools (dense seawater filling ground depres-

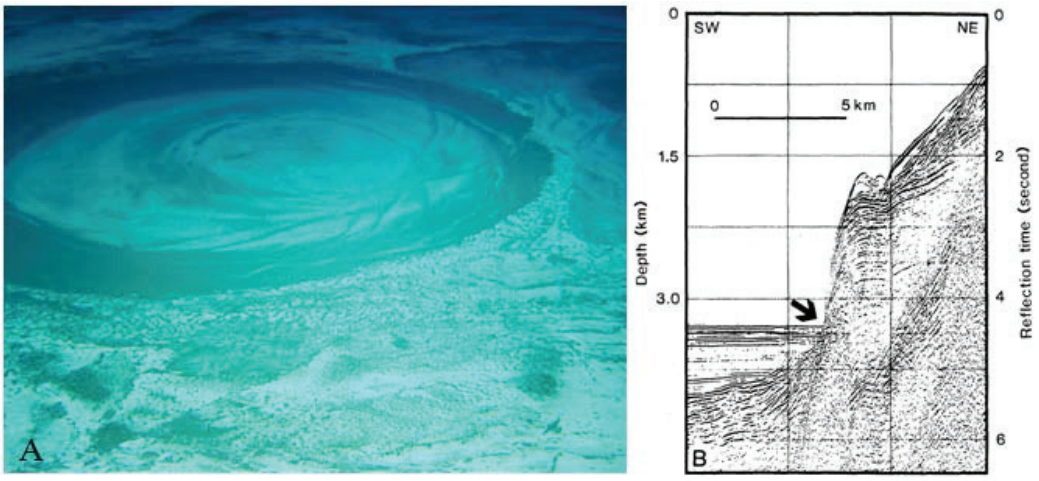

Figure 2.5: The discovery of cold seeps - Cold seeps can resemble lakes on the bottom of the ocean (A, brine lake at the summit of the Cheops mud volcano in the western province of the NDSF from Ifremer/Victor 6000/Medeco 2007) and were first discovered during submersible dives to the Gulf of Mexico. By using bathymetry mapping, the scientists found a saline seep (arrow in B) with abundant life at the Florida Escarpment (Paull *et al.*, 1984).

sions), oil and asphalt seeps (outflow of hydrocarbons) (Jorgensen & Boetius, 2007). Cold seeps are globally distributed, from shallow to deep waters, active margins (e.g. the Peruvian margin) and passive margins (e.g. the Gulf of Mexico) to remote place like Antarctica (Domack *et al.*, 2005).

2.2.1 Cold seep geology and fluid chemistry

Similar to hydrothermal vents, cold seeps feature a flow of reduced chemicals from below the surface of the ocean to the sea floor, but they are not associated with a heat source or located at spreading centers. At the margin of the oceanic crust, subduction compresses the accumulated sediment (active margin) or the oceanic crust is pushed towards the continental crust (passive margins), leading to an outflow of fluids and gases from deeper reservoirs. These fluids have a lower temperature than the hot vent fluids and also a lower flow velocity (Boetius & Suess, 2004). As with the hydrothermal vents, the chemical composition of these fluids is determined by the local geological settings and many cold seeps are associated with gas hydrates and hydrocarbon reservoirs. Additionally, the chemical

environment of the cold seeps can be influenced by present microorganisms. The presence of methane at these ecosystems for example can either be related to (ancient) archaeal methanogenesis (biogenic production of methane by archaea) or abiogenic deposits of methane at subduction zones (Birgel *et al.*, 2006).

2.2.2 Cold seeps at the Gulf of Mexico

The Gulf of Mexico (GoM) connects with the Caribbean Sea via the Yucatan Channel between Mexico and Cuba and with the Atlantic Ocean through the Florida Strait between the U.S.A. and Cuba. The geological evolution of the GoM includes the deposition of sediment over thick Jurassic salt deposits (Figure 2.6). These sediment deposits compress the salt bodies and force gases and fluids to leak up to the ocean floor, often forming distinct seafloor features: the seeps and mud volcanoes of the Gulf of Mexico (Fisher *et al.*, 2007).

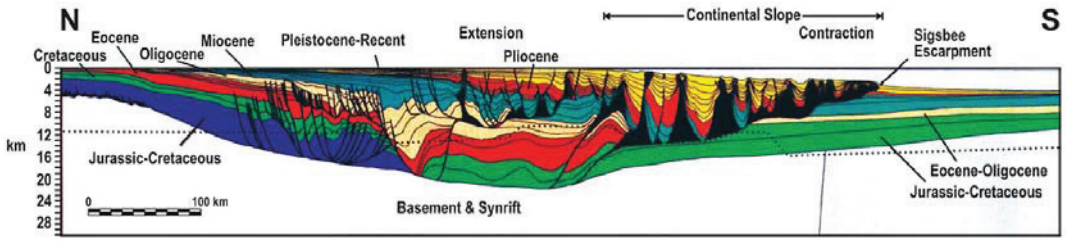

Figure 2.6: North-South cross section of the northern Gulf of Mexico - Illustration of the geological setting at the GoM, with sediments from different geological eras interacting with salt deposits (black), resulting in faults and a leakage of subsurface gases and fluids to the surface. From Peel *et al.* (1995).

This unique geological setting makes the deep waters of the Gulf of Mexico the world's largest site for oil and gas production. Mainly due to the presence and investigations of the petroleum industry (regarding the extraction of hydrocarbons), the entire northern Gulf's continental slope has been imaged more intensely with seismic and acoustic systems than any other similar deep-sea ecosystem. The result is a growing database since the 1990's on seismic 3D-data that can be used by the scientists to study the expulsion sites and migration pathways of the cold seep fluids in high detail (Roberts *et al.*, 2010).

Part I - Introduction

3

The bathymodiolin symbiosis

3.1 The dominance of bathymodiolin mussels at vents and seeps

The most dominant molluscs at vents and seeps are the ones of the genus *Bathymodiolus* (Mytilidae, Bathymodiolinae)(Dover & Trask, 2000). Since the original characterization of the genus, 22 *Bathymodiolus* species have been described (Miyazaki *et al.*, 2010). The ancestors of this genus, that is today so prevailing at deep-sea ecosystems in the Atlantic, might have used sunken wood and whale falls as 'stepping stones' in order to conquer habitats in the deep-sea (Samadi *et al.*, 2007). Bathymodiolin mussels maintain symbiotic relationships with sulfur- and/or methane-oxidizing symbionts (Table 3.1) and harbor their symbionts in specialized gill cells, the bacteriocytes. The bacteria are located close to the flow of fresh seawater inside the mussel, being very close to their nutrition source that comes with the hydrothermal vent fluids filtered by the mussel (Wendeberg *et al.*, 2011). All bathymodiolin species show a reduction of their digestive system and gain most of their nutrition from their symbionts. For some, an additional nutrition by filter-feeding of particles from the environment has been shown (Page *et al.*, 1991).

3.2 Diversity of bathymodiolin mussels in the Atlantic

The mussels sampled for this work came from the Atlantic and include the bathymodiolin species *B. azoricus*, *B. puteoserpentis*, *B. brooksi* and *B. childressi*. The first three have a dual symbiosis with sulfur- and methane-oxidizing bacteria, whereas *B. childressi* has a single symbiosis with methane oxidizers (Table 3.1).

Table 3.1: Bathymodiolin species from the Atlantic. Adapted from Duperron *et al.* (2009).

Species name	Habitat	Sites	Symbionts	Symbiont reference	Host reference
B. azoricus	Vent	Menez Gwen, Lucky Strike, Rainbow, Broken Spur (MAR)	SOX, MOX	(Duperron *et al.*, 2006)	(Jones *et al.*, 2006)
B. puteoserpentis	Vent	Broken Spur, Snake Pit, Logatchev (MAR)	SOX, MOX	(Duperron *et al.*, 2006)	(Jones *et al.*, 2006)
B. cf. boomerang	Seep	Regab (Gulf of Guinea, Gabon continental margin)	SOX, MOX	(Duperron *et al.*, 2005)	(Olu-Le Roy *et al.*, 2007)
B. heckerae	Seep	West Florida Escarpment (GoM), Blake Ridge	SOX, MOX, MY	(Duperron *et al.*, 2007a)	(Jones *et al.*, 2006)
B. brooksi	Seep	Alaminos Canyon, Atwater Canyon (GoM)	SOX, MOX	(Duperron *et al.*, 2007a)	(Jones *et al.*, 2006)
B. childressi	Seep	Alaminos canyon, Lousiana Slope (GoM)	MOX	(Distel & Cavanaugh, 1994)	(Jones *et al.*, 2006)

SOX = sulfur oxidizers, **MOX** = Methane oxidizers, **MY** = Methylotrophs.

Bathymodiolin mussels form a monophyletic group within the mytilid family, but the phylogeny of this genus is still under debate. Major questions have not been solved yet, as for example when and how these animals invaded the deep-sea environment and how they become isolated to speciate. The high dispersal capacities of some bathymodiolin species might be explained by their reproductive strategies, e.g. *B. childressi*, featuring spawning with a synchronized gametogenesis and the production of numerous planctonic larvae (Tyler & Young, 1999).

3.3 Diversity and acquisition of bathymodiolin symbionts

The transmission of the symbionts is proposed to be horizontally, which means they are taken up by the mussel host from the environment. The proof of a

horizontal transmission mode for the symbionts (e.g. by showing a symbiont-free larval stage) has not been done yet, but there are indications for a life-long uptake of the symbionts by the mussel: a study by Won *et al.* (2003) presents genetic and cytological evidence for the environmental acquisition of thiotrophic endosymbionts in *Bathymodiolus* mussels from the MAR. Also, the symbiont-containing mussel gills grow continuously throughout the mussel life (Neumann & Kappes, 2003, Cannuel *et al.*, 2009) and preliminary results indicate that the youngest areas of the gills (the budding-zones) are symbiont-free and get colonized by symbionts as they grow (unpublished data by C. Wentrup). Additionally, the genome of the sulfur-oxidizing symbiont of *B. puteoserpentis* (that has recently been sequenced) resembles a mixture of genomes rather than belonging to one genome (unpublished data by Dr. J. Petersen, Dr. E. Pelletier and Dr. V. Barbe). This also supports the theory of horizontal symbiont transmission, because if the symbionts get transmitted vertically (from adult to offspring) the bottleneck of vertical transmission might only allow a few symbionts to colonize the new mussel (and then spread out by propagation). The mixture of genomes rather points in the direction of an environmental uptake of many (slightly) different symbiont phylotypes. Still, how the symbionts are recognized by the invertebrate immune system and finally get incorporated into the gill cells, is not known.

3.3.1 Sulfur-oxidizing symbionts in bathymodiolin mussels

Most sulfur-oxidizing symbionts of bathymodiolin mussels belong to the γ-proteobacteria and recent phylogenetic studies showed at least 9 different clades, mixed with free-living bacteria (Dubilier *et al.*, 2008b). This phylogeny most probably reflects an independent evolution of sulfur-oxidizing symbioses at multiple time points and in different places. The thiotrophic bacteria of the bathymodiolin mussels were first discovered in the gills of *B. thermophilus* (Fiala-Medioni *et al.*, 1986), characterized by a rod or coccoid shape, a mean diameter of 0.35 μm and the presence of the RubisCO enzyme (indicating the use of the CBB cycle for carbon fixation) (Fiala-Medioni *et al.*, 2002). The symbionts are generally host specific, but based on 16S rRNA gene analysis, there are cases where symbiont

phylotypes from different host species are highly similar to identical. That is the case for the symbiosis of the two mussels *B. azoricus* and *B. puteoserpentis*, both containing sulfur- and methane-oxidizing symbionts. The northern provinces of the MAR (Menez Gwen, Rainbow and Lucky Strike) are colonized by *B. azoricus*, whereas *B. puteoserpentis* is found at the vent sites closest to the equator (Snake Pit and Logatchev). Both host species co-occur at one site called Broken Spur and the similarity of their thiotrophic symbionts is very high at the 16S rRNA level (Duperron *et al.*, 2006). By using a faster evolving marker like the ITS (internal transcribed spacer) region instead of 16S rRNA, more genetic differences between the two thiotrophic symbionts became apparent (Won *et al.*, 2003).

Investigations on the metabolic potential of the bathymodiolin thiotrophic symbiont have recently shown that it can oxidize hydrogen additionally to sulfur (see Section 1.5). By harboring a symbiont species that uses at least two different electron donors for chemosynthetic primary production, the flexibility of the mussel hosts in respect to local energy source concentrations is increased.

3.3.2 Methane-oxidizing symbionts in bathymodiolin mussels

Methane-oxidizing symbionts were first discovered in *B. childressi* (Childress *et al.*, 1986) and represent a monophyletic group inside the γ-proteobacteria, forming a sister group to the type I methanotrophic bacteria *Methylobacter* and *Methylomicrobium* (Dubilier *et al.*, 2008b). They are bigger in size compared to the thiotrophs (1.5-2 μm), have centrally stacked intracytoplasmic membranes (typical for type I and X methanotrophs) and have the key enzyme for aerobic methane oxidation (shown for *B. azoricus* symbionts by Pernthaler & Amann (2004)). The biogeography of the methanotrophs shows no geographically-based trend. Methanotrophs from the *Bathymodiolus* mussel from the Gabon Margin for example are more closely related to the symbionts of *B. heckerae* (from the Gulf of Mexico) than to the ones from *B. puteoserpentis* (from the Mid-Atlantic Ridge), which are geographically much closer (Petersen & Dubilier, 2009).

3.3.3 Other bathymodiolin symbionts

Some bathymodiolin mussels harbor additional phylotypes of sulfur and methane oxidizers or other bacteria inside their gill cells. In the cold seep mussel *B. heckerae*, Duperron *et al.* (2007b) found a methanotroph, two phylogenetically distinct thiotrophs and a methylotroph-related symbiont inside the gills. Other studies at Gulf of Mexico asphalt seeps have found a heterotrophic γ-proteobacterium, related to *Psychromonas*, inside *B. brooksi* mussels. At the same site, the co-occurring *B. heckerae* mussels were shown to harbor *Cycloclasticus*-related bacteria in the same bacteriocytes with sulfur and methane oxidizers. This hydrocarbon degrading bacterium could extend the potential energy sources for the symbiosis by the degradation of heavy hydrocarbons at asphalt seeps (unpublished data by L. Raggi).

The intracellular bacteria in bathymodiolin mussels found so far are not all beneficial. The intranuclear parasite '*Candidatus* Endonucleobacter bathymodioli' infects the nuclei of mussel gill cells (Zielinski *et al.*, 2009). The invaded cells get disrupted during the life cycle of the parasite and the infection spreads throughout the whole mussel gill. There is evidence that those gill cells occupied by sulfur- and methane-oxidizing symbionts are protected against the infection of the parasite, but this remains to be proven.

3.4 Bacterial response to environmental changes in bathymodiolin symbioses

The symbionts harbored in the gills of bathymodiolin mussels benefit from the symbiosis by access to a constant supply of energy sources. The water filtered by the mussel delivers the energy sources from the hydrothermal vent fluids directly to the gills. Bacterial distribution and activity seems to be highest in the gill area, where fluids enter the mussel (Wendeberg *et al.*, 2011). The mussel acquires organic matter from the symbionts, most probably by digesting them in lysosomal bodies of the gill cells. This was shown by lysosomic enzyme tests (Boetius & Felbeck, 1995) and microscopic observations (Barry *et al.*, 2002).

The dissolved gases (e.g. methane, sulfide and oxygen) entering the mussel

with the water flow can only be transported via advection and diffusion processes towards the symbiont-hosting bacteriocytes (Childress & Fisher, 1992). Together with the fact that the bathymodiolin mussels can neither store nutrients in their blood (like *Calyptogena* clams (Zal *et al.*, 2000)) nor that their symbionts are able to store sulfur inside their cells (like the lucinid symbionts (Caro *et al.*, 2007)) makes the bathymodiolin symbiosis an ideal system for investigating the response to gradients in energy sources. That means, the availability of energy sources from the environment is directly reflected in the symbionts nutrition and from the moment on when the concentration of the energy sources changes, the symbionts have to cope with this change. Their response to nutritional gradients can either be a change in metabolic activity (short-term response) or a change in bacterial numbers (long-term response). In case of a decrease in energy source concentrations, this long-term response can either lead to the lysis of bacteria inside host bacteriocytes (Boetius & Felbeck, 1995) or to a release of bacteriocytes from the gill tissue. The later could lead to a transport of emitted bacteriocytes to the digestive system of the mussel, but has not been investigated yet. In case of elevated concentrations of energy sources, bacterial numbers inside the host can increase either by an uptake of bacteria from the environment, or by a multiplication of the residual bacteria inside the mussel.

Another advantage of the bathymodiolin system is, that the mussels are very robust and they survive the retrieval from kilometers down in the deep-sea. At vent sites such as Menez Gwen at 38'N on the MAR, where the animals live in about 800 m depth, they can even be kept alive for months in aquaria (Colaço *et al.*, 2011).

4

Thesis goal

The work presented in this thesis was accomplished within the Priority Program SPP 1144 (*From Mantle to Ocean*) funded by the German Research Foundation (Deutsche Forschungsgesellschaft, DFG). In the scope of this program, I joined three research cruises and conducted work on board the research vessels and in the laboratories of the MPI in Bremen. This PhD project aimed to:

1. Establish a reliable and fast quantification protocol for the methane- and sulfur-oxidizing endosymbionts of bathymodiolin mussels. This included the optimization and validation of two different quantitative methods (qPCR and 3D-FISH).

2. Investigate the response of bathymodiolin vent symbionts to a temporal change in energy source concentrations by *in situ* transplantation of the mussel hosts.

3. Study the link between spatial geochemical gradients in the environment and symbiont abundance in bathymodiolin symbioses at Gulf of Mexico cold seeps.

After explaining the methods used for this study, the results for the goals above start with the realization of a suitable quantification protocol for bathymodiolin endosymbionts. Further results describe the use of this quantification protocol in the field. First, to investigate the link between symbiont abundance

and the availability of their energy sources. In a radical approach, we removed mussels from Mid-Atlantic Ridge vents from the hydrothermal fluids and starved them *in situ*, simulating a temporal change in energy source concentrations. The next field application was the investigation of symbiont abundances from different habitats in relation to local concentrations of sulfide and methane. This experiment has been conducted on bathymodiolin mussels from two cold seeps at the Gulf of Mexico and was accompanied by fine-scale measurements of sulfide and methane with *in situ* mass spectrometry at the same sampling sites.

Part II

Material and Methods

5

Experiments on bathymodiolin vent and seep symbioses

5.1 Sampling sites

Bathymodiolus mussels sampled for this work came from two hydrothermal vent fields (Logatchev and Menez Gwen) on the MAR and two cold seeps (Garden Banks and Green Canyon) from the GoM (Figure 5.1).

The ultramafic-hosted Logatchev hydrothermal vent field, discovered in 1994 by the Russians, is located at 3000 m depth at the Fifteen-Twenty fracture zone (14°45'N, 44°58'W) (Batuev *et al.*, 1995). It contains smoking craters, diffuse flow sites and inactive areas covered with sulfide sediments (Sudarikov & Roumiantsev, 2000).The *Bathymodiolus puteoserpentis* mussels were collected in May 2005 at the Logatchev vent field during cruise M64/2 on the research vessel Meteor. The mussels were sampled from a diffuse venting site in 3035 m depth (Irina II).

The slow-spreading Menez Gwen ridge segment is about 55 km long and reaches from 38°03'N to 37°35'N. It is characterized by the presence of a large central volcano that is 700 m high and 15 km in diameter and several small volcanoes located at its northern end. Two areas of hydrothermal activity are known along the small volcanoes: A cluster of four vents within the Marine Protected Area and those studied during the M82-3 cruise. The White Flames site (37°50'40 N, 31°31'8 W, -832 m) was sampled for *Bathymodiolus azoricus* mussels.

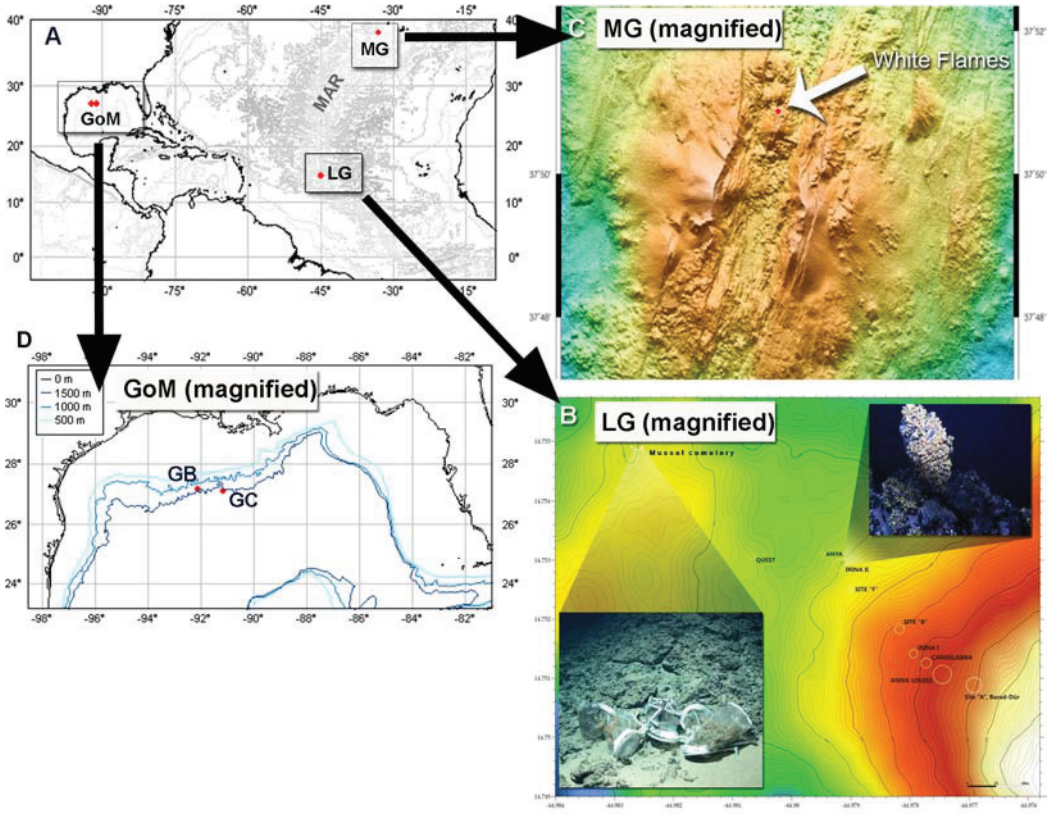

Figure 5.1: Sampling sites for *Bathymodiolus* mussels - A: Overview of all sampling sites with three areas highlighted (red dots) that have been sampled for this study. The *B. puteoserpentis* mussels used for the method development and optimization have been collected at the Logatchev hydrothermal vent field (LG). The mussels were displaced away from their hydrothermal fluid source (B, vent site Irina II) to a remote area with no access to vent fluids (B, mussel cemetery). The hydrothermal vent field Menez Gwen (C) has been sampled for *B. azoricus* mussels, and these mussels were transplanted back and forth from active venting (Figure 5.2). D: Cold seeps at the Gulf of Mexico (Garden Banks (GB) and Green Canyon (GC)), where *B. brooksi* and *B. childressi* mussels were collected. The underwater images are courtesy of the MARUM Bremen and the map is modified from Nico Augustin (IFM Geomar). The bathymetric map in (C) is courtesy of Christian Ferreira (Seeflor imaging Group of the MARUM Bremen).

Bathymodiolus mussels from several cold seep habitats were sampled in May 2006 during the CHEMCO III cruise in the northern Gulf of Mexico. The *B.*

brooksi and *B. childressi* mussel samples for this study come from two locations on the middle continental slope region, namely Garden Banks (GB829, 27°06'6 N, 91°09'9 E at -1407 m) and Green Canyon (GC852, 27°11'1 N, 92°07'5 E at -1258 m).

5.2 Transplantation experiments at MAR hydrothermal vents

Transplantation experiments have been conducted *in situ* with MAR vent mussels, taking place at the hydrothermal vent fields of Logatchev and Menez Gwen (Figure 5.1B, C). At Logatchev, *Bathymodiolus puteoserpentis* mussels were collected with scoop nets from the active vent site Irina II and were either brought directly on board (samples from here on called 'day 0') or displaced to a designated area with no hydrothermal activity (Mussel cemetery, 14°45.20'N, 44°58.78'W) for starvation. These mussels have been left without access to hydrothermal vent fluids for 10 days and were then brought up to the ship (samples from here on called 'day 10').

Another transplantation experiment was conducted with *Bathymodiolus azoricus* mussels from the Menez Gwen hydrothermal vent field. At the White Flames site (37°50'40 N, 31°31'8 W, -832 m) *B. azoricus* mussels were collected with scoop nets by the ROV QUEST and transferred into mesh cages. These cages were placed 80 m apart from the active vent to an inactive area with pillow lava, where no hydrothermal activity was detected by the ISMS (mussel purgatory, 37°50'40 N, 31°31'6 W, -843 m). After a maximum starvation time of 14 days, some cages containing starved vent mussels were returned into hydrothermal fluid flow at the White Flames vent site. The sampling procedure (see Figure 5.2) resulted in three different kinds of *B. azoricus* samples:

- Mussels from the native habitat.
- Transplanted mussels that were sampled at 1-14 days after starvation at the mussel purgatory.

- Retransplanted mussels that have been returned after transplantation to the active vent fluids, where they were allowed to recover for 1-7 days before sampling.

The mussels that have been retransplanted had starved for different times. For example, one batch of mussels starved for 10 days and recovered for 2 days, another batch starved for 13 days and recovered for 4 days, resulting in a different starvation and recovery time for each mussel batch.

All mussels used for transplantation experiments in Logatchev and Menez Gwen were collected from single locations. Even so, it is highly impossible that all mussels from one patch had experienced the same geochemical conditions. Furthermore, the cages that were used to transplant the mussels had to contain enough individuals for more than one sampling event, e.g. the sampling of mussels before and after the retransplantation. As a consequence, the ROV had to find mussel patches with a sufficient amount of animals, leading to a collection of samples from different patches at the vent sites.

5.3 Sampling at cold seep habitats at the GoM

B. brooksi and *B. childressi* mussels were collected at Garden Banks and Green Canyon cold seeps with metal-ringed nets and mussel pots (Cordes *et al.*, 2010), or by grabbing several individuals by using the submersible's manipulator arm. Environmental data at the sampling sites was recorded with an *in situ* mass spectrometer (ISMS, described by Wankel *et al.* (2010)). The ISMS delivers online readings of the concentration of diluted volatiles in the seawater from the location of the remotely operated vehicle (ROV) in the deep-sea. Scientists on board the research vessel can therefore acquire chemical information on sampling sites in real-time and plan their sampling strategy accordingly. Using the ROV manipulator, the ISMS sample inlet was positioned and held in place on the mussel sampling spots until the ISMS response reached steady state. A total of five independent sets of measurements were made at each sampling spot prior to the mussel sampling.

5.4 Dissection and fixation of mussel gills

The mussels were kept in chilled seawater (8°C) until they were processed (latest 4 hours after recovery by the ROV). The immediate dissection of the mussels included subsamples of gill tissue for FISH and qPCR analyses. Subsamples for FISH consisted of a small part of dissected gill tissue from the middle of the gill demibranches. Each gill subsample contained several gill filaments still attached to the mantle to decrease damage during the transport to the home lab. These gill samples were fixed for FISH in 1 x phosphate buffered saline (PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4) containing 2% paraformaldehyde at 4°C for up to 18 hours. Samples were washed three times in 1 x PBS for 10 minutes each and transferred to cold PBS/ethanol solution containing 1 x PBS and pure ethanol in equal parts. The FISH-fixed samples were kept at 4°C on board the research vessel, air-freighted back to the laboratory at 4°C and finally stored at -20°C. Subsamples for qPCR consisted of small gill pieces (∼200 mg), dissected from several parts throughout the whole gill filament. These samples were stored immediately at -20°C in 1.5 ml eppendorf vials.

A unique identifier was given to every sample in the form of barcoded permanent stickers. Furthermore, the background information of every sample (e.g. sampling time, location, fixation procedure, storage) was collected in spreadsheets. After the cruise, these spreadsheets were submitted to an online database, making the contextual data of all samples available to the researchers (Hankeln *et al.*, 2010).

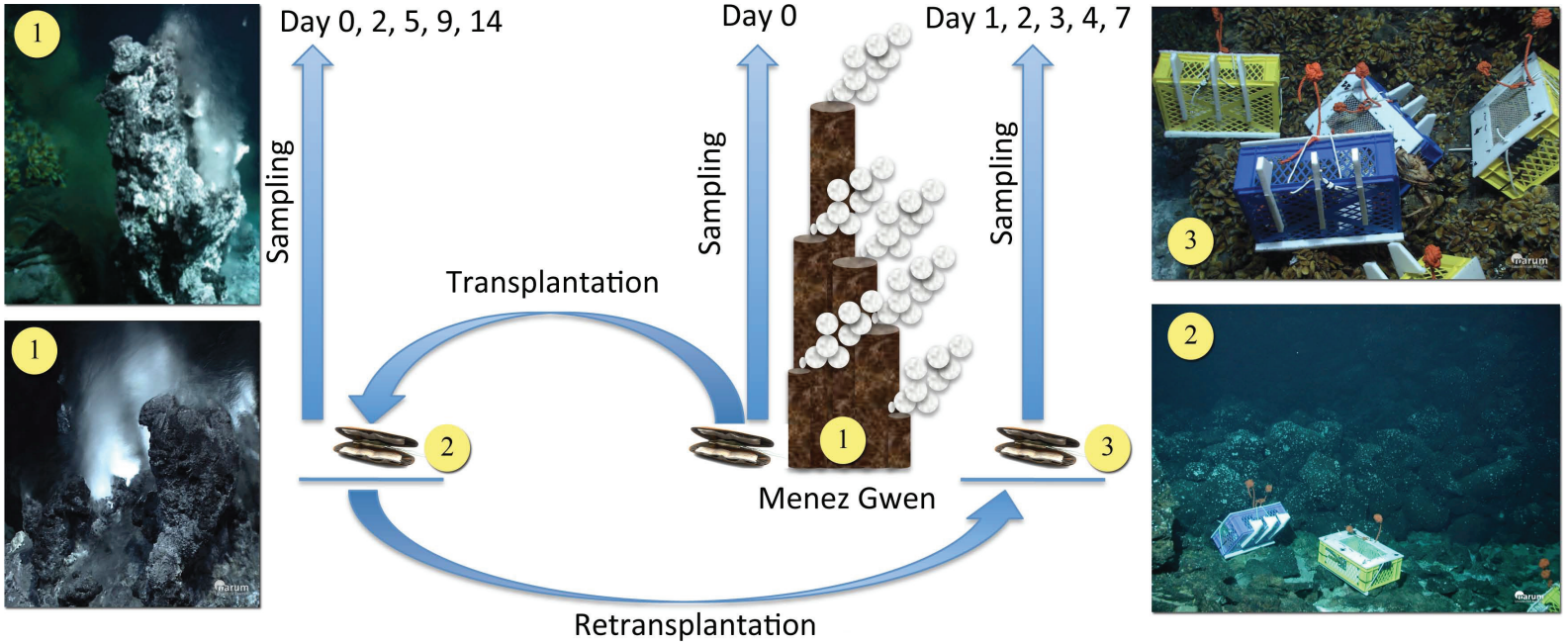

Figure 5.2: Scheme of the *in situ* transplantation experiment at Menez Gwen - *Bathymodiolus azoricus* mussels were displaced in cages from the active vent site White Flames (1) to an area 80 m east with no hydrothermal activity (2). After a maximum of 14 days, some cages have been brought back into hydrothermal venting at the White Flames site and put on top of mussel patches (3). Samples were brought on board the research vessel at several time points ("Sampling"). All ROV images are courtesy of the MARUM Bremen.

6

Quantification of bathymodiolin symbionts

6.1 Background information

6.1.1 3-dimensional fluorescence *in situ* hybridization (3D-FISH)

6.1.1.1 Quantitative FISH

The FISH method is targeting the 16S ribosomal RNA in cells with specifically designed oligonucleotide probes. These probes (derived from comparative sequence analysis, see Amann *et al.* (1995)) are linked to fluorescent dyes, entering the cells and binding specifically to the 16S rRNA of the cell. Subsequent excitation of samples under the fluorescent microscope leads to fluorescent emissions that are detected by the operator. The result is a specific staining of microbial cells, with the possibility to distinguish between many phylotypes in a given sample (Valm *et al.*, 2011).

The conventional FISH application for the localization and identification of microbial cells is based on 2D-imaging of fluorescent signals from a single focal plane, using wide field microscopy. For microbial quantification with FISH inside animal tissue, 2D-imaging is not sufficient. Since microbial cells are distributed 3-dimensionally within animal tissue (like the symbiont-host system), microbial

quantification based on 2D-imaging leads to an underestimation of the bacterial abundance. If target cells are distributed heterogeneously (which is mostly the case in biological systems), FISH-based quantification of fluorescent signals in 2D-space is not representative for the whole sample. Laser-scanning microscopy (LSM) can be coupled with FISH, acquiring fluorescent signals of bacterial targets throughout the sample planes (3D-FISH) in order to deliver image data that is representative. The additional advantage of laser-scanning over wide field microscopy is an adjustable pinhole in the optical path, which allows light collection almost exclusively from the focal plane and prevents most light causing blur. This leads to sharper images and better data prior to the software-based image analysis. Therefore, the possibility to acquire images from many focal planes together with a high resolution of these images makes the LSM the adequate tool for studying the abundance and distribution of bacteria in 3-dimensional space.

The procedure of image analysis consists of the following steps for most approaches (Landmann, 2002):

1. Acquisition of an image or stack of images of the region of interest (ROI).
2. Adjustment of the threshold intensity value to separate features of interest from the background.
3. Segmentation of the input images to generate the output images.
4. Computation of the data.

The output of the LSM-based FISH is a 3D-image stack that consists of fluorescently labeled target cells, therefore this method was termed 3D-FISH. With suitable software for the post-processing of these image stacks, the quantitative information for different bacterial cells in the sample can be assessed.

6.1.1.2 Application and data analysis

Before its application in microbiology, 3D-FISH has been used to investigate higher order chromatin architecture in cell nuclei (Solovei *et al.*, 2002). Nowadays, microbiologists use FISH-based methods to quantify bacteria in many environments. In this respect, 2D-FISH has been used for bacterial quantification in

marine pelagic (Schattenhofer *et al.*, 2009) and benthic (Mussmann *et al.*, 2005) ecosystems, or artificial habitats like activated sludge (Daims *et al.*, 2001). Also bacteria inside plants have been investigated by FISH (Ruppel *et al.*, 2006) and recently, 3D-FISH has been applied to quantify the symbionts of the *B. azoricus* mussel (Halary *et al.*, 2008). These studies show the quantitative potential of this method, but up to now the quantification of bacteria with 3D-FISH is not conducted by one accepted protocol but rather solved by each scientist with different approaches.

Since imaging becomes increasingly important in biological research, with more and more visual information being contained in the actual image, the trend has been to replace the subjective human operator with software-based post-processing of visual data. Software-based image processing is faster, more reproducible and therefore more reliable than visual inspections by human operators (Meijering & Cappellen, 2007). The challenge in quantitative 3D-FISH lies in the software-based quantification of bacterial signals, because until now there is no optimal solution at hand. Many different programs exist to detect fluorescent signals in 3D-space but the software-based segmentation does not always work sufficiently. Segmentation refers to the process that analyses the given image and judges over positive and negative signals (instead of the human operator choosing those). The output of a 'segmented image' contains only signals that surpass a certain intensity threshold. All signals above this threshold are recognized as positive signals (and will be quantified during the post-processing) and all signals below this threshold (negative signals) will not be taken into account for the quantification. The level of signal intensity chosen for the threshold is defined by the algorithms of the software. This crucial segmentation step can not yet be accomplished with all kinds of samples with the same efficiency and is therefore done by each scientist with a software of choice. There is a wide range of open source and licensed software, and three of them have been evaluated in depth for this thesis work and are introduced here.

ImageJ (Image Processing and Analysis in Java) is an open source and platform-independent image analysis software developed by the National Institute of Health (NIH) in the USA. It offers great flexibility in image processing

with over 400 plugins that are available now (Collins, 2007). One of those plugins (SymbiontJ) was used to quantify the endosymbionts in 3D-FISH image stacks from *B. azoricus* (Halary *et al.*, 2008). SymbiontJ is a combination of several other ImageJ plugins including thresholding, filtering, segmentation and 3D-object counting in one application. The operator can load all image stacks in one folder and SymbiontJ will quantify the bacteria as batch-process, giving out all results in a spreadsheet.

Daime (Digital Image Analysis in Microbial Ecology) is an open source software package developed by Dr. Holger Daims from the Microbial Ecology Group of the University of Vienna. The version 1.0 used during this evaluation automatically finds 2D and 3D objects in images and confocal image stacks and offers special functions for quantifying microbial populations and evaluating new FISH probes. The tool was successfully applied on quantitative analyses of activated sludge and provided quantitative evidence that functionally linked ammonia and nitrite oxidizers cluster together in their habitat (Daims *et al.*, 2006). Daime has a manual and automated threshold option, giving the operator the opportunity to segment the images manually, if the automated function is not adequate.

PHLIP (Phobia Laser scanning microscopy Image Processor) is a MATLAB toolbox (Mathworks, 1994) for the automated analysis of multichannel LSM data, which has been developed by the EU/FP5 funded project PHOBIA and is free-available. PHLIP has already been used for the quantitative analysis of biofilms, regarding for example the biovolume and surface-to-volume ratio of bacteria (Mueller *et al.*, 2006). In addition, PHLIP offers manual and automated thresholding and batch processing of all image stacks at once. The advantage is an intermediate step, at which the operator can check the threshold that has been set by the automated function and can adjust it easily for the whole batch.

6.1.2 Quantitative polymerase chain reaction (qPCR)

Quantitative polymerase chain reaction (qPCR) has been described as 'the most powerful tool for quantitative nucleic acids analysis' and is another method that is widely used to quantify microorganisms (Kubista *et al.*, 2006). The PCR protocol was inspired by nature and employs the mechanisms of natural DNA

amplification, using the enzyme DNA polymerase to copy the desired piece of single-strained oligonucleotides in consecutive cycles (Mullis *et al.*, 1986). Higuchi *et al.* (1993) improved the basic PCR method with the analysis of PCR kinetics and the construction of a system that detects PCR products as they accumulate. This was the birth of qPCR, a method to quantify specific nucleic acids as they accumulate during PCR.

6.1.2.1 QPCR steps

The possibility of real-time monitoring of the PCR product (and the difference to common PCR) comes with the use of a fluorescent reporter that binds to the product formed and reports its presence by fluorescence. SYBR Green (Molecular Probes) is a fluorogenic dye that emits a strong fluorescent signal when bound to double-stranded DNA. It is inexpensive, easy to use, and sensitive. It intercalates into double-stranded DNA and can be excited by fluorescent light at this stage (Figure 6.1A). In the qPCR, this leads to an accumulation of fluorescence with exponential increase. During the initial qPCR cycles, the fluorescent signal produced is weak and cannot be distinguished from the background but in the growth phase of the qPCR reaction, the signal increases exponentially. The point, at which the signal enters this phase, depends on the amount of template DNA. Therefore, the higher the DNA concentration, the earlier the signal reaches a particular threshold fluorescence level. The number of cycles required to reach this threshold is called the C_T-value (Figure 6.1B). The intensity of the fluorescence signal is linearly correlated with the amount of the target DNA product. By using a DNA standard with a known concentration of the target nucleotide sequence (e.g. a piece of the 16S rRNA gene), this linear relationship can be used to calculate the concentration of target molecules in the unknown sample by creating a standard curve (Kubista *et al.*, 2006)). After a certain amount of cycles, the signal saturates, which is due to a decreasing amount of reaction components, like the polymerase, dNTP's, primers, etc.
The disadvantage of using SYBR Green as a reporter molecule is, that it will bind to any double-stranded DNA in the reaction, including primer dimers and other non-specific reaction products, which can result in an overestimation of the

target concentration (Dharmaraj, 2006). Because primer dimers are much shorter than the specific double-stranded DNA product, they melt at lower temperature. If primer dimers occur, they employ a part of the polymerase during the qPCR runs for their amplification and therefore lower the amplification efficiency of the specific product (resulting in a lower amount of product and a smaller melting curve peak due to lower fluorescence). In order to check for these unspecific products, a melting curve analysis is programmed at the end of the qPCR reaction to visualize the dissociation of double-stranded qPCR products (Figure 6.1C). After a qPCR run, it is additionally recommended to sequence the final product to ensure that the right target sequence was amplified.

6.1.2.2 Application and quality control

Quantitative PCR is today the most sensitive and accurate method for the quantification of nucleic acids (Klein, 2002), is 1000fold more sensitive then dot blot hybridization (Malinen *et al.*, 2003) and can even detect a single copy of a specific transcript (Palmer *et al.*, 2003). This method for the quantification of microorganisms is especially popular in the clinical practice of infectious disease, where the rapid and reliable quantification of pathogens is critical for patient treatment (Yang & Rothman, 2004). Also in the microbiological field, qPCR has been established in the quantification of microorganisms, e.g. in bentic (Leloup *et al.*, 2007) and pelagic (Labrenz, 2009) marine habitats, mixed-organism assemblages (Mayfield *et al.*, 2009) and human nutrition, quantifying the bacterial load in probiotic yogurts (Masco *et al.*, 2007) or pathogens in drinking water (Sen *et al.*, 2007).

This popularity is reflected in an exponentially increasing amount of publications containing qPCR-derived data (25.000 publications in 10 years (Taylor *et al.*, 2010)) linked with decreasing costs for the qPCR chemicals and machines. The major downside of this publicity is that many researchers are nowadays expected to deliver qPCR-data additionally to their main research and due to the diversity in protocols (and lack of experience), this data of different sources is hard or impossible to judge by reviewers. To unite the scientific community in the way they do and interpret qPCR, Stephen Bustin and colleagues therefore

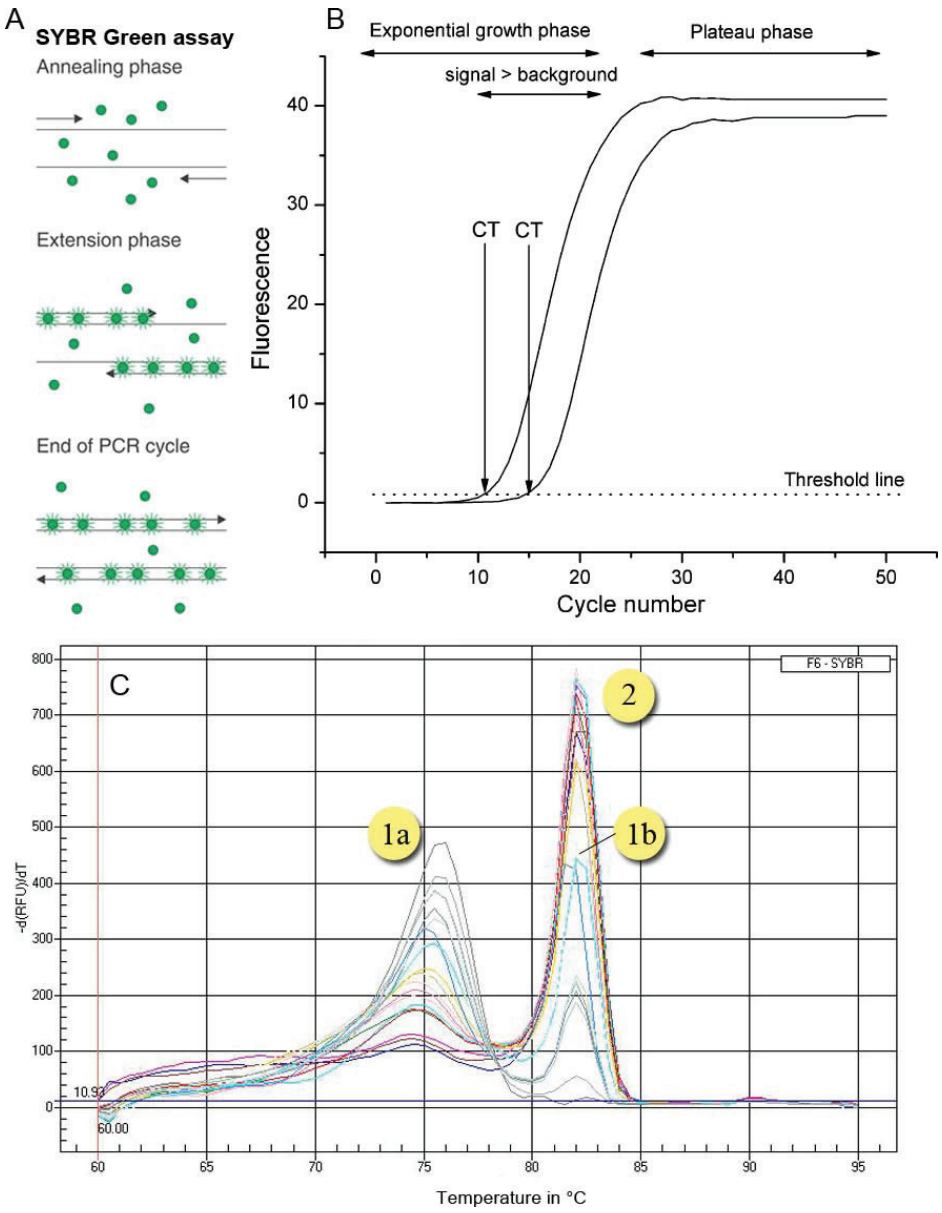

Figure 6.1: The SYBR Green assay and visualization of the qPCR - A: The SYBR Green assay employs a fluorescent reporter molecule (SYBR Green, green dots) that binds to double-stranded DNA during the repetitive qPCR cycles and results in a linear relationship of fluorescence signal and DNA product concentration. B: The cycle in which this fluorescence crosses a certain threshold and enters an exponential phase is called C_T-value. The reaction curve enters a plateau phase when the amount of crucial reaction components (e.g. primers, polymerase, dNTP's) is too low. C: The melting curve analysis after a qPCR run helps to identify unspecific products like primer dimers (1a), which melt at a lower temperature then the specific product and result in a lower qPCR amplification efficiency for the product (1b). If no primer dimers are formed, the specific product is amplified with high efficiency, resulting in more product and a higher fluorescence signal (2). Images from (A) van der Velden *et al.* (2003) and (B) Kubista *et al.* (2006).

try to establish guidelines for evaluating PCR experiments. These *Minimum Information for Publication of Quantitative Real-Time PCR Experiments* (MIQE) guidelines ask the operator of the experiments to provide a list of all chemicals, sequences, protocols and assay characteristics used. These details, submitted along with the publication, should help the reviewers in order to better validate the qPCR effort of the researcher (Bustin *et al.*, 2009, Bustin, 2010). Following these guidelines, all details concerning qPCR experiments for this thesis work are included in the Material and Methods chapter.

6.1.2.3 Marker genes for bathymodiolin symbioses

The quantification of microorganisms via qPCR is based on marker genes, which are the targets that get amplified during the repetitive cycles. These genes have to be present in all microorganisms studied, and at the same time, contain variable regions that allow the differentiation between closely related microorganisms.

The most prominent genetic marker used to identify microorganisms is the ribosomal RNA (rRNA). Being a part of the ribosomes in each living cell, the rRNA structure and sequence is conserved throughout all domains of live. Additionally, the rRNA contains enough variable regions that change at different rates, allowing phylogenetic analyses from close to distant relationships (Woese, 1987). The small subunit (SSU) of the prokaryotic rRNA contains the 16S rRNA (1541 nucleotides in *E. coli*), which has a high information content and parts with variable sequences (Thiele & Amann, 2011). This marker gene was chosen to create large phylogenetic databases (like the National Center for Biotechnology Information, NCBI), giving the researchers the possibility to design primers for yet unknown organisms.

The 16S rRNA gene can be present in multiple copies in the bacterial chromosome, enabling fast-growing cells to build up their protein factory with higher efficiency than other cells (Nomura *et al.*, 1977). In case of the bathymodiolin symbionts, the 16S rRNA operon number is not known yet. This means, that the 16S rRNA gene as a target for bacterial quantification can only be used in a relative quantification approach. Because of the unknown number of 16S rRNA operons in the bathymodiolin symbionts (and also to increase the reli-

ability of the quantification method itself) I used another phylogenetic marker gene in the qPCR-based quantification approach, the *recA* gene. The *recA* is a single-copy gene (352 nucleotides in *E. coli*) that encodes for a protein with crucial roles for homologous DNA recombination, SOS induction, and mutagenesis (Kowalczykowski *et al.*, 1994). It has been found in all free-living organisms, but some vertically transmitted symbionts (like the symbionts of the *Calyptogena* clam) have undergone reductive genome evolution (RGE) and lost the *recA* gene (Kuwahara *et al.*, 2008). The popularity of the *recA* gene as a phylogenetic marker has lead to an increased amount of sequences in public databases (almost 9000 sequences available from the functional gene database by now) that can be used to reconstruct phylogenetic relationships between organisms in addition to the common 16S rRNA-based approach. While the amount of available sequence data for the *recA* gene is still below that of the 16S rRNA gene, initial comparisons of their phylogenies suggest congruence (Eisen, 1995).

Additionally to the quantification of symbiont marker genes, I quantified the 18S rRNA gene of the mussel hosts. Comparable to the prokaryotic 16S rRNA gene, the eukaryotic 18S rRNA is part of the small eukaryotic subunit and a component of the ribosomes. The 18S rRNA gene has about 1800 nucleotides in Bivalvia (Meyer *et al.*, 2010) and includes highly conserved as well as variable regions. The reason for the quantification of this host gene additionally to the symbiont marker genes was the need of an internal standard for the qPCR-based quantification. The weight of the mussel gill tissue used for DNA extraction and subsequent gene quantification could not be determined sufficiently. By amplifying a marker gene of the host as well as symbiont marker genes in the same sample, the relative amount of symbiont marker genes could be normalized with the relative amount of the host marker gene, making the procedure independent of the weight of the gill tissue. For example, if twice the volume of mussel tissue was used for DNA extraction, the ratio of bacterial to host genes would still be the same, providing a relative value for the bacterial gene abundance.

6.2 Overview of quantification protocols used in this study

Two different approaches based on qPCR and 3D-FISH for the quantification of bathymodiolin endosymbionts have been evaluated for this study (Figure 6.2). Originating from an environmental sample (here: the mussel *B. puteoserpentis*), gill tissue samples were taken and processed for qPCR and 3D-FISH analyses. The aim was, to optimize each protocol and to compare their potential in quantifying a proxy for the bacterial abundance, either in terms of symbiont genes (based on qPCR) or symbiont biovolume (based on 3D-FISH).

6.3 3D-FISH

The goal of the 3D-FISH approach is to quantify the biovolume of the thiotrophic and methanotrophic symbionts of *B. puteoserpentis* inside the bacteriocytes. The necessary steps are described in the following sections and involve the hybridization of bacterial 16S rRNA with specific oligonucleotide probes, the data acquisition at the confocal laser-scanning microscope and the software-based biovolume quantification.

6.3.1 Specimen preparation

The FISH-fixed gill pieces were embedded in a wax matrix and cut with a microtome to allow access of the probes to the symbiont-containing tissue. I used Steedman's polyester wax (Steedman, 1957) for the embedding, which is a general-purpose histological embedding medium that combines the advantages of a low melting temperature (37°C), negligible tissue shrinkage and is ready to be cut after one day. Mussel gill tissue consisting of 10-20 filaments was dehydrated in ascending ethanol concentrations (70% - 80% - 96%), 30 min at room temperature (RT) for each concentration. The filaments were then placed in a mixture of Steedman's wax and ethanol (50/50 v/v) for 60 min at 38°C followed by three steps in pure wax for 60 min at 38°C.

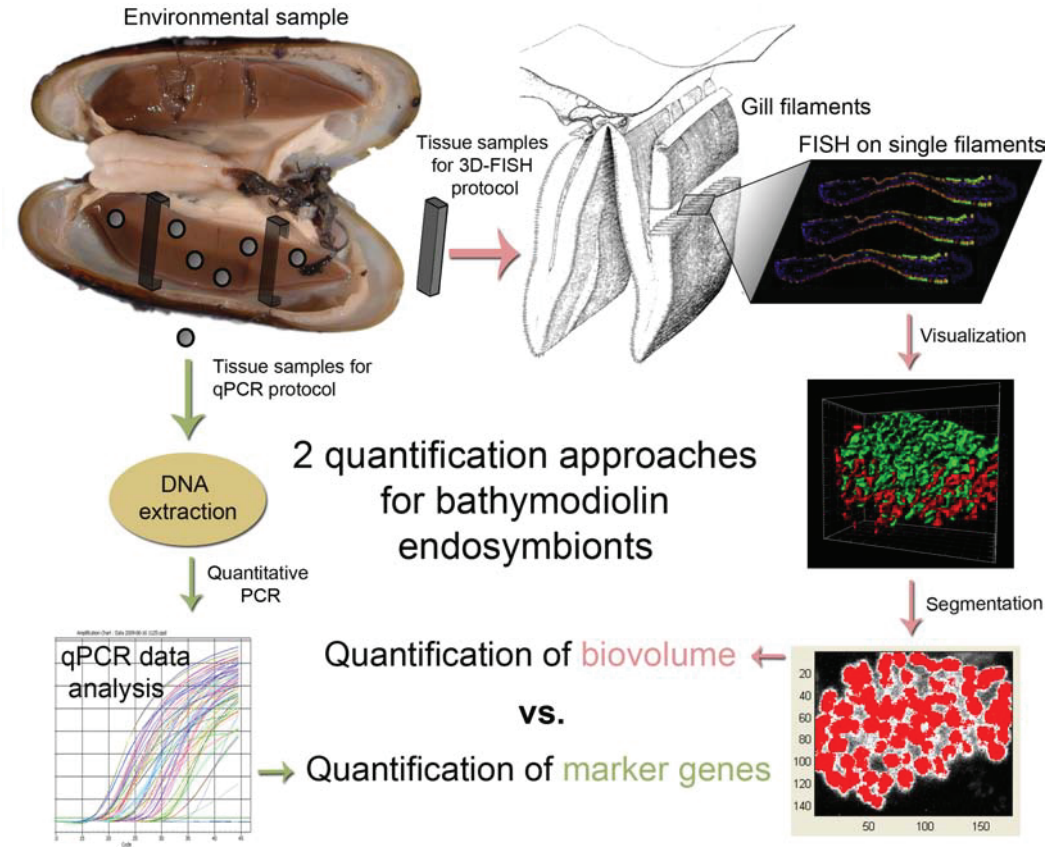

Figure 6.2: Quantification protocols used in this study - Two protocols for the quantification of bacteria inside animal tissue. Based on the environmental sample (here: *B. puteoserpentis*), tissue samples are processed for the qPCR-based protocol (green arrows) and the 3D-FISH-based protocol (pink arrows). For the qPCR-based approach, genomic DNA was extracted from mussel gill tissue and qPCR was used to quantify symbiont and host marker genes. For the 3D-FISH approach, gill filaments were dissected and hybridized with specific FISH probes for the bacterial targets. Confocal laser-scanning microscopy was used to acquire image stacks from hybridized sections of gill tissue. Post-processing software was used to visualize the symbionts and to quantify their biovolume.

The block was left to harden over night and was cut with a microtome (Leica, Germany) in 2-20 micrometer thick sequential sections, which were mounted on SuperFrost Plus slides (Menzel, Germany). Sections were allowed to dry over night and then stored at 4°C.

6.3.2 Fluorescence *in situ* hybridization

Specific oligonucleotide probes for the symbionts of *B. puteoserpentis* have been designed by Duperron *et al.* (2006). For dual hybridizations, the probe for the thiotrophic symbiont BMARt-193 (CGAAGGTCCTCCACTTTA), and the probe for the methanotrophic symbiont BMARm-845 (GCTCCGCCACTAAGCCTA) were labeled with the fluorescent dyes Cy3 and Cy5 (ThermoHybaid, Germany). Both probes hybridize specifically with their target at 35% formamide (Duperron *et al.*, 2006). EUB338 I-III (Amann *et al.*, 1990, Daims *et al.*, 1999) and NON 338 (Wallner *et al.*, 1993) were used as positive and negative controls. The slides with gill sections were put in glass trays and rehydrated in three baths of descending ethanol concentrations (96% - 80% - 70%) for 5 min at RT each. The air-dried sections were encircled with a rubber pen (pap pen, Kisker Biotechnology, Germany). 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 at an end concentration of 5 ng 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 in a 1% DAPI solution, left for 3 min, rinsed with distilled water and air-dried. Sections were mounted in a mixture of Citifluor and Vectashield and stored at -20°C until examination.

6.3.3 Stereological data acquisition

The 3D-FISH quantification approach is based on acquiring image stacks of bacteriocytes from the mussel gill and quantifying the symbionts inside these cells. Recent studies already point out, that the distribution of the symbionts across the gill is not random, but follows a distinct pattern. This pattern could for example be due to different energy source concentrations and their accessibility for the symbionts (Wendeberg *et al.*, 2011). In order to account for this potentially heterogeneous distribution of target cells inside the animal tissue, we applied the rules of sampling in ordered biological systems as proposed by Weibel *et al.* (1966). As the representative structure here (out of 5 patterns that Weibel described) we selected the polarized structure, because the symbionts show a gra-

dient like distribution pattern from the ciliated part towards the abfrontal part of the gill (Figure 7.6). A stereological sampling procedure known as 'multistage sampling' (Weibel, 1979) was applied in order to quantify the volume of the bacteria. For the microscopical fields of view, the cross section of the mussel gill filament was divided into three parts: Ciliated edge, middle, and abfrontal edge (see Figure 7.5).

From each mussel gill, two filaments have been screened in the ciliated, middle and abfrontal area of the cross section. For each area, 12 image stacks of bacteriocytes were acquired, leading to 36 image stacks per filament and 72 image stacks per mussel. All images were taken with the confocal laser-scanning-microscope LSM 510 (Carl Zeiss, Oberkochen, Germany). To minimize the noise and to keep the photobleaching low, I selected an acquisition time of 4 s per focal plane and summed the signal of four scans per focal plane to produce 1024 x 1024 pixel RGB 8 bit images as a standard procedure for all the applications. The vertical distance between the focal planes was 0.2-0.3 μm and the image stacks contained an average of 60 consecutive images.

6.3.4 Biovolume quantification

In the scope of this study I tested three different softwares that are capable of post-processing fluorescent images in respect to their applicability in reliable symbiont quantification. After the acquisition of image stacks, these softwares were used to measures the voxels (colored pixels in 3-dimensional space) of multiple channels in image stacks. In case of the symbionts of *B. puteoserpentis*, each image stack contained the signals of the methanotrophic and thiotrophic symbionts in two different channels. Depending on the software, the abundance of each symbiont species was either given in biovolume (PHLIP and SymbiontJ) or percentage (Daime).

6.3.5 Visualization of bacteria in 3D

Additionally to the quantification of bacteria, I used 3D-FISH data to visualize the bathymodiolin symbionts in 3D. LSM-data consisting of image stacks of bacteriocytes was processed with the Imaris software package (Bitplane, Switzerland),

using the signal information of serial sections to recreate the spatial distribution of the symbionts in 3-dimensional space. Furthermore, the software was used to coat the fluorescent signals of the bacteria with isosurfaces, resulting in maximum signal intensity of all positive signals and a suitable 3D-image for studying the bacterial distribution.

6.4 QPCR

6.4.1 Total nucleic acid extraction

The efficiency of the DNA extraction method used to produce the DNA that is amplified during the qPCR has a direct influence on the results (Smith *et al.*, 2006). Therefore, three common DNA extraction protocols have been tested in respect of DNA yield and purity of the product. These protocols were the Qiagen DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany) and the standard protocol for DNA extraction after Zhou and colleagues (Zhou *et al.*, 1996). The later protocol has been tested using hexadecylmethylammonium bromide (CTAB) or polyvinylpolypyrrolidone (PVPP) to remove contaminants from the DNA. Gill tissue from two different samples has been homogenized in liquid nitrogen to ensure efficient cell lysis and afterwards 25 mg of the homogenate have been used for DNA extraction with the different protocols. The concentration and purity of the DNA were determined spectrophotometrically by measuring the absorbance of the DNA extract at 230, 260 and 280 nm with NanoDrop (NanoDrop Technologies, Montchanin, USA). Based on this preliminary test, all samples for this study have been processed with the Qiagen Kit for DNA extraction. DNA was stored in aliquots à 50 μl at -20 °C.

6.4.2 *recA* sequence analysis

Prior to this study, *recA* gene sequences of the bathymodiolin symbionts were not known. One *recA* gene sequence was available for the sulfur oxidizer *Thiomicrospira crunogena*, which is considered closely related to the thiotrophic *Bathymodiolus* symbiont group based on the 16S rRNA gene. The only available

methanotrophic *recA* sequence was that of *Methylococcus capsulatus*, which is only distantly related to methanotrophic *Bathymodiolus* symbionts and other free-living methanotrophs.

Prior to the *recA* primer design, I therefore analyzed the *recA* sequences of bathymodiolin symbionts in respect to their phylogenetic affiliation to known *recA* sequences of free-living bacteria with sequence data from the functional gene database (http://fungene.cme.msu.edu). For this purpose, *recA* genes of symbionts from different mussel hosts were amplified with the universal *recA* primer set, cloned in *E. coli* and sequenced. These mussel hosts were *B. puteoserpentis*, *B. thermophilus*, *B. childressi* and a new *Bathymodiolus* species that lives in dual symbiosis with sulfur and methane oxidizers at the southern Mid-Atlantic Ridge (SMAR). *B. thermophilus* harbors only sulfur oxidizers and *B. childressi* has only methane oxidizers. With the *recA* sequences of these two mussel species, the *recA* sequences obtained from the same dual symbiotic host should be allocated either to the methanotrophic or thiotrophic co-occurring symbiont type (all in co-work with K. van der Heijden, unpublished).

6.4.3 Primer design for qPCR

The qPCR requires special primers, because the desired amplicons must be short and the melting temperature of all primers must be the same because they are used simultaneously in the same qPCR run. All oligonucleotide primers were designed with the probe design function in ARB (Ludwig *et al.*, 2004) or by visually identifying a suitable target in the ARB alignment, using the SILVA small subunit alignment (Pruesse *et al.*, 2007). The sequences in the ARB alignment originate from the amplification of the symbiont and host marker genes with universal primers (see Table 6.1). Specific primers for the 16S RNA and 18S rRNA marker genes were designed based on these alignments with the SILVA ribosomal RNA database. Specific primers for the *recA* genes of the symbionts were designed by sequence comparison in the ARB alignment of bathymodiolin *recA* sequences and known *recA* sequences of free-living bacteria (Section 6.4.2). After choosing a suitable oligonucleotide sequence, the NetPrimer software (PREMIER Biosoft International, www.premierbiosoft.com) was used to determine possible primer

dimers and calculate melting temperatures. The result of the whole primer design effort were 3-4 primer pairs per gene target (18S rRNA of the host, 16S rRNA and *recA* of the symbionts). In the process of method optimization (tests in qPCR efficiency, melting curve analysis and gel electrophoresis for each primer pair), one primer pair for each target gene was chosen for the qPCR-based quantification of the symbionts.

6.4.4 Quantitative polymerase chain reaction

The DNA extract of the bathymodiolin gill tissue was quantified for marker genes of the symbionts (16S rRNA and *recA*) and the host (18S rRNA). 100-fold dilutions from DNA extracts of these environmental samples as well as standards and negative controls (NTC) were amplified in triplicate in a SYBR Green qPCR assay using the specific primers designed for this study. Each 25 μl qPCR reaction contained 5 μl of template, 12.5 μl of POWER SYBR Green PCR Master Mix (Applied BioSystems), 0.25 μl of each primer (50 μM stock) and 0.5 μl of fluorescein (0.5 μM stock, BioRad) as well as 6.5 μl of sterile water. Polymerase chain reaction and fluorescence signal detection was performed using the iQ5 multicolor real-time PCR detection system (BioRad) and iCycler DNase and RNase free 96 well plates (BioRad). An initial denaturation step of 95°C was followed by 40 cycles of 95°C for 15 seconds (denaturation) and 60°C for 1 minute (annealing and elongation). The subsequent melting curve was recorded by raising the temperature in 0.5°C steps from 60°C to 95°C with a dwell time of 10 seconds per step.

6.4.5 QPCR standards

For the quantification of marker genes of the mussel and its symbionts with qPCR, standards of known amounts of template had to be created. Therefore, the symbiont and host marker genes (16S rRNA, *recA* and 18S rRNA) were amplified via common PCR, using DNA extracted from *B. puteoserpentis* gill tissue and universal qPCR primers (see Table 6.1).

Table 6.1: Universal primers for qPCR - These universal primers have been used to amplify known sequences of the 16S rRNA, 18S rRNA and *recA* target genes in *B. puteoserpentis* host and symbionts in order to create standards for the qPCR.

Target gene	Primer	Organisms used for primer design	Source
18S rRNA	3F 5R	*Macrobiotus hufelandi*	Giribet *et al.* (1996)
16S rRNA	GM3F GM4R	*Escherichia coli*	Muyzer *et al.* (1995)
recA	RecAF RecAR	*Sodalis glossinidius*	Dale *et al.* (2003)

The reaction mixture for the PCR amplifications contained 50 pmol of each primer, 2.5 μmol of each deoxynucleotide triphosphate, 1x Eppendorf buffer, 1 U of Eppendorf Taq and approximately 200 ng of the genomic DNA template. The final volume was adjusted to 20 μl with sterile water. The PCR program consisted of an initial denaturation step at 95°C for 5 min, followed by 34 cycles of 95°C for 1 min (denaturation), 60°C for 1 min (annealing) and 72°C for 3 min (elongation), with a final elongation step at 72°C for 10 min. PCR products for each target gene were pooled from 6 separate reactions and purified with the QIAquick PCR purification Kit (QIAGEN, Hilden, Germany).

Because amplification with the universal primers for 16S rRNA and *recA* in mussels with dual symbiosis results in a mixture of these genes from sulfur and methane oxidizers, the PCR products were cloned in *E. coli*. The ligation was performed using the pGEM-T Easy vector (Promega, USA). The transformation of *E. coli* with the ligation product was conducted with the TOPO-TA Kit (Invitrogen, USA). After partial sequencing of 96 clones, the phylogenetic affiliation of the 16S rRNA gene was checked with BLAST (Altschul *et al.*, 1990). The identification of clones containing the *recA* gene sequences was done in the same way as the primer design: by relating the *recA* clone sequences to other known bacterial *recA* sequences (Section 6.4.3). Plasmid DNA of the clones containing the symbiont gene sequence inserts was extracted using the QiAprep Spin Miniprep Kit (Quiagen, Hilden, Germany) and the insert size was controlled by PCR screening with the vector primers M13F and M13R.

The amount and purity of DNA for all standards was measured with NanoDrop and the copy number for each marker gene was calculated, assuming a molecular mass (MW) of 660 Da for double-stranded DNA, using the following formula including Avogrado's number (*):

$$\text{Copy Number} = \frac{^{*}6.023\times10^{23}\,(copies\cdot mol^{-1}) \times concentration\,of\,standard\,(g\cdot\mu l^{-1})}{MW(g\cdot mol^{-1})}$$

Aliquots of these standards were frozen at -20°C and only used once for qPCR runs to minimize errors due to DNA degradation. Serial 10-fold dilutions of each standard were checked in qPCR runs with consecutive melting curve analysis for their consistency and reproducibility. Only standards with a consistent amplification rate of 3.2-3.4 cycles from one 10-fold dilution to the next (resulting in an amplification efficiency of 95-100%) were used in the quantification of marker genes.

6.4.6 QPCR data analysis

The marker genes in the environmental samples were quantified against the standard curve using the iQ5 optical system software v.2.0 (BioRad). The iCycler analysis mode of *PCR base line subtracted curve line* was chosen to enable an automatic determination of the baseline and C_T-values. The C_T value marks the cycle in which the sample fluorescence crosses a certain baseline and is used for the relative quantification calculations. For each standard curve, the slope, y-intercept and the coefficient of determination (R^2) were determined by the software. The amplification efficiencies (E) of dilution series of the standard were calculated by using the difference of the C_T-values (ΔC_T = C_T dilution 2 - C_T dilution 1) in the following equation:

$$E = (10^{\frac{1}{\Delta C_T}} - 1) \times 100$$

The C_T-values of the symbiont genes (16S rRNA & *recA*) were normalized against the C_T-values of the host gene (18S rRNA). For this, the comparative C_T method (2-ΔC_T) has been used to calculate the relative abundance of the

target gene for each sample (Kubista *et al.*, 2006). Therefore, the result of this quantification effort was the amount of each symbiont marker gene (16S rRNA and *recA*) relative to the amount of the host marker gene (18S rRNA) in the mussel samples.

6.5 Validation of the qPCR-based quantification of symbionts

In order to validate the optimized qPCR protocol, I quantified two different target genes (16S rRNA & *recA*) of the endosymbionts in *B. puteoserpentis* mussels and compared their relative abundance. The hypothesis was, that the relative abundance of the 16S rRNA gene in a sample represents a multiple of the relative abundance of the single copy *recA* gene, and that the multiple corresponds to the number of 16S rRNA operons. The ratio of the relative abundances of the 16S rRNA and *recA* gene for one symbiont species should not differ in between samples, because this ratio is independent of the symbiont abundance. This experiment should therefore (1) validate the qPCR method itself and (2) yield information about the 16S rRNA operon number of the two different symbiont species of *B. puteoserpentis*. Four individuals of *B. puteoserpentis* were used for this experiment and represent mussel samples from the Logatchev hydrothermal vent field (day 0 of the starvation experiment).

6.6 Method comparison - qPCR vs. 3D-FISH

In order to test which method is most suitable for the reliable quantification of the mussel endosymbionts, I applied quantification approaches based on qPCR and 3D-FISH on samples from the Logatchev starvation experiment. This experiment was conducted to test the accuracy and reliability of qPCR and 3D-FISH in the quantification of bathymodiolin endosymbionts. Furthermore, different regions of the symbiont-containing mussel gill were compared with 3D-FISH in respect to symbiont biovolume, in order to study the distribution pattern of the endosymbionts across the gill section.

The symbionts 16S rRNA gene abundance (measured by qPCR) and their biovolume per bacteriocyte (measured by 3D-FISH) were studied in six mussels (three individuals from day 0 and three individuals from day 10 of the starvation experiment). For the 3D-FISH quantification, cross sections of *B. puteoserpentis* gill filaments were hybridized with general and specific FISH-probes for both endosymbiont species.

6.6.1 Statistical analyses

In order to check whether qPCR and 3D-FISH data can be correlated, the correlation coefficient (R^2) was used to explain a linear regression model. In addition, the Spearman's rank correlation (ρ) was used to correlate the data. This is a non-parametric measure of correlation that describes the relationship between two variables without making any assumptions about the nature of this relationship and is therefore more robust than the parametric correlation coefficient. The Spearman's rank correlation is given by the equation:

$$\text{Spearman's Rank correlation } (\rho) = 1 - \frac{6 \cdot \sum d_i^2}{n \cdot (n^2 - 1)}$$

All statistics have been checked with SigmaStat (Version 3.5).

Part III

Results

7

Fast and reliable quantification of bathymodiolin symbionts

This first part of the thesis results deals with the evaluation of two methods (qPCR and 3D-FISH) for the quantification of bathymodiolin endosymbionts. It includes the assessment of a suitable software to process the 3D-FISH results and optimizations on qPCR. In the end, the most suitable quantification approach is presented and tested on the *B. puteoserpentis* symbiosis from the MAR.

7.1 Software evaluation for post-processing 3D-FISH data

After acquiring image data with the confocal laser-scanning microscope, the image stacks of symbiont-containing signals had to be post-processed by suitable software. For reliable symbiont quantification, the software processes had to be transparent enough for the operator to understand the different steps that lead to the results. To reduce subjectivity of the operator, the software also had to offer automated thresholding of the image stacks. The steps of thresholding and segmenting during image quantification are crucial, because they separate positive signals (here: symbionts) from negative ones (background noise) and therefore determine the outcome of the whole quantification process. The three softwares ImageJ (with plugin SymbiontJ), Daime and PHLIP were evaluated in these re-

spects (see Table 7.1) and used to quantify the thiotrophic and methanotrophic symbionts in *B. puteoserpentis*.

Table 7.1: Evaluation of post-processing software for 3D-FISH - Comparison of three different open-source software packages for the quantification of fluorescent signals in 2D- and 3D-images.

Software feature	**SymbiontJ**	**Daime**	**PHLIP**
Open source	✓	✓	✓
Image file type	.lsm	.lsm	.tif stack
2D/3D image processing	✓/✓	✓/✓	–/✓
Batch processing	✓	–	✓
Single processes are comprehensible for the operator	–	–	✓
Automated thresholder	✓	✓	✓
Manual thresholder	–	✓	✓
Biovolume output	μm^3	percent	μm^3
Additional features	Independent of operating system	Testing of FISH probes, intuitive handling	Surface-to-volume ratio, colocalization analysis
Reference	Halary *et al.* (2008)	Daims *et al.* (2006)	Mueller *et al.* (2006)

In **SymbiontJ**, the setting of a threshold to separate positive signals from background is done automatically for each stack and only visual control is given. This means that it is not possible to compare or change the thresholds that have been chosen by the plugin to segment the images. For this reason, this software was not applied further in the quantification approach. With the software **Daime**, automatic thresholding was not sufficient, because the software was too inaccurate in the detection of the small hybridization signals of the sulfur oxidizers. The manual thresholder was not an option due to time constrains. Also, a batch processing of many image stacks at a time was not available. Due to these restrictions, the Daime software was not used.

The software package **PHLIP** offers manual and automated thresholding and batch processing of all image stacks at once. The advantage is an intermediate step, at which the operator can check the threshold that has been set by the

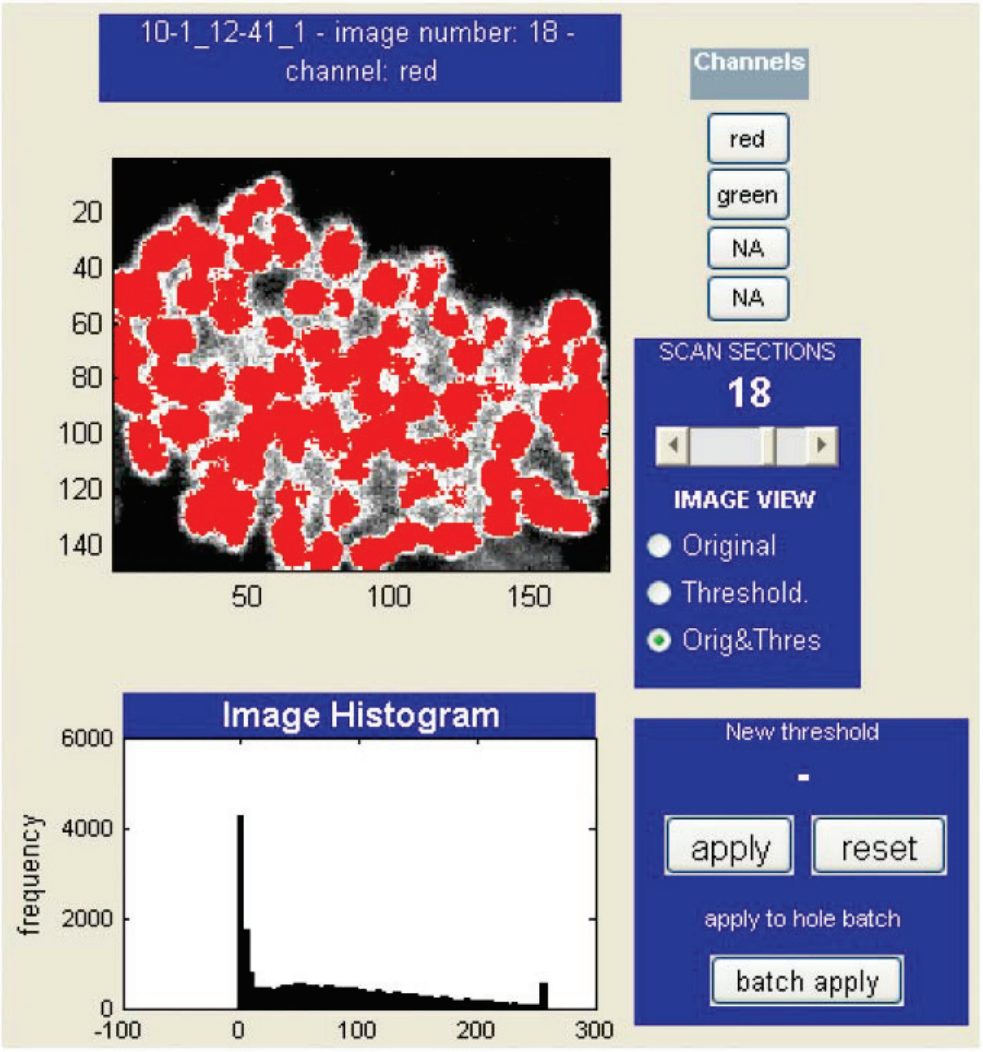

Figure 7.1: The MATLAB package PHLIP - Segmentation results of the methanotrophic symbiont fraction in a bacteriocyte of *B. puteoserpentis*, using the automated threshold option of PHLIP. Areas in red are positive signals that are taken into account by the software for the bacterial quantification in 3D.

automated function and can adjust it easily for the whole batch (see Figure 7.1). Additionally, PHLIP informs the user about the exact threshold value and therefore enables comparisons between image stacks. This transparency of processes for the operator and the efficient quantification of the symbiont signals in a batch qualified this software for post-processing of the bathymodiolin symbionts in this thesis work.

7.2 QPCR optimizations

7.2.1 Evaluation of DNA extraction efficiency

Three common DNA extraction protocols have been used to extract genomic DNA from *B. puteoserpentis* mussels and were evaluated in regard to DNA yield and purity. The purity of the DNA extract is given by the ratio of A260/A280 (protein contaminations) and A260/A230 (humic acid contaminations). Whereas the purity of all DNA extractions is comparable (Table 7.2), the extraction of genomic DNA with the QuiaQuick Blood & Tissue Kit yielded the highest amount of DNA per amount of gill tissue and was therefore used for all DNA extractions during this study.

Table 7.2: Evaluation of DNA extraction methods - Results of the DNA extraction evaluation, including the protocols of the QuiaQuick Blood & Tissue Kit and two different version of the Zhou *et al.* protocol.

DNA extraction method	DNA yield	A260/A280	A260/A230
QuiaQuick Blood & Tissue Kit	Sample 1: 7.8 ng per μg tissue	2.08	2.27
	Sample 2: 4.1 ng per μg tissue	2.03	2.19
Zhou *et al.* (CTAB 2%)	Sample 1: 1.6 ng per μg tissue	2.04	2.12
	Sample 2: 1.1 ng per μg tissue	2.01	2.09
Zhou *et al.* (PVPP)	Sample 1: 2.4 ng per μg tissue	2.05	2.37
	Sample 2: 1.8 ng per μg tissue	2.05	2.26

7.2.2 *recA* sequence analysis of *Bathymodiolus* symbionts

Amplification and cloning of the *recA* gene of *Bathymodiolus* symbionts using universal *recA* primers yielded sequences of 690-715bp length. The alignment of these sequences to other *recA* sequences of free-living bacteria and subsequent maximum-likelihood analysis resulted in 5 well separated clone groups for the *recA* sequences from 4 *Bathymodiolus* hosts (Figure 7.2). These groups separated in two sister groups: one sister group contained two clone groups separated according to the host species *B. puteoserpentis* (one group) and *B. childressi* and the new *Bathymodiolus* species from the SMAR (one group). The other sister group also contained well separated clone groups that were derived from the host species *B. thermophilus*, *B. puteoserpentis* and the new *Bathymodiolus* species

from the SMAR. The symbionts of *B. thermophilus* and *B. childressi* are separated in two sister groups. Because *B. thermophilus* symbionts are thiotrophic, while *B. childressi* symbionts are methanotrophic, we suggest that one of the two sister groups represents thiotrophic *Bathymodiolus* symbionts while the other one represents methanotrophic symbionts.

According to this phylogenetic tree, all *Bathymodiolus* symbionts form a monophyletic group that is closest related to the sulfur oxidizer *Thiomicrospira crunogena*. This seems to contradict the 16S rRNA phylogeny of these groups, however, the data basis for *recA* genes is still small and this picture may change in the future, when more *recA* sequences will be available.

7.2.3 Primer evaluation

The evaluation of the designed primer pairs for qPCR included qPCR efficiency tests, melting curve analysis and product inspection with gel electrophoresis. For each target gene (16S rRNA, *recA* and 18S rRNA) a minimum of 3 primer pairs were designed and tested in this respect. The qPCR runs indicated, that even qPCR primer pairs designed for the same target gene can differ in their amplification efficiency (Figure 7.3a). This difference could be due to the formation of primer dimers, as it is indicated by the melting curves (Figure 7.3c). A decrease of the amplification efficiency due to primer dimers was apparent for primer pair 1 in the example given in Figure 7.3c (red curves). Furthermore, the decreased amplification efficiency of primer pair 1 lead to a lower amount of product and a weaker band in the gel electrophoresis, compared to primer pair 2 (Figure 7.3b).

Only primer sets that showed a clean single peak in their melting curve (indicating the amplification of one specific product only), possessed a repeatable qPCR efficiency of $\geq$95% and amplified the right target (proved by direct sequencing) were considered for qPCR application. The primers used for the amplification of target genes in this study are listed in Table 7.3. Bathymodiolin hosts and symbionts which can be targeted by the 18S rRNA and 16S rRNA primers used in this study are listed in Table 7.4.

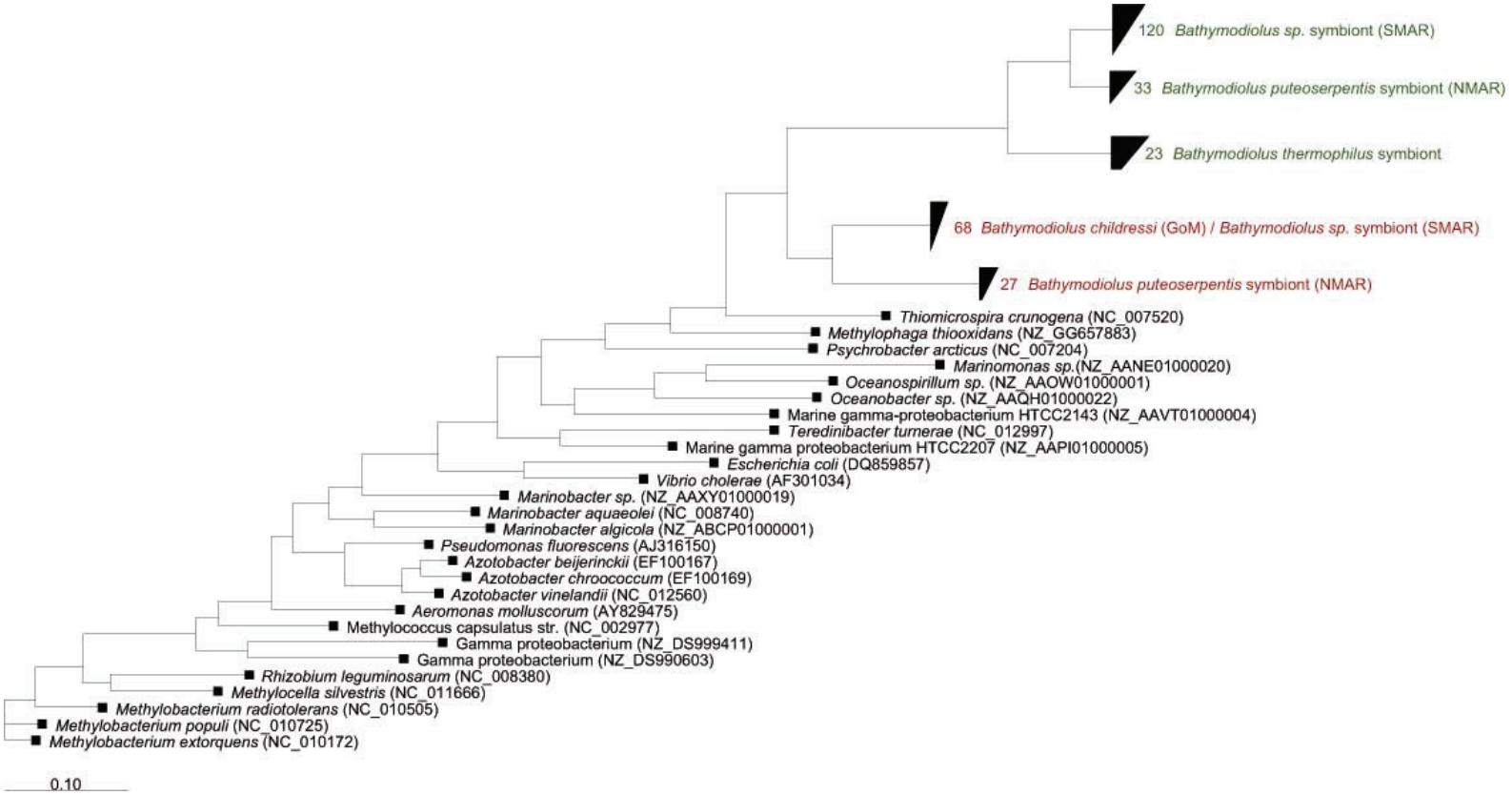

Figure 7.2: Phylogenetic reconstruction of the *recA* gene of free-living bacteria including the new sequences of *Bathymodiolus* symbionts

- Clone libraries of *recA* genes amplified from different bathymodiolin mussel hosts were aligned with known *recA* sequences of free-living bacteria. A maximum-likelihood phylogenetic tree was constructed (HKY nucleotide substitution, scale bar represents 10% genetic distance). Sister groups of bathymodiolin clones are marked with the number of clones they contain. According to the clustering, we identified thiotrophic (green) and methanotrophic (red) symbionts and used the respective sequences for specific primer design. The *recA* phylogeny was produced in co-work with K. van der Heijden (unpublished data).

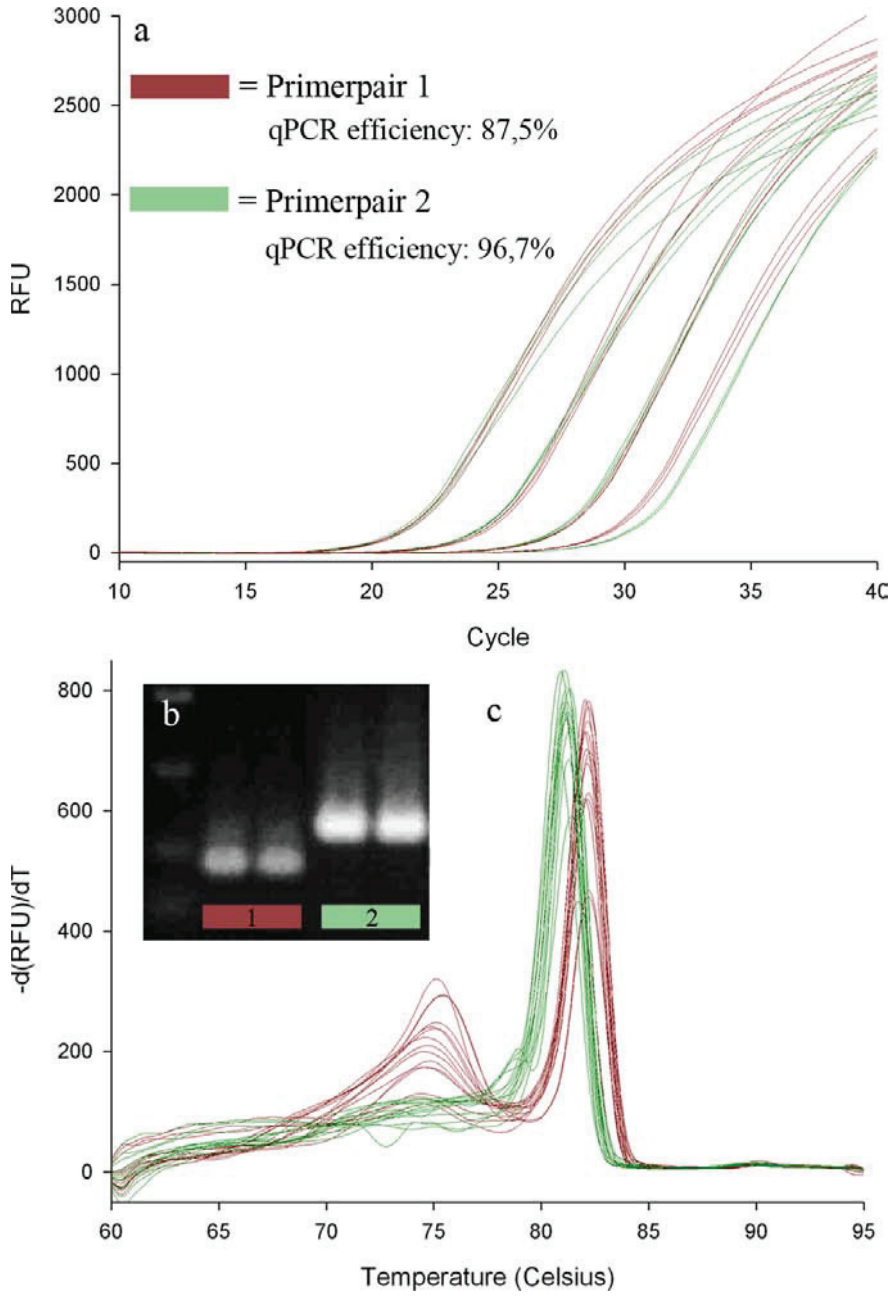

Figure 7.3: Primer evaluation for qPCR - Two primer pairs are compared: (a) in their qPCR efficiency, (b) in the sizes and product yields of their products and (c) in their melting curve behavior. Primer pair 1 shows a lower qPCR efficiency compared to primer pair 2 (a), less amount of product after PCR amplification (b) and an additional small peak in the melting curve which is an indication for primer dimer formation (c).

Table 7.3: Results of primer design for qPCR - Primers used for the quantification of each target gene with qPCR. The primer positions relate to [1,3]positions on the nucleotide sequence of aligned environmental sample sequences (*B. puteoserpentis*) for 18S rRNA and *recA*, and [2]the respective position on the 16S rRNA gene of *E. coli*.

Target organism	**Target gene**	**Primer name**	**Sequence 5'→3'**	**Primer position**	**Product size**
Mussel host	18S rRNA	Bputeo_1_for	GAAGGGCACCACCAGGAGTG	780-799[1]	123 bp
		Bputeo_1_rev	CCACCACCCACCGAATCAAG	883-902[1]	
Thiotrophic symbiont	16S rRNA	qThio_4_for	AAGGAACATCAATGGCGAAG	711-730[2]	134 bp
		qThio_4_rev	AAATCCTCCCAACGGCTAGT	825-844[2]	
	recA	recA_thio_1_for	GTGCGCTCAGGTGGTGTGGA	213-232[3]	134 bp
		recA_thio_1_rev	TGAGACATTAAACGCGCCTGC	326-346[3]	
Methanotrophic symbiont	16S rRNA	qMeth_2_for	ATACCGCATACGCCCTACGG	172-191[2]	107 bp
		qMeth_2_rev	CGCCTTGGTGAGCCTTTACC	259-278[2]	
	recA	recA_Meth_1_for	AGTTGGTGGTACGGCTGCTTT	71-91[3]	146 bp
		recA_Meth_1_rev	CGCACCAGCATATCAGCGAT	197-216[3]	

Table 7.4: Gene targets of the specific qPCR primers designed for this work - Data on south Mid-Atlantic Ridge symbioses (SMAR) from K. van der Heijden, unpublished data. Species and sequences printed in bold are organisms used in this study.

Organisms & target gene	Sequence accession number
Mussel host, 18S rRNA gene	
Bathymodiolus azoricus	**AY649822**
Bathymodiolus brooksi	**AY649825**
Bathymodiolus puteoserpentis	**AF221640**
Bathymodiolus childressi	**AF221641**
Bathymodiolus brevior	AY649824
Bathymodiolus tangaroa	AY649820
Bathymodiolus mauritanicus	AY649828
Bathymodiolus marisindicus	AY649818
Bathymodiolus aff. thermophilus	AY649823
Bathymodiolus heckerae	AY649830
Bathymodiolus thermophilus	AY649829
Bathymodiolus sp.	AY649819
Thiotrophic symbiont, 16S rRNA gene	
***Bathymodiolus azoricus* / *B. puteoserpentis* symbiont**	**AM083974**
***Bathymodiolus brooksi* symbiont**	**AM236331**
Bathymodiolus mauritanicus	FM213421
Bathymodiolus heckerae	AM236328
Bathymodiolus thermophilus	M99445
Bathymodiolus sp. (SMAR 5°S / 9°S)	unpublished data
Methanotrophic symbiont, 16S rRNA gene	
***Bathymodiolus azoricus* / *B. puteoserpentis* symbiont**	**AM083950**
***Bathymodiolus brooksi* symbiont**	**AM236330**
***Bathymodiolus childressi* symbiont**	**AM236329**

7.3 Relative quantification of 16S rRNA and *recA* genes in *B. puteoserpentis* symbionts

The relative quantification of the 16S rRNA and *recA* genes in 4 samples of *B. puteoserpentis* showed a consistency of the relative abundance of the 16S rRNA to *recA* ratio (Figure 7.4). The thiotrophic symbionts (SOX) display a ratio of of 16S rRNA to *recA* of about 1 (R^2=0.9987, p<0.001), whereas the methanotrophic symbionts (MOX) show a ratio of about 8 (R^2=0.999, p<0.001). Furthermore, the relative abundances of both marker genes show consistently, that the abundance of sulfur and methane oxidizers is different from animal to animal (intra-species variability). The standard deviation of the qPCR abundances quantified in triplicates for the same sample was below 8% for all samples.

7.4 Method comparison - qPCR vs. 3D-FISH

In the hybridized gill cross sections of *B. puteoserpentis* mussels, signals of the specific FISH probes were restricted to the bacteriocytes. The probe specific for the methanotrophic symbiont hybridized with larger cells whereas the thiotrophic probe bound to smaller cells.

The FISH-signals showed a heterogeneous distribution of symbionts throughout the gill filaments of mussels from the active vent site (day 0, Figure 7.5B). Whereas the methanotrophic symbionts were distributed with no distinct pattern across the filament section, the sulfur-oxidizing bacteria were concentrated towards the ciliated edge. This distribution pattern of the sulfur-oxidizing symbiont was congruent with earlier observations on bathymodiolin endosymbiont distribution (Fiala-Medioni *et al.*, 1986). Gill filaments of starved mussels displayed a reduction in host tissue and symbiont abundance and no structured distribution of bacteria could be observed (day 10, Figure 7.5B). The mean biovolume of both endosymbionts in ciliated edge, middle part and abfrontal edge of the gill filament cross section showed huge variations between the individual mussels. This was reflected by large error bars and non-significant results (p>0.1) for all comparisons between day 0 and day 10 (Figure 7.5A). The most promi-

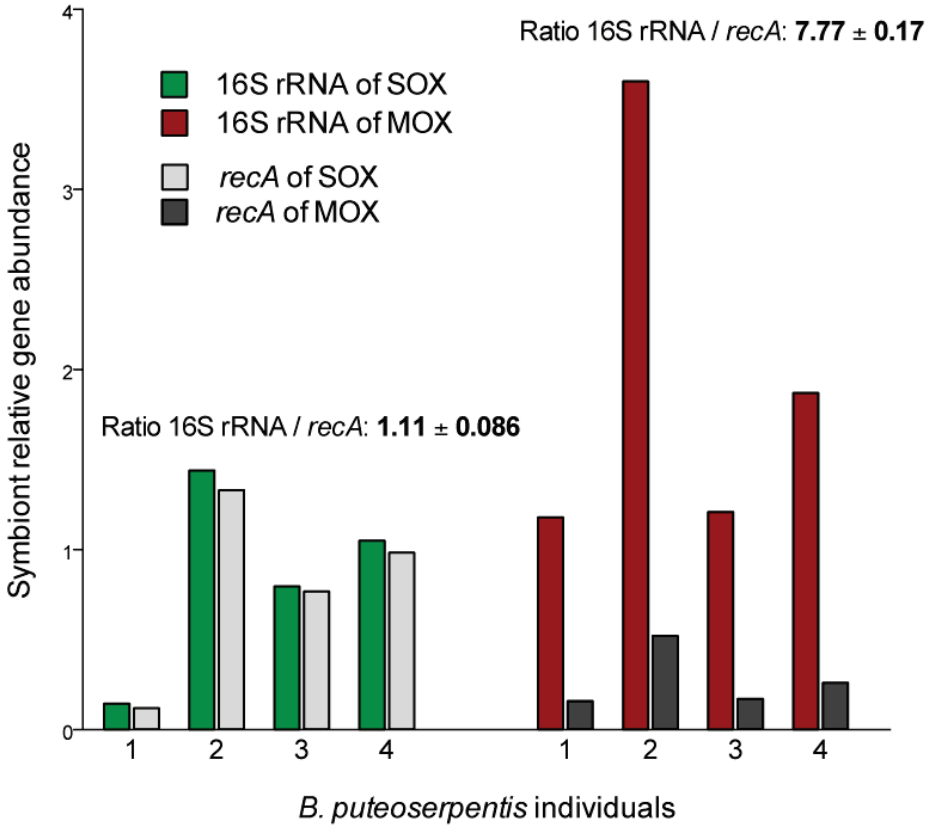

Figure 7.4: Relative abundances of the 16S rRNA and *recA* genes in *B. puteoserpentis* symbionts - Comparison of the relative abundance of two target genes (16S rRNA & *recA*) of endosymbionts in *B. puteoserpentis* individuals. Each pair of bars represents the abundance of the 16S rRNA and *recA* gene in one mussel gill sample, showing in total the gene abundances in 4 samples.

nent though insignificant differences were seen for the biovolumes of thiotrophic (p=0.1153) and methanotrophic (p=0.1310) symbionts at the ciliated edge region of the gill filament cross section.

In contrast to the 3D-FISH results, the relative 16S rRNA abundance of the endosymbionts (Figure 7.5C) showed a significant decrease from day 0 to day 10 of the starvation experiment ($p<0.05$ for both symbionts). Correlating the 3D-FISH and qPCR results, that means analyzing how the symbionts biovolume is coupled to the relative 16S rRNA gene abundance in the same samples, showed the following picture: the gene abundance and the biovolume for all stereological

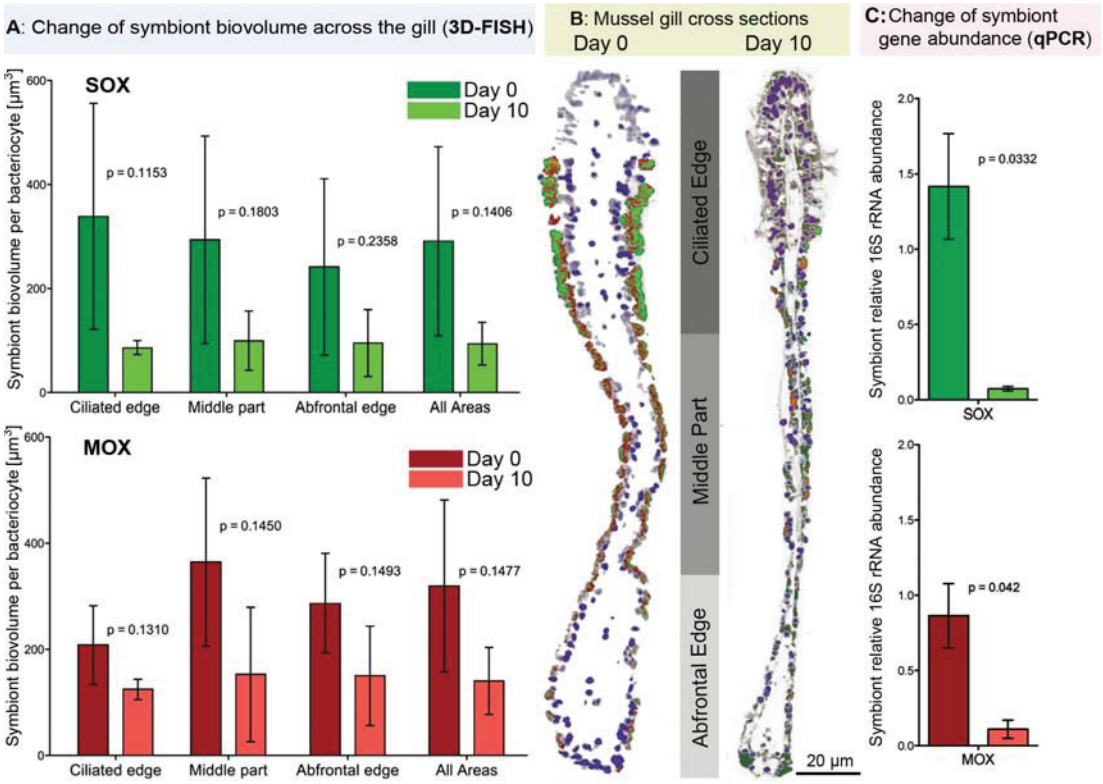

Figure 7.5: QPCR vs. 3D-FISH - Quantification of *B. puteoserpentis* methanotrophic (MOX) and thiotrophic (SOX) endosymbionts with 3D-FISH and qPCR (n=3, error bar = SD). A: Mean biovolume of SOX (green) and MOX (red) symbionts in bacteriocytes from different parts of mussel gills from Logatchev starvation samples. B: Mussel gill cross sections from day 0 and 10 of the starvation experiment, hybridized with specific FISH-probes for SOX (green) and MOX (red). Nuclei are stained with DAPI (blue).C: Relative quantification of the symbionts 16S rRNA gene with qPCR on the same samples as (A). The corresponding p-values for the t-tests done on the significant difference between day 0 and day 10 samples are plotted above the bars.

sampling areas displayed a positive correlation, pointing out that individuals with a higher number of 16S rRNA gene copies for a symbiont species also have a higher biovolume of this symbiont in their bacteriocytes. The most significant correlation (correlation coefficient and non-parametric Spearman's rho) was given when correlating the biovolume determined in the ciliated edge region of the mussel gill with the 16S rRNA gene abundance of the symbionts (Table 7.5).

Table 7.5: Correlation between qPCR and 3D-FISH results - Correlation between the abundance of the bacterial 16S rRNA gene and the biovolume of the thiotrophic (SOX) and methanotrophic (MOX) symbionts inside the bacteriocytes of certain areas of the gill filament has been statistically evaluated using the correlation coefficient (R^2) and the Spearman's correlation (ρ). The p-values relate to the Spearman's correlation.

Mussel gill area	Symbiont	Correlation coefficient (R^2)	Spearman's correlation (ρ)	p-value
Whole filament	SOX	0.8663	0.829	0.059
	MOX	0.5302	0.657	0.175
Ciliated edge	SOX	0.9359	0.943	0.002
	MOX	0.5682	0.714	0.136
Middle part	SOX	0.8644	0.829	0.005
	MOX	0.5180	0.657	0.175
Abfrontal edge	SOX	0.5244	0.771	0.103
	MOX	0.5343	0.657	0.175

7.5 The Logatchev starvation experiment

Having identified the ciliated edge as the most suitable area of the gill filament for the quantification of the endosymbionts via 3D-FISH, the effect of 10 day *in situ* starvation on these bacteria was investigated. Quantification of two symbiont marker genes (16S rRNA and *recA*) displayed significant loss of abundance over the course of the 10 days while the ratios between those genes within mussel individuals remained constant (Figure 7.6). The thiotrophic symbionts gene abundance decreased with 96,6% (16S rRNA) and 97,5% (*recA*) and the methanotrophic symbiont even showed a 99% decrease in relative gene abundance (16S rRNA and *recA*) after 10 days of starvation. In terms of biovolume, the thiotrophic symbionts lost 65,3% and the methanotrophic symbionts 74,2% of their biovolume inside the bacteriocytes after the starvation experiment. The overall distribution of the endosymbionts inside the host bacteriocytes was not influenced by the starvation, displaying the thiotrophic symbionts at the apical edge and the methanotrophic symbionts on the basal edge of the bacteriocytes (Figure 7.7).

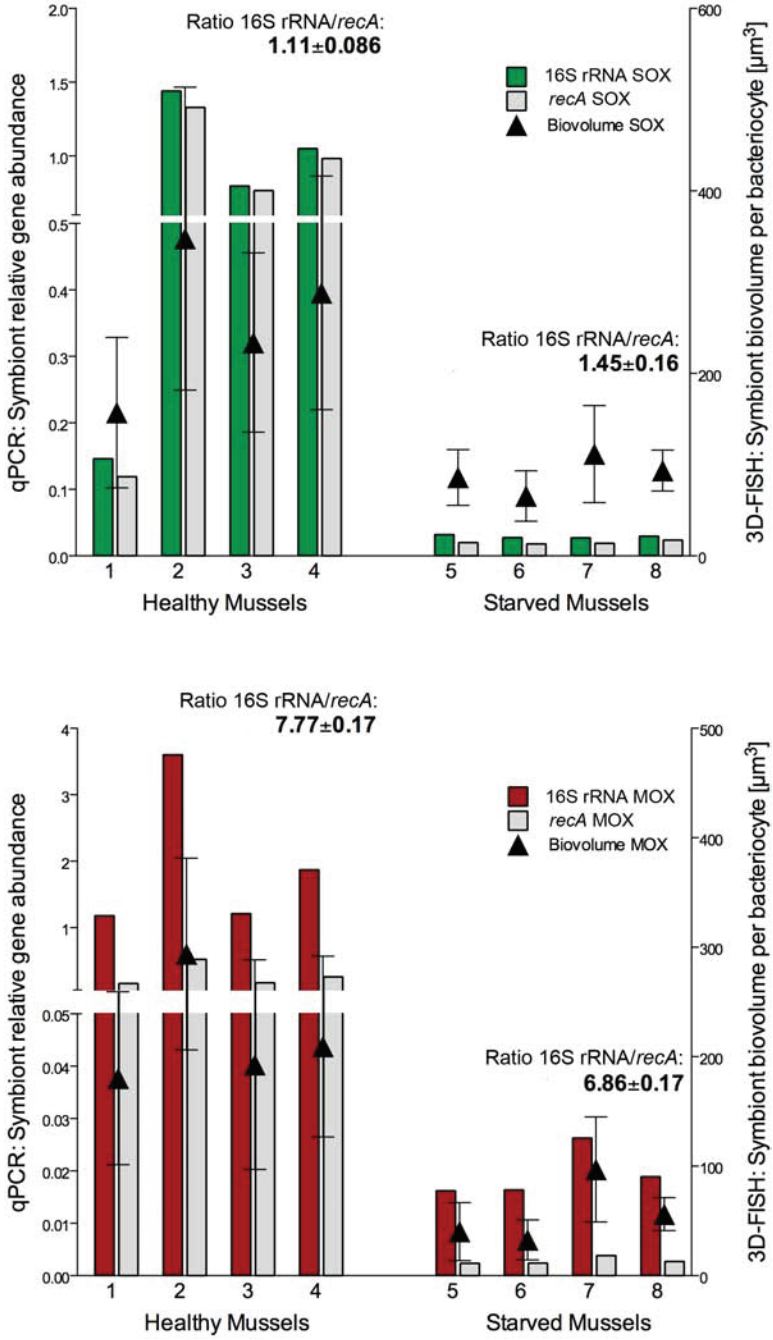

Figure 7.6: Logatchev starvation experiment - The effect of *in situ* starvation on endosymbionts of *B. puteoserpentis*. The relative gene abundance of the symbionts 16S rRNA and *recA* genes dropped down to 1% after 10 days of starvation (bars) whereas the biovolume (triangles) decreased only to 35% (sulfur oxidizers, SOX) and 26% (methane oxidizers, MOX).

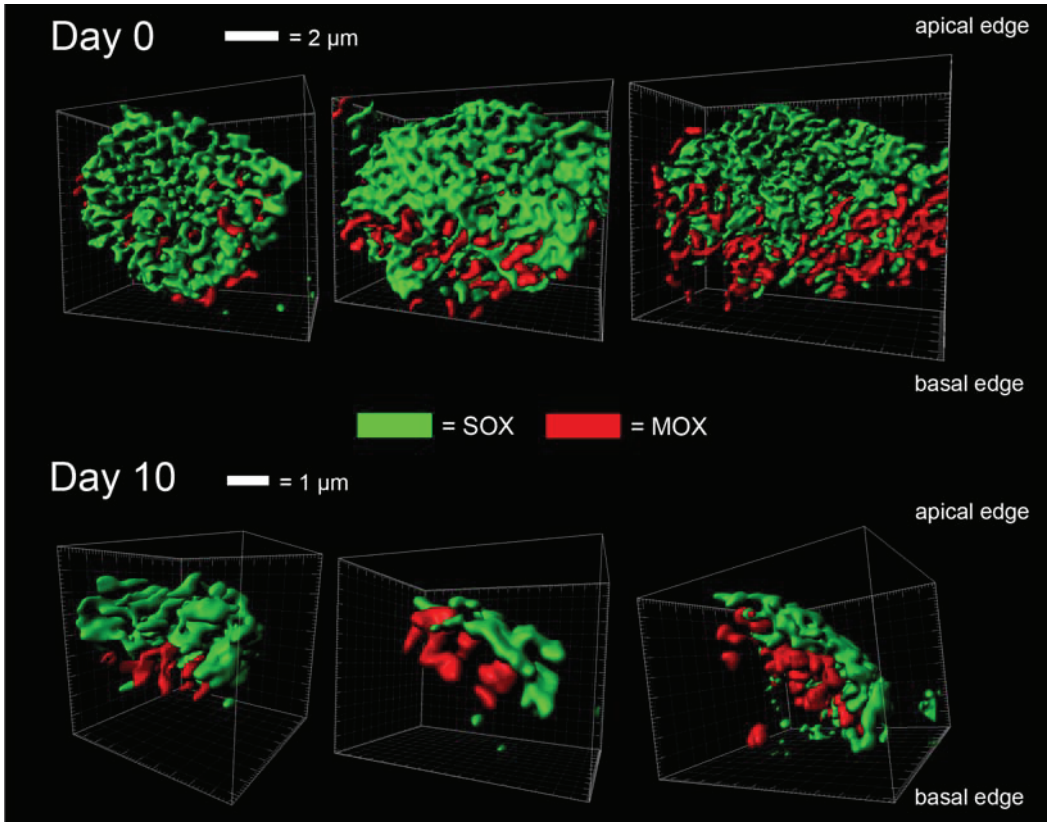

Figure 7.7: Symbiont distribution inside host bacteriocytes - Representative examples for bacteriocytes from *B. puteoserpentis* mussels of day 0 and day 10 of the Logatchev starvation experiment. The symbionts were stained with specific FISH probes for thiotrophic (green) and methanotrophic (red) bacteria. Image stacks of hybridized mussel gill sections were processed with Imaris software, covering positive FISH-signals in 3D with isosurfaces for better visualization.

8

Quantitative response of vent symbioses to *in situ* starvation and recovery

The transplantation experiment (see scheme in Figure 5.2) at the Menez Gwen hydrothermal vent field at the MAR yielded *B. azoricus* mussel samples from the active vent site (day 0), starved mussels (transplantation day 2-14) and retransplanted mussels (retransplantation day 1-7).

Displayed in Figure 8.1 are the results of this transplantation experiment in respect to the relative abundance of the symbiont 16S rRNA gene. Transplantation resulted in a significant drop of the thiotrophic (down to 26,6%) and methanotrophic (down to 35,0%) symbiont abundance after 14 days of *in situ* starvation ($p<0.05$ for methanotrophs and thiotrophs). The group of retransplanted mussels showed no significant trend of symbiont number recovery 7 days after retransplantation. In one case, mussels have been collected from the same cage, meaning that there is quantitative information on the symbiont abundance before and after retransplantation for these individuals (see Figure 8.1). Mussels sampled from this cage showed a significant increase of methanotrophic symbionts after 4 days of recovery ($p<0.05$), while the thiotrophic population did not recover.

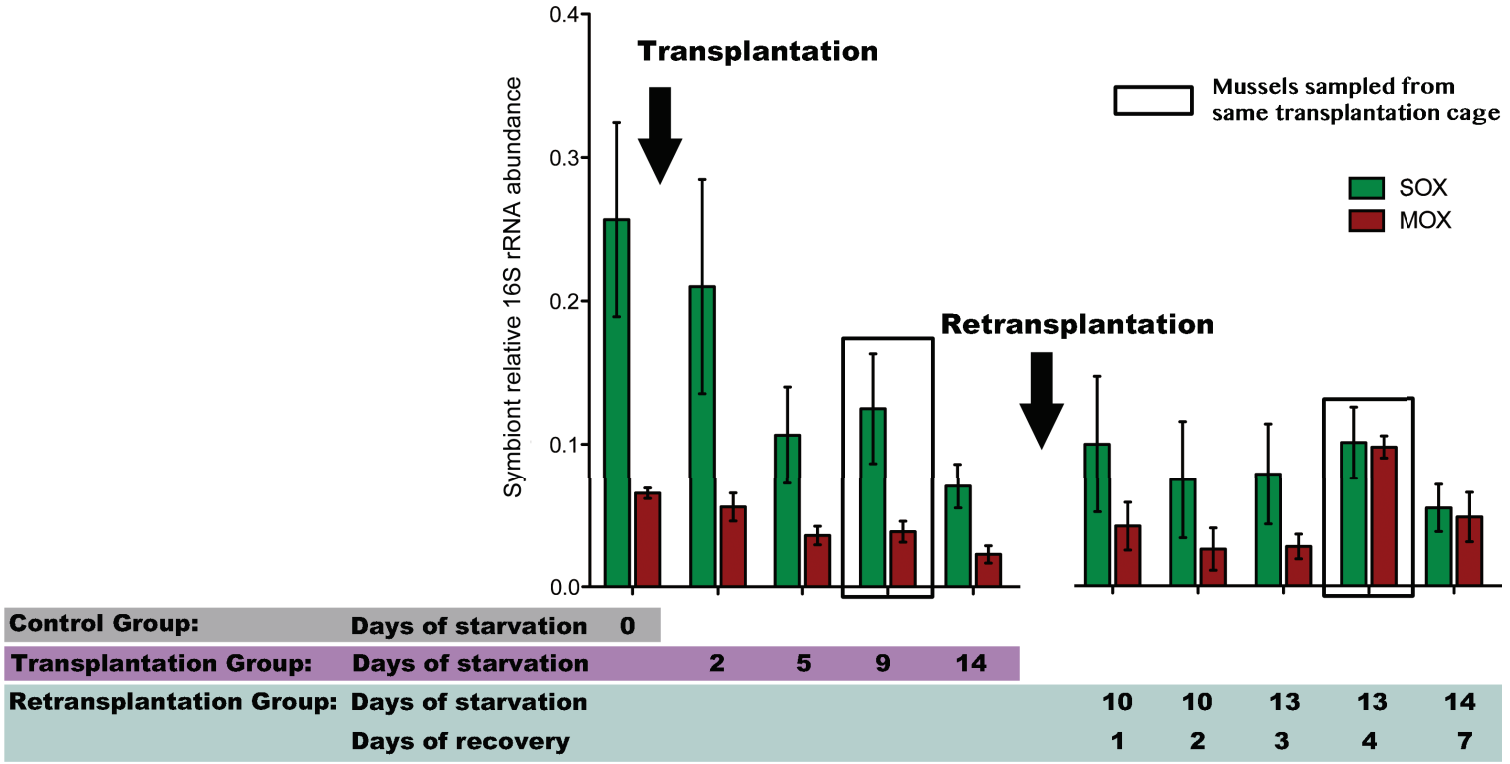

Figure 8.1: Transplantation experiment on Menez Gwen vent mussels - Relative 16S rRNA abundance of methanotrophic (MOX, red) and thiotrophic (SOX, green) symbionts in *B. azoricus* mussels from the Menez Gwen hydrothermal vent field. After sampling at the hydrothermal vent site (day 0) the mussels have been displaced to an area with no hydrothermal fluid access and sampled at various time points (see transplantation scheme in Figure 5.2). Latest 14 days after the transplantation, the mussels have been retransplanted to the active vent site and left to recover for up to 7 days. The mussels that have been retransplanted have been starved for different days (see different groups on the x-axis). Each bar represents the mean relative 16S rRNA abundance of 4 animals from the respective sampling point (error bars = SD).

9

Linking the abundance of cold seep endosymbionts to geofuel availability

Prior to the mussel sampling, sulfide and methane concentrations have been measured in two GoM habitats Green Canyon and Garden Banks with the *in situ* mass spectrometer (ISMS). The quantification of each diluted volatile was different in respect to the accuracy. The results for the methane measurements were rather precise, while the sulfide concentrations were given in a range (see Table 9.1). These measurements showed 10 times more sulfide in the sampled fluids from Green Canyon than at the Garden Banks seep. In contrast to that, the methane concentrations at the Garden Banks site were up to twice as high as at Green Canyon. Results of the relative quantification of methanotrophic symbionts in *B. childressi* and methano- and thiotrophic symbionts in *B. brooksi* gill tissue are displayed in Figure 9.1.

The relative abundance of the symbionts 16S rRNA gene displayed a high heterogeneity in all mussel samples, even from within the same sampling spot. Comparing the symbiont abundances of the two habitats, the *B. brooksi* mussels showed on average a higher abundance of sulfur oxidizers at the Green Canyon site (p=0.101) and a lower abundance of methane oxidizers (p=0.073). In *B. childressi* mussels, this difference in symbiont abundances between the two habitats

Table 9.1: Location of GoM samples and ISMS data[a] - Sampling locations at the GoM with information on how many mussels have been sampled and the amount of methane and sulfide measured with the ISMS at the sampling sites.

Location	Lat./Long.	Depth	*B. brooksi* samples	*B. childressi* samples	Methane conc.	Sulfide conc.
Green Canyon (GC852)	27°06'6 N/ 91°09'9 E	-1407 m	8	11	179±96 μM	~100 μM - 1.21 mM
Garden Banks (GB829)	27°11'1 N/ 92°07'5 E	-1258 m	6	6	341±107 μM	~10 μM - 150 μM

[a] ISMS data from Prof. Peter Girguis (Harvard University), mussel sample reference from Cordes *et al.* (2010).

was even more distinct, with a significant higher abundance of methanotrophic symbionts at the Garden Banks site (p=0.018). Despite the heterogeneity of symbiont abundances from the same sampling area, mussels from the Green Canyon site (high sulfide, low methane) contained on average more thiotrophic symbionts and less methanotrophic symbionts than mussels from the Garden Banks site (high methane, low sulfide). Comparing the methanotrophic abundances in *B. brooksi* and *B. childressi*, a higher abundance of methanotrophic symbionts in the *B. brooksi* mussels from Green Canyon (p=0.006) and Garden Banks (p=0.0519) was visible.

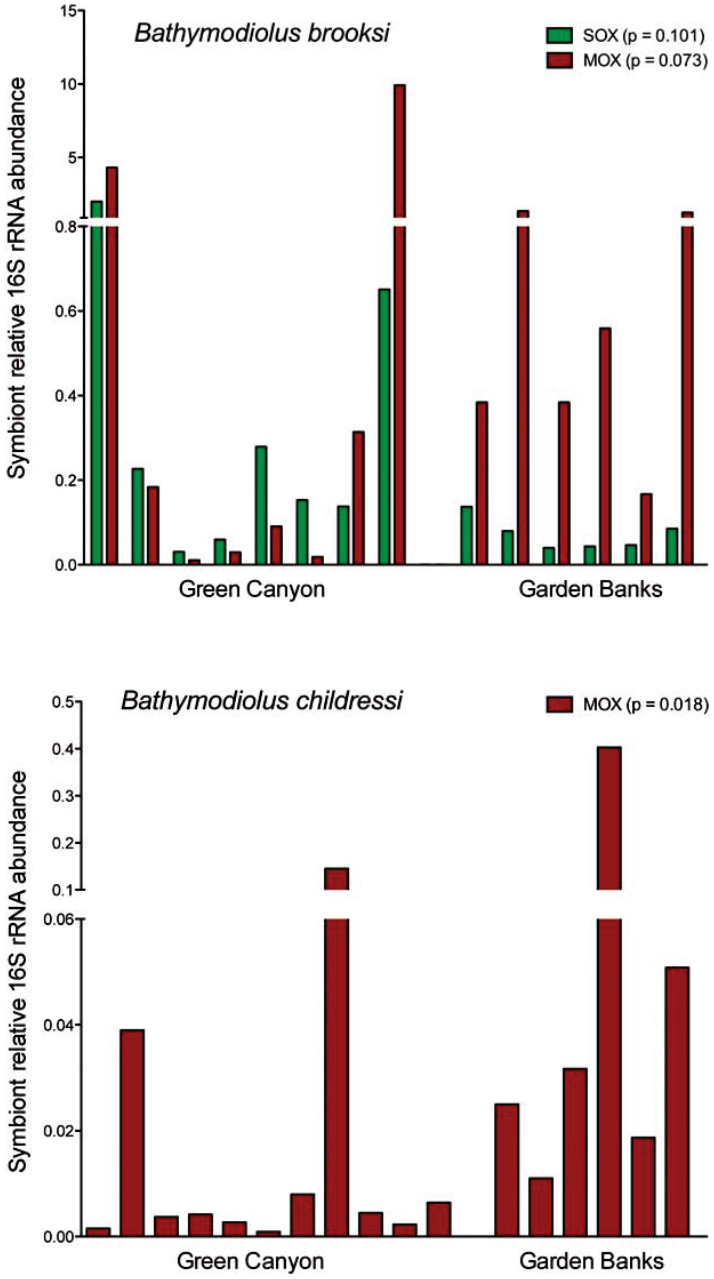

Figure 9.1: Mussel symbioses from the Gulf of Mexico - Relative symbiont abundance in two co-occurring seep mussel species from the Gulf of Mexico. The dual symbiosis of *B. brooksi* and the single symbiosis of *B. childressi* mussels were studied at the Green Canyon and Garden Banks cold seeps. Each bar (*B. childressi*) or each pair of bars (*B. brooksi*) represents the relative symbiont abundance from a single mussel individual.

Part IV

Discussion

9.1 Fast and reliable quantification of bathymodiolin symbionts

9.1.1 Distribution of symbionts in the *Bathymodiolus* gill

3D-FISH was successfully applied in this thesis work to study the distribution of the endosymbionts inside the *B. puteoserpentis* mussel. The thiotrophic symbionts were located in a gradient-like distribution across the gill cross section, showing the highest abundance at the ciliated edge and the lowest abundance at the abfrontal edge of the gill filament. The ciliated edges of the filaments represent the region of the mussel gills that is in closest contact to the water that is freshly inhaled by the mussel and that is rich in oxygen and energy-rich volatiles. In contrast to this, the methanotrophic symbionts were distributed more evenly across the gill section (Figure 7.5B). Additionally, the thiotrophic symbionts were located in the apical part of the bacteriocytes, which is in close contact to the respiratory surface of the gill filament. The methanotrophic symbionts on the other hand occured closer to the basal edge (Figure 7.7). Both observations indicate, that the thiotrophic symbionts prefer a short distance to the fresh water that contains oxygen and electron donors for chemosynthetic primary production. No studies exist so far on the accessibility of methane and sulfide for the symbionts at different locations in the gill and how efficiently the energy sources are taken up. Still, these results are congruent with studies on the bacterial activity inside the *B. puteoserpentis* gills that showed a higher symbiont activity at the ciliated edge of the gills (Wendeberg *et al.*, 2011). The location-dependent activity and abundance of symbionts across the mussel gill indicates differences in energy source availability for different areas of the gill. This could mean that symbionts, once they have been taken up by the mussel host and were incorporated in bacteriocytes, grow and multiply inside the bacteriocytes according to the spatial availability of energy sources inside the gill, leading to the heterogeneous distribution of symbionts across gill filament sections observed in this study.

9.1.2 Symbiont biovolume in *B. puteoserpentis*

Additionally to the study of bacterial distribution, 3D-FISH was used to quantify the biovolume of bathymodiolin symbionts inside the bacteriocytes of the mussel gill tissue. This was a relative quantification approach. First of all, there is no software yet that can reliably detect the shape of bacterial cells in 3D-space. As outlined in Section 6.1.1, even the segmentation of images and the differentiation of positive and negative signals in 3D is already a huge challenge. For the detection and absolute quantification of bacterial cells, the software would additionally need to recognize the shape of (potentially overlapping) targets in 3D. More over, the fluorescent signals of the bathymodiolin sulfur oxidizers were below the resolution limit of the LSM 510, leading to an inability to resolve single thiotrophic cell structures. Therefore, I used the 3D-FISH method for the relative quantification of bacterial cells (in terms of biovolume), rather than to attempt to detect absolute cell numbers. In theory, knowledge of the bacterial biovolume could be used to calculate cell numbers, but only if the bacterial cell size is homogeneous. This is not the case for the sulfur-oxidizing symbionts and the quantification based on biovolume is therefore a common approach, which was already applied in the quantification of thiotrophic symbionts in *B. azoricus* mussels (Halary *et al.*, 2008).

The advantage of having a local visualization of the bacterial targets is at the same time the biggest disadvantage of this method for bacterial quantification. Microscopic cross sections of individual gill filaments represent a spatially very limited fraction of the whole gill (Figure 6.2). Many gill filaments sampled in various regions of the gill and a large amount of image stacks would be necessary to ensure statistically significant sampling. In addition, the heterogeneous distribution of symbionts in bacteriocytes across the gill sections accounted for huge differences of quantified symbiont biovolume within the same sample, represented in the large error bars of the bacterial biovolume in Figure 7.5. The number of 72 image stacks per animal that was used in this study was not enough to resolve the large variance of biovolume in bacteriocytes and can hardly stand as representative quantification unit for the whole mussel. The application of the 3D-FISH quantification on starved mussels showed a decrease in biovolume of both

endosymbionts (Figure 7.6), but not with the same magnitude as the decrease of the marker genes (quantified with qPCR). Together with the comparatively large variance of the biovolume in the bacteriocytes, this demonstrated the lower level of resolution of the 3D-FISH data compared with the qPCR data and the drawback of this stereological method when applied on heterogeneous distributed targets inside a biological system.

9.1.3 Quantification of symbiont marker genes with qPCR

The qPCR-based protocol for the quantification of bathymodiolin symbionts was validated by the quantification of two symbiont marker genes (see Section 7.3). The consistent ratio between the relative abundances of the 16S rRNA and *recA* genes for each symbiont species in individual mussels demonstrated the reliability of this quantification protocol. After this validation, the quantification of bathymodiolin symbionts with only one marker gene was suffient. I've therefore chosen to use the 16S rRNA gene as target for further quantitative analyses with qPCR, because future primer design based on 16S rRNA is facilitated by quality checked and aligned rRNA databases with many sequences (compared to *recA* databases).

Independent of which target gene is used for qPCR-based quantification of the symbionts, this approach does not yield absolute cell numbers. The absolute quantification of target genes or the relation to cell numbers is hindered, if multiple operons are present. The ribosomal RNA database (rrnDB) contains estimates of the number of 16S rRNA operons for 785 bacterial and 69 archaeal strains and shows a range of 1 to 15 16S rRNA copies per genome (Lee *et al.*, 2009). From the bacteria present in the database, 40% have either one or two 16S rRNA operons and 11% percent have eight or more 16S rRNA operons. The latter group consists exclusively of bacteria in the phylum *Firmicutes* and the class γ-proteobacteria. The closest relative in this database to the bathymodiolin thiotrophic symbiont is *Thiomicrospira crunogena*, a sulfur-oxidizing γ-proteobacterium which has 3 operons of the 16S rRNA gene. This low amount of 16S rRNA operon numbers is comparable with the detection of only one copy of the 16S rRNA gene in the thiotrophic symbiont of *B. puteoserpentis*. The closest relative to the methane-oxidizing symbiont in the database is *Methylococ-*

cus capsulatus, a methylotrophic γ-proteobacterium that has 2 copies of the 16S rRNA gene. This seems to contradict my results of multiple 16S rRNA operons in the methanotrophic symbiont, but *Methylococcus capsulatus* is only very distantly related to the *B. puteoserpentis* methanotrophic symbiont (shown with phylogenetic analyses on bathymodiolin methanotrophic symbionts by Petersen & Dubilier (2009)). The database for 16S rRNA operon numbers in bacteria is still small and knowledge on the 16S rRNA gene copies in bathymodiolin symbionts may increase in the future with the availability of more genome sequences.

A second factor that can lead to an overestimation of cell numbers is the presence of multiple genomes. The occurrence of multiple operons has for example been shown in the human pathogen *Neisseria gonorrhoeae*, probably ensuring a rapid propagation of the pathogen in its host (Tobiason & Seifert, 2006). Furthermore, the sulfur-oxidizing symbiont of *Codakia orbicularis* (Bivalvia, Lucinidae) shows an increase in genome copies per cell as a reaction to changing concentrations in energy sources (Caro *et al.*, 2009). In case of the bathymodiolin symbionts, no knowledge exists on the number of genome copies per cell. This might change in the future, because there is ongoing effort in the Symbiosis Group at the MPI for the isolation and separation of the symbionts of *B. puteoserpentis*. Isolated symbiont fractions could be analyzed with flow cytometry (FCM) to determine their size and nucleic acid content and investigate the occurrence of multiple operons (Caro *et al.*, 2007).

The quantitative information on the 16S rRNA operon number of the endosymbionts can be used to draw conclusions about the life style of both species. In general, organisms with one or a few ribosomal operons tend to be slow-growing and organisms with more operons tend to be fast-growing species (Lee *et al.*, 2009). An example for the latter is the fast growing γ-proteobacterium *E.coli*, containing 7 operons for the 16S rRNA gene (source: rrnDB) and growing with a generation time of about 20 minutes (Cooper & Helmstetter, 1968). The methanotrophic symbiont of *Bathymodiolus* (possibly containing 8 copies of 16S rRNA operons) could therefore be able to grow faster than the thiotophic symbiont. Compared to the symbiont distribution in the mussel gills (Section 9.1.2), this could explain why the methanotrophic symbiont inhabits bacteriocytes of the whole gill cross section. After the process of symbiont incorporation into

the mussel gill cells, the methanotrophic symbionts could be able to propagate faster than the co-occuring thiotrophic symbiont (possibly containing only one 16S rRNA operon). Only at the ciliated edge of the gill filaments, where energy sources are most optimal for the thiotrophic symbiont, this species might be able to dominate the space given by the bacteriocytes.

9.1.4 Method of choice: QPCR

The qPCR assay developed in the scope of this study has proven to be a suitable tool for the relative quantification of bacteria inside animal tissue. After the necessary evaluation of this method for the *B. puteoserpentis* symbiosis, the symbiont and host marker genes could be quantified and used as a proxy for the bacterial abundance. The reliability of this method was proven by three factors: high analytical specificity, good accuracy in determining gene abundances and repeatability of the results. The analytical specificity characterizes the specific detection of target sequences versus non-target sequences, which was checked by sequencing the end products of the qPCR amplifications. By strictly following the guidelines for qPCR by Bustin *et al.* (2009), I was regularly able to directly sequence the desired genes successfully after qPCR runs, indicating that the analytical specificity of my symbiont quantifications was always high. By employing two different target genes, I validated the abundance results of the qPCR experiments. The stability of the ratio between 16S rRNA and *recA* gene abundances in all samples demonstrated very high accuracy of the method. Finally, repeatability of the measurements was evidenced by only 8% variation of the abundance results between triplicate measurements, which is lower than other published qPCR variations (e.g. 20% variation in triplicate qPCR measurements by Zemanick *et al.* (2010)).

Compared to the 3D-FISH quantification protocol, the tissue sampling for qPCR is much easier and faster, because the morphology of the tissue does not have to be preserved. And due to the high-throughput nature of the qPCR assay, multiple sampling points across the animal tissue (Figure 6.2) result in a good representation of the subsamples in respect to the whole animal. With qPCR, the total nucleic acid content of many different areas across the animal

tissue is sampled, whereas with 3D-FISH, comparatively very small and only a few sampling points have to represent the quantitative symbiont information for the whole animal. Concluding from these results, qPCR is recommended as the method of choice for the relative quantification of bathymodiolin symbionts.

9.2 Quantitative response of vent symbioses to *in situ* starvation and recovery

A key question of this study was, how the bathymodiolin symbiont abundance is related to the availability of energy sources in the environment. We first addressed this question by investigating, if we could trigger any response at all of the bathymodiolin symbionts by starvation, how fast this response takes and how severe it is. For this purpose, we simulated a radical temporal decrease in the energy sources of the symbionts by removing mussels from sites of active venting to sites without hydrothermal influence. These experiments were conducted on *B. puteoserpentis* mussels from Logatchev and *B. azoricus* mussels from Menez Gwen, both hydrothermal vent fields at the MAR.

The results showed for both mussels species a significant drop in symbiont abundance upon starvation. The thiotrophic and methanotrophic populations of *B. puteoserpentis* were almost completely depleted after 10 days of starvation (Section 7.5) and both symbiont species in the *B. azoricus* mussels were left with about 30% of their original abundance after 14 days of starvation (Section 8). The depletion of the endosymbionts can happen 1) via cell lysis in the bacteriocytes of the gill or 2) the release of whole bacteriocytes from the gill tissue. The lysosomal digestion of symbionts in host bacteriocytes has been observed in different studies with cytochemistry and electron microscopy, e.g. on *B. thermophilus* mussels (Boetius & Felbeck, 1995) and on *B. azoricus* mussels (Kádár *et al.*, 2007). In contrast to that, there is no scientific data on the release of host bacteriocytes. In some cases, detached bacteriocytes have been observed in FISH-fixed samples of bathymodiolin gill tissue (unpublished data by A. Wendeberg) but this could also be an artifact of the fixation process itself. In this study, the distribution pattern of the two bacterial species inside *B. puteoserpentis* bacteriocytes

remained the same after starvation (Figure 7.7) and there were no indications on whole bacteriocyte loss across the gill sections. Therefore, I conclude that the mussel host digested the endosymbionts with high efficiency, in case of the methanotroph symbiont with 99% on the total abundance after 10 days of starvation. This is congruent with findings of earlier starvation experiments done in aquaria on methanotrophic symbionts of *B. childressi* (Kochevar *et al.*, 1992) and on thiotrophic symbionts of *B. thermophilus* (Raulfs *et al.*, 2004). These were *ex situ* experiments and the abundance of the symbionts has only been assessed by microscopy. For *B. childressi*, a decrease of the methanotroph population of 90% after 24h in methane-free water was shown.

The *B. puteoserpentis* samples from this study showed, that a starvation for 10 days is enough to almost completely deplete the bacterial community in this mussel species. In case of the *B. azoricus* symbiosis, the mussels retained a higher population of symbionts after a longer period of starvation. This difference in symbiont abundance after starvation between mussel species could be due to (1) different symbiont digestion efficiencies of the hosts or (2) different tolerances of the endosymbionts towards sudden changes in energy source availability. Although it has been shown in different studies, that there is a high variation of symbiont abundances within a host species and between sites (Trask & Van Dover, 1999, Fiala-Medioni *et al.*, 2002, Duperron *et al.*, 2007b), the digestion efficiencies between hosts and energy source-related tolerances between symbiont species have not been addressed yet. Both mussel hosts investigated for this study apparently do not distinguish between different symbiont species during digestion/release, because thiotrophic and methanotrophic populations decrease to the same extend in the respective host. This indicates, that if the host digests symbionts via lysosomes in the bacteriocytes, that this process is not specific for the symbiont species.

Additionally to the starvation, the *B. azoricus* mussels have been retransplanted to the active vent site and recovered for up to 7 days before collection. Conclusions on the recovery potential of the symbionts are difficult, because the symbiont abundance of the retranplanted mussel group prior to the retransplantation was not known. Due to steep gradients, the availability of energy-rich volatiles can change within centimeters in hydrothermal habitats. This has for

example been shown for temperature gradients in mussel beds at the Rose Garden hydrothermal vent field (Johnson *et al.*, 1988) and temperature, sulfide and oxygen gradients in diffuse fluids from the Logatchev hydrothermal vent field (Zielinski *et al.*, 2011). Because mussel samples of the transplantation experiment might originate from different mussel patches (Section 5.2), symbiont abundances might be different in animals that co-occur at the same habitat. Nevertheless, there was one case where mussels have been sampled from the same transplantation cage before and after retransplantation. This means, that these individuals experienced similar conditions in respect to changes in energy source concentrations. The methanotrophic symbiont numbers in this case have more than doubled after 4 days of recovery compared to the abundance at 9 days of starvation, whereas the thiotrophic populations did not recover (Figure 8.1). Differences in the recovery potential of *B. azoricus* symbionts have already been observed in *ex situ* experiments in aquaria by Riou *et al.* (2008). Their results refer to observations in one individual and they detected a stronger increase in methanotrophic symbionts upon a methane pulse (35% increase of abundance) compared to the increase of thiotrophic bacteria upon a sulfide pulse (25% increase of abundance). The differences of symbiont recovery in the host gills can either be due to (1) differences in the symbiont-specific uptake by the host or (2) a different propagation speed of residual symbiont species inside the host bacteriocytes. The *ex situ* studies with *B. azoricus* mussels have not addressed this question yet and therefore, no knowledge is available on the recovery mode of methanotrophic and thiotrophic symbionts yet. The results from my study in respect to 16S rRNA operon number indicate, that the methanotrophs might have the ability to multiply faster than the thiotrophs (see Section 9.2), which might explain the faster recovery of methanotrophic symbionts during *in situ* and *ex situ* experiments.

9.3 Abundance of cold seep endosymbionts and geofuel availability

The transplantation experiments on *Bathymodiolus* mussels at vents have shown, that temporal changes in energy source concentrations influence the abundance

of the mussel symbionts. The next step was, to investigate this link in more detail with the aim to study symbioses in their natural environment. The question was, if the same mussel species would harbor different amounts of symbiont species according to the concentration of energy sources in the environment. Therefore, symbionts of the two mussel species *B. brooksi* and *B. childressi* (co-ocurring at habitats at the GoM) were quantified. The collection of the mussels was coupled to the fine-scale measurement of methane and sulfide at the same sampling spot. Independent of the investigated mussel species, the results prove a positive correlation of the symbionts energy source concentration in the environment and symbiont abundance. In case of the methanotrophic symbionts, less than 100 micromolar difference in methane concentration at the two cold seep habitats (see Table 9.1) already lead to different amounts of symbionts in the mussels, demonstrating how tight this symbiosis is coupled to environmental conditions. Comparable to the observations of the transplantation experiments, it is not clear if the favorable conditions in energy source concentrations lead to a higher uptake of symbionts from the environment, or if the symbiont populations inside the host propagate faster. More over, if the environmental conditions already favor one population of free-living bacteria (e.g. by elevated concentrations of methane compared to sulfide), this could already lead to a dominance of one free-living species compared to the other. The occurrence of free-living populations of symbiont phylotypes in the vicinity of their hosts has already been shown, e.g. for symbionts of vestimentiferan tube worms (Harmer *et al.*, 2008) and the lucinid *Codakia orbicularis* (Gros *et al.*, 2003). We also found free-living relatives of the thiotrophic and methanotrophic symbiont populations in the ambient seawater of the vent fields visited for this study (unpublished data), supporting the hypothesis of an environmental uptake of symbionts from the environment of the bathymodiolin host. Because there is no quantitative data available of the abundances of free-living symbiont species, it is not known if the dominance of an endosymbiont species is related to the dominance of the respective free-living relative.

The significantly higher number of methanotrophic symbionts in *B. brooksi* than in *B. childressi* is unexpected, because *B. childressi* lives in a single symbiosis and relies entirely on its methanotrophic symbionts. *B. brooksi* on the

other hand benefits from a dual symbiosis with methanotrophic and thiotrophic symbionts. It is known, that the majority of *B. brooksi* symbionts are methanotrophic (85% of total bacterial rRNA according to Duperron *et al.*) but no knowledge on the absolute numbers of methanotrophs in *B. brooksi* is available. Genetic analyses of the *B. childressi* methanotrophs revealed six different *pmoA* sequences (Duperron *et al.*, 2007b), which might be due to the presence of several methanotrophic strains (which are not resolved on the 16S rRNA level). If several phylotypes exist, the specific 16S rRNA qPCR primers used in this study may not match with all strains. In that case, the abundance of methanotrophs in *B. childressi* could be underestimated. This comparison of the abundance of *B. childressi* and *B. brooksi* methanotrophs is based on the assumption, that both symbiont species have the same number of 16S rRNA operons. Even if this remains to be proven, the ribosomal operons are very conserved among related species and mutations are not very likely (Itoh *et al.*, 1999). These first hints of an inter-species difference in the amount of bathymodiolin symbionts have to be confirmed by further investigations, for example the visual confirmation of symbiont abundance with FISH and genetic comparisons of different symbiont phylotypes in respect to 16S rRNA operon number and strain variability.

Part V

Conclusions

10

Summary and conclusions

Before this thesis work, analyses of symbiont abundance in bathymodiolin research have either been conducted by (quantitative) FISH (Halary *et al.*, 2008) or rRNA slot blot hybridization (Duperron *et al.*, 2007b). These techniques have the disadvantages of time-consuming protocols and are not suitable for processing many different samples at once. The desire to improve the current status of symbiont quantification by using state of the art stereological and molecular methods went along with the demand of a high-throughput method, which is fast and reliable.

3D-FISH and qPCR methods have been chosen for the establishment of a suitable quantification protocol for bathymodiolin symbionts and most of the work for this thesis went into the method optimization and validation of these two methods. Having validated the qPCR-based protocol by the quantification of different symbiont target genes (16S rRNA & *recA*), the comparison of qPCR with 3D-FISH results demonstrated the advantage of qPCR over 3D-FISH in the quantification of bathymodiolin symbionts. Due to the heterogeneity of symbiont distribution inside the hosts, the 3D-FISH protocol was less accurate in the symbiont quantification than qPCR. Furthermore, the qPCR protocol was faster, had a higher sample throughput and was more reliable than 3D-FISH in the quantification of the endosymbionts. Therefore, further quantitative studies on the symbionts have been carried out by qPCR (see Section 9.1.4).

This optimized symbiont quantification protocol based on qPCR has been applied on bathymodiolin symbioses of hydrothermal vents and cold seeps. First

of all, the quantitative reaction of symbionts to starvation was assessed, investigating the *in situ* response of vent symbioses to a sudden and extreme temporal change of energy source concentrations. Results of this starvation experiment on *B. puteoserpentis* and *B. azoricus* mussels showed, that the vent symbiosis reacts within days to the cessation of energy sources by a decrease of both symbiotic partners. For the *B. puteoserpentis* mussels at the Logatchev vent field, 10 days of *in situ* starvation were enough to almost completely deplete the symbiotic populations in the hosts. The *B. azoricus* mussels at the Menez Gwen vent field still retained 26 - 35% of their symbionts even after 14 days of starvation. An *in situ* recovery time for *B. azoricus* mussels of 7 days lead to a significant increase of the methanotrophic symbiont population in some mussels. This reaction of the symbiosis to decreasing an increasing concentrations of energy sources proved, that a link exists between symbiont abundance and energy source availability and that drastic temporal changes in energy source concentrations trigger a quantifiable reaction of the symbiont populations within days.

Quantification of the GoM mussel symbionts (Section 9) showed that environmental concentrations in sulfide and methane were reflected in the symbiont abundance. More methanotrophic symbionts have been detected in *B. brooksi* and *B. childressi* host species at the seep habitat with high methane concentrations (Garden Banks) compared to the habitat with lower methane concentrations (Green Canyon). In contrast, *B. brooksi* mussels at Green Canyon (high sulfide) contained more sulfur oxidizers than individuals at Garden Banks (low sulfide).

Concluding from these results, qPCR has been established successfully in the reliable quantification of at least 4 different bathymodiolin symbioses at vents and seeps. The results characterize the link between symbiont abundance and availability of energy sources and in this respect extend the knowledge that has only been gained by *ex situ* experiments so far. The dominance of bathymodiolin symbioses at vents and seeps is emphasized by showing the flexibility of maintaining different symbiont species at different environmental conditions.

11

Outlook

The quantification of different bathymodiolin symbioses gave the most reliable quantitative data so far in this field. But this data can not stand alone when investigating the complex dynamics of deep-sea symbioses and needs to be combined with additional methods to clarify why some bathymodiolin species harbor more symbionts than others and why certain symbiont species react differently to changes in energy source concentrations than others. This includes the investigation of possible filter-feeding in those bivalves as well quantitative stereological and molecular analyses.

The ongoing advance in laser-scanning microscopy will enable the operator to scan more samples in a faster way, giving symbiosis researchers the possibility to automate the acquisition of many images on whole tissues. In contrast to the current status, this could yield representative samples for whole animals and elucidate the distribution of symbionts even better. Together with qPCR, this would yield a more complete picture not only of the abundance, but also the distribution of all symbionts in animal tissue. The current efforts in sequencing the symbiont genomes will give additional insight into the genetic potential of the symbionts, including the operon numbers of their 16S rRNA genes. With this knowledge, the primers designed for this study can be used to assess the abundance of many bathymodiolin symbionts (see primer targets in Table 7.4) with even higher accuracy and enables the comparison of the abundance of different symbiont species (e.g. thiotrophic vs. methanotrophic symbionts).

The fast protocol for the qPCR and the minor demands on technical equip-

ment give the researchers the opportunity to conduct these analyses during their sampling in the field. This way, the samples would not have to be frozen and carried to the home-lab (preventing any DNA damage) and deliver reliable quantitative data on fresh samples. The qPCR protocol can be handled in the field on a research vessel. The data would be at hand in a few hours, giving the researchers quantitative feedback on field status, e.g. conditions of symbioses at different habitats. This would enable the researchers to adjust their field work program according to their own results.

This study laid the groundwork for future quantitative analyses of deep-sea symbionts and adds to the current molecular methods which help to unravel the dynamics of symbiont life style and their linkage to environmental conditions.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410. 57

Amann, R., Ludwig, W. & Schleifer, K. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.*, **59**, 143–169. 41

Amann, R.I., Binder, B.J., Olson, R.J., Chisholm, S.W., Devereux, R. & Stahl, D.A. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.*, **56**, 1919–1925. 52

Amend, J.P. & Shock, E.L. (2001). Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. *FEMS Microbiology Reviews*, **25**, 175–243. 10

Baker, E.T. (1996). Geological indexes of hydrothermal venting. *Journal of Geophysical Research*, **101**, PP. 13,741–13,753. 18

Barry, J.P., Buck, K.R., Kochevar, R.K., Nelson, D.C., Fujiwara, Y., Goffredi, S.K. & Hashimoto, J. (2002). Methanebased symbiosis in a mussel, Bathymodiolus platifrons, from cold seeps in Sagami Bay, Japan. *Invertebrate Biology*, **121**, 47–54. 29

Bates, A.E. (2007). Persistence, morphology, and nutritional state of a gastropod hosted bacterial symbiosis in different levels of hydrothermal vent flux. *Marine Biology*, **152**, 557–568. 9

Batuev, B.N., Krotov, A.G., Markov, V.F., Krasnov, S.G., Lisitsyn, E.D. & Cherkashev, G.A. (1995). A New Hydrothermal Field at 14-Degrees-45, Mid-Atlantic Ridge. *Doklady Akademii Nauk*, **343**, 75–79. 35

Bayer, C., Heindl, N.R., Rinke, C., Lücker, S., Ott, J.A. & Bulgheresi, S. (2009). Molecular characterization of the symbionts associated with marine nematodes of the genus Robbea. **1**, 136–144. 9

Berg, I.A. (2011). Ecological Aspects of the Distribution of Different Autotrophic CO2 Fixation Pathways. *Appl. Environ. Microbiol.*, **77**, 1925–1936. 11

Birgel, D., Thiel, V., Hinrichs, K.U., Elvert, M., Campbell, K.A., Reitner, J., Farmer, J.D. & Peckmann, J. (2006). Lipid biomarker patterns of methane-seep microbialites from the Mesozoic convergent margin of California. *Organic Geochemistry*, **37**, 1289–1302. 23

Boetius, A. & Felbeck, H. (1995). Digestive enzymes in marine invertebrates from hydrothermal vents and other reducing environments. *Marine Biology*, **122**, 105–113. 29, 30, 92

Boetius, A. & Suess, E. (2004). Hydrate Ridge: a natural laboratory for the study of microbial life fueled by methane from near-surface gas hydrates. *Chemical Geology*, **205**, 291–310. 22

Bowman, J. (2006). The Methanotrophs The Families Methylococcaceae and Methylocystaceae. 266–289, Springer New York. 14

Bustin, S.a. (2010). Why the need for qPCR publication guidelines?–The case for MIQE. *Methods (San Diego, Calif.)*, **50**, 217–26. 48

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J. & Wittwer, C.T. (2009). The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clin Chem*, **55**, 611–622. 48, 91

Cannuel, R., Beninger, P.G., McCombie, H. & Boudry, P. (2009). Gill Development and its functional and evolutionary implications in the blue mussel Mytilus edulis (Bivalvia: Mytilidae). *The Biological Bulletin*, **217**, 173–188. 27

Caro, A., Gros, O., Got, P., De Wit, R. & Troussellier, M. (2007). Characterization of the Population of the Sulfur-Oxidizing Symbiont of Codakia orbicularis (Bivalvia, Lucinidae) by Single-Cell Analyses. *Appl. Environ. Microbiol.*, **73**, 2101–2109. 8, 30, 90

Caro, A., Got, P., Bouvy, M., Troussellier, M. & Gros, O. (2009). Effects of Long-Term Starvation on a Host Bivalve (Codakia orbicularis, Lucinidae) and Its Symbiont Population. *Applied and Environmental Microbiology*, **75**, 3304–3313. 90

Cary, S.C., Cottrell, M.T., Stein, J.L., Camacho, F. & Desbruyeres, D. (1997). Molecular identification and localization of filamentous symbiotic bacteria associated with the hydrothermal vent annelid Alvinella pompejana. *Applied and Environmental Microbiology*, **63**, 1124–1130. 8

Cavanaugh, C.M., Gardiner, S.L., Jones, M.L., Jannasch, H.W. & Waterbury, J.B. (1981). Prokaryotic cells in the hydrothermal vent tube worm Riftia pachyptila Jones: possible chemoautotrophic symbionts. *Science*, **213**. 6

Charlou, J.L., Donval, J.P., Douville, E., Jean-Baptiste, P., Radford-Knoery, J., Fouquet, Y., Dapoigny, A. & Stievenard, M. (2000). Compared geochemical signatures and the evolution of Menez Gwen and Lucky Strike hydrothermal fluids, south of the Azores Triple Junction on the Mid-Atlantic Ridge. *Chemical Geology*, **171**, 49–75. 21

REFERENCES

Childress, J.J. & Fisher, C.R. (1992). The Biology of Hydrothermal Vent Animals - Physiology, Biochemistry, and Autotrophic Symbioses. *Oceanography and Marine Biology*, **30**, 337–441. 30

Childress, J.J., Fisher, C.R., Brooks, J.M., Kennicutt II, M.C., Bidigare, R. & Anderson, A.E. (1986). A methanotrophic marine molluscan (Bivalvia, Mytilidae) symbiosis: mussels fueled by gas. *Science*, **233**, 1306–1308. 28

Colaço, A., Bettencourt, R., Costa, V., Lino, S., Lopes, H., Martins, I., Pires, L., Prieto, C. & Serrão Santos, R. (2011). LabHorta: a controlled aquarium system for monitoring physiological characteristics of the hydrothermal vent mussel Bathymodiolus azoricus. *ICES Journal of Marine Science: Journal du Conseil*, **68**, 349 –356. 30

Collins, T.J. (2007). ImageJ for microscopy. *Biotechniques*, **43**, 25. 44

Cooper, S. & Helmstetter, C.E. (1968). Chromosome replication and the division cycle of Escherichia coli Br. *Journal of Molecular Biology*, **31**, 519–540. 90

Cordes, E.E., Becker, E.L., Hourdez, S. & 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 Research Part II: Topical Studies in Oceanography*, **57**, 1870–1881. 38, 82

Corliss, J.B. & Ballard, R.D. (1977). Oases of Life in the cold Abyss. *National Geographic*, **152**, 441–453. 5

Daims, H., Brühl, A., Amann, R., Schleifer, K.H. & Wagner, M. (1999). The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Systematic and Applied Microbiology*, **22**, 434–444. 52

Daims, H., Ramsing, N.B., Schleifer, K.H. & Wagner, M. (2001). Cultivation-Independent, Semiautomatic Determination of Absolute Bacterial Cell Numbers in Environmental Samples by Fluorescence In Situ Hybridization. *Appl. Environ. Microbiol.*, **67**, 5810–5818. 43

Daims, H., Lucker, S. & Wagner, M. (2006). daime, a novel image analysis program for microbial ecology and biofilm research. *Environmental Microbiology*, **8**, 200–213. 44, 64

Dale, C., Wang, B., Moran, N. & Ochman, H. (2003). Loss of DNA Recombinational Repair Enzymes in the Initial Stages of Genome Degeneration. *Mol Biol Evol*, **20**, 1188–1194. 57

Dalton, H. (2005). The Leeuwenhoek Lecture 2000 The natural and unnatural history of methane-oxidizing bacteria. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1207 –1222. 13

DeChaine, E.G. & Cavanaugh, C.M. (2006). Symbioses of methanotrophs and deep-sea mussels (Mytilidae: Bathymodiolinae). *Progress in Molecular and Subcellular Biology*, **41**, 227–249. 12

Desbruyères, D., Almeida, A., Biscoito, M., Comtet, T., Khripounoff, A., Le Bris, N., Sarradin, P.M. & Segonzac, M. (2000). A review of the distribution of hydrothermal vent communities along the northern Mid-Atlantic Ridge: dispersal vs. environmental controls. *Hydrobiologia*, **440**, 201–216. 20

Dharmaraj, S. (2006). RT-PCR: The basics. 46

Distel, D.L. & Cavanaugh, C.M. (1994). Independent Phylogenetic Origins of Methanotrophic and Chemoautotrophic Bacterial Endosymbioses in Marine Bivalves. *Journal of Bacteriology*, **176**, 1932–1938. 26

Domack, E., Ishman, S., Leventer, A., Sylva, S., Willmott, V. & Huber, B. (2005). A Chemotrophic Ecosystem Found Beneath Antarctic Ice Shelf. *Eos*, **86**, P. 269. 22

Douville, E., Charlou, J.L., Oelkers, E.H., Bienvenu, P., Jove Colon, C.F., Donval, J.P., Fouquet, Y., Prieur, D. & Appriou, P. (2002). The rainbow vent fluids (3614'N, MAR): the influence of ultramafic rocks and phase separation on trace metal content in Mid-Atlantic Ridge hydrothermal fluids. *Chemical Geology*, **184**, 37–48. 20

Dover, C.L.V. & Trask, J.L. (2000). Diversity at deep-sea hydrothermal vent and intertidal mussel beds. *Marine Ecology Progress Series*, **195**, 169–178. 25

Dubilier, N. (1986). Association of filamentous epibacteria with Tubificoides benedii (Oligochaeta: Annelida). *Marine Biology*, **92**, 285–288. 8

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

Dubilier, N., Bergin, C. & Lott, C. (2008b). Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nature reviews. Microbiology*, **6**, 725–40. 27, 28

Duperron, S., Nadalig, T., Caprais, J.C., Sibuet, M., Fiala-Médioni, A., Amann, R. & Dubilier, N. (2005). Dual symbiosis in a Bathymodiolus sp. mussel from a methane seep on the Gabon continental margin (Southeast Atlantic): 16S rRNA phylogeny and distribution of the symbionts in gills. *Applied and Environmental Microbiology*, **71**, 1694–1700. 26

Duperron, S., Bergin, C., Zielinski, F., Blazejak, A., Pernthaler, A., McKiness, Z.P., DeChaine, E., Cavanaugh, C.M. & Dubilier, N. (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. *Environmental Microbiology*, **8**, 1441–1447. 12, 26, 28, 52

Duperron, S., Fiala-Medioni, A., Caprais, J.C., Olu, K. & Sibuet, M. (2007a). Evidence for chemoautotrophic symbiosis in a Mediterranean cold seep clam (Bivalvia: giribetLucinidae) comparative sequence analysis of bacterial 16S rRNA, APS reductase and RubisCO genes. *FEMS Microbiology Ecology*, **59**, 64–70. 26

References

Duperron, S., Sibuet, M., MacGregor, B.J., Kuypers, M.M.M., Fisher, C.R. & Dubilier, N. (2007b). Diversity, relative abundance and metabolic potential of bacterial endosymbionts in three Bathymodiolus mussel species from cold seeps in the Gulf of Mexico. *Environmental Microbiology*, **9**, 1423–1438. 29, 93, 96, 99

Duperron, S., Lorion, J., Lopez, P., Samadi, S., Gros, O. & Gaill, F. (2009). Symbioses between deep-sea mussels (Mytilidae: Bathymodiolinae) and chemosynthetic bacteria: diversity, function and evolution. *C. R. Acad. Sci. Biol.*, **332**, 298–310. 26

Eisen, J.A. (1995). The RecA protein as a model molecule for molecular systematic studies of bacteria: Comparison of trees of RecAs and 16S rRNAs from the same species. *Journal of Molecular Evolution*, **41**, 1105–1123. 49

Felbeck, H. (1981). Chemoautotrophic Potential of the Hydrothermal Vent Tube Worm, Riftia pachyptila Jones (Vestimentifera). *Science (New York, N.Y.)*, **213**, 336–8. 5

Fiala-Medioni, A., Metivier, C., Herry, A. & Pennec, M. (1986). Ultrastructure of the gill of the hydrothermal-vent mytilid Bathymodiolus sp. *Marine Biology*, **92**, 65–72. 27, 72

Fiala-Medioni, A., McKiness, Z.P., Dando, P., Boulegue, J., Mariotti, A., Alayse-Danet, A.M., Robinson, J.J. & Cavanaugh, C.M. (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. *Marine Biology*, **141**, 1035–1043. 27, 93

Fisher, C., Roberts, H., Cordes, E., Bernard, B. & Charles (2007). Cold Seeps and Associated Communities of the Gulf of Mexico. *Oceanography*, **20**, 118–129. 23

Giribet, G., Carranza, S., Baguñà, J., Riutort, M. & Ribera, C. (1996). First molecular evidence for the existence of a Tardigrada and Arthropoda clade. *Molecular Biology and Evolution*, **13**, 76 –84. 57

Godet, L., Zelnio, K.A. & Van Dover, C.L. (2011). Scientists as Stakeholders in Conservation of Hydrothermal Vents. *Conservation Biology*, **25**, 214–222. 15, 16

Goffredi, S.K., Waren, A., Orphan, V.J., Van Dover, C.L. & Vrijenhoek, R.C. (2004). Novel forms of structural integration between microbes and a hydrothermal vent gastropod from the Indian Ocean. *Applied and Environmental Microbiology*, **70**, 3082–3090. 6, 9

Goffredi, S.K., Jones, W.J., Erhlich, H., Springer, A. & Vrijenhoek, R.C. (2008). Epibiotic bacteria associated with the recently discovered Yeti crab, Kiwa hirsuta. *Environmental Microbiology*, **10**. 8

Gros, O., Liberge, M., Heddi, A., Khatchadourian, C. & Felbeck, H. (2003). Detection of the free-living forms of sulfide-oxidizing gill endosymbionts in the lucinid habitat (Thalassia testudinum environment). *Applied and Environmental Microbiology*, **69**, 6264–6267. 95

Gruber-Vodicka, H.R., Dirks, U., Leisch, N., Baranyi, C., Stoecker, K., Bulgheresi, S., Heindl, N.R., Horn, M., Lott, C., Loy, A., Wagner, M. & Ott, J. (2011). Paracatenula, an ancient symbiosis between thiotrophic Alphaproteobacteria and catenulid flatworms. *Proceedings of the National Academy of Sciences*, **108**, 12078 –12083. 8

Halary, S., Riou, V., Gaill, F., Boudier, T. & Duperron, S. (2008). 3D FISH for the quantification of methane- and sulphur-oxidizing endosymbionts in bacteriocytes of the hydrothermal vent mussel Bathymodiolus azoricus. *ISME Journal*, **2**, 284–292. 43, 44, 64, 88, 99

Hankeln, W., Buttigieg, P., Fink, D., Kottmann, R., Yilmaz, P. & Glockner, F. (2010). MetaBar - a tool for consistent contextual data acquisition and standards compliant submission. *BMC Bioinformatics*, **11**, 358. 39

Hanson, R. & Hanson, T. (1996). Methanotrophic bacteria. *Microbiol. Rev.*, **60**, 439–471. 13

Harada, M., Yoshida, T., Kuwahara, H., Shimamura, S., Takaki, Y., Kato, C., Miwa, T., Miyake, H. & Maruyama, T. (2009). Expression of genes for sulfur oxidation in the intracellular chemoautotrophic symbiont of the deep-sea bivalve Calyptogena okutanii. *Extremophiles*, **13**, 895–903. 10, 11

Harmer, T.L., Rotjan, R.D., Nussbaumer, A.D., Bright, M., Ng, A.W., DeChaine, E.G. & Cavanaugh, C.M. (2008). Free-Living Tube Worm Endosymbionts Found at Deep-Sea Vents. *Appl. Environ. Microbiol.*, **74**, 3895–3898. 95

Haymon, R., Fornari, D., Von Damm, K., Lilley, M., Perfit, M., Edmond, J., Shanks III, W., Lutz, R., Grebmeier, J., Carbotte, S., Wright, D., McLaughlin, E., Smith, M., Beedle, N. & Olson, E. (1993). Volcanic eruption of the mid-ocean ridge along the East Pacific Rise crest at 945-52'N: Direct submersible observations of seafloor phenomena associated with an eruption event in April, 1991. *Earth and Planetary Science Letters*, **119**, 85–101. 18

Higuchi, R., Fockler, C., Dollinger, G. & Watson, R. (1993). Kinetic Pcr Analysis - Real-Time Monitoring of DNA Amplification Reactions. *Bio-Technology*, **11**, 1026–1030. 45

Itoh, T., Takemoto, K., Mori, H. & Gojobori, T. (1999). Evolutionary instability of operon structures disclosed by sequence comparisons of complete microbial genomes. *Molecular Biology and Evolution*, **16**, 332 –346. 96

Johnson, K.S., Beehler, C.L., Sakamoto-Arnold, C.M. & Childress, J.J. (1986). In Situ Measurements of Chemical Distributions in a Deep-Sea Hydrothermal Vent Field. *Science*, **231**, 1139 –1141. 19

Johnson, K.S., Kennicutt, M.C., Lutz, R.A., Macko, S.A., Newton, A., Powell, M.A., Somero, G.N., Soto, T., Fisher, C.R., Childress, J.J., Arp, A.J., Brooks, J.M., Distel, D.L., Dugan, J.A., Felbeck, H., Fritz, L.W. & Hessler, R.R. (1988). Variation in the hydrothermal vent clam, Calyptogena magnifica, at the Rose Garden vent on the Galapagos Spreading-Center. *Deep-Sea Research Part I-Oceanographic Research Papers*, **35**, 1811–1831. 94

Jones, W.J., Won, Y.J., Maas, P.A.Y., Smith, P.J., Lutz, R.A. & Vrijenhoek, R.C. (2006). Evolution of habitat use by deep-sea mussels. *Marine Biology*, **148**, 841–851. 26

REFERENCES

Jorgensen, B.B. & Boetius, A. (2007). Feast and famine - microbial life in the deep-sea bed. *Nat Rev Micro*, **5**, 770–781. 22

Juniper, S.K. & Tunnicliffe, V. (1997). Crustal Accretion and the Hot Vent Ecosystem. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, **355**, 459–474. 18

Kádár, E., Davis, S.A. & Lobo-da Cunha, A. (2007). Cytoenzymatic investigation of intracellular digestion in the symbiont-bearing hydrothermal bivalve Bathymodiolus azoricus. *Marine Biology*, **153**, 995–1004. 92

Klein, D. (2002). Quantification using real-time PCR technology: applications and limitations. *Trends in Molecular Medicine*, **8**, 257–260. 46

Kochevar, R.E., Childress, J.J., Fisher, C.R. & Minnich, E. (1992). The methane mussel: roles of symbiont and host in the metabolic utilization of methane. *Marine Biology*, **112**, 389–401. 93

Kowalczykowski, S.C., Dixon, D.A., Eggleston, A.K., Lauder, S.D. & Rehrauer, W.M. (1994). Biochemistry of homologous recombination in Escherichia coli. *Microbiological Reviews*, **58**, 401–465. 49

Krueger, D.M., Dubilier, N. & Cavanaugh, C.M. (1996). Chemoautotrophic symbiosis in the tropical clam Solemya occidentalis (Bivalvia: Protobranchia): ultrastructural and phylogenetic analysis. *Marine Biology*, **126**, 55–64. 8

Kubista, M., Strömbom, L., Stå hlberg, A., Zoric, N., Andrade, J.M., Bengtsson, M., Forootan, A., Jonák, J., Lind, K., Sindelka, R., Sjöback, R. & Sjögreen, B. (2006). The real-time polymerase chain reaction. *Molecular Aspects of Medicine*, **27**, 95–125. 44, 45, 47, 59

Kuwahara, H., Takaki, Y., Yoshida, T., Shimamura, S., Takishita, K., Reimer, J.D., Kato, C. & Maruyama, T. (2008). Reductive genome evolution in chemoautotrophic intracellular symbionts of deep-sea Calyptogena clams. *Extremophiles*, **12**, 365–374. 49

Labrenz, M. (2009). Development and Application of a Real-Time PCR Approach for Quantification of Uncultured Bacteria in the Central Baltic Sea. 46

Landmann, L. (2002). Deconvolution improves colocalization analysis of multiple fluorochromes in 3D confocal data sets more than filtering techniques. *Journal of Microscopy-Oxford*, **208**, 134–147. 42

Lee, Z.M.P., Bussema, C. & Schmidt, T.M. (2009). rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucl. Acids Res.*, **37**, D489–493. 89, 90

Leloup, J., Loy, A., Knab, N.J., Borowski, C., Wagner, M. & Jø rgensen, B.B. (2007). Diversity and abundance of sulfate-reducing microorganisms in the sulfate and methane zones of a marine sediment, Black Sea. *Environmental Microbiology*, **9**, 131–142. 46

Lonsdale, P. (1977). Structural geomorphology of a fast-spreading rise crest: The East Pacific Rise near 325?S. *Marine Geophysical Researches*, **3**, 251–293. 16

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Förster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., König, A., Liss, T., Lüss mann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A. & Schleifer, K. (2004). ARB: a software environment for sequence data. *Nucleic Acids Research*, **32**, 1363 –1371. 55

Malinen, E., Kassinen, A., Rinttilä, T. & Palva, A. (2003). Comparison of real-time PCR with SYBR Green I or 5-nuclease assays and dot-blot hybridization with rDNA-targeted oligonucleotide probes in quantification of selected faecal bacteria. *Microbiology*, **149**, 269 –277. 46

Margulis, L. (1999). *Symbiotic planet: a new look at evolution*. Basic Books. 3

Markert, S., Arndt, C., Felbeck, H., Becher, D., Sievert, S.M., Hügler, M., Albrecht, D., Robidart, J., Bench, S., Feldman, R.A., Hecker, M. & Schweder, T. (2007). Physiological Proteomics of the Uncultured Endosymbiont of Riftia pachyptila. *Science*, **315**, 247 –250. 12

Masco, L., Vanhoutte, T., Temmerman, R., Swings, J. & Huys, G. (2007). Evaluation of real-time PCR targeting the 16S rRNA and recA genes for the enumeration of bifidobacteria in probiotic products. *International Journal of Food Microbiology*, **113**, 351–357. 46

Mayfield, A.B., Hirst, M.B. & Gates, R.D. (2009). Gene expression normalization in a dual-compartment system: a real-time quantitative polymerase chain reaction protocol for symbiotic anthozoans. *Molecular Ecology Resources*, **9**, 462–470. 46

Meijering, E. & Cappellen, G. (2007). Quantitative Biological Image Analysis. 45–70. 43

Meyer, A., Todt, C., Mikkelsen, N.T. & Lieb, B. (2010). Fast evolving 18S rRNA sequences from Solenogastres (Mollusca) resist standard PCR amplification and give new insights into mollusk substitution rate heterogeneity. *BMC Evolutionary Biology*, **10**, 70. 49

Miyazaki, J.I., Martins, L.d.O., Fujita, Y., Matsumoto, H. & Fujiwara, Y. (2010). Evolutionary Process of Deep-Sea Bathymodiolus Mussels. *PLoS ONE*, **5**, e10363. 25

Mueller, L., de Brouwer, J., Almeida, J., Stal, L. & Xavier, J. (2006). Analysis of a marine phototrophic biofilm by confocal laser scanning microscopy using the new image quantification software PHLIP. *BMC Ecology*, **6**, 1. 44, 64

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, **51 Pt 1**, 263–273. 45

References

Musat, N., Giere, O., Gieseke, A., Thiermann, F., Amann, R. & Dubilier, N. (2007). Molecular and morphological characterization of the association between bacterial endosymbionts and the marine nematode Astomonema sp. from the Bahamas. *Environmental Microbiology*, **9**, 1345–1353. 8

Mussmann, M., Ishii, K., Rabus, R. & Amann, R. (2005). Diversity and vertical distribution of cultured and uncultured Deltaproteobacteria in an intertidal mud flat of the Wadden Sea. *Environmental Microbiology*, **7**, 405–418. 43

Muyzer, G., Teske, A., Wirsen, C.O. & 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. *Archives of Microbiology*, **164pull**, 165–172. 57

Neumann, D. & Kappes, H. (2003). On the Growth of Bivalve Gills Initiated from a Lobule-Producing Budding Zone. *Biological Bulletin*, **205**, 73–82. 27

Nomura, M., Morgan, E.A. & Jaskunas, S.R. (1977). Genetics of Bacterial Ribosomes. *Annual Review of Genetics*, **11**, 297–347. 48

Olu-Le Roy, K., von Cosel, R., Hourdez, S., Carney, S.L. & Jollivet, D. (2007). Amphi-Atlantic cold-seep Bathymodiolus species complexes across the equatorial belt. *Deep Sea Research Part I: Oceanographic Research Papers*, **54**, 1890–1911. 26

Page, H., Fiala-Medioni, A., Fisher, C. & Childress, J. (1991). Experimental evidence for filter-feeding by the hydrothermal vent mussel, Bathymodiolus thermophilus. *Deep Sea Research Part A. Oceanographic Research Papers*, **38**, 1455–1461. 25

Palmer, S., Wiegand, A.P., Maldarelli, F., Bazmi, H., Mican, J.M., Polis, M., Dewar, R.L., Planta, A., Liu, S., Metcalf, J.A., Mellors, J.W. & Coffin, J.M. (2003). New Real-Time Reverse Transcriptase-Initiated PCR Assay with Single-Copy Sensitivity for Human Immunodeficiency Virus Type 1 RNA in Plasma. *J. Clin. Microbiol.*, **41**, 4531–4536. 46

Paull, C.K., Hecker, B., Commeau, R., Freemanlynde, R.P. & Neumann, C. (1984). Biological Communities at the Florida Escarpment resemble hydrothermal Vent Taxa. *Science*, **226**, 965–967. 22

Peel, F.J., Travis, C.J. & Hossack, J.R. (1995). Genetic structural provinces and salt tectonics of the Cenozoic offshore US Gulf of Mexico: A preliminary analysis. vol. 65 of *AAPG MEMOIRS*, 153–175, Amer Assoc. Petroleum Geologists, TULSA. 23

Pernthaler, A. & Amann, R. (2004). Simultaneous Fluorescence In Situ Hybridization of mRNA and rRNA in Environmental Bacteria. *Appl. Environ. Microbiol.*, **70**, 5426–5433. 28

Petersen, J.M. (2009). *Diversity and Ecology of Chemosynthetic Symbioses in Deep-Sea Invertebrates*. Ph.D. thesis, Max Planck Institut for Marine Microbiology. 9

Petersen, J.M. & Dubilier, N. (2009). Methanotrophic symbioses in marine invertebrates. *Environmental Microbiology Reports*, **1**, 319–335. 14, 28, 90

Petersen, J.M., Ramette, A., Lott, C., Cambon-Bonavita, M.A., Zbinden, M. & Dubilier, N. (2010). Dual symbiosis of the vent shrimp Rimicaris exoculata with filamentous gamma- and epsilonproteobacteria at four Mid-Atlantic Ridge hydrothermal vent fields. *Environmental Microbiology*, **9999**. 8

Petersen, J.M., Zielinski, F.U., Pape, T., Seifert, R., Moraru, C., Amann, R., Hourdez, S., Girguis, P.R., Wankel, S.D., Barbe, V., Pelletier, E., Fink, D., Borowski, C., Bach, W. & Dubilier, N. (2011). Hydrogen is an energy source for hydrothermal vent symbioses. *Nature*, **476**, 176–180. 10

Petersen, S., Kuhn, K., Kuhn, T., Augustin, N., Hékinian, R., Franz, L. & Borowski, C. (2009). The geological setting of the ultramafic-hosted Logatchev hydrothermal field (1445'N, Mid-Atlantic Ridge) and its influence on massive sulfide formation. *Lithos*, **112**, 40–56. 20

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J. & Glöckner, F.O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, **35**, 7188 –7196. 55

Raulfs, E., Macko, S. & Van Dover, C. (2004). Tissue and Symbiont Condition of Mussels (Bathymodiolus Thermophilus) Exposed to Varying Levels of Hydrothermal Activity. *Journal of the Marine Biological Association of the United Kingdom*, **84**, 229–234. 93

Rinke, C., Lee, R., Katz, S. & Bright, M. (2007). The effects of sulphide on growth and behaviour of the thiotrophic Zoothamnium niveum symbiosis. *Proceedings of the Royal Society B: Biological Sciences*, **274**, 2259–2269. 9

Riou, V., Halary, S., Duperron, S., Bouillon, S., Elskens, M., Bettencourt, R., Santos, R.S., Dehairs, F. & Colaço, A. (2008). Influence of CH4 and H2S availability on symbiont distribution, carbon assimilation and transfer in the dual symbiotic vent mussel Bathymodiolus azoricus. *Biogeosciences*, **5**, 1681–1691. 94

Roberts, H., Shedd, W. & Hunt Jr., J. (2010). Dive site geology: DSV ALVIN (2006) and ROV JASON II (2007) dives to the middle-lower continental slope, northern Gulf of Mexico. *Deep Sea Research Part II: Topical Studies in Oceanography*, **57**, 1837–1858. 23

Robidart, J.C., Bench, S.R., Feldman, R.A., Novoradovsky, A., Podell, S.B., Gaasterland, T., Allen, E.E. & Felbeck, H. (2008). Metabolic versatility of the Riftia pachyptila endosymbiont revealed through metagenomics. *Environmental Microbiology*, **10**, 727–737. 12

Robigou, V., Delaney, J.R. & Stakes, D.S. (1993). Large massive sulfide deposits in a newly discovered active hydrothermal system, The HighRise Field, Endeavour Segment, Juan De Fuca Ridge. *Geophysical Research Letters*, **20**, 1887. 19

Ruppel, S., Rühlmann, J. & Merbach, W. (2006). Quantification and Localization of Bacteria in Plant Tissues Using Quantitative Real-Time PCR and Online Emission Fingerprinting. *Plant and Soil*, **286**, 21–35. 43

REFERENCES

Samadi, S., Quemere, E., Lorion, J., Tillier, A., von Cosel, R., Lopez, P., Cruaud, C., Couloux, A. & Boisselier-Dubayle, M.C. (2007). Molecular phylogeny in mytilids supports the wooden steps to deep-sea vents hypothesis. *Comptes Rendus Biologies*, **330**, 446–456. 25

Schattenhofer, M., Fuchs, B.M., Amann, R., Zubkov, M.V., Tarran, G.A. & Pernthaler, J. (2009). Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environmental Microbiology*, **11**, 2078–2093. 43

Schmidt, K., Koschinsky, A., Garbe-Schönberg, D., de Carvalho, L.M. & Seifert, R. (2007). Geochemistry of hydrothermal fluids from the ultramafic-hosted Logatchev hydrothermal field, on the Mid-Atlantic Ridge: Temporal and spatial investigation. *Chemical Geology*, **242**, 1–21. 20, 21

Schulte, M., Blake, D., Hoehler, T. & McCollom, T. (2006). Serpentinization and Its Implications for Life on the Early Earth and Mars. *Astrobiology*, **6**, 364–376. 20

Sen, K., Schable, N.A. & Lye, D.J. (2007). Development of an Internal Control for Evaluation and Standardization of a Quantitative PCR Assay for Detection of Helicobacter pylori in Drinking Water. *Appl. Environ. Microbiol.*, **73**, 7380–7387. 46

Smith, C.J., Nedwell, D.B., Dong, L.F. & Osborn, A.M. (2006). Evaluation of quantitative polymerase chain reaction-based approaches for determining gene copy and gene transcript numbers in environmental samples. *Environmental Microbiology*, **8**, 804–815. 54

Smith, U., Ribbons, D.W. & Smith, D.S. (1970). The fine structure of Methylococcus capsulatus. *Tissue and Cell*, **2**, 513–520. 13

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M. & Herndl, G.J. (2006). Microbial diversity in the deep sea and the underexplored rare biosphere. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 12115–12120. 19

Solovei, I., Cavallo, A., Schermelleh, L., Jaunin, F., Scasselati, C., Cmarko, D., Cremer, C., Fakan, S. & Cremer, T. (2002). Spatial Preservation of Nuclear Chromatin Architecture during Three-Dimensional Fluorescence in Situ Hybridization (3D-FISH). *Experimental Cell Research*, **276**, 10–23. 42

Somero, G.N. (1984). Physiology and biochemistry of the hydrothermal vent animals. *Oceanus*, **27**, 67–72. 4

Spiridonova, E.M., Kuznetsov, B.B., Pimenov, N.V. & Tourova, T.P. (2006). Phylogenetic characterization of endosymbionts of the hydrothermal vent mussel Bathymodiolus azoricus by analysis of the 16S rRNA, cbbL, and pmoA genes. *Microbiology*, **75**, 694–701. 12

Steedman, H.F. (1957). Polyester Wax: A New Ribboning Embedding Medium for Histology. *Nature*, **179**, 1345. 50

Stewart, F.J. & Cavanaugh, C.M. (2006). Symbiosis of Thioautotrophic Bacteria with Riftia pachyptila. vol. 41, 197–225, Springer-Verlag, Berlin/Heidelberg. 8

Stewart, F.J., Newton, I.L. & Cavanaugh, C.M. (2005). Chemosynthetic endosymbioses: adaptations to oxic-anoxic interfaces. *Trends in Microbiology*, **13**, 439–448. 6

Stewart, F.J., Dmytrenko, O., DeLong, E.F. & Cavanaugh, C.M. (2011). Metatranscriptomic Analysis of Sulfur Oxidation Genes in the Endosymbiont of Solemya Velum. *Frontiers in Microbiology*, **2**. 10

Sudarikov, S.M. & Roumiantsev, A.B. (2000). Structure of hydrothermal plumes at the Logatchev vent field, Mid-Atlantic Ridge: evidence from geochemical and geophysical data. *Journal of Volcanology and Geothermal Research*, **101**, 245–252. 35

Suzuki, Y., Suzuki, M., Tsuchida, S., Takai, K., Horikoshi, K., Southward, A.J., Newman, W.A. & Yamaguchi, T. (2009). Molecular Investigations of the Stalked Barnacle Vulcanolepas Osheai and the Epibiotic Bacteria from the Brothers Caldera, Kermadec Arc, New Zealand. *Journal of the Marine Biological Association of the United Kingdom*, **89**, 727–733. 8

Taylor, S., Wakem, M., Dijkman, G., Alsarraj, M. & Nguyen, M. (2010). A practical approach to RT-qPCR–Publishing data that conform to the MIQE guidelines. *Methods*, **50**, S1–S5. 46

Theisen, A.R., Ali, M.H., Radajewski, S., Dumont, M.G., Dunfield, P.F., McDonald, I.R., Dedysh, S.N., Miguez, C.B. & Murrell, J.C. (2005). Regulation of methane oxidation in the facultative methanotroph Methylocella silvestris BL2. *Molecular Microbiology*, **58**, 682–692. 13

Thiele, S. & Amann, R. (2011). Identification of Microorganisms Using the Ribosomal RNA Approach and Fluorescence In Situ Hybridization. 48

Tobiason, D.M. & Seifert, H.S. (2006). The Obligate Human Pathogen, Neisseria gonorrhoeae, Is Polyploid. *PLoS Biol*, **4**, e185. 90

Trappe, J.M. (2004). A.B. Frank and mycorrhizae: the challenge to evolutionary and ecologic theory. *Mycorrhiza*, **15**, 277–281. 3

Trask, J.L. & Van Dover, C.L. (1999). Site-specific and ontogenetic variations in nutrition of mussels (Bathymodiolus sp.) from the Lucky Strike hydrothermal vent field, Mid-Atlantic Ridge. *Limnology and Oceanography*, **44**, 334–343. 93

Turekian, K. (2010). *Marine chemistry and geochemistry*. Academic Press, Amsterdam. 20

Tyler, P. & Young, C. (1999). Reproduction and Dispersal at Vents and Cold Seeps. *Journal of the Marine Biological Association of the United Kingdom*, **79**, 193–208. 26

Vacelet, J., Fiala-Médioni, A., Fisher, C.R. & BouryEsnault, N. (1996). Symbiosis between methane-oxidizing bacteria and a deep-sea carnivorous cladorhizid sponge. *Marine Ecology Progress Series*, **145**, 77–85. 8

References

Valm, A.M., Mark Welch, J.L., Rieken, C.W., Hasegawa, Y., Sogin, M.L., Oldenbourg, R., Dewhirst, F.E. & Borisy, G.G. (2011). Systems-level analysis of microbial community organization through combinatorial labeling and spectral imaging. *Proceedings of the National Academy of Sciences*. 41

van der Velden, V.H.J., Hochhaus, A., Cazzaniga, G., Szczepanski, T., Gabert, J. & van Dongen, J.J.M. (2003). Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. *Leukemia*, **17**, 1013–1034. 47

Van Dover, C.L. & Trask, J.L. (2000). Diversity at deep-sea hydrothermal vent and intertidal mussel beds. *Marine Ecology-Progress Series*, **195**, 169–178. 16

Van Dover, C.L., German, C.R., Speer, K.G., Parson, L.M. & Vrijenhoek, R.C. (2002). Marine biology - Evolution and biogeography of deep-sea vent and seep invertebrates. *Science*, **295**, 1253–1257. 20

Verna, C. (2010). *Phylogeny and diversity of symbionts from whale fall invertebrates*. Ph.D. thesis. 8

Wallner, G., Amann, R. & Beisker, W. (1993). Optimizing fluorescent in situ hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. *Cytometry*, **14**, 136–143. 52

Wankel, S.D., Joye, S.B., Samarkin, V.A., Shah, S.R., Friederich, G., Melas-Kyriazi, J. & Girguis, P.R. (2010). New constraints on methane fluxes and rates of anaerobic methane oxidation in a Gulf of Mexico brine pool via in situ mass spectrometry. *Deep Sea Research Part II: Topical Studies in Oceanography*, **57**, 2022–2029. 38

Weibel (1979). Multiple stage sampling. 53

Weibel, E.R., Weibel, R. & Scherle, F. (1966). Practical stereological methods for morphometric cytology. 52

Wendeberg, A., Zielinski, F.U., Borowski, C. & Dubilier, N. (2011). Expression patterns of mRNAs for methanotrophy and thiotrophy in symbionts of the hydrothermal vent mussel Bathymodiolus puteoserpentis. *ISME J*. 25, 29, 52, 87

Williamson, S.J., Cary, S.C., Williamson, K.E., Helton, R.R., Bench, S.R., Winget, D. & Wommack, K.E. (2008). Lysogenic virus-host interactions predominate at deep-sea diffuse-flow hydrothermal vents. *ISME J*, **2**, 1112–1121. 20

Windoffer, R. & Giere, O. (1997). Symbiosis of the Hydrothermal Vent Gastropod Ifremeria nautilei (Provannidae) With Endobacteria-Structural Analyses and Ecological Considerations. *Biol Bull*, **193**, 381–392. 8

Woese, C.R. (1987). Bacterial evolution. *Microbiological Reviews*, **51**, 221–271. 48

Won, Y.J., Hallam, S.J., O'Mullan, G.D., Pan, I.L., Buck, K.R. & Vrijenhoek, R.C. (2003). Environmental acquisition of thiotrophic endosymbionts by deep-sea mussels of the genus Bathymodiolus. *Applied and Environmental Microbiology*, **69**, 6785–6792. 27, 28

Yang, S. & Rothman, R.E. (2004). PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. *The Lancet Infectious Diseases*, **4**, 337–348. 46

Zal, F., Leize, E., Oros, D.R., Hourdez, S., Van Dorsselaer, A. & Childress, J.J. (2000). Haemoglobin structure and biochemical characteristics of the sulphide-binding component from the deep-sea clam Calyptogena magnifica. *Cahiers De Biologie Marine*, **41**, 413–423. 30

Zemanick, E.T., Wagner, B.D., Sagel, S.D., Stevens, M.J., Accurso, F.J. & Harris, J.K. (2010). Reliability of Quantitative Real-Time PCR for Bacterial Detection in Cystic Fibrosis Airway Specimens. *PLoS ONE*, **5**, e15101. 91

Zhou, J., Bruns, M. & Tiedje, J. (1996). DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.*, **62**, 316–322. 54

Zielinski, F., Gennereich, H.H., Borowski, C., Wenzhöfer, F. & Dubilier, N. (2011). In situ measurements of hydrogen sulfide, oxygen, and temperature in diffuse fluids of an ultramafic-hosted hydrothermal vent field (Logatchev, Mid-Atlantic Ridge): implications for chemosymbiotic bathymodiolin mussels. *Geochem. Geophys. Geosyst.*. 94

Zielinski, F.U., Pernthaler, A., Duperron, S., Raggi, L., Giere, O., Borowski, C. & Dubilier, N. (2009). Widespread occurrence of an intranuclear bacterial parasite in vent and seep bathymodiolin mussels. *Environmental Microbiology*, **11**, 1150–1167. 29

Acknowledgments

Many thanks to...

Dr. Nicole Dubilier, for opening my eyes to the crazy world of science, for showing me my strengths and weaknesses and guiding me through an awesome PhD. A hard boss, kind and caring women, and inspiring mentor.

Dr. Christian Borowski, for taking me to the ocean three times and bringing me back in one piece, for giving me a hand when I needed it and advices when I lost my track.

Prof. Dr. Ulrich Fischer, for enriching my thesis committee with fruitful discussions and for reviewing my thesis.

Prof. Dr. Rudolf Amann, for joining my defense committee and having an open door at all time.

The members of my thesis committee and all the people helping me to finish this thesis work in time.

The Symbiosis Group, for being an extraordinary team over the years, for not being only great colleagues, but friends.

My MPI friends, for the coffee breaks, office chats, lab work, party times, scientific discussions, procrastination and much much more.

My homies Ivo, Wolle, Dennis and Thorben, for all the poker nights, climbing hours, tram parties, and just every day life.

Bernd Stickfort, a man that can find any scientific publication written by human kind and has a good music taste.

The Microbial Genomics Group, for adopting me each day for lunch as one of their kind.

Christiane Glöckner and Karl-Heinz Blotevogel, for doing such a great job in coordinating PhD's at this institute.

Olaf Gunderman and the EDV, for their excellent technical support and help.

My band Emil, Morten, Arne and Kolja, for sharing with me the energy and power that only music can give you and for counterbalancing the every day lunacy of science.

My flatmates Helge, Robert, Henning, Sebastian, Annika, Colette and Susi, for dragging me out of my world and putting me into our living room, a perfect shelter for me when I needed it. For brightening up each day with coffee and cigarette breaks, movie nights and warm beers. Best flatmates ever! - Playstation 3 and Beamer included.

All my friends and people I forgot to mention here. You kept me going all the time and I hope to see you again.

My family in Cologne, for their endless support over the years.

Michi, my love, my best friend and my anchor in this turbulent ocean called life.

Part VI

Appendix

Contributions to publications are included as appendix:

1. **Petersen, J.M., Zielinski, F.U., Pape, T., Seifert, R., Moraru, C., Amann, R., Hourdez, S., Girguis, P.R., Wankel, S.D., Barbe, V., Pelletier, E., Fink, D., Borowski, C., Bach, W. Dubilier, N.** (2011). Hydrogen is an energy source for hydrothermal vent symbioses. Nature, 476, 176-180.
   D.F. participated in two cruises that contributed to this paper, during which he collected and prepared specimens, and did on-board experiments showing uptake of hydrogen in mussel gill tissues.

2. **Hankeln, W., Buttigieg, P., Fink, D., Kottmann, R., Yilmaz, P. Glöckner, F.** (2010). MetaBar - a tool for consistent contextual data acquisition and standards compliant submission. BMC Bioinformatics, 11, 358.
   D.F. tested the software on cruises and provided feedback for design improvements.

ARTICLE

doi:10.1038/nature10325

Hydrogen is an energy source for hydrothermal vent symbioses

Jillian M. Petersen[1]*, Frank U. Zielinski[1,2]*, Thomas Pape[3], Richard Seifert[4], Cristina Moraru[1], Rudolf Amann[1], Stephane Hourdez[5], Peter R. Girguis[6], Scott D. Wankel[6], Valerie Barbe[7], Eric Pelletier[7], Dennis Fink[1], Christian Borowski[1], Wolfgang Bach[8] & Nicole Dubilier[1]

The discovery of deep-sea hydrothermal vents in 1977 revolutionized our understanding of the energy sources that fuel primary productivity on Earth. Hydrothermal vent ecosystems are dominated by animals that live in symbiosis with chemosynthetic bacteria. So far, only two energy sources have been shown to power chemosynthetic symbioses: reduced sulphur compounds and methane. Using metagenome sequencing, single-gene fluorescence *in situ* hybridization, immunohistochemistry, shipboard incubations and *in situ* mass spectrometry, we show here that the symbionts of the hydrothermal vent mussel *Bathymodiolus* from the Mid-Atlantic Ridge use hydrogen to power primary production. In addition, we show that the symbionts of *Bathymodiolus* mussels from Pacific vents have *hupL*, the key gene for hydrogen oxidation. Furthermore, the symbionts of other vent animals such as the tubeworm *Riftia pachyptila* and the shrimp *Rimicaris exoculata* also have *hupL*. We propose that the ability to use hydrogen as an energy source is widespread in hydrothermal vent symbioses, particularly at sites where hydrogen is abundant.

Deep-sea hydrothermal vents and their associated chemosynthetic communities were discovered in 1977[1]. Since then, two energy sources have been shown to fuel primary production by the symbiotic bacteria that form the basis of the food chain in marine chemosynthetic ecosystems. In 1981, chemolithoautotrophic bacteria that use reduced sulphur compounds as an energy source were discovered in the gutless *Riftia pachyptila* tubeworms from hydrothermal vents in the Pacific[2,3]. Five years later, the first symbionts that use methane as an energy source were discovered in mussels from hydrocarbon seeps in the Gulf of Mexico[4,5]. Since then, a vast array of chemosynthetic ecosystems has been explored, and novel symbioses of phylogenetically diverse hosts and symbionts are constantly being described[6]. Despite this, no other source of energy for metazoan chemosynthetic symbioses has yet been found. This is remarkable, given that many potential sources of energy for chemosynthesis are available, such as hydrogen, ammonium, ferrous iron and manganese(II)[7], and free-living vent microbes able to use these energy sources are well known[8–10].

Some hydrothermal vents produce fluids with very high hydrogen concentrations due to the interaction of seawater with mantle-derived ultramafic rocks. Fluids originating from ultramafic-hosted vents are also characterized by high methane concentrations, whereas H_2S concentrations are rather low. In contrast, basalt-hosted vents produce fluids comparably high in H_2S but low in H_2 and CH_4. The Logatchev vent field at 14° 45′ N on the Mid-Atlantic Ridge (MAR) is located in a ridge segment characterized by ultramafic outcrops (Supplementary Fig. 2). Fluids venting at this site have the highest hydrogen concentrations ever measured in hydrothermal systems (19 mM in the end-member fluids)[11], which would provide a rich source of energy for chemosynthetic microbes. Hydrogen is a particularly favourable electron donor as the energy yield from hydrogen oxidation is much higher than from methane oxidation, sulphur oxidation and all other potential electron donors for chemolithoautotrophic growth under standard conditions[12]. In fact, our thermodynamic model predicts that at the Logatchev vent field, aerobic hydrogen oxidation could provide up to 7 times more energy per kilogram of vent fluid than methane oxidation, and up to 18 times more energy per kilogram of vent fluid than sulphide oxidation (Supplementary Fig. 1).

At Logatchev, *Bathymodiolus puteoserpentis* mussels are by far the most abundant macrofauna[13]. They live in a dual symbiosis with methane-oxidizing and chemoautotrophic sulphur-oxidizing bacteria that are hosted in their gills[14–16]. Here we show that the sulphur-oxidizing symbiont of *B. puteoserpentis* uses hydrogen as an energy source. We also show that *Bathymodiolus* symbionts from basalt-hosted MAR vent fields can oxidize hydrogen. This ability is therefore not limited to the ultramafic-hosted vent fields and could also be powering chemosynthetic symbioses at basalt-hosted vent fields.

Uptake hydrogenase genes in mussels

The key enzymes involved in hydrogen metabolism are hydrogenases, which catalyse the reaction: $H_2 \leftrightarrow 2H^+ + 2e^-$ (ref. 17). Enzymes of the group 1 NiFe hydrogenases are membrane-bound respiratory enzymes that channel electrons from hydrogen into the quinone pool, providing the link between hydrogen oxidation and energy production[17,18]. The large subunit of the membrane-bound uptake hydrogenase is encoded by the *hupL* gene. We amplified and sequenced this gene from symbiont-containing *B. puteoserpentis* gill tissues from Logatchev (Supplementary Fig. 3). The closest related sequence was the large subunit of the NiFe hydrogenase from the alphaproteobacterium *Oligotropha carboxidovorans* (79.6% amino acid identity). *O. carboxidovorans* can grow chemolithoautotrophically under aerobic conditions using either CO or H_2 as an electron donor[19]. Phylogenetic analyses based on multiple treeing methods placed the enzyme from *B. puteoserpentis* in a well-supported cluster with other group 1 NiFe hydrogenases, showing the genetic potential for hydrogen oxidation

[1]Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, 28359 Bremen, Germany. [2]Helmholtz Centre for Environmental Research - UFZ, Permoserstrasse 15, 04318 Leipzig, Germany. [3]MARUM - Center for Marine Environmental Sciences and Department of Geosciences, University of Bremen, Klagenfurter Strasse, 28334 Bremen, Germany. [4]University of Hamburg, Institute for Biogeochemistry and Marine Chemistry, Bundesstrasse 55, 20146 Hamburg, Germany. [5]Equipe Genetique des Adaptations aux Milieux Extremes, CNRS-UPMC UMR 7144, Station Biologique, BP74, 29682 Roscoff, France. [6]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [7]Commissariat à l'Energie Atomique/Genoscope, 91000 Évry, France, Centre National de la Recherche Scientifique, UMR8030, 91000 Évry, France, and Université d'Evry Val d'Essone 91000 Évry, France. [8]Department of Geosciences, University of Bremen, PO Box 33 04 40, 28334 Bremen, Germany. *These authors contributed equally to this work.

linked to energy generation in *B. puteoserpentis* endosymbionts. To examine whether the endosymbionts of mussels from hydrogen-poor habitats also have the genetic potential for hydrogen uptake, we tried to amplify the *hupL* gene from gill tissues of mussels from basalt-hosted vents and cold seeps that have fluids with low hydrogen concentrations. Indeed, we could amplify the *hupL* gene from mussels from basalt-hosted vents, including undescribed *Bathymodiolus* mussels (*Bathymodiolus* spp.) from vents on the southern MAR (Wideawake at 4° 48′ S, 5,600 km from Logatchev, and Lilliput at 9° 33′ S, 6,500 km from Logatchev), and *B.* aff. *thermophilus* from the Axial Dome vent on the Pacific–Antarctic Ridge at 37° 47′ S (Supplementary Figs 2, 3 and Supplementary Table 7), showing that the genetic potential for hydrogen oxidation is not restricted to mussel symbionts from ultramafic-hosted vent fields that have high hydrogen concentrations. Intriguingly, we could not amplify this gene from any of the cold seep mussels investigated (see Supplementary Discussion).

B. puteoserpentis symbionts use H_2

Given the genetic potential for hydrogen uptake in *Bathymodiolus* mussels, we incubated *B. puteoserpentis* mussel gill tissues from the ultramafic-hosted Logatchev vent field on the MAR with hydrogen at partial pressures of ~100 p.p.m. in the headspace (0.08 μM H_2 dissolved in the medium) and measured its consumption over time. Hydrogen was taken up rapidly by the symbiont-containing gill tissues at a rate of ~650 ± 200 nmol h^{-1} (g wet weight)$^{-1}$ ($n = 7$). In contrast, symbiont-free foot tissue did not consume hydrogen at rates above the negative controls—boiled gill tissue and seawater (Fig. 1 and Supplementary Table 3).

Because hydrogen uptake by bacteria is not necessarily coupled to CO_2 fixation[17,20], we incubated mussel gill tissues with hydrogen in seawater containing ^{14}C-bicarbonate to determine whether hydrogen is an energy source for autotrophic CO_2 fixation by *B. puteoserpentis* symbionts. Control gill tissues were incubated in the presence of sulphide or without an electron donor. ^{14}C uptake was stimulated by sulphide, known to be an energy source for the sulphur-oxidizing symbionts of *Bathymodiolus* mussels[21], and also by hydrogen (Fig. 1). The rates of carbon fixation with hydrogen were comparable to those supported by sulphide oxidation, which suggests that hydrogen could be fuelling autotrophy to the same extent as sulphide. We therefore conclude that hydrogen provides energy for the production of mussel biomass at the Logatchev hydrothermal vent field.

Symbionts from basalt-hosted vents use H_2

We examined whether the symbionts of *Bathymodiolus* mussels from basalt-hosted vent fields on the southern MAR at Comfortless Cove and Lilliput with low *in situ* hydrogen concentrations could also consume hydrogen. Hydrogen concentrations measured in discrete samples of the diffuse fluids from these vents were typically below 0.1 μM (refs 22,23) as opposed to Logatchev, which had up to 154 μM, on the basis of discrete sampling (Supplementary Table 3). We incubated gill tissues of *B.* spp. mussels and measured hydrogen consumption over time as described earlier. Our results show that mussels from these vents were also able to take up hydrogen (Fig. 1 and Supplementary Table 3). However, the rates at which the symbiotic gill tissues of the southern MAR mussels consumed hydrogen were 20- to 30-fold lower than those of mussels from Logatchev.

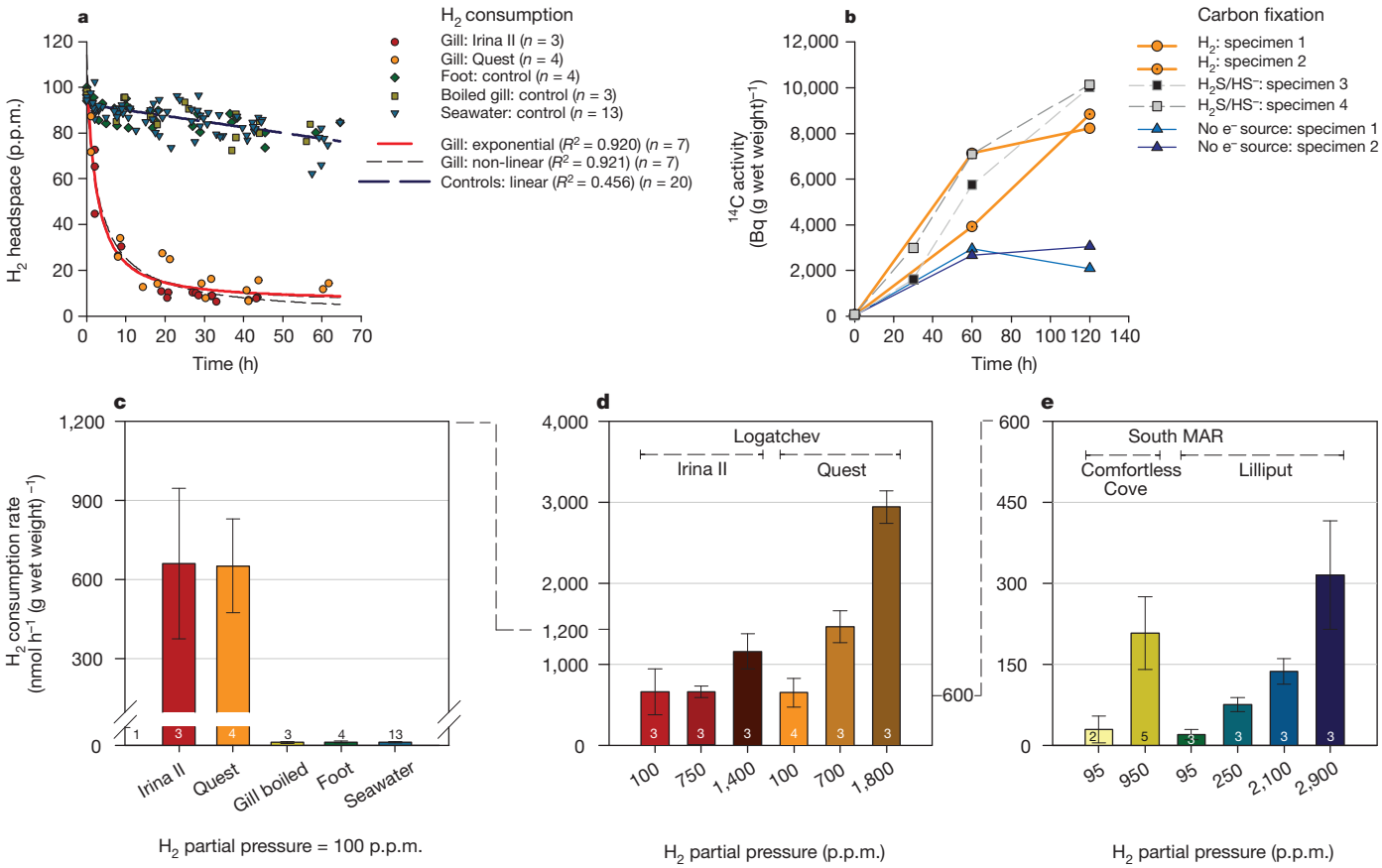

Figure 1 | Hydrogen consumption in *Bathymodiolus* gills. **a**, **b**, Hydrogen was consumed rapidly by *B. puteoserpentis* gills (Logatchev) and carbon fixation was stimulated after incubation with ^{14}C-bicarbonate in the presence of hydrogen or sulphide. Each data point is one distinct measurement. **c**, Consumption rates of gills from two sites at Logatchev were in the same range. **d**, **e**, Rates from Logatchev and the southern MAR fields (*B.* spp.; Comfortless Cove, 4° 48′ S and Lilliput, 9° 33′ S) increased with increasing hydrogen partial pressures. Rates were roughly 20–30-fold higher at the hydrogen-rich Logatchev field than at the hydrogen-poor southern MAR vents. Error bars represent one standard deviation.

Effect of H_2 concentration on consumption

Some hydrogen-oxidizing microorganisms are known to induce hydrogenase expression only in the presence of hydrogen, and some express this enzyme constitutively at a low level, but increase its expression upon incubation with hydrogen[24]. To investigate the effect of dissolved hydrogen concentration on hydrogen consumption rates, we incubated gill tissues from the Logatchev and southern MAR venting sites at partial pressures of up to 3,000 p.p.m. (~2.3 μM dissolved H_2). Hydrogen consumption rates in mussels from Logatchev, Comfortless Cove and Lilliput increased by 135, 21 and 8 nmol h^{-1} (g wet weight)$^{-1}$, respectively, for each 100 p.p.m. increase in partial pressure (Fig. 1). Thus, hydrogen uptake is clearly stimulated by increasing hydrogen concentrations.

The sulphur-oxidizing symbiont uses H_2

Many phylogenetically diverse microorganisms can grow autotrophically on hydrogen. *Bathymodiolus* mussels from the MAR host two types of chemosynthetic gammaproteobacteria: a sulphur oxidizer and a methane oxidizer[14] (Fig. 2). In addition, these mussels have a gammaproteobacterial parasite that infects gill nuclei[25]. Hydrogen is known to provide energy for chemo- and photolithoautotrophic growth of some free-living sulphur-oxidizing bacteria[26,27]. Some free-living methane-oxidizing bacteria can also oxidize hydrogen in addition to methane[28]. Lastly, hydrogen has been shown to be a key determinant of the virulence of pathogenic *Helicobacter pylori*[29]. Accordingly, it was unclear which of the three types of bacteria associated with *Bathymodiolus* mussels could be using the hydrogen. To investigate this further, we used three molecular methods allowing us to link identity with function at the DNA and protein level: genome sequencing of the mussel symbionts, single-gene fluorescence *in situ* hybridization (geneFISH) and immunohistochemistry, the latter two combined with 16S rRNA FISH.

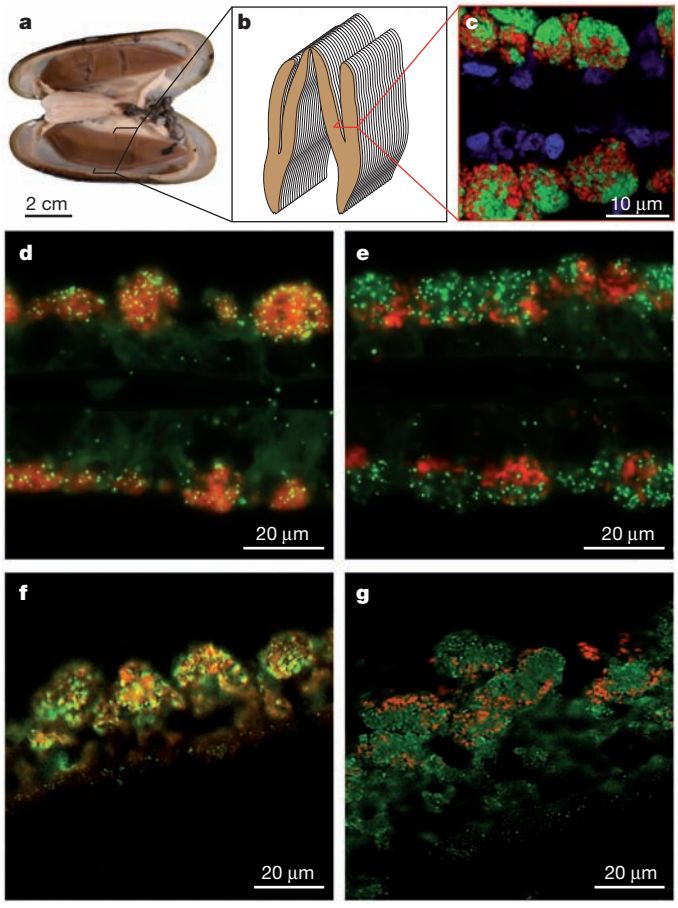

Figure 2 | The sulphur-oxidizing symbiont has the gene for hydrogen uptake, which it expresses. **a**, *B. puteoserpentis* with brown, symbiont-containing gills. **b**, Morphology of a single gill, each composed of many filaments (modified from ref. 48). **c**, FISH of a cross-section through a single gill filament (host cell nuclei, blue; sulphur-oxidizing symbiont, green; methane-oxidizing symbiont, red). **d**, **e**, *hupL* geneFISH (signals in green) with 16S rRNA FISH (signals in red) for the sulphur-oxidizing symbiont (**d**) and the methane-oxidizing symbiont (**e**). The negative control gave no gene signals (Supplementary Fig. 7). **f**, **g**, HupL immunohistochemistry (signals in green) with 16S rRNA FISH (signals in red) for the sulphur-oxidizing symbiont (**f**) and the methane-oxidizing symbiont (**g**). Yellow colours in **d** and **f** appear where green and red signals overlap.

In the genomes of hydrogen-oxidizing bacteria, the genes for the large and small subunits of membrane-bound NiFe hydrogenases are often found clustered together with the genes necessary for biosynthesis, maturation and processing[17]. We used metagenomics to investigate hydrogen oxidation in the genome of *Bathymodiolus* symbionts from the southern MAR (Lilliput). We found a hydrogenase operon with 19 open reading frames on a 726-kb genome fragment (Supplementary Fig. 4). The large subunit gene from this genome fragment was 99.6% identical to the *hupL* gene amplified from the Wideawake (southern MAR) mussels using polymerase chain reaction (PCR), and 83% identical to the gene from *B. puteoserpentis* from Logatchev on the northern MAR. Homologues of all structural genes and those necessary for hydrogenase synthesis, assembly and function in the chemolithoautotrophic hydrogen oxidizer *Cupriavidus necator*[30] were present on our genome fragment. The southern MAR *Bathymodiolus* symbiont therefore has all of the genetic components required for hydrogen uptake. On the same genome fragment, we found the key genes for sulphur oxidation via the reverse dissimilatory sulphite reductase (rDsr) and Sox pathways, as well as for CO_2 fixation by the Calvin–Benson–Bassham (CBB) cycle (Supplementary Fig. 4), indicating that this fragment is from the sulphur-oxidizing symbiont of *Bathymodiolus*.

To establish further that it is the sulphur-oxidizing symbiont that uses hydrogen, we used geneFISH, a method in which simultaneous detection of single genes and rRNA enables the linking of function and identity at the single-cell level[31]. We detected the *hupL* gene in the symbiont-containing gill tissue of *B. puteoserpentis* (Fig. 2). GeneFISH signals from probes designed to target the *hupL* gene amplified by PCR from *B. puteoserpentis* overlapped with the 16S rRNA signals from the previously described sulphur-oxidizing symbiont (Fig. 2). This further suggests that the *hupL* gene we amplified is from the sulphur-oxidizing symbiont of *B. puteoserpentis*.

To show that the sulphur-oxidizing symbiont expresses the uptake hydrogenase, we combined immunohistochemistry with FISH of the 16S rRNA of the symbionts of *B. puteoserpentis*, thus linking identity and function at the level of gene expression. Using a polyclonal antiserum against the membrane-bound hydrogen-uptake NiFe hydrogenase from *C. necator* (74% amino acid sequence identity to the *B. puteoserpentis* symbiont HupL; Supplementary Fig. 5), we were able to detect hydrogenase expression in single symbiont cells in *B. puteoserpentis* (Fig. 2). Signals from the anti-hydrogenase antibody were seen in gill bacteriocytes of *B. puteoserpentis* where they overlapped with the FISH signals from the sulphur-oxidizing symbiont, but not with those from the methane-oxidizing symbiont.

Environmental significance of H_2 use

One of the major challenges in hydrothermal vent research is to link physiological experiments, performed on board the research vessel at atmospheric pressure, with processes occurring under *in situ* temperature and hydrostatic pressure. We deployed an *in situ* mass spectrometer (ISMS) to measure hydrogen concentrations in the mussel habitat at Logatchev, and compared these with temperature as a conservative tracer for mixing of the hydrothermal fluids with seawater. Simultaneous measurements of hydrogen concentrations and temperature were made in two different settings; first, in an area of

focused flow, directly at the source where the fluids exit the seafloor. Here, the fluids had not been exposed to *B. puteoserpentis* mussels or other macrofauna. Second, we measured hydrogen concentrations and temperature in an area where the fluids had passed through a *B. puteoserpentis* mussel bed. The slope of the regression line calculated for hydrogen concentration versus temperature was significantly lower in the mussel bed compared to the source fluid ($P < 0.0001$), indicating that fluids in the mussel bed are hydrogen-depleted compared to the source fluid (Fig. 3). This difference is probably due to hydrogen consumption by the sulphur-oxidizing symbionts of *B. puteoserpentis*. We observed the same effect with measurements of methane and temperature ($P < 0.0001$; Supplementary Fig. 6). These measurements confirm that significant amounts of hydrogen are being consumed in Logatchev mussel beds.

The terrestrial hydrogen biogeochemical cycle has been the topic of numerous studies (summarized in ref. 32), and interest in the global hydrogen cycle has recently grown with the prospect of a hydrogen economy[33]. In contrast to the terrestrial environment, there is a paucity of data on hydrogen turnover in the oceans. Conservative calculations based on our measurements of uptake rates in Logatchev mussels show that at 50 μM dissolved hydrogen a single mussel with a gill weight of 5 g could oxidize up to 435 μmol H_2 h^{-1} (Supplementary Information). On the basis of our own estimates and those of others, the mussel population at Logatchev spreads in layers over at least 200 m^2 and accounts for most of the invertebrate biomass[13]. One square meter is covered by approximately 1,100–2,500 adult individuals corresponding to a total population of 250,000 to 500,000 individuals. This population could consume between 0.5 and 1.0 mol H_2 m^{-2} h^{-1}, that is, in total up to 200 mol H_2 h^{-1}. Previous studies have reported hydrogen uptake rates in the range of several nmol ml^{-1} h^{-1} in hydrothermal fluids[34], and several nmol m^{-3} h^{-1} in coastal waters[35]. On the basis of the limited data available for comparison, the symbionts of *Bathymodiolus* mussels from the MAR probably have a role as a significant hydrogen sink at these hydrothermal vents.

We do not know if the chemoautotrophic symbionts of vent mussels from other mid-ocean ridges are actively consuming hydrogen, although we show here that *B.* aff. *thermophilus* from the Pacific–Antarctic Ridge has the key gene necessary. Hydrogen is clearly present in the fluids at many basalt-hosted vents[36,37], and could well be powering biomass production at these sites as well. In addition, use of hydrogen as an energy source may also have a role in other chemoautotrophic symbioses; the episymbionts of the hydrothermal vent shrimp *Rimicaris exoculata* from MAR vent fields have *hupL* genes[38] (Supplementary Fig. 3), as does the endosymbiont of the giant tubeworm *Riftia pachyptila* from basalt-hosted vents in the East Pacific, which it expresses *in situ* (S. Markert, personal communication). This indicates that hydrogen use could be widespread in chemosynthetic symbioses.

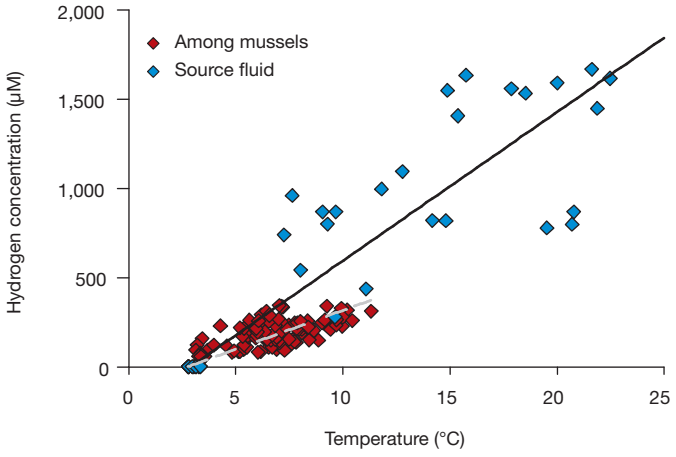

Figure 3 | Hydrogen is consumed in mussel beds of *B. puteoserpentis*. Plot of hydrogen concentrations versus temperature, measured by *in situ* mass spectrometry in the source fluid and in a *B. puteoserpentis* mussel bed. The unbroken line shows linear regression analysis for the source fluid ($y = 83.308x - 240$, $R^2 = 0.63$), the broken line shows linear regression analysis for the mussel bed fluid ($y = 43.576x - 120$, $R^2 = 0.56$). The difference in the slopes of the regression lines between the source fluid and in the mussel bed is due to the consumption of hydrogen in the mussel bed.

METHODS SUMMARY

A detailed description of the sampling sites (Supplementary Tables 1 and 2) and all methods used in this study can be found in the Supplementary Information.

For amplification of the *hupL* gene, we extracted DNA from gill samples[39] and used published primers[40]. Cloning and sequencing were done as described elsewhere[41]. Maximum likelihood and parsimony HupL phylogenies were calculated in ARB[42] using a MAFFT[43] alignment. For incubation experiments, *Bathymodiolus* mussels were collected from hydrothermal vent fields on the MAR by remotely operated vehicles and incubated immediately after recovery on board the research vessel. Mussel gill and foot tissues were incubated in sterile-filtered seawater in glass serum vials (Supplementary Table 4). Hydrogen was added to the vials, and the change in headspace concentration was measured over time by gas chromatography. CO_2 fixation was measured by incubation of gill tissues with ^{14}C-bicarbonate after adding H_2, Na_2S (for H_2S/HS^-), or no electron donor. The incorporation of labelled ^{14}C uptake over time was measured by liquid scintillation counting in separate tissue pieces from the same individual that were incubated for 0, 30, 60 and 120 min. Mussels for genome sequencing were collected from the Lilliput vent site on the MAR. Gill tissue was immediately homogenized and fixed on board. We enriched for symbiont cells by filtering sequentially through 8, 5, 3 and 2 μm GTTP filters. We extracted DNA[39] and constructed a 6 kb paired-end library, which was sequenced by 454-Titanium and Illumina. Titanium reads were assembled using Newbler (v.2.3-PreRelease-10/19/2009–454; Life Sciences Corporation). One 726 kb scaffold was identified that contained the hydrogenase operon, and this scaffold was annotated by RAST[44]. Polynucleotide probe design and geneFISH were done as previously described[31,45], with modifications listed in Supplementary Information (Supplementary Tables 5 and 6). Immunohistochemistry was done by microwaving gill sections of *B. puteoserpentis*, blocking in western blocking reagent (Roche), incubating in primary antibody, then in HRP-conjugated secondary antibody. Signal amplification was done by CARD[46]. The ISMS was deployed as previously described[47] with modifications listed in Supplementary Information. Statistical analysis was done with the software JMP5.

Received 15 April; accepted 20 June 2011.

1. Corliss, J. B. *et al.* Submarine thermal springs in the Galapagos Rift. *Science* **203,** 1073–1083 (1979).
2. Cavanaugh, C. M., Gardiner, S. L., Jones, M. L., Jannasch, H. W. & Waterbury, J. B. Prokaryotic cells in the hydrothermal vent tube worm *Riftia pachyptila* Jones: possible chemoautotrophic symbionts. *Science* **213,** 340–342 (1981).
3. Felbeck, H. Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila* Jones (Vestimentifera). *Science* **213,** 336–338 (1981).
4. Childress, J. J. *et al.* A methanotrophic marine molluscan (Bivalvia, Mytilidae) symbiosis: mussels fueled by gas. *Science* **233,** 1306–1308 (1986).
5. Cavanaugh, C. M., Levering, P. R., Maki, J. S., Mitchell, R. & Lidstrom, M. E. Symbiosis of methylotrophic bacteria and deep-sea mussels. *Nature* **325,** 346–348 (1987).
6. Dubilier, N., Bergin, C. & Lott, C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nature Rev. Microbiol.* **6,** 725–740 (2008).
7. Tivey, M. K. Generation of seafloor hydrothermal vent fluids and associated mineral deposits. *Oceanography (Wash. D.C.)* **20,** 50–65 (2007).
8. Fisher, C. R., Takai, K. & Le Bris, N. Hydrothermal vent ecosystems. *Oceanography (Wash. D.C.)* **20,** 14–23 (2007).
9. Takai, K., Nakagawa, S., Reysenbach, A.-L. & Hock, J. In *Back-Arc Spreading Systems—Geological, Biological, Chemical, and Physical Interactions* (eds Christie, D. M. *et al.*) 185–213 (American Geophysical Union, 2006).
10. Perner, M. *et al.* The influence of ultramafic rocks on microbial communities at the Logatchev hydrothermal field, located 15° N on the Mid-Atlantic Ridge. *FEMS Microbiol. Ecol.* **61,** 97–109 (2007).
11. Schmidt, K., Koschinsky, A., Garbe-Schönberg, D., de Carvalho, L. M. & Seifert, R. 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 (2007).
12. Amend, J. P. & Shock, E. L. Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. *FEMS Microbiol. Rev.* **25,** 175–243 (2001).
13. Gebruk, A. V., Chevaldonné, P., Shank, T., Lutz, R. A. & Vrijenhoek, R. C. Deep-sea hydrothermal vent communities of the Logatchev area (14° 45′ N, Mid-Atlantic Ridge): diverse biotopes and high biomass. *J. Mar. Biol. Assoc. UK* **80,** 383–393 (2000).
14. Duperron, S. *et al.* 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 (2006).
15. Petersen, J. M. & Dubilier, N. Methanotrophic symbioses in marine invertebrates. *Environ. Microbiol. Rep.* **1,** 319–335 (2009).
16. Wendeberg, A., Zielinski, F. U., Borowski, C. & Dubilier, N. Expression patterns of mRNAs for methanotrophy and thiotrophy in symbionts of the hydrothermal vent mussel *Bathymodiolus puteoserpentis. ISME J.* doi:10.1038/ismej.2011.81 (7 July 2011).
17. Vignais, P. M. & Billoud, B. Occurrence, classification, and biological function of hydrogenases: an overview. *Chem. Rev.* **107,** 4206–4272 (2007).
18. Bernhard, M., Schwartz, E., Rietdorf, J. & Friedrich, B. The *Alcaligenes eutrophus* membrane-bound hydrogenase gene locus encodes functions involved in maturation and electron transport coupling. *J. Bacteriol.* **178,** 4522–4529 (1996).
19. Meyer, O. & Schlegel, H. G. Reisolation of the carbon monoxide utilizing hydrogen bacterium *Pseudomonas carboxydovorans* (Kistner) comb. nov. *Arch. Microbiol.* **118,** 35–43 (1978).
20. Schwartz, E. & Friedrich, B. in *The Prokaryotes: A Handbook on the Biology of Bacteria* (eds Dworkin, M. *et al.*) Vol. 2, 496–563 (Springer, 2006).
21. Nelson, D. C., Hagan, K. D. & Edwards, D. B. The gill symbiont of the hydrothermal vent mussel *Bathymodiolus thermophilus* is a psychrophilic, chemoautotrophic, sulfur bacterium. *Mar. Biol.* **121,** 487–495 (1995).
22. Haase, K. M. *et al.* Diking, young volcanism and diffuse hydrothermal activity on the southern Mid-Atlantic Ridge: the Lilliput field at 9° 33′ S. *Mar. Geol.* **266,** 52–64 (2009).
23. Haase, K. M. *et al.* Young volcanism and related hydrothermal activity at 5° S on the slow-spreading southern Mid-Atlantic Ridge. *Geochem. Geophys. Geosys.* **8,** Q11002 (2007).
24. Friedrich, B. & Schwartz, E. Molecular biology of hydrogen utilization in aerobic chemolithotrophs. *Annu. Rev. Microbiol.* **47,** 351–383 (1993).
25. Zielinski, F. U. *et al.* Widespread occurrence of an intranuclear bacterial parasite in vent and seep bathymodiolin mussels. *Environ. Microbiol.* **11,** 1150–1167 (2009).
26. Ohmura, N., Sasaki, K., Matsumoto, N. & Saiki, H. Anaerobic respiration using Fe^{3+}, S^0, and H_2 in the chemolithoautotrophic bacterium *Acidithiobacillus ferrooxidans. J. Bacteriol.* **184,** 2081–2087 (2002).
27. Imhoff, J. F., Hiraishi, A. & Sühling, J. in *Bergey's Manual of Systematic Bacteriology* (eds Brenner, D. J. *et al.*) Vol. 2, part A, 119–132 (Springer, 2005).
28. DiSpirito, A. A., Kunz, R. C., Choi, D.-W. & Zahn, J. A. in *Respiration in Archaea and Bacteria.* (ed. Zannoni, D.) Vol. 2, 149–168 (Springer, 2004).
29. Olson, J. W. & Maier, R. J. Molecular hydrogen as an energy source for *Helicobacter pylori. Science* **298,** 1788–1790 (2002).
30. Bowien, B. & Schlegel, H. G. Physiology and biochemistry of aerobic hydrogen-oxidizing bacteria. *Annu. Rev. Microbiol.* **35,** 405–452 (1981).
31. Moraru, C., Lam, P., Fuchs, B. M., Kuypers, M. M. M. & Amann, R. GeneFISH—an *in situ* technique for linking gene presence and cell identity in environmental microorganisms. *Environ. Microbiol.* **12,** 3057–3073 (2010).
32. Constant, P., Piossant, L. & Villemur, R. Tropospheric H_2 budget and the response of its soil uptake under the changing environment. *Sci. Total Environ.* **407,** 1809–1823 (2009).
33. Tromp, T., Shia, R.-L., Allen, M., Eiler, J. M. & Yung, Y. L. Potential environmental impact of a hydrogen economy on the stratosphere. *Science* **300,** 1740–1742 (2003).
34. Perner, M., Petersen, J. M., Zielinski, F., Gennerich, H. H. & Seifert, R. Geochemical constraints on the diversity and activity of H_2-oxidizing microorganisms in diffuse hydrothermal fluids from a basalt- and an ultramafic-hosted vent. *FEMS Microbiol. Ecol.* **74,** 55–71 (2010).
35. Punshon, S., Moore, R. M. & Xie, H. Net loss rates and distribution of molecular hydrogen (H_2) in mid-latitude coastal waters. *Mar. Chem.* **105,** 129–139 (2007).
36. Welhan, J. A. & Craig, H. in *Hydrothermal Processes at Seafloor Spreading Centers* (eds Rona, P. A. *et al.*) 391–410 (Plenum, 1983).
37. Lilley, M. D., DeAngelis, M. A. & Gordon, L. I. CH_4, H_2, CO and N_2O in submarine hydrothermal vent waters. *Nature* **300,** 48–50 (1982).
38. Hügler, M., Petersen, J. M., Dubilier, N., Imhoff, J. F. & Sievert, S. M. Pathways of carbon and energy metabolism of the epibiotic community associated with the deep-sea hydrothermal vent shrimp *Rimicaris exoculata. PLoS ONE* **6,** e16018 (2011).
39. Zhou, J. Z., Bruns, M. A. & Tiedje, J. M. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62,** 316–322 (1996).
40. Csáki, R., Hanczár, T., Bodrossy, L., Murrell, J. C. & Kovács, K. L. Molecular characterization of structural genes coding for a membrane bound hydrogenase in *Methylococcus capsulatus* (Bath). *FEMS Microbiol. Lett.* **205,** 203–207 (2001).
41. Petersen, J. M. *et al.* 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 (2010).
42. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32,** 1363–1371 (2004).
43. Katoh, K., Asimenos, G. & Toh, H. Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* 39–64 (2009).
44. Aziz, R. K. *et al.* The RAST server: Rapid annotations using subsystems technology. *BMC Genomics* **9,** 75 (2008).
45. Moraru, C., Moraru, G., Fuchs, B. M. & Amann, R. Concepts and software for a rational design of polynucleotide probes. *Environ. Microbiol. Rep.* **3,** 69–78 (2011).
46. Pernthaler, A., Pernthaler, J. & Amann, R. Fluorescence *in situ* hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl. Environ. Microbiol.* **68,** 3094–3101 (2002).
47. Wankel, S. D. *et al.* Influence of subsurface biosphere on geochemical fluxes from diffuse hydrothermal fluids. *Nature Geosci.* **4,** 461–468 (2011).
48. Le Pennec, M. & Hily, A. Anatomie, structure et ultrastructure de la branchie d'un Mytilidae des sites hydrothermeaux du Pacifique oriental. *Oceanol. Acta* **7,** 517–523 (1984).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the chief scientists, and the captains and crews of the research vessels and remotely operated vehicles involved in sampling and analyses at sea. Thank you to D. Garbe-Schönberg, K. van der Heijden and J. Stecher for on-board sampling and analysis, and S. Duperron, M.-A. Cambon-Bonavita and M. Zbinden for providing samples. We acknowledge B. Friedrich and O. Lenz for the antiserum against the *C. necator* uptake hydrogenase, J. Milucka for help with western blots and T. Holler for culturing *C. necator*. S. Wetzel provided technical assistance. This work was supported by the German Science Foundation (DFG) Priority Program 1144 "From Mantle to Ocean: Energy-, Material- and Life Cycles at Spreading Axes" (publication number 60), the DFG Cluster of Excellence "The Ocean in the Earth System" at MARUM (Center for Marine Environmental Sciences), and the Max Planck Society.

Author Contributions J.M.P., F.U.Z., T.P., R.S., C.B., D.F. and N.D. did the on-board experiments during the research cruises. F.U.Z. analysed the data from physiology experiments. J.M.P. amplified and sequenced *hupL*, analysed the genome data, did western blots and immunohistochemistry. C.M. and R.A. did the geneFISH. S.H., S.D.W. and P.R.G. did the *in situ* mass spectrometry and analysed the data. V.B. and E.P. did the genome sequencing and assembly. W.B. did the thermodynamic modelling. J.M.P., F.U.Z. and N.D. conceived the study and wrote the paper.

Author Information All *hupL* sequences have been deposited at NCBI under accession numbers FR851255–FR851274. The sequences that make up the genome fragment and the RAST annotation can be found at NCBI under project identification 65421 (accession numbers CAEB01000001–CAEB01000078). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.D. (ndubilie@mpi-bremen.de).

Hankeln *et al. BMC Bioinformatics* 2010, **11**:358
http://www.biomedcentral.com/1471-2105/11/358

BMC Bioinformatics

SOFTWARE Open Access

MetaBar - a tool for consistent contextual data acquisition and standards compliant submission

Wolfgang Hankeln[1,2], Pier Luigi Buttigieg[1,2], Dennis Fink[3], Renzo Kottmann[1], Pelin Yilmaz[1,2] and Frank Oliver Glöckner*[1,2]

Abstract

Background: Environmental sequence datasets are increasing at an exponential rate; however, the vast majority of them lack appropriate descriptors like sampling location, time and depth/altitude: generally referred to as metadata or contextual data. The consistent capture and structured submission of these data is crucial for integrated data analysis and ecosystems modeling. The application MetaBar has been developed, to support consistent contextual data acquisition.

Results: MetaBar is a spreadsheet and web-based software tool designed to assist users in the consistent acquisition, electronic storage, and submission of contextual data associated to their samples. A preconfigured Microsoft® Excel® spreadsheet is used to initiate structured contextual data storage in the field or laboratory. Each sample is given a unique identifier and at any stage the sheets can be uploaded to the MetaBar database server. To label samples, identifiers can be printed as barcodes. An intuitive web interface provides quick access to the contextual data in the MetaBar database as well as user and project management capabilities. Export functions facilitate contextual and sequence data submission to the International Nucleotide Sequence Database Collaboration (INSDC), comprising of the DNA DataBase of Japan (DDBJ), the European Molecular Biology Laboratory database (EMBL) and GenBank. MetaBar requests and stores contextual data in compliance to the Genomic Standards Consortium specifications. The MetaBar open source code base for local installation is available under the GNU General Public License version 3 (GNU GPL3).

Conclusion: The MetaBar software supports the typical workflow from data acquisition and field-sampling to contextual data enriched sequence submission to an INSDC database. The integration with the megx.net marine Ecological Genomics database and portal facilitates georeferenced data integration and metadata-based comparisons of sampling sites as well as interactive data visualization. The ample export functionalities and the INSDC submission support enable exchange of data across disciplines and safeguarding contextual data.

Background

The technological advancement in molecular biology facilitates investigations of biodiversity and functions on a temporal and geospatial scale. Improved sampling and laboratory methods, together with fast and affordable sequencing technologies [1], provide the framework to create a network of data points capable to answer basic ecological questions such as: 'Who is out there?' and 'What are these organisms doing?' To shed light on the complex interplay, adaptation and survival mechanisms of organisms in times of global change, contextual data describing the surrounding environment of sampling locations are of crucial importance [2]. At the very least, the latitude and longitude (x, y), the depth/altitude (z) in relation to sea level, and the sampling date and time (t) must be provided to allow anchoring molecular sequence data to their environmental context. If every sequence entry in the INSDC databases, comprising of DDBJ, EMBL and GenBank, would be thus georeferenced, researchers would have the post factum opportunity to contextualize these sequences with environmental data [3]. The power of contextual data enriched sequence data sets for the environmental and medical field has been recently documented [4-12].

* Correspondence: fog@mpi-bremen.de
[1] Microbial Genomics Group, Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, D-28359 Bremen, Germany
Full list of author information is available at the end of the article

Unfortunately, a survey in the EMBL sequence repository has shown that only a minor set of sequences are accompanied by a relevant amount of contextual data. For example, latitude, longitude (INSDC: lat_lon), and time (INSDC: collection_date), elements of the key contextual data tuple (x,y,z,t), are only reported in 7.3% and 7.2% of all submissions [Guy Cochrane, personal communication, October 2009]. But even if these data are available, correctness is not guaranteed.

The paucity of sequence associated contextual data has been recognized by the primary database providers and biocuration efforts are currently underway for specific subsets. The National Center for Biotechnology Information (NCBI), for example, curates the Reference Sequence (RefSeq) database which aims to provide a comprehensive, non-redundant, well-annotated set of sequences, including genomic DNA, transcripts and proteins http://www.ncbi.nlm.nih.gov/RefSeq/. The European Molecular Biology Laboratory (EMBL) provides the UniProt/Swiss-Prot Knowledgebase which focuses on high quality protein sequence annotations http://www.ebi.ac.uk/uniprot/ [13]. However, the common aim of these efforts is to enhance the quality of the sequence or protein data and annotations rather than to provide more information on the data processing or the environment where the sample or organism has been taken.

To improve the quantity and quality of contextual data describing the environment of a sample is currently addressed by several projects which systematically collect georeferenced sequence data, environmental parameters, and further curated metadata [14]. SILVA [15] or RDP II [16] are examples for specialized databases that offer users curated and quality checked ribosomal RNA sequences that are often enriched with more reliable contextual and taxonomic information than originally annotated by the sequence submitters.

Furthermore, there are projects which curate the contextual data associated to the primary sequence data to facilitate specific analysis purposes. For example the Genomes OnLine Database (GOLD) collects metadata for ongoing and completed genome sequencing projects [17]. The Visualization and Analysis of Microbial Population Structures (VAMPS) project, with its integrated collection of tools for researchers, aims to visualize and analyze data for microbial population structures and distributions. All the contextual data in VAMPS comes from the MICROBIS database management system of the International Census of Marine Microbes (ICoMM: http://icomm.mbl.edu/microbis/). The megx.net portal http://www.megx.net[18] systematically integrates environmental parameters and sequence data of marine microbial genomes and metagenomes using georeferencing as an anchor.

In 2005 the international Genomic Standards Consortium (GSC) introduced checklists to promote standardized contextual data acquisition and storage. So far the Minimum Information about a Genome (Metagenome) Sequence (MIGS/MIMS) has been published [2] and the Minimum Information about an Environmental Sequence (MIENS) is in development http://gensc.org/gc_wiki/index.php/MIENS. For data exchange, the Genomic Contextual Data Markup Language (GCDML) [19] has been developed. A corollary of these ongoing efforts is the need to support field scientists in the consistent capture, storage and submission of both contextual and sequence data. Handlebar, a lightweight Laboratory Information Management System (LIMS) for the management of barcoded samples, in part addresses this issue by supporting the acquisition and processing of contextual data compliant with GSC standards [20]. The Barcode of Life Database (BOLD) initiative, which aims to identify and classify all eukaryotic life on Earth [21], also includes an advanced data acquisition and submission system. Unfortunately the system only supports phylogenetic markers which serve the eukaryotic domain e.g. the cytochrome c oxidase I (COI), which is only present in Eukarya and absent in the other domains of life, and so far exclude Archaea and Bacteria. Furthermore, it does not support the printing of database identifiers as barcodes to label collected samples.

Even though initiatives and tools exist to enhance the quantity and quality of contextual data subsequent to sequence submission, the amount of contextual data in the INSDC databases remains an issue. In summary, the most likely reasons for the persisting scarcity of consistent contextual data are: (1) Contextual data that are recorded in the field are often not stored electronically in structured databases. Consequently contextual data get rapidly unlinked from the sequence data and finally 'forgotten' in the sequence submission process. (2) There is a lack of automatic quality checking mechanisms active before data submission. Unfortunately, the flood of data entering the public databases prevents any manual curation process. (3) The sheer amount of potential contextual data with respect to the different fields of research ranging from textual data to images or even videos would rapidly exceed the capacities of the INSDC databases. Consequently, only a commonly agreed and standardized subset of data can be stored and made available.

Here the user-centric, web-based tool MetaBar is presented. MetaBar offers all the required features for sample identification and barcode labeling allowing robust sample tracking and inventorying. MetaBar is focused on the acquisition of contextual data recorded during sampling in the field 'offline' using spreadsheets. All recorded contextual data can be subsequently uploaded and consistently stored in an underlying database. The web

Graphical User Interface (GUI) provides advanced user management and access to data and barcodes. Vitally, the tool captures GSC standards compliant data and it is integrated into a set of tools to facilitate further data usage such as integration, visualization and analysis available from the Marine Ecological Genomics database and portal, megx.net. Finally, MetaBar supports contextual data enriched sequence submission to the INSDC databases. The tool is not restricted to any given research field or domain of life, but can universally be applied to capture the contextual data of any biological sample. It is designed to support the complete workflow from the sampling event up to the sequence submission to an INSDC database.

Implementation

Programming languages, tools and frameworks

MetaBar is programmed in the object-oriented, platform-independent programming language, Java 1.5 http://www.java.com/en/. MetaBar is a multiuser web application using Apache Tomcat http://tomcat.apache.org/, the open source Spring framework http://www.springsource.org/about, jasig CAS http://www.jasig.org/cas, which is used as a central authentication service to implement the user management, and Apache POI http://poi.apache.org/ to parse the Microsoft® Excel® spreadsheets used for data input. Any Java objects generated are stored in a PostgreSQL database using the iBATIS persistence framework http://ibatis.apache.org/. The input fields in the Microsoft® Excel® spreadsheet are validated using Visual Basic (VBA) macros. The web interface has been continuously tested during development using Selenium IDE http://seleniumhq.org/projects/ide/ and JUnit http://www.junit.org/. The source code of the MetaBar application is available under the GNU GPL3 http://www.megx.net/metabar/data/metabar-1.0.tar.gz and as additional file 1 to this publication.

Core software components

The MetaBar application consists of (1) the Microsoft® Excel® acquisition spreadsheet which is used to capture and auto-correct the contextual data, (2) the MetaBar server which generates and receives the acquisition spreadsheets, parses the data and stores them in (3) MegDB, a PostgreSQL database which is the central database of the megx.net portal [18].

External software components

MetaBar is integrated into a set of external tools directly accessible from the web interface. The interpolation of environmental physical and chemical parameters of the oceans can be initiated via the WOA05 data extractor of the megx.net portal. On the fly visualization of sampling sites on a world map can be performed using the Genes Mapserver http://www.megx.net/gms and in Google Earth® via the KML export function. The data can be exported prior to sequence submission as a structured comment block for submission to INSDC by the Sequin tool http://www.ncbi.nlm.nih.gov/Sequin/index.html. MetaBar also includes a data export to GCDML [19] for report creation and data exchange.

Results

The core application can be best explained by describing the workflow across the different MetaBar components (Figure 1). First, users log on to the MetaBar web server (Figure 1, step 1). Upon entry, users can allocate a certain range of sample identifiers before, during or after a sampling campaign. The identifier consists of a six digit [sample-id] that is incremented with every new sample, a six digit [project-id] that is incremented with every new project and a two digit [institute-id] that is fixed and identifies a certain institute. The combination of these three parts in one identifier assures the unique identification of each sample. The identifiers can be printed (Figure 1, step 2) as barcodes onto labels (Figure 2) that can be placed on sample containers and pasted into laboratory notebooks for consistency. Users can download (Figure 1, step 3) the acquisition spreadsheet containing the allocated identifiers and the empty contextual data fields in the first worksheet. As the user fills these fields, VBA validation macros check the inputs and users are prompted to use, for example, correct formats in the correct numerical range, where applicable. New worksheets can be

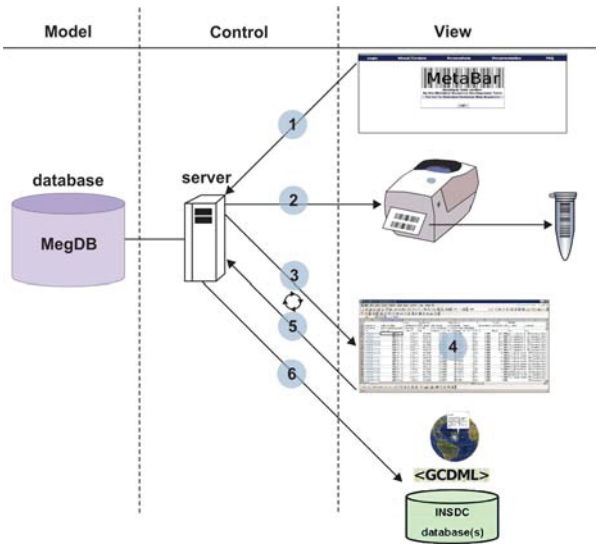

Figure 1 Scheme of the MetaBar workflow. Users are allowed to create barcodes, print them onto physical labels, and to capture, upload, and update contextual data for the barcodes using Excel® spreadsheets. The contextual data can be exported to various formats and submitted to the INSDC databases.

Part:	[institute-ID]	[project-ID]	[sample-ID]
Length:	2 digits	6 digits	6 digits
Example:	01	000001	000001

Figure 2 The barcode identifier. Barcodes consist of three parts ([institute-ID], [project-ID] and [sample-ID]) which in combination uniquely identify a sample. The barcodes are printed onto physical labels that can be placed on sample containers.

Figure 4 Screenshot of the graphical user interface II. Contextual data entries can be extended with GSC parameters. All contextual data can be exported and submitted to the INSDC databases.

added to the spreadsheet. Thus, any additional data outside of the MetaBar model can be added to the same file. Once the worksheets are filled (Figure 1, step 4) they can be uploaded (Figure 1, step 5) to the MetaBar web server. After the upload is finished, the file is parsed and the values in the first worksheet are stored in the respective relational fields of the central database. The additional worksheets in the file are not lost, but stored as binary data in the database. The latter three steps can be repeated whenever it is necessary to edit and update the data. Users can log in to the system at any time to search and browse their data via the web GUI (Figure 3).

Additionally, the MetaBar core set of contextual data fields can be extended for each sample with further GSC compliant parameters. These additional fields are organized into different types of report and environmental packages, each containing further parameters. The parameters can be directly selected and updated via the web interface (Figure 4).

MetaBar can also be used as an inventory e.g. for freezer contents. The database may be queried using sample identifiers by scanning their barcodes with an appropriate device, by manually entering their corresponding numeric code, or by text search on a metadata field. The query then retrieves all corresponding contextual data stored in the system.

MetaBar is integrated into a set of external tools with direct access from the web interface. The interpolation of physical and chemical parameters such as temperature, nitrate, phosphate, salinity, silicate, dissolved oxygen, oxygen saturation, apparent oxygen utilization and chlorophyll of a marine sampling site can be initiated via the WOA05 data extractor of the megx.net portal. On the fly visualization of sampling sites on a world map can be performed using the Genes Mapserver http://www.megx.net/gms. Furthermore, four export functions (Figure 1, step 6) are currently supported: (1) an export to KML to visualize sampling sites including their contextual data in Google Earth®, (2) an export to GCDML [19] for report creation and for data exchange, (3) an export to a GSC compliant MIGS/MIMS/MIENS spreadsheet, and (4) an export as a structured comment for sequence data submission to the INSDC databases using the Sequin tool http://www.ncbi.nlm.nih.gov/Sequin/index.html.

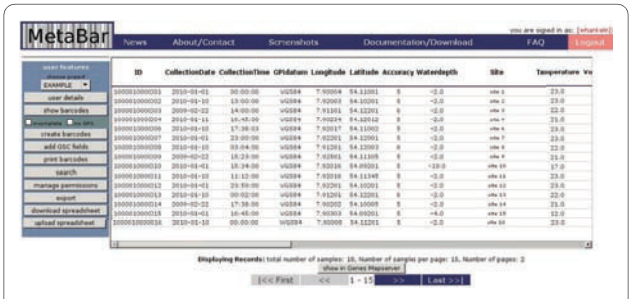

Figure 3 Screenshot of the graphical user interface I. Uploaded contextual data can be browsed and queried online.

Role, ownership and permission concept

MetaBar's user management provides a "MetaBar admin", a "project admin" and a "MetaBar user" mode. These modes depend on the role that is assigned to a certain user account. The web GUI possesses a cascading menu on the left which contains the "MetaBar admin features", the "project admin features", and the "MetaBar user features", respectively. Furthermore, a sophisticated ownership and permission concept offers the users to share their data with other users in the same project giving them read or write permissions, or to prevent access for others. Project admins have the possibility to transfer the ownership of a set of samples to another user, create projects, assign users to a project and to remove users from a project. For a given MetaBar installation there is only one MetaBar admin who can create users, assign or dismiss project admins and delete samples or whole projects.

A quick reference guide describing the general workflow of MetaBar from the acquisition to the submission of data is available on the website. Examples for metadata enriched INSDC database entries, created with MetaBar, are available through the accession numbers: [GenBank:GU949561 and GenBank:GU949562].

Discussion

MetaBar can be used whenever it is necessary to capture contextual data that describe the environmental origin of a sample. The system has been tested in several studies in close collaboration with biologists taking samples in the field. By integrating their feedback MetaBar should qualify as user-friendly, scientist-centric software tool.

Case study

The above mentioned studies allow the typical MetaBar workflow to be generalized as follows. A scientist acting as the project administrator (PA) plans a sampling campaign together with two members in his research team (PM1, PM2). The project EXAMPLE is created in MetaBar and the users PM1 and PM2 are added to EXAMPLE. It is anticipated that PM1 will collect five push core samples of sediment while PM2 will collect five water samples. Thus, PM1 and PM2 create five barcodes each and download their acquisition spreadsheets. These barcodes are printed in multiple copies to label sample containers, appear in field notebooks and for contingency.

During sample collection, PM1 and PM2 log contextual data such as latitude, longitude, depth and sampling time next to the barcode labels pasted in their field notebooks. The sample containers are labeled with the corresponding barcodes and transported back to the laboratory. The link between sample and environment is thus established. At the end of the sampling campaign the contextual data gathered in the field are transferred to PM1 and PM2's acquisition spreadsheets. Barcodes pasted on the sample, the field records and present in the spreadsheet ensure fidelity and the data are then uploaded to the MetaBar server over the internet. PM1 and PM2 may enter further contextual data specific to their sampling environments by selecting the relevant GSC-compliant metadata packages (e.g. "sediment" and "water", respectively) through the web GUI. The PA and both members of the project can now review the consolidated contextual data for errors or missing values. Corrective action at this stage improves the quality of the data prior to submission.

During laboratory processing, every new subsample is labeled with a copy of the original sample's barcode, preserving the link to the *in situ* sampling event. Native laboratory protocols and practices are otherwise unaffected and are documented in laboratory books. PM1 sequences the genomes of several sediment sample isolates and PM2 sequences microbial metagenomes from the community in the water sample. Congruent to the environmental extensions, GSC packages corresponding to various study types are available. PM1 and PM2 may use the "MIENS culture (miens_c)" and "metagenome (me)" packages, respectively, to record data specific to their study type (Figure 4). PM1 and PM2 receive their genome and metagenome sequences as FASTA files with automatically generated sequence identifiers in the header. The researchers enter these identifiers into the "seqID" field in the acquisition spreadsheet and export the data to a format for submission to INSDC. With this mapping, these contextual datasets can easily be combined with one or more FASTA sequences using a suitable submission tool. The researchers then submit their metadata-enriched sequences from the EXAMPLE project to an INSDC database.

MetaBar implements a neat trade-off between universality and specificity. The export functions assure that the collected data can be publicly stored and shared with the scientific community.

Comparison of MetaBar and Handlebar

The idea of uniquely identifying samples and storing data about these samples in databases is not new and is widely used in many applications and disciplines. However, tools able to capture the contextual data of environmental samples combined with barcode labeling are rare. To our knowledge with the exception of MetaBar, the only open source tool using barcoding to identify georeferenced samples from the environment is Handlebar [20]. A tabular comparison of the programs' general features can be found in Table 1.

HandleBar, as a lightweight LIMS, not only covers contextual data that are recorded during sampling, but also aims to document subsequent sample processing steps in the laboratory. In this respect, MetaBar is a simplification focusing only on the capture of contextual data in the field. MetaBar does not seek to replace well established laboratory bookkeeping or professional LIM systems, but rather aims to complement this process to ensure that contextual data are electronically accessible. Nevertheless, users may choose to use the tool as a storage inventory manager or to store intermediate results of sample processing because it is possible to store additional data in the spreadsheets. It is important to note that coupling contextual data with sequence data before submission to the INSDC databases is a unique feature of MetaBar.

The barcodes are, by concept, solely used to link environmental samples to contextual and, if available, sequence and species data derived from a labeled sample, thus, no hierarchy or processing method is encoded in the identifiers. Also, sample hierarchies and complex identifier schemes are avoided. This concept does not interfere with native laboratory sample tracking methods, yet ensures consistency in environmental contextual data capture.

It is important that users have the flexibility to cover different sample types. MetaBar offers a single template in which a restricted part is parsed to the database and an unrestricted part of the spreadsheet can be changed to

Table 1: Features of Handlebar and MetaBar

	Handlebar	MetaBar
Focus	Web-based lightweight LIMS for handling barcoded samples	Web-based tool for consistent contextual data acquisition with barcoded samples
System requirements	Operating system: Windows® or GNU Linux Apache, Perl, PostgreSQL, OpenOffice or Microsoft® Excel®	Without local MetaBar server installation: Operating system: Windows® Internet connection, web browser (e.g. Firefox), Microsoft® Excel® 2003 or higher Optional: EPL barcode printer (e.g. a Zebra® TLP 2824) With local MetaBar server installation: Operating system: Windows® or GNU Linux Apache, Java, Spring, jasig CAS, PostgreSQL, Microsoft® Excel® 2003 or higher
Coverage	Metadata that emerges during sampling events and subsequent processing step data	Contextual data that emerges during sampling events (other data optional)
Sample type templates	Various	One generic and extensible template
Input validation	Done by the server	Done by VBA® macros in the acquisition spreadsheet and on the server
Integration into data analysis tool set	GenQuery	http://www.megx.net
Export functions	-	GCDML, KML (for Google Earth)
Contextual data enriched sequence submission support	-	Export to MIGS/MIMS/MIENS and structured comment

contain sample specific additional data. HandleBar offers a set of non-constrained sample templates depending on the sample type and also individual templates can be created. In MetaBar each sample can be extended with further parameters organized into types of report and environmental packages suggested by the GSC.

In contrast to HandleBar, data entered into MetaBar's acquisition spreadsheet is validated on input, ensuring correct format before upload to the MetaBar server. This avoids frequent rejection of the acquisition sheet. In HandleBar the validation is done by the web server and erroneous sheets have to be corrected retrospectively by the uploading user. The variety of export features are currently unique to MetaBar.

MetaBar is integrated into the megx.net tool set and connected to MegDB. This offers opportunity to work with the data and to analyze them alone, or in the context of other research project data stored in the megx.net database. This level of integration necessitated a user authentication and authorization management system and SSL encryption. Consequently, the local installation of MetaBar requires modification of the open source code base. The software and a detailed installation manual are available at http://www.megx.net/metabar. However, accounts on the MetaBar installation hosted at the MPI for Marine Microbiology in Bremen can easily be given to interested users and an "anonymous" project exists where data of external users can be stored anonymously. It is the intention of the Microbial Genomics and Bioinformatics Group at the MPI-Bremen to support this tool as open source in the future.

Applicability

MetaBar has been developed at the Max Planck Institute for Marine Microbiology; however, the tool may be readily applied to a wide range of research fields outside the marine sciences. Contextual data fields relevant to air, host associated, human associated, sediment, soil, waste-

water sludge or water samples are available via the "add GSC fields" function. The parameters in each of these environmental packages have been selected based on community usage and consensus http://gensc.org/gc_wiki/index.php/MIENS. For example, fields requesting data on barometric pressure, carbon dioxide, carbon monoxide, chemical administration, humidity, methane, organism count, oxygen, oxygenation status of sample, perturbation, pollutants, respirable particulate matter, sample salinity, sample storage duration, sample storage location, sample storage temperature, solar irradiance, temperature, ventilation rate, ventilation type, volatile organic compounds, wind direction, and wind speed would be presented to users using the air environmental package. Users may easily add new, custom fields as columns using standard Microsoft® Excel® operations. Combined, the GSC extensions and freedom for customization generalize MetaBar's applicability to any scenario necessitating the capture of contextual data describing a sample's environmental origin.

Conclusion

MetaBar offers an integrated contextual data acquisition, storage, and submission solution to the INSDC system. The impact of better contextual data availability and correctness in the primary sequence databases will greatly improve the possibilities to reach a higher level of data integration and interpretation to address basic ecological questions. MetaBar's integration into the megx.net tool set and its export mechanisms offer extended analysis possibilities via comparison to other scientific studies and with complementary interpolated environmental data. The visualization of the sampling sites on the Genes Mapserver and in Google Earth® offers the users a simple way to show sampling events on the globe and to relate them to other publicly available scientific studies.

Statistical analysis of phylogenetic and functional biodiversity in their environmental context will reveal new insights into the biogeography and habitat adaptation of organisms. In the medical field, for example, it will be possible to create detailed disease maps which reveal mutation patterns of a certain pathogenic organism over time [5,11]. Such maps might help to predict the dispersal of epidemics and pandemics around the globe. For marine microbiology, Ed DeLong and coworkers have successfully shown that there is a stratification of genomic variability along the depth continuum in the water column at a specific sampling location [4]. It has also been demonstrated that specific diversity patterns are annually recurring [7]. A dense network of data points, enriched with contextual data, will lead to new insights into the complex interplay of organisms by comparing different sampling sites around the globe and over time. The denser this network of data points, the more will be revealed about the influence of the biotic factor in the elementary nutrient cycles that profoundly affect Earth's climate.

Availability and Requirements

Project name: MetaBar

Software

Project homepage: http://www.megx.net/metabar

Operating systems: Linux and Windows

Programming language: Java JRE 1.5 or higher

Other requirements: Microsoft® Excel® 2003 or higher, Google Earth® (optional)

License: GNU General Public License version 3 (GNU GPL3)

Hardware

At least 1024 Mb of RAM

EPL barcode printer (e.g. a Zebra TLP 2824) (optional)

Barcode handscanner (optional)

The software can be tested anonymously using the login: "anonymous" with the password: "testmetabar".

Additional material

Additional file 1 MetaBar open source code base under GNU GPL3 license. Zipped open source code base for local installation. A general installation manual can be found in the README.txt.

Authors' contributions

WH developed and implemented MetaBar and wrote the manuscript. RK advised programming design and helped with the integration of MetaBar with MegDB and megx.net. DF tested the software on cruises and provided feedback for design improvements. PY assured the MIGS/MIMS/MIENS standard compliance in MetaBar. PLB critically revised the manuscript and took care of EnvO integration. PY, PLB and RK tested the tool in the field. FOG supervised the work and helped with writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Thanks to Jens Harder, Christine Klockow, Mirja Meiners and all the testers. Thanks to Dawn Field and Tim Booth at NERC, UK for feedback and useful input during the design phase of MetaBar. Thanks to Norma Wendel for her help in finalizing the MetaBar source code. The study was supported by the Max Planck Society.

Author Details

[1]Microbial Genomics Group, Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, D-28359 Bremen, Germany, [2]Jacobs University Bremen gGmbH, D-28759 Bremen, Germany and [3]Symbiosis Group, Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, D-28359 Bremen, Germany

Received: 11 March 2010 Accepted: 30 June 2010
Published: 30 June 2010

References

1. Hall N: **Advanced sequencing technologies and their wider impact in microbiology.** *J Exp Biol* 2007, **210:**1518-1525.
2. Field D, *et al.*: **The minimum information about a genome sequence (MIGS) specification.** *Nat Biotechnol* 2008, **26:**541-547.
3. Wieczorek J: **The point-radius method for georeferencing locality descriptions and calculating associated uncertainty.** *International journal of geographical information science* 2004, **18:**745-767.
4. DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N.-U, Martinez A, Sullivan MB, Edwards R, Brito BR, Chisholm SW, Karl DM:

Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 2006, **311:**496-503.
5. Janies D, Hill AW, Guralnick R, Habib F, Waltari E, Wheeler WC: **Genomic analysis and geographic visualization of the spread of avian influenza (H5N1).** *Syst Biol* 2007, **56:**321-329.
6. Ramette A: **Multivariate analyses in microbial ecology.** *FEMS Microbiol Ecol* 2007, **62:**142-160.
7. Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown MV, Naeem S: **Annually reoccurring bacterial communities are predictable from ocean conditions.** *Proc Natl Acad Sci USA* 2006, **103:**13104-13109.
8. Pommier T, Canbäck B, Riemann L, Boström KH, Simu K, Lundberg P, Tunlid A, Hagström A: **Global patterns of diversity and community structure in marine bacterioplankton.** *Mol Ecol* 2007, **16:**867-880.
9. Parks DH, Porter M, Churcher S, Wang S, Blouin C, Whalley J, Brooks S, Beiko R: **GenGIS: A geospatial information system for genomic data.** *Genome Res* 2009, **10:**1896-1904.
10. Green JL, Bohannan BJM, Whitaker RJ: **Microbial biogeography: from taxonomy to traits.** *Science* 2008, **320:**1039-1043.
11. Schriml LM, Arze C, Nadendla S, Ganapathy A, Felix V, Mahurkar A, Phillippy K, Gussman A, Angiuoli S, Ghedin E, White O, Hall N: **GeMInA, Genomic Metadata for Infectious Agents, a geospatial surveillance pathogen database.** *Nucleic Acids Res* 2009, **38:**D754-D764.
12. Field D: **Working together to put molecules on the map.** *Nature* 2008, **453:**978.
13. Consortium, U: **The Universal Protein Resource (UniProt) in 2010.** *Nucleic Acids Res* 2010, **38:**D142-D148.
14. Howe D, Costanzo M, Fey P, Gojobori T, Hannick L, Hide W, Hill D, Kania PR, Schaeffer M, Pierre SS, Twigger S, White O, Rhee SY: **Big data: The future of biocuration.** *Nature* 2008, **455:**47-50.
15. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO: **SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB.** *Nucleic Acids Res* 2007, **35:**7188-7196.
16. Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, McGarrell DM, Garrity GM, Tiedje JM: **The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis.** *Nucleic Acids Res* 2005, **33:**D294-D296.
17. Liolios K, Chen IMA, Mavromatis K, Tavernarakis N, Hugenholtz P, Markowitz VM, Kyrpides NC: **The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata.** *Nucleic Acids Res* 2009, **38:**D346-D354.
18. Kottmann R, Kostadinov I, Duhaime MB, Buttigieg PL, Yilmaz P, Hankeln W, Waldmann J, Glöckner FO: **Megx.net: integrated database resource for marine ecological genomics.** *Nucleic Acids Res* 2009, **38:**D391-D395.
19. Kottmann R, Gray T, Murphy S, Kagan L, Kravitz S, Lombardot T, Field D, Glöckner FO, Consortium GS: **A standard MIGS/MIMS compliant XML Schema: toward the development of the Genomic Contextual Data Markup Language (GCDML).** *OMICS* 2008, **12:**115-121.
20. Booth T, Gilbert J, Neufeld JD, Ball J, Thurston M, Chipman K, Joint I, Field D: **Handlebar: a flexible, web-based inventory manager for handling barcoded samples.** *Biotechniques* 2007, **42:**300-302.
21. Ratnasingham S, Hebert PDN: **bold: The Barcode of Life Data System.** *Mol Ecol Notes* 2007, **7:**355-364 [http://www.barcodinglife.org].

doi: 10.1186/1471-2105-11-358
Cite this article as: Hankeln *et al.*, MetaBar - a tool for consistent contextual data acquisition and standards compliant submission *BMC Bioinformatics* 2010, **11**:358