Metagenomic analyses of a deep-sea mussel symbiosis

Dissertation

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

– Dr. rer. nat. –

dem Fachbereich Biologie / Chemie

der Universität Bremen

vorgelegt von

Merle Ücker

Bremen

März 2021

Die vorliegende Doktorarbeit wurde von Juni 2017 bis März 2021 in der Abteilung Symbiose am Max-Planck-Institut für Marine Mikrobiologie unter der Leitung von Prof. Dr. Nicole Dubilier und direkter Betreuung von Dr. Rebecca Ansorge und Dr. Yui Sato angefertigt.

The research presented in this thesis was performed from June 2017 until March 2021 at the Max Planck Institute for Marine Microbiology in the Department of Symbiosis under leadership of Prof. Dr. Nicole Dubilier and direct supervision of Dr. Rebecca Ansorge and Dr. Yui Sato.

Gutachter*innen | reviewers

Prof. Dr. Nicole Dubilier

Prof. Dr. Thorsten Reusch

Assist. Prof. Dr. Roxanne Beinart

Tag des Promotionskolloquiums | date of doctoral defence

21 April 2021

Diese zur Veröffentlichung erstellte Version der Dissertation enthält Korrekturen.

To the ocean.

Contents

Summary

Symbiosis is ubiquitous across all domains of life. Deep-sea hydrothermal vents are home to extraordinary examples of symbiosis. Fascinating symbiotic communities are fuelled by reduced chemical compounds released from fissures in the oceanic crust. Chemosynthetic symbionts use the energy of reduced chemicals energy to produce biomass and to support their animal hosts to thrive in environments where nutrients are scarce. *Bathymodiolus* mussels are among the most successful fauna in such habitats. Within their gills, they host sulphur- and methane-oxidising symbionts, among others. These symbiotic bacteria are acquired from the environment, suggesting the existence of a free-living stage. The mussels are well studied for their symbionts' physiology, host-symbiont interaction and the host's immune system. Some interesting questions, however, have remained unresolved. I used metagenomics, a versatile and cultivation-independent approach, to address some of these questions:

What can a mussel hybrid zone reveal about factors driving symbiont composition? Hybrid zones provide an opportune system to study evolutionary processes in their natural context. Analysis of symbionts from co-occurring hybrid and parental mussels at the Broken Spur vent field allowed me to identify whether host genetics, geography, or the environment, is driving the symbiont community composition. Phylogenomics revealed the presence of a new location-specific symbiont subspecies. Symbionts of hybrids and parental mussels could not be distinguished genetically. Thus, host genetics seem to have little influence on the symbiont community. Instead, geography explained much of the observed symbiont variation. Whether the symbiont population structure results from a geographical structuring of the free-living pool of symbionts remains to be elucidated.

Are free-living symbionts present in the water column? Knowledge about the free-living stage of horizontally transmitted symbionts can give insights into the symbiont uptake and the specificity of the association. To investigate the presence of symbionts in the free-living stage, I screened for symbiont marker genes in water metagenomes. While symbiont-related genes were detected, they always co-occurred with host DNA. This raises the question whether the symbionts in my data are free-living or still associated with their hosts. The results suggest that transmission via host particles may be more important than anticipated. To further future research based on the experiences of this work, I suggest sampling schemes to learn more about the free-living stage.

Are mitochondrial and nuclear genomes congruent in Bathymodiolus*?* Species assignment is often performed using mitochondrial marker genes. Mitochondrial inheritance in bivalves is often complex, and incongruent nuclear and mitochondrial genomes have been described for the *Bathymodiolus* hybrid zone. I compared mitochondrial clades to clustering based on the nuclear genome and found incongruences for 10 % of the 175 analysed mussels. The high-resolution analysis further revealed a lack of subpopulation structure in conspecific mussels from different sites. Both findings suggest a strong genetic connectivity of populations at the Mid-Atlantic Ridge, probably enabled by long-distance migration of planktotrophic larvae. The biological processes underlying the mitonuclear discordance are exciting topics for future analyses.

Altogether, the research presented in this thesis enhances our understanding of the symbiotic association in *Bathymodiolus* mussels and provides the basis for further population genomic studies of host, symbionts and the free-living bacterial community.

Zusammenfassung

Symbiosen sind allgegenwärtig. Besondere Beispiele finden sich an Tiefsee-Hydrothermalquellen, wo reduzierte chemische Stoffe aus Rissen der Ozeankruste entweichen und die Grundlage für Gemeinschaften faszinierender symbiotischer Tiere bilden. *Bathymodiolus* Muscheln gehören zu den wenigen Tieren, die in diesem Lebensraum erfolgreich sind. Ihre chemosynthetische Symbionten nutzen chemische Energie, um Biomasse herzustellen. So unterstützen sie ihren tierischen Wirt dabei, in einem Lebensraum zu gedeihen, der sonst lebensfeindlich wäre. Die Sulfat- und Methan-oxidierende Symbionten werden aus der Umwelt aufgenommen, was die Existenz eines freilebenden Stadiums voraussetzt. Viele Studien haben sich mit der *Bathymodiolus* Symbiose auseinandergesetzt, doch einige Fragen blieben bisher ungelöst. Mithilfe von Metagenomik, einem vielseitigen Ansatz, der unabhängig von Kultivierung ist, habe ich mich mit folgenden Fragen beschäftigt:

Welche Faktoren bestimmen die Zusammensetzung von Symbiontengemeinschaften? Hybridzonen bieten die Möglichkeit, evolutionäre Prozesse in ihrem natürlichen Kontext zu erforschen. Anhand einer Analyse der Symbionten von zusammen vorkommenden Hybriden und Eltern im Broken Spur Hydrothermalfeld habe ich untersucht, ob Wirtsgenetik, Geographie oder Umweltfaktoren die Symbiontenzusammensetzung bestimmen. Mit Phylogenomik habe ich eine neue, ortsspezifische Symbiontenunterart entdeckt. Die Symbionten von Hybriden und Eltern unterscheiden sich genetisch nicht. Scheinbar ist die Wirtsgenetik für die Zusammensetzung der Symbiontengemeinschaft weniger wichtig als die Geographie, die einen Großteil der Symbionten-Variation erklärt. Ob die Populationsstruktur der Symbionten durch die geographische Struktur von freilebenden Symbionten bestimmt wird, muss in zukünftigen Studien geklärt werden.

Gibt es freilebende Symbionten in der Wassersäule? In horizontal übertragenen Symbiosen ermöglicht Wissen über das freilebende Stadium Rückschlüsse auf die Symbiontenaufnahme und die Spezifität der Symbiose. Um die Existenz von Symbionten im freilebenden Stadium zu erforschen, habe ich Metagenome von Wasserproben auf Symbionten Markergene untersucht. Dabei wurden Symbionten-bezogene Gene gefunden. Sie traten allerdings immer zusammen mit der DNA des Wirtes auf. Dies lässt vermuten, dass eine Symbiontenübertragung über Wirtspartikel eine wichtigere Rolle spielen könnte als bisher vermutet. Für zukünftige Forschungsprojekte habe ich Vorschläge für die Beprobung erarbeitet, um dem freilebenden Symbiontenstadium weiter auf den Grund zu gehen.

Stimmen mitochondriale und nukleare Genome von Bathymodiolus *überein?* Mitochondrien werden häufig zur Artenbestimmung herangezogen, ihre Vererbung in Muscheln ist aber zum Teil sehr komplex. Außerdem passen nukleare und mitochondriale Genome in der Hybridzone nicht zusammen. Ich habe das mitochondriale mit dem nuklearen Genom verglichen und dabei entdeckt, dass diese in 10 % der untersuchten Muscheln nicht zusammenpassten. Zudem hat die Analyse gezeigt, dass sich Muscheln der gleichen Art von verschiedenen Orten kaum unterscheiden. Beide Ergebnisse deuten darauf hin, dass die Muschelpopulationen am Mittelatlantischen Rücken stark genetisch verbunden sind. Die biologischen Prozesse, die für die mitonuklearen Unterschiede sorgen, bieten ein spannendes Forschungsfeld für die Zukunft.

Zusammenfassend trägt diese Doktorarbeit zu einem besseren Verständnis der *Bathymodiolus* Symbiose bei und bildet die Grundlage für zukünftige populationsgenomische Studien des Wirtes, der Symbionten und der freilebenden Bakteriengemeinschaft.

symbiosis

[ˌsɪmbɪˈəʊsɪs, ˌsɪmbʌɪˈəʊsɪs] *noun*

interaction between two different organisms living in close physical association, typically to the advantage of both

Chapter I | Introduction

Why deep-sea research?

The deep sea represents ~98.5 % of our Earth's volume that can be inhabited by animals [1]. It is the largest but one of the least studied environments on this planet. Although often perceived as an inhospitable environment characterised by darkness, high pressure, and nutrient limitation, the deep sea sustains diverse and abundant forms of life, especially at hydrothermal vents [2–4]. Advances in technology did not only foster scientific deep-sea exploration but also increased economic interests. While the environment in its complexity is still poorly understood, its exploitation by humans through deep-sea fishing, mining, and hydrocarbon extraction is already ongoing [1,5–7]. Basic research on deep-sea animal populations, their evolutionary history, and their contemporary connectivity is now more important than ever to grow our knowledge of this underexplored ecosystem [8]. In this thesis, I investigate aspects of the symbiosis between deep-sea *Bathymodiolus* mussels, the bacterial symbionts and their free-living stage. The aim of my research is to add a piece of knowledge to our understanding of the mysterious deep-sea environment.

## Bacterial communities in chemosynthetic vent environments

Deep-sea hydrothermal vents are globally distributed regions where hot and chemical-rich seawater is released from the seafloor into the deep ocean (Figure 1 A, B). Clusters of vent fields occur at seafloor spreading centres, e.g. along mid-ocean ridges but also in other geological settings [9,10]. Seawater penetrates to the subsurface through cracks in the oceanic crust and reacts with hot rocks. Through this process, water is enriched with reduced chemicals, heavy metals, and dissolved gases, and is released from fissures and orifices in the seafloor [11]. Upon contact with the colder oxic seawater, oxic-anoxic boundaries are created that fuel life in these environments [12]. Life at hydrothermal vents is challenged with fluctuating conditions such as varying concentrations of reduced chemicals in addition to a high probability of habitat extinction and renewal depending on plate tectonics [13–16].

In an environment deprived of light, such as the deep sea, the production of biomass using light energy, i.e. photosynthesis, is not possible. Life has found other ways to sustain itself, namely via chemosynthesis [17]. Microorganisms in these environments are able to use chemical instead of light energy for carbon fixation and biomass production [18]. These microorganisms live along thermal and chemical gradients between the anoxic hydrothermal

fluids and the colder and oxic seawater that generate distinct niches for the organisms and their different metabolisms [12,19–21]. While the ambient seawater in the deep has a temperature of around 2 °C, the fluids emerging from the vents can reach more than 400 °C [22]. Described energy sources for bacterial chemosynthesis include reduced sulphur compounds, hydrogen, and methane [23,24]. Oxygen is often used as terminal electron acceptor. This comes with the challenge that reduced energy sources and oxygen rarely co-occur in nature [25]. However, some chemotrophs are able to respire nitrate [26].

The composition of bacterial communities in hydrothermal vent habitats varies between sites based on the physical and chemical parameters at each site. Yet, some general patterns can be observed in the bacterial communities (reviewed in [27,28]). Gammaproteobacteria, especially those from the SUP05 clade and *Beggiatoa*, dominate the vent community in colder, oxic seawater (2-10 °C). Different taxa from the Epsilonproteobacteria are abundant at cold (*Arcobacter*, *Sulfurimonas* and *Sulfurovum*) to medium temperatures (up to 70 °C, *Caminibacter* and *Nautila*). Hot hydrothermal fluids (> 400 °C) are dominated by (hyper-) thermophiles, e.g. *Thermococcus*, different archaea, and bacteria involved in methanogenesis, such as *Methanococcus*, *Methanocaldococcus,* and members of the Methanopyri and Methanosarcinales. These generalised patterns might lead to the assumption that hydrothermal vent communities are similar between sites. When investigated on a lower taxonomic level, species and strains are often distinct from each other depending on their specific habitat, which is reflected by traits related to their specific niche. For example, *Thermococcus* from sulphide and subseafloor habitats were phylogenetically distinct and had different temperature maxima although exchange of organisms between the habitats was not restricted [29].

One challenge in microbial ecology is to understand the biogeography of microbes and to differentiate between processes leading to similar patterns, e.g. isolation-by-distance compared to isolation-by-environment [30–32]. For marine microorganisms, both environmental selection and dispersal limitation have been described to shape their biogeography depending on which microorganisms are studied [33–38]. Regarding the hydrothermal vent microbiome, it becomes more eminent that it is probably "distributed globally [and] shaped locally" [27], i.e. members of the same taxonomic group are present at hydrothermal vents across the world's oceans but their metabolism is well adapted to their specific local niche.
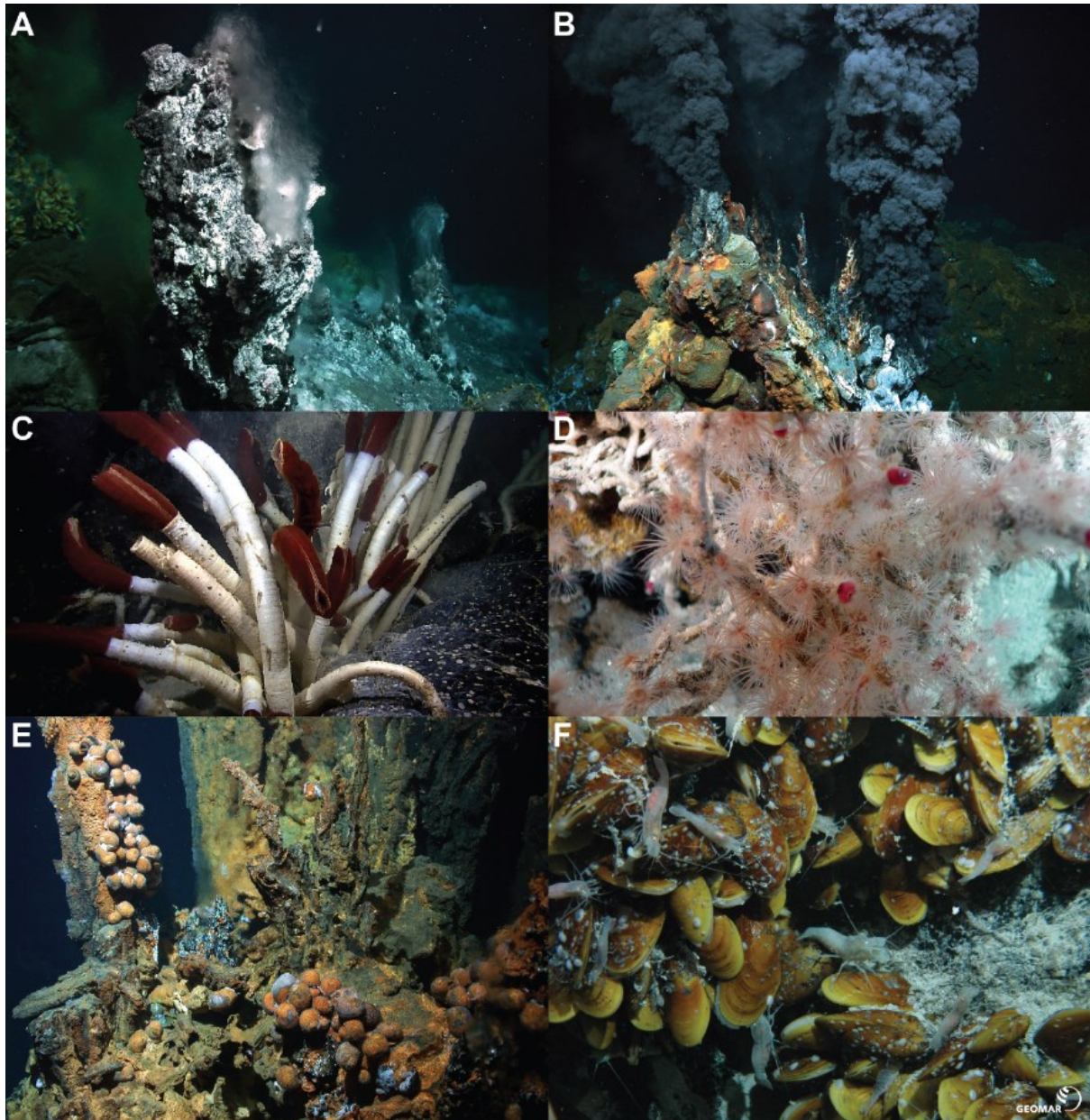
Figure 1 | Hydrothermal vents and associated symbiotic fauna. A: Hydrothermal vent at the Menez Gwen vent field in 860 m water depth. B: Black smoker at almost 3000 m water depth at the Mid-Atlantic Ridge. C: *Riftia* tubeworms from the Galapagos Rift. D: *Lamellibrachia* tubeworm bushes get up to 1 meter tall at the Mariana Arc in the Pacific Ocean. E: *Alviniconcha* snails at a hydrothermal vent in the Bismarck Sea. F: *Bathymodiolus* mussels and *Rimicaris* shrimps at the Logatchev vent field. Image courtesy: A, B & E: MARUM, Centre for Marine Environmental Sciences, University of Bremen (CC-BY 4.0). C: NOAA Okeanos Explorer Program, Galapagos Rift Expedition 2011 (CC-BY 2.0). D: Pacific Ring of Fire 2004 Expedition. NOAA Office of Ocean Exploration; Dr. Bob Embley, NOAA PMEL, Chief Scientist (CC-BY 2.0). F: GEOMAR (CC-BY 4.0).

Symbionts support fascinating animal communities at hydrothermal vents

Life in chemosynthetic environments such as hydrothermal vents is sustained by microorganisms [23]. Nevertheless, fascinating animal communities (Figure 1 C-F) were discovered at hydrothermal vents at the Galapagos Rift in 1977, which caught the scientists on board the manned submersible Alvin by surprise [39,40]. Contrasting the widespread believes that the deep sea resembled a vast desert and that only photosynthesis could support life, they encountered communities of highly specialised animals that can grow to huge sizes and large abundances [41]. Among these species, tubeworms and molluscs serve as foundation fauna that provide habitats for diverse grazers, predators, and decomposers [42]. The reason for the animal's existence in these environments are symbioses with chemosynthetic bacteria.

Symbiosis is a ubiquitous phenomenon occurring across all domains of life [43]. Greek for "living together", symbiosis describes a close interaction between unlike organisms that occurs over a long period of time [44]. Symbiotic associations can be divided into three main forms: (1) mutualism, referring to a symbiosis that benefits both partners, (2) commensalism, in which only one partner benefits but the other is not harmed, or (3) parasitism, where one partner benefits at the cost of the other partner. In this thesis, I will focus only on mutualistic interactions.

Symbiotic association with bacteria can enable a host to thrive in new habitats that would not have been favourable without the symbionts [45–47]. One example are nutritional symbioses. In these symbioses, hosts gain nutrition through their symbionts in environments depleted of carbon sources that are accessible by animals [48] so that energy is transferred between trophic levels [16,49]. In addition, symbiotic bacteria in general can have great impact on animal health and development among other important factors (Box 1). Chemosynthetic symbioses were first discovered at hydrothermal vents, but also occur in a variety of other environments, such as cold seeps, whale and wood falls, and shallow-water environments [50]. These symbioses between animal hosts and chemosynthetic bacteria have evolved convergently in multiple events. Chemosynthetic symbioses are found across different animal phyla including molluscs, nematodes, annelids, flatworms, arthropods, and sponges [25,50,51]. Also on the symbiont side, a range of taxa can be involved in the association. Symbiotic lifestyles of chemosynthetic bacteria has been shown in Proteobacteria, mostly Gamma-, but also Epsilon-, Alpha- and potentially Delta- and Zetaproteobacteria [50,52–57]. For bacterial symbionts in eukaryotic hosts, reduced sulphur compounds, methane, hydrogen,
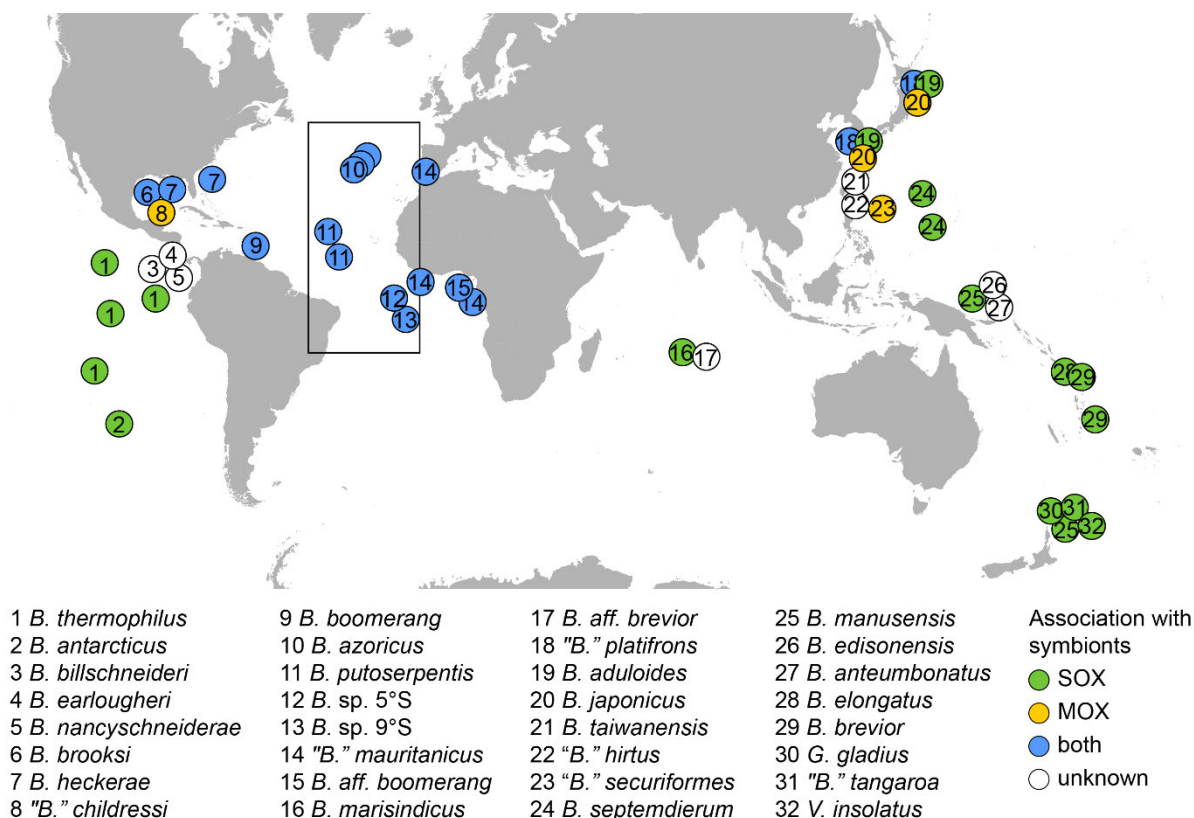
and carbon monoxide are common energy sources [17,46,50,55,58,59]. Some of these symbioses have evolved to a state of complete dependency, in which the hosts have lost mouth and digestive system, e.g. in deep-sea vestimentiferan tubeworms or in shallow-water flatworms [49,60–63]. The benefit for the hosts is obvious: their nutritional requirements are fulfilled by the symbionts. In return, the symbionts are assumed to benefit from a habitat sheltered from the surrounding environment, resulting in reduced competition with other bacteria, a constant supply of energy sources from the host, and protection from predators [48,64,65]. These excellent conditions might improve the symbiont's overall metabolic performance resulting in significant contribution of the symbiosis to ecosystem processes [66].

Box 1 | Animal–microbe symbioses or the power of microbes

Sometimes it is easy to forget how big the influence is that even the smallest organisms have. Microorganisms do no only make the world go round when it comes to global cycling of nitrogen or carbon, or the generation of oxygen on this planet. Microorganisms such as bacteria that are associated with animals also play a crucial role in their host's development, speciation, and evolution. One of the most famous examples for impactful animal–microbe associations is the evolution of mitochondria and chloroplasts that provided the basis for complex and multicellular life [67,68]. Symbioses provide a viable source of evolutionary innovation such as new metabolic capabilities or increased dispersal and mobility [43]. Chemosynthetic bacteria can gain access to both anoxic and oxic conditions for their energy sources and terminal electron acceptors when they are associated to mobile hosts that can bridge greater redox zones [12]. Besides this, many examples have been described, in which infection with bacteria is crucial for animal development, e.g. the development of a functioning immune system [69]. In tsetse flies, bacterial infection during gestation ensures the correct development of the gut lining [70]. In mice, a model system for microbial colonisation in mammals, the gut microbiome has been reported to impact the brain development and behaviour [71]. Throughout an animal's lifetime, its bacterial partners can influence hormone signalling [72], weight changes [73], and health condition [71,74], among others. Moreover, the role of microbiomes in animal speciation has received increasing attention in recent years. A study on *Drosophila* demonstrated how diet-induced changes of the gut microbiome influences the mating preference probably due to altered sex pheromone production [75]. In jewel wasps, incompatibilities between the host genotype and the microbiome composition even led to complete hybrid breakdown that could be cured by antibiotic treatment [76]. Such examples remind us how even the smallest organisms can have a great impact on the animal world.

Distribution, phylogeny, and dispersal of *Bathymodiolus* mussels

Bathymodiolus mussels in the family Mytilidae [77] are one of the key players in hydrothermal vent environments. Their large mussel beds often dominate the vent community biomass and provide habitat and settlement substrate for other organisms, making *Bathymodiolus* mussels important foundation fauna for their ecosystem [42,65]. The distribution of *Bathymodiolus* mussels covers the Atlantic, Pacific, and Indian Oceans, the Gulf of Mexico and the Caribbean (Figure 2). *Bathymodiolus* mussels are present in all environments with reducing characteristics such as wood and whale falls, cold seeps and hydrothermal vents [78–80]. The colonisation of deep-sea environments was probably enabled via stepping stones such as sunken wood or whale carcasses [78,81,82]. The separation from their shallow-water relatives *Mytilus* spp. has been estimated to have occurred around 89 million years ago followed by rapid adaptive divergence [81,82].


1 *B. thermophilus*	9 *B. boomerang*	17 *B. aff. brevior*	25 *B. manusensis*
2 *B. antarcticus*	10 *B. azoricus*	18 *"B." platifrons*	26 *B. edisonensis*
3 *B. billschneideri*	11 *B. putoserpentis*	19 *B. aduloides*	27 *B. anteumbonatus*
4 *B. earlougheri*	12 *B. sp. 5°S*	20 *B. japonicus*	28 *B. elongatus*
5 *B. nancyschneiderae*	13 *B. sp. 9°S*	21 *B. taiwanensis*	29 *B. brevior*
6 *B. brooksi*	14 *"B." mauritanicus*	22 *"B." hirtus*	30 *G. gladius*
7 *B. heckerae*	15 *B. aff. boomerang*	23 *"B." securiformes*	31 *"B." tangaroa*
8 *"B." childressi*	16 *B. marisindicus*	24 *B. septemdierum*	32 *V. insolatus*

Association with symbionts
- SOX
- MOX
- both
- unknown

Figure 2 | Distribution of *Bathymodiolus* and *Gigantidas* species and their associated symbionts. Species in the black rectangle are the focus of this thesis. Colours of the circles represent the association with SOX, MOX, or both symbiont types. White circles indicate that the exact association with symbionts is unknown. *B.*: *Bathymodiolus*, *"B."*: *"Bathymodiolus"*; all of these species were reclassified as *Gigantidas* spp., *G.*: *Gigantidas*, *V.*: *Vulcanidas*; *V. insolatus* was previously known as *B.* sp. NZ3, SOX: sulphur-oxidising symbiont, MOX: methane-oxidising symbiont. Figure modified after [83] and [84] with updates to current knowledge.

To date, 30 species of the genus *Bathymodiolus* have been accepted based on their morphological and molecular characteristics [85], some of them being extinct fossil species, with the last species description in 2020 [86]. Some mussel species were first described as *Bathymodiolus* spp. but later reclassified as *Gigantidas* spp. based on molecular data. One example is *"B." childressi* [87].

Bathymodiolus species in the Atlantic Ocean that occur at hydrothermal vents along the Mid-Atlantic Ridge (MAR) are the focus of this thesis (Figure 3 A). Four lineages of *Bathymodiolus* mussels have been described in this geographic region [88,89]. At the northern MAR, *B. azoricus* occurs between 38°N and 36°N and *B. puteoserpentis* between 29°N and 13°N. While *B. azoricus* mussels occur at shallower vents at around 800 m to 2200 m water depth, *B. puteoserpentis* is present at vents deeper than 2500 m water depth [90,91]. At the southern MAR, two unnamed lineages populate vents at either 5°S or 9°S. The four lineages are monophyletic and probably radiated after divergence from the next closest relatives from the north Atlantic *B. boomerang* complex including *B. heckerae* [89,92] (Figure 3 B). *Bathymodiolus* mussels most likely colonised the MAR starting at the northern ridge segments and expanding southwards in the Atlantic [89]. Great geographical barriers along the equatorial belt, including large transform faults such as the Romanche Transform Fault [93], separate the northern and southern lineages of mussels. Despite these barriers, mussel populations north and south of the equator have been shown to be genetically connected [89]. The analysis was based on the mitochondrial cytochrome c oxidase I (COI) gene, which is a common marker gene for species determination in animals [94]. A recent population genomic study based on mitochondrial and 100 nuclear single-nucleotide polymorphism (SNP) markers, i.e. substitutions of single nucleotides in the genome, further demonstrated a widespread introgression between the two northern species emphasising the ongoing gene flow between the mussel populations in this region [95].

Bathymodiolus mussels typically settle on the seafloor after a planktonic larval stage [96,97]. As for many vent species, dispersal during the larval stage most likely enables and maintains the genetic connectivity between geographically separated populations [98,99]. Larval dispersal is influenced by factors such as water currents, water temperatures, time of dispersal, larval predators, planktonic larval duration, and many others [98–100]. Most of these factors are still unknown for *Bathymodiolus* species at the MAR. Breusing et al. (2016) used a biophysical modelling approach suggesting a median maximum distance of 150 km for drifting larvae at the MAR. How far larvae can reach is also determined by their

behaviour, their feeding mode, and the time of dispersal [98–100]. Some of these factors were not directly included in the modelling mentioned above but might play a role for *Bathymodiolus* larvae. The common behavioural mode of shallow-water mytilids is vertical migration of larvae through active swimming [101,102]. Such behaviour has also been suggested for deep-sea *"B." childressi* larvae in the Gulf of Mexico [103] and other deep-sea species [104,105]. *Bathymodiolus* larvae rely on their lipid storage and probably filter feeding for nutrition [97].
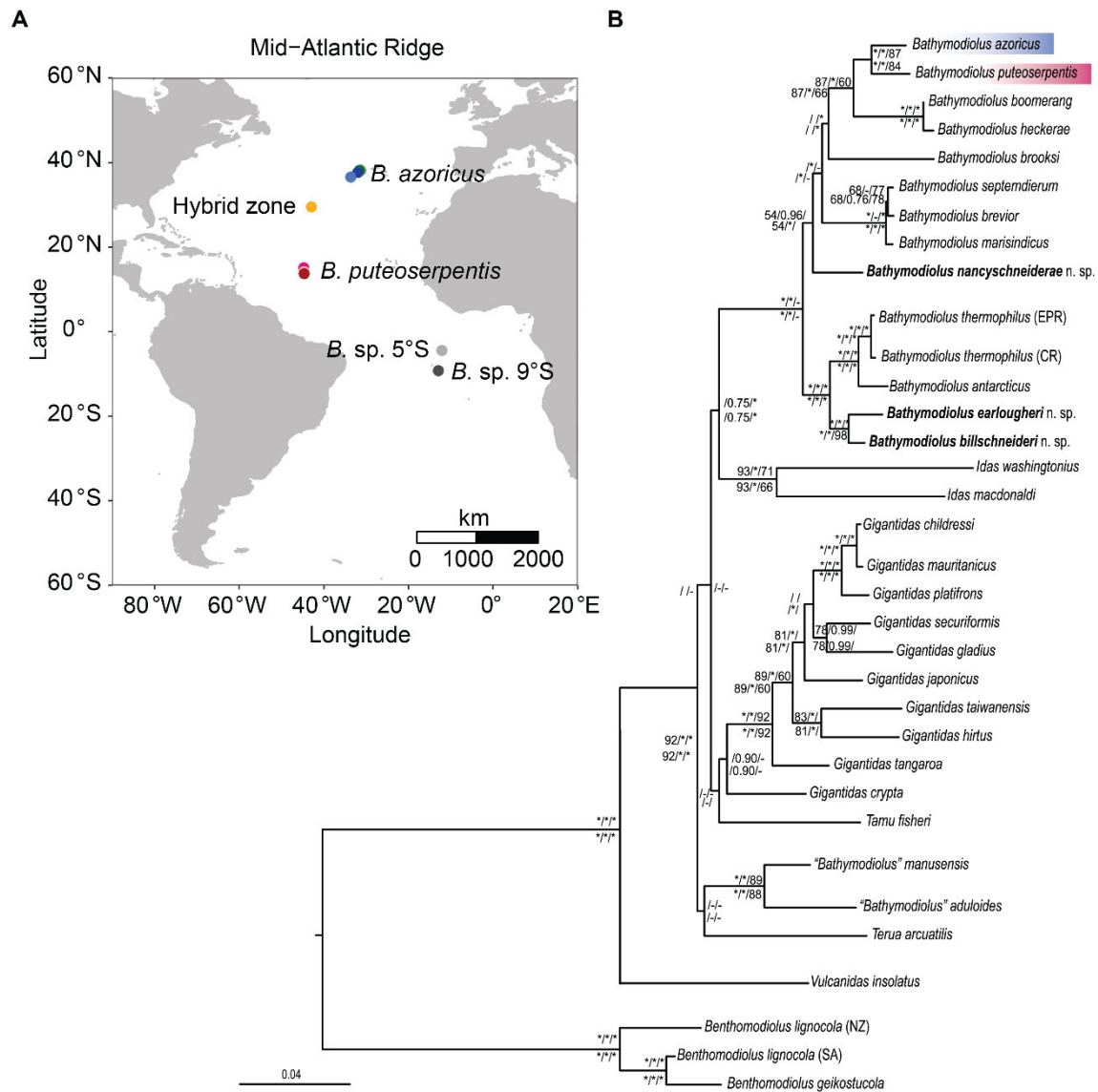

Figure 3 | Distribution and phylogeny of *Bathymodiolus* species. A: Geographic distribution of *Bathymodiolus* along the Mid-Atlantic Ridge (MAR). B: Phylogeny of *Bathymodiolus* and close relatives based on four mitochondrial (COI, 16S, ND4, HSP70) and four nuclear (18S, 28S, HH3, ANT) genes. The tree was reconstructed with RAxML with best-fit model GTR+G+I. Northern MAR species are highlighted: blue: *B. azoricus*, pink: *B. puteoserpentis*. Southern MAR species are not included in the tree, but would fall into a monophyletic clade with *B. azoricus* and *B. puteoserpentis*. Figure modified after [86].

In case they migrate vertically and have access to phototrophic carbon, their dispersal time would be extended, which may increase potential dispersal distance. Higher temperatures at the sea surface might further influence the larval dispersal besides water currents. Thus, the mode of migration and the geographic range of *Bathymodiolus* larvae dispersal remain uncertain.

Association of *Bathymodiolus* mussels with sulphur- and methane-oxidising symbionts

Bathymodiolus mussels owe their ecological success to the chemosynthetic symbionts that they harbour in their gills [50] (Figure 2, Figure 4). Morphological and phylogenetic evidence suggests that symbionts are transmitted horizontally, meaning that they are acquired from the environment with each generation [106–110]. Starvation experiments demonstrated that, after induced loss of symbionts, re-acquisition of symbionts from the water column occurs, providing further evidence for horizontal transmission [111]. The mode of transmission strongly influences the genetic diversity of symbiont populations and the stability of the association [65,112] (Box 2). A recent microscopy study showed that symbionts are absent in *Bathymodiolus* larvae and only colonise after settlement [97]. The symbionts are present in different tissues during early stages of colonisation and later restricted to the symbiotic organ, the gills. The gills are the respiratory and feeding organ of the mussel [113], so the ideal location for the symbionts to ensure a constant supply of vent-derived energy sources [114]. They appear thick and fleshy in *Bathymodiolus* compared to their shallow water relatives, e.g. *Mytilus* [115], and their surface is increased about 20-fold [116]. The mussels have not completely lost their ability to filter-feed and seem to obtain a small proportion of organic matter through feeding themselves [111,117]. However, their digestive system gets reduced during metamorphosis and they mostly rely on their symbionts for nutrition [97,111], especially in the carbon-limited deep-sea environment. Most mussels are associated with sulphur-oxidising (SOX) or methane-oxidising (MOX) symbionts, or both [118], depending on the environment (Figure 2). Besides the two main symbionts, additional symbionts and parasites have been described in these mussels. Examples are epsilonproteobacterial epibionts that potentially oxidise sulphur [119] or parasitic bacteria that invade host nuclei in symbiont-free gill cells of various *Bathymodiolus* species [120,121]. These associations can still be considered "low diversity" systems compared to complex microbiomes with hundreds or thousands of bacterial species such as the human gut microbiome [122]. The low species diversity makes *Bathymodiolus* mussels an ideal study system to investigate the individual

partners of this mutualistic animal–microbe symbiosis at a high resolution and zoom in on within-species variation of the SOX symbionts.
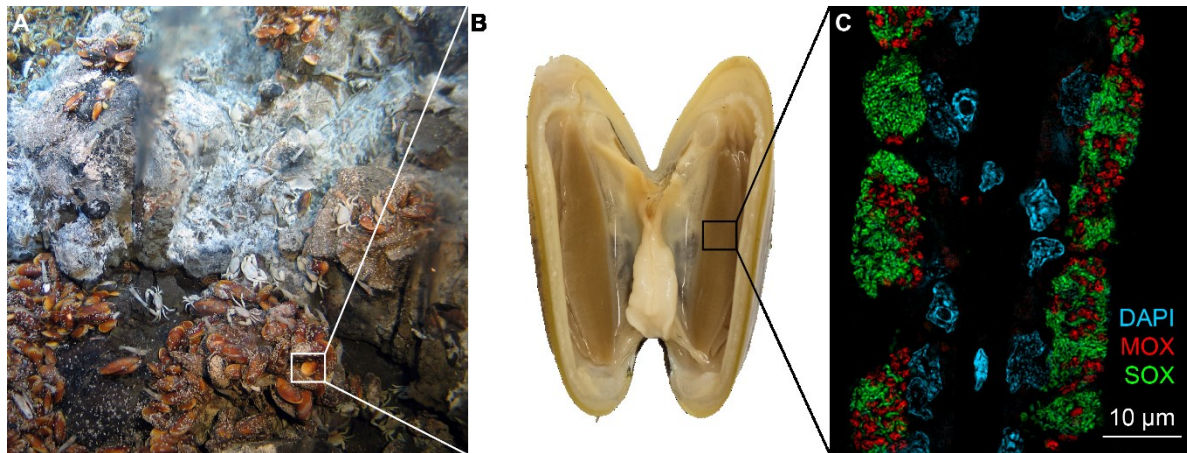
Figure 4 | The *Bathymodiolus* symbiosis. A: Habitat of *Bathymodiolus* mussels at a small hydrothermal vent at the Golden Valley vent site at 5°S. B: Opened *Bathymodiolus* sp. mussel. C: Fluorescence *in situ* hybridisation of *Bathymodiolus* gill filaments. The colours correspond to the DAPI stain (blue) and the specific symbiont probes for methane-oxidising (MOX, red) and sulphur-oxidising symbionts (SOX, green). Image courtesy: A: GEOMAR (CC-BY 4.0). B: Max Planck Institute for Marine Microbiology. C: M. Á. González Porras, Max Planck Institute for Marine Microbiology.

The association with MOX symbionts in metazoans is less common than the association with SOX symbionts [123]. Still, the MOX symbiont is present in at least 10 *Bathymodiolus* species from vent and seep environments, in most cases in dual symbiosis with a SOX symbiont [82]. In dual symbiosis, as found in the Atlantic *Bathymodiolus* lineages, the SOX symbiont is often more abundant, probably due to higher concentrations of reduced sulphur in the environment [116]. *Bathymodiolus* SOX symbionts use reduced sulphur compounds or hydrogen as energy sources to fix carbon [46,124–126]. Oxygen and nitrate serve as terminal electron acceptors [124,125]. The MOX symbionts use methane as carbon and energy source for biomass assimilation [125,127,128].
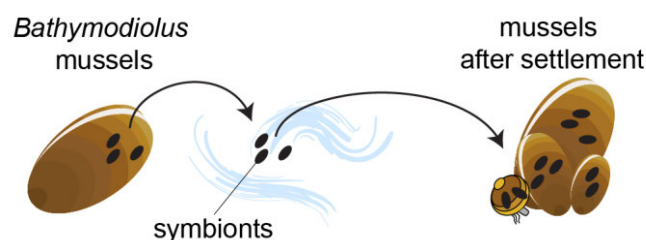
Box 2 | Symbiont transmission

Every symbiotic system comes with a main challenge: **How are symbionts passed on from one generation to the next?** There are two main modes of symbiont transmission: vertical transmission, in which symbionts are directly transmitted from parent to offspring through the germline, and horizontal transmission, where symbionts are taken up from the environment by each new generation. By now, a wide range of transmission modes between the two extremes have been described [112,129], e.g. vertical transmission with events of horizontal transmission in solemyid bivalves [130]. The mode of symbiont transmission has important implications for the host dependence, the diversity of the symbiont community, and the symbiont genome evolution [65,112,131].

Vertical transmission can lead to strong host fidelity as these symbionts stay within the germline and egg [132]. Phylogenetic congruence of hosts and vertically transmitted symbionts is usually high and can result from a process referred to as "co-speciation" [133]. With regard to the genetic diversity of the symbionts, vertical transmission represents a strong bottleneck that can lead to decreased genetic diversity and over evolutionary time frames to symbiont genome reduction [134,135].
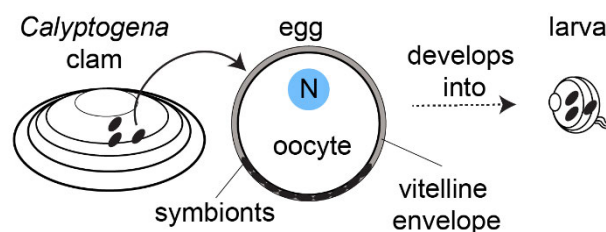
Box 2 Figure 1 | Vertical transmission in deep-sea clams. Symbionts are transferred in the vitelline envelope of the egg directly to the offspring [136]. Parts of the figure are modified after [137].

Horizontal transmission allows for genetic exchange of the symbiotic and free-living bacterial community, which can enhance genetic diversity within the symbiont population [138]. The degree of diversity also depends on the timeframe in which the host can be colonised. If the uptake is restricted to only a short period during development, e.g. in the squid–*vibro* symbiosis [139], the number of strains that colonise the host is limited. If hosts are colonised throughout their lifetime as shown for *Bathymodiolus* mussels, they are likely to host a more diverse symbiont population [126,140]. Phylogenies of hosts and horizontally transmitted symbionts often display little congruence [112].

Box 2 Figure 2 | Horizontal transmission in deep-sea mussels. Symbionts are acquired from the environment after larval settlement [97]. The uptake continues throughout the mussel's lifetime [140].

Hybridisation in the marine environment and in *Bathymodiolus* mussels

Two *Bathymodiolus* species from the MAR, *B. azoricus* and *B. puteoserpentis*, have been described to hybridise at the Broken Spur vent field [90,95] (Figure 3 A). Hybridisation describes the mixing of individuals from different species. When the purebred parental species interbreed, they form first generation (F1) hybrids with mixed genetic features. If the hybrids are fertile, they can further reproduce either among themselves forming the next generations of hybrids (F2 to FX) or with their parental species forming backcrosses. Hybrid zones are valuable study systems as they can serve as natural experiments to investigate the dynamics of evolutionary processes. If the hybridising species harbour symbionts, it becomes even more interesting as animal microbiomes have been shown to influence such processes by promoting speciation or enhancing reproductive barriers [76,141] (Box 1). The interbreeding species can either become one or evolve mechanisms against hybridisation, e.g. hybrid incompatibilities or pre-mating isolation [142]. These mechanisms can ensure that parental species remain separated or promote speciation in cases where hybrids only interbreed with other hybrids and eventually form a new species. Hybridisation is widely accepted to enable faster evolution of genetic novelties compared to mutation alone [143,144]. An expansion of sequencing-based studies over the last years revealed that hybridisation in natural populations is a common phenomenon and of evolutionary importance [145,146]. While terrestrial hybrid zones have been studied with great interest for decades [147,148], marine hybrid zones have received less attention. Initially, they were assumed less frequent as the ocean was perceived as homogenous environment with little gradients that are needed to stabilise hybrid zones [149]. Moreover, pelagic larvae were thought to enable genetic connectivity to a degree that does not allow for hybridisation [150,151]. However, hybridisation in marine environments has been studied in some species such as shallow-water *Mytilus* spp. in the Atlantic and Pacific Ocean [152,153].

In this thesis, I investigate the *Bathymodiolus* hybrid zone at the Broken Spur vent field [90,95] (Figure 3 A). At this site, parental *B. puteoserpentis* mussels and hybrids of *B. azoricus* and *B. puteoserpentis* have been described to co-occur. Detected hybrids were in the F2 to F4 generation, indicating that the hybrids are viable and produce offspring themselves [95]. In addition, Breusing et al. (2017) described *B. puteoserpentis* backcrosses, i.e. offspring from a hybrid mating with parental *B. puteoserpentis*, indicating ongoing introgression [95]. While the symbioses in both parental species have been well characterised over the years, it is unknown how symbionts in Broken Spur are related to other symbionts at

the MAR, whether hybrids harbour the same symbionts as co-occurring *B. puteoserpentis*, and if the host genotype is relevant for the symbiont composition.

Within-species diversity of *Bathymodiolus* SOX symbionts in the Atlantic

When *Bathymodiolus* symbioses were first described, symbionts with different taxonomic affiliation and function were differentiated as morphotypes based on microscopic evidence [154,155]. Since then, advances in sequencing technologies and research on natural symbioses completely changed our view on how complex these symbiont communities are. Analysis of symbiont 16S rRNA genes and the internal transcribed spacer region (ITS), that is more variable than the conserved 16S rRNA gene, delivered the first indication that several 16S rRNA and ITS types of the SOX symbionts co-occur in the mussels [108,156]. This first evidence of an unexpected microdiversity of the SOX symbionts was further supported in 2016, when Ikuta et al. discovered co-existing and functionally diverse symbionts strains in *B. septemdierum* based on functional genes such as hydrogenases [157]. Most recently, high-resolution analyses of single-nucleotide polymorphisms (SNPs) across whole metagenome-assembled symbiont genomes revealed the presence of up to 16 strains of the same symbiont species within single *Bathymodiolus* individuals [126,158]. The co-existence of so many bacterial strains in the same habitat, the gill, challenged theories on competition in natural communities. Evolutionary theory would predict that competition leads to extinction of one competitor, or separation between the competitors, either physically or by niche adaptation [159,160]. Ansorge et al. (2019) showed that fine-scale niche partitioning may occur in *Bathymodiolus* mussels as strains with and without the hydrogenase gene were observed in different bacteriocytes [126]. However, the authors also argued that the co-existence of functionally diverse symbiont strains might not be detrimental to the stability of the symbiosis but rather an asset that allows increased potential for environmental adaptation, among other benefits. Moreover, this discovery highlighted that we need to apply the highest resolution available to us if we want to understand the dynamics of the SOX symbionts at the strain level in *Bathymodiolus* mussels.

When analysing the SOX symbionts in metagenomes, we can compare them at different levels of resolution (Figure 5). In the following paragraph, I introduce these levels and define the terms species, phylotype, subspecies, and strain. The definitions are used throughout the thesis. Four mussel lineages from the MAR are analysed in this thesis, *B. azoricus* and
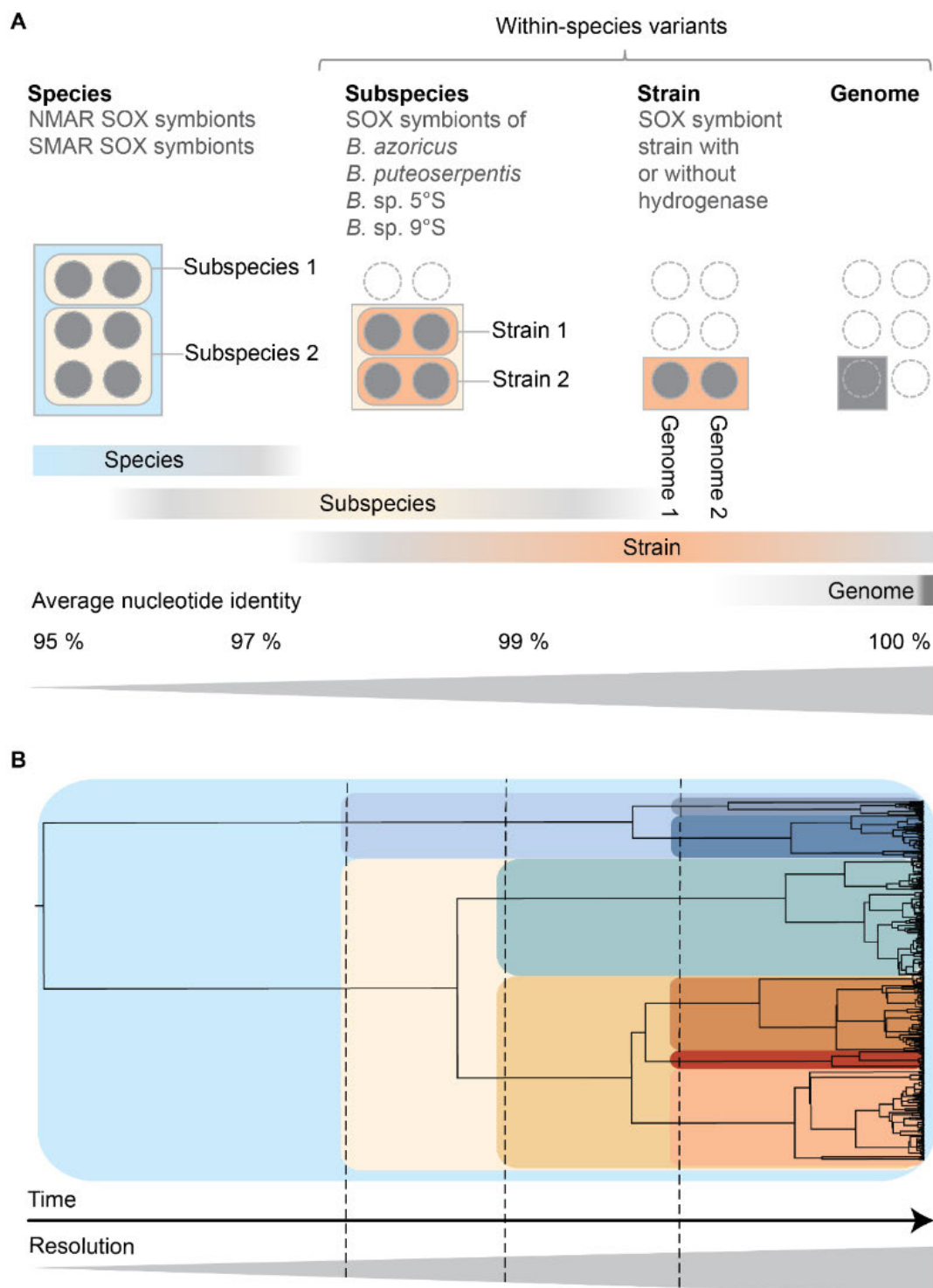
Figure 5 | Within-species variation in *Bathymodiolus* SOX symbionts. A: Definitions for the different levels of variation within bacterial species down to single nucleotide differences in individual genomes. Pictograms show the relation between the different levels to each other. Bars indicate the range of resolution in which the definition is used. Grey portions show the common but often unspecific range of use while coloured portions indicate the recommended, specific use. Approximate cut-offs of average nucleotide identity are shown along the grey triangle. Corresponding examples of *Bathymodiolus* SOX symbionts for the levels of resolution are written in grey. B: Phylogenetic tree depicting the relationship between diversification, time, and resolution needed to analyse the within-species variation. Colours light blue, beige and orange correspond to the definitions in A. Figure modified after [162] and slides from R. Ansorge and A. Kupczok. NMAR: Northern Mid-Atlantic Ridge, SMAR: Southern Mid-Atlantic Ridge, SOX: Sulphur-oxidising symbionts, *B.*: *Bathymodiolus*.

B. puteoserpentis from the northern MAR (NMAR) and *B.* sp. 5°S and *B.* sp. 9°S from the southern MAR (SMAR), and with them, their SOX symbionts. In contrast to their hosts, the symbionts fall into only two bacterial species with above 95 % average nucleotide identity (ANI) [161]: one at the NMAR and one at the SMAR. While the northern symbiont species shares an identical 16S rRNA gene sequence, i.e. one phylotype [118], the southern symbionts can be distinguished based on one consistent, location-specific nucleotide substitution, i.e. they are two phylotypes [89]. Symbionts from all four host lineages belong to different subspecies, defined as bacteria with ANI values above 97 % [108,161,162]. Within the species and subspecies, the symbiont cells are not clonal but represent different strains that can encode for strain-specific genes. These genes have been shown to be involved in metabolic processes, e.g. hydrogenases, host-symbiont recognition such as genes involved in cell surface modification, or CRISPR-Cas mediated phage defence [126,157].

Even below subspecies level, the term strain does not necessarily refer to the same level of resolution (Figure 5 B) and should be defined based on the applied analysis. Symbionts of the same species within mussel individuals can be considered a population of strains. This means that we can extend analyses of the symbionts from phylogenomics or ANI comparisons that capture symbiont subspecies by applying population genomic measures to the symbiont strains. One of these measures is the population differentiation index F_{ST}. F_{ST} is based on SNP analysis and indicates whether two populations are the same ($F_{ST} = 0$) or distinct ($F_{ST} = 1$).

Bathymodiolus SOX symbionts in the free-living stage and their relatives

Bathymodiolus SOX symbionts are part of the *Thioglobaceae* family in the order of Thiomicrospirales. Within the *Thioglobaceae* family, SOX symbionts belong to the widespread SUP05/Arctic96BD-19 clade. 16S rRNA nucleotide identity within the clade is quite high with 94 % [163]. Despite the close phylogenetic relatedness, bacteria in this clade have diverse metabolisms and both symbiotic and free-living lifestyles [163]. The free-living members of the *Thioglobaceae* occur in a variety of marine habitats including oxygen minimum or anoxic marine zones and hydrothermal vents [19,26,164]. Due to their high cell abundance, *Thioglobaceae* are now considered important players for the sulphur and nitrogen cycling in the ocean [164–166]. Two free-living relatives of the *Bathymodiolus* SOX symbionts could be isolated: *Candidatus* Thioglobus autotrophicus EF1 [167] and *Ca.* Thioglobus singularis PS1 [168]. Besides *Bathymodiolus* symbionts, other symbiotic bacteria

from the SUP05 clade have been studied, e.g. *Ca.* Ruthia magnifica [169] and *Ca.* Vesicomyosocious okutanii [170]. Both are endosymbionts of deep-sea vesicomyid clams.

Bathymodiolus symbionts are horizontally transmitted [106–110], suggesting a free-living stage. Recent comparative genomics of the *Thioglobaceae* family indicate that *Bathymodiolus* SOX symbionts do not have reduced genomes and are probably able to survive outside their hosts [161]. During the free-living stage, the symbionts are potentially more strongly influenced by conditions in the water column than during their host-associated stage, and could be able to exchange genetic content with the free-living pool of bacteria. To date, only few studies investigated the free-living stage of *Bathymodiolus* symbionts [109,157,171]. Symbiont-related genes such as 16S rRNA or the Calvin-Benson-Cycle gene *cbbL* have been detected in seawater samples or microbial mats with PCR [109,157,171]. However, how different symbiont subspecies are distributed geographically during their free-living stage remains to be elucidated.

Aims of this thesis

The aim of this thesis is to gain a better understanding of the population structure of *Bathymodiolus* mussels and their symbionts, factors driving the SOX symbiont composition as well as the free-living symbiont stage in seawater surrounding the mussels. When I began my studies, knowledge about the free-living stage of symbionts was scarce but of great interest because of its implications for symbiont transmission and the specificity of the animal–microbe association. With the discovery of the *Bathymodiolus* hybrid zone, scientists found the ideal natural experiment to study animal hybridisation in the presence of symbiosis. Yet, only one study had been performed on the symbionts from the Broken Spur hybrid zone [108]. In addition, population genomic studies of non-model organisms started to gain more attraction thanks to advances in sequencing technologies and bioinformatics tools. Studies on the population structure of *Bathymodiolus* mussels on the Mid-Atlantic Ridge had been performed, but the resolution of the methods applied was limited and left some open questions concerning the subpopulation structure of *Bathymodiolus* from different vent sites in the same geographic region.

Throughout my studies, I used metagenomics to approach the questions at hand. Investigation of the host and symbiont population structure would have been feasible with targeted DNA-sequencing of marker genes. However, only metagenomes provide the resolution high enough to investigate the symbionts on strain level and to assess subpopulation structure of

the hosts with confidence. When analysing highly similar organisms, such as the SOX symbiont subspecies of *Bathymodiolus* mussels in this thesis, marker genes do not provide useful information, as they are limited to the genus or species level. Additionally, metagenomes capture the whole symbiotic association including host and symbionts. Any targeted approach for the bacterial partner would not allow studying the eukaryotic partner in case questions arise. By now, metagenomics can be considered a cost-efficient and high throughput method that does not require time-consuming laboratory preparations. With more efficient bioinformatics tools and increases in computing power, there are little limitations when it comes to metagenomic analyses. One last advantage is that metagenomes are versatile and can be used for a wide range of analyses as demonstrated in this thesis. In the following paragraphs, I introduce the specific research questions that I addressed in the next three chapters.

Do hybrid and parental *Bathymodiolus* mussels harbour different symbionts?
A hybrid zone of the parental species *B. azoricus* and *B. puteoserpentis* had been described at the Broken Spur vent field. While the symbiosis in the parental species has been the subject of many studies until now, most studies on the hybrids and co-occurring *B. puteoserpentis* from Broken Spur focused on the host [90,91,95,172]. Only one study investigated the SOX symbionts, however, the hybrid status of the corresponding host individuals remained unclear [91].

The symbiont composition plays an important role in determining the metabolic potential of the symbiotic animal. What factors are driving the composition is a topic of intense debate and varies between symbiotic systems [173–176]. The Broken Spur hybrid zone is an opportune system to study the influence of host genetics on the symbiont composition as hybrid and parental mussels co-occur at the same site. I took a stepwise approach to resolve the question whether hybrid and parental mussels harbour different symbionts. First, I determined the hybrid status of the mussels based on Fluidigm genotyping [88]. In the second step, I analysed the relatedness of symbionts from Broken Spur to other *Bathymodiolus* symbionts using phylogenomics and comparisons of average nucleotide identity. Lastly, I used high-resolution SNP analysis to detect potential differences between symbionts of hybrids and parentals on strain level. The manuscript in chapter II has been accepted for publication in ISME Journal.

Are free-living *Bathymodiolus* symbionts present in the water column and how are they geographically distributed?

Bathymodiolus mussels acquire their symbionts from the environment, likely throughout their lifetime [108,126,140]. This mode of transmission suggests the existence of symbionts in a free-living stage in the environment. During the free-living stage, the symbionts are potentially more strongly influenced by environmental parameters than during their host-associated stage. In addition, they could be able to exchange genetic content with the free-living pool of bacteria. Whenever the symbionts within mussel gills are analysed, the results also raise the question how findings, e.g. high strain diversity or location-specificity, relate to the free-living symbiont population (Chapter II, [84,177]). For instance, geographical structuring of symbionts associated with mussel hosts could result from a geographical structuring of symbionts in the free-living stage. Knowledge about the geographical distribution of symbionts in the water column could therefore advance our understanding of the specificity of the symbiont uptake and whether the host or the environment is selecting suited symbionts. For *Riftia* tubeworms, the existence of a free-living stage has been shown based on PCR of single genes and fluorescence *in situ* hybridisation of ribosomal genes in surface-attached biofilms and filtered water samples [178]. Few studies targeted *Bathymodiolus* symbionts and discovered symbiont-related genes (16S rRNA, *cbbL*, and functional genes such as hydrogenase) in microbial mats or seawater samples using PCR [109,157,171]. In contrast to the tubeworm example, most studies in *Bathymodiolus* mussels did not screen for host tissue or DNA. Thus, it remains unclear whether detected sequences represented the free-living or host-associated stage. In addition, marker genes such as 16S rRNA cannot distinguish between symbiont subspecies and each study only analysed samples from a single site.

Understanding biogeographical patterns of microorganisms in the marine environment is challenging. Assessment of the distribution of *Bathymodiolus* symbionts in the water column is the first step to clarify whether these symbionts are widely distributed across the ocean and selected by their host or the environment, or whether they are geographically structured based on dispersal barriers. I analysed filtered water samples from several vent sites along the Mid-Atlantic Ridge in chapter III to investigate the presence and distribution of *Bathymodiolus* symbionts in the free-living stage. For better comparison with previous studies, I first screened for the symbiont 16S rRNA and the host 18S rRNA and cytochrome oxidase c subunit I (COI) in the metagenomes. Further, I reconstructed a 16S rRNA phylogeny of

water-derived Thiomicrospirales, the group that contains *Bathymodiolus* SOX symbionts, to assess the diversity of symbiont relatives in the deep-sea environment. Lastly, I used symbiont subspecies specific markers to analyse their distribution in mussel and seawater metagenomes across different sites.

How do *Bathymodiolus* mitochondrial and nuclear genomes relate?

Species determination in *Bathymodiolus* mussels is often performed based on the mitochondrial COI gene. Cytonuclear disequilibrium, i.e. the non-random association of genotypes at a nuclear locus with cytoplasmic haplotypes, has been described for *Bathymodiolus* mussels from the Broken Spur hybrid zone [90,91,95]. This finding indicates that mitochondrial haplotypes are not necessarily linked to the nuclear genotype in these mussels. Beyond the hybrid zone, such incongruence between mitochondrial and nuclear genome information has not been observed. However, most studies focused on either one or the other genome type, or the number of samples was limited. I wanted to know whether mitochondrial and nuclear genomes provide congruent information, with the exception of the hybrid zone, in *Bathymodiolus*. This is important if we want to understand the phylogenetic patterns based on either mitochondrial or nuclear genome information. In chapter IV, I compared mitochondrial clades to genetic clusters based on exome-wide SNP analysis to investigate how mitochondrial and nuclear genomes relate.

Can high-resolution population genomic analysis reveal subpopulation structure in *Bathymodiolus* from the MAR?

Previous population genomic studies of *Bathymodiolus* mussels at the MAR revealed four mussel lineages based on COI and 100 nuclear marker genes [89,95]. Within the northern lineages, Breusing et al. (2017) detected introgression but no subpopulation structure between individuals from different sampling sites in the same region [95]. If we consider the chemical and physical differences and the geographical barriers and distances between the vent sites together with the modelled larvae dispersal distance of 150 km [88], the lack of subpopulation structure is rather surprising. While the 100 nuclear markers were a great method improvement moving beyond COI and allozymes, the authors cautioned that the chosen markers might not be sufficient to resolve contemporary population structure [95]. In addition, the study solely focused on the northern MAR although subpopulation structure might also be present at 5°S on the southern MAR, where mussels of the same lineage are distributed across various vent sites.

I used exome-wide SNP analysis to investigate the population genomics of *Bathymodiolus* mussels at the MAR. I wanted to resolve whether there is a subpopulation structure in conspecific mussels from different sites. For the analysis, I made use of previously as well as newly sequenced metagenomes to increase the sample size for reliable results. Low coverage host reads were mapped to transcriptome assemblies to bypass the need for a host genome that is not available. Between 200 000 and 600 000 SNP sites were used to calculate pairwise genetic distances, admixture proportions, and to determine nucDNA clusters revealing potential subpopulation structure. Results of the population genomic study of the host are presented and discussed in chapter IV.

References

1. Thurber AR, Sweetman AK, Narayanaswamy BE, Jones DOB, Ingels J, Hansman RL. Ecosystem function and services provided by the deep sea. Biogeosciences. 2014; 11: 3941–63.

2. Hessler RR, Sanders HL. Faunal diversity in the deep-sea. Deep-Sea Res Oceanogr Abstr. 1967; 14: 65–78.

3. Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. How many species are there on earth and in the ocean? PLoS Biol. 2011; 9: e1001127.

4. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored "rare biosphere." PNAS. 2006; 103: 12115–20.

5. Ramirez-Llodra E, Tyler PA, Baker MC, Bergstad OA, Clark MR, Escobar E, et al. Man and the last great wilderness: human impact on the deep sea. PLOS ONE. 2011; 6: e22588.

6. Benn AR, Weaver PP, Billet DSM, Hove S van den, Murdock AP, Doneghan GB, et al. Human activities on the deep seafloor in the north east Atlantic: an assessment of spatial extent. PLOS ONE. 2010; 5: e12730.

7. Morato T, Watson R, Pitcher TJ, Pauly D. Fishing down the deep. Fish Fish. 2006; 7: 24–34.

8. Mengerink KJ, Dover CLV, Ardron J, Baker M, Escobar-Briones E, Gjerde K, et al. A call for deep-ocean stewardship. Science. 2014; 344: 696–8.

9. Beaulieu SE, Baker ET, German CR. Where are the undiscovered hydrothermal vents on oceanic spreading ridges? Deep Sea Res Part II Top Stud Oceanogr. 2015; 121: 202–12.

10. Vent Fields | InterRidge Vents Database Ver. 3.4. n.d. https://vents-data.interridge.org/ (accessed February 24, 2021).

11. Von Damm KL. Controls on the chemistry and temporal variability of seafloor hydrothermal fluids. In: Humphris SE, Zierenberg RA, Mullineaux LS, Thomson RE, editors. Seafloor Hydrothermal Systems: Physical, Chemical, Biological, and Geological Interactions, vol. 91, American Geophysical Union (AGU); 1995, p. 222–47.

12. Stewart FJ, Newton ILG, Cavanaugh CM. Chemosynthetic endosymbioses: adaptations to oxic–anoxic interfaces. Trends Microbiol. 2005; 13: 439–48.

13. Lalou C, Brichet E. Ages and implications of East Pacific Rise sulphide deposits at 21°N. Nature. 1982; 300: 169–71.

14. Lalou C, Thompson G, Arnold M, Brichet E, Druffel E, Rona PA. Geochronology of TAG and Snakepit hydrothermal fields, Mid-Atlantic Ridge: Witness to a long and complex hydrothermal history. Earth Planet Sci Lett. 1990; 97: 113–28.

15. Juniper SK, Tunnicliffe V. Crustal accretion and the hot vent ecosystem. Phil Trans R Soc A. 1997; 355: 459–74.

16. Van Dover CL. The Ecology of Deep-Sea Hydrothermal Vents. Princeton University Press; 2000.

17. Martin W, Baross J, Kelley D, Russell MJ. Hydrothermal vents and the origin of life. Nat Rev Microbiol. 2008; 6: 805–14.

18. Sievert S, Vetriani C. Chemoautotrophy at deep-sea vents: Past, present, and future. Oceanography. 2012; 25: 218–33.

19. Meier DV, Pjevac P, Bach W, Hourdez S, Girguis PR, Vidoudez C, et al. Niche partitioning of diverse sulfur-oxidizing bacteria at hydrothermal vents. ISME J. 2017; 11: 1545–58.

20.   Reysenbach A-L, Banta AB, Boone DR, Cary SC, Luther GW. Microbial essentials at hydrothermal vents. Nature. 2000; 404: 835–835.

21.   Orcutt BN, Sylvan JB, Knab NJ, Edwards KJ. Microbial ecology of the dark ocean above, at, and below the seafloor. Microbiol Mol Biol Rev. 2011; 75: 361–422.

22.   Haase KM, Petersen S, Koschinsky A, Seifert R, Devey CW, Keir R, et al. Young volcanism and related hydrothermal activity at 5°S on the slow-spreading southern Mid-Atlantic Ridge. Geochem Geophys Geosyst. 2007; 8: Q11002.

23.   Jannasch HW. Review lecture - the chemosynthetic support of life and the microbial diversity at deep-sea hydrothermal vents. Proc Royal Soc B. 1985; 225: 277–97.

24.   Jannasch HW, Wirsen CO. Chemosynthetic primary production at east Pacific sea floor spreading centers. BioScience. 1979; 29: 592–8.

25.   Cavanaugh CM, McKiness ZP, Newton ILG, Stewart FJ. Marine chemosynthetic symbioses. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. The Prokaryotes: Prokaryotic Biology and Symbiotic Associations, Berlin, Heidelberg: Springer; 2013, p. 579–607.

26.   Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG, et al. Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. Science. 2009; 326: 578–82.

27.   Dick GJ. The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. Nat Rev Microbiol. 2019; 17: 271–83.

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

29.   Summit M, Baross JA. A novel microbial habitat in the mid-ocean ridge subseafloor. PNAS. 2001; 98: 2158–63.

30.   Wright S. Isolation by distance. Genetics. 1943; 28: 114–38.

31.   Wang IJ, Bradburd GS. Isolation by environment. Mol Ecol. 2014; 23: 5649–62.

32.   Chase AB, Arevalo P, Brodie EL, Polz MF, Karaoz U, Martiny JBH. Maintenance of sympatric and allopatric populations in free-living terrestrial bacteria. MBio. 2019; 10: e02361-19.

33.   Whitaker RJ, Grogan DW, Taylor JW. Geographic barriers isolate endemic populations of hyperthermophilic archaea. Science. 2003; 301: 976–8.

34.   Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, et al. Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol. 2006; 4: 102–12.

35.   Baas Becking LGM. Geobiologie of inleiding tot de milieukunde. Den Haag: W.P. Van Stockum & Zoon; 1934.

36.   Wit RD, Bouvier T. 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? Environ Microbiol. 2006; 8: 755–8.

37.   Mino S, Nakagawa S, Makita H, Toki T, Miyazaki J, Sievert SM, et al. Endemicity of the cosmopolitan mesophilic chemolithoautotroph *Sulfurimonas* at deep-sea hydrothermal vents. ISME J. 2017; 11: 909–19.

38.   Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. Beyond biogeographic patterns: Processes shaping the microbial landscape. Nat Rev Microbiol. 2012; 10: 497–506.

39.   Weiss RF, Lonsdale P, Lupton JE, Bainbridge AE, Craig H. Hydrothermal plumes in the Galapagos Rift. Nature. 1977; 267: 600–3.

40.   Corliss JB, Dymond J, Gordon LI, Edmond JM, von Herzen RP, Ballard RD, et al. Submarine thermal springs on the Galapagos Rift. Science. 1979; 203: 1073–83.

41.   Jones ML. *Riftia pachyptila*, new genus, new species, the vestimentiferan worm from the Galapagos Rift geothermal vents (Pogonophora). Proc Biol Soc Wash. 1981; 93: 1295–313.

42.   Govenar B. Shaping vent and seep communities: Habitat provision and modification by foundation species. In: Kiel S, editor. The Vent and Seep Biota: Aspects from Microbes to Ecosystems, Dordrecht: Springer Netherlands; 2010, p. 403–32.

43.   Douglas A. The significance of symbiosis. The Symbiotic Habit, vol. 61, Princeton, New Jersey: Princeton University Press; 2010, p. 1–214.

44.   de Bary A. Die Erscheinung der Symbiose: Vortrag gehalten auf der Versammlung Deutscher Naturforscher und Aerzte zu Cassel. Trübner; 1879.

45.   Fisher CR, Childress JJ, Oremland RS, Bidigare RR. The importance of methane and thiosulfate in the metabolism of the bacterial symbionts of two deep-sea mussels. Mar Biol. 1987; 96: 59–71.

46.   Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R, et al. Hydrogen is an energy source for hydrothermal vent symbioses. Nature. 2011; 476: 176–80.

47.   Beinart RA, Sanders JG, Faure B, Sylva SP, Lee RW, Becker EL, et al. Evidence for the role of endosymbionts in regional-scale habitat partitioning by hydrothermal vent symbioses. PNAS. 2012; 109: E3241–50.

48.   Moran NA. Symbiosis. Curr Biol. 2006; 16: R866–71.

49.   Cavanaugh CM, Gardiner SL, Jones ML, Jannasch HW, Waterbury JB. Prokaryotic cells in the hydrothermal vent tube worm *Riftia pachyptila* Jones: possible chemoautotrophic symbionts. Science. 1981; 213: 340–2.

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

51.   Ott J, Rieger G, Rieger R, Enderes F. New mouthless interstitial worms from the sulfide system: Symbiosis with prokaryotes. Mar Ecol. 1982; 3: 313–33.

52.   Goffredi SK, Johnson SB, Vrijenhoek RC. Genetic diversity and potential function of microbial symbionts associated with newly discovered species of *Osedax* polychaete worms. Appl Environ Microbiol. 2007; 73: 2314–23.

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

54.   Jan C, Petersen JM, Werner J, Teeling H, Huang S, Glöckner FO, et al. The gill chamber epibiosis of deep-sea shrimp *Rimicaris exoculata*: an in-depth metagenomic investigation and discovery of Zetaproteobacteria. Environ Microbiol. 2014; 16: 2723–38.

55.   Kleiner M, Petersen JM, Dubilier N. Convergent and divergent evolution of metabolism in sulfur-oxidizing symbionts and the role of horizontal gene transfer. Curr Opin Microbiol. 2012; 15: 621–31.

56.   Petersen JM, Ramette A, Lott C, Cambon-Bonavita M-A, Zbinden M, Dubilier N. Dual symbiosis of the vent shrimp *Rimicaris exoculata* with filamentous Gamma- and Epsilonproteobacteria at four Mid-Atlantic Ridge hydrothermal vent fields. Environ Microbiol. 2010; 12: 2204–18.

57.   Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO, et al. Symbiosis insights through metagenomic analysis of a microbial consortium. Nature. 2006; 443: 950–5.

58.   Kleiner M, Wentrup C, Holler T, Lavik G, Harder J, Lott C, et al. Use of carbon monoxide and hydrogen by a bacteria–animal symbiosis from seagrass sediments. Environ Microbiol. 2015; 17: 5023–35.

59. Stewart FJ, Young CR, Cavanaugh CM. Lateral symbiont acquisition in a maternally transmitted chemosynthetic clam endosymbiosis. Mol Biol Evol. 2008; 25: 673–87.

60. Felbeck H. Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila* Jones (Vestimentifera). Science. 1981; 213: 336–8.

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

62. Kiers ET, West SA. Evolving new organisms via symbiosis. Science. 2015; 348: 392–4.

63. Sterrer W, Rieger R. Retronectidae - a new cosmopolitan marine family of Catenulida (Turbellaria). In: Riser NW, Morse P, editors. Biology of the Turbellaria, New York: McGraw-Hill; 1974, p. 63–92.

64. Ott J, Bright M, Bulgheresi S. Marine microbial thiotrophic ectosymbioses. In: Gibson RN, Atkinson RJA, Gordon J, editors. Oceanography and Marine Biology, vol. 42. 1st ed., CRC Press; 2004, p. 95–118.

65. Vrijenhoek RC. Genetics and evolution of deep-sea chemosynthetic bacteria and their invertebrate hosts. In: Kiel S, editor. The Vent and Seep Biota, Springer Netherlands; 2010, p. 15–49.

66. Beinart RA. The significance of microbial symbionts in ecosystem processes. MSystems. 2019; 4: e00127-19.

67. Margulis L. Origin of eukaryotic cells: Evidence and research implications for a theory of the origin and evolution of microbial, plant, and animal cells on the Precambrian earth. New Haven: Yale University Press; 1970.

68. Margulis L, Fester R. Bellagio conference and book. Symbiosis as source of evolutionary innovation: Speciation and morphogenesis. Symbiosis. 1991; 11: 93–101.

69. Weinstein N, Garten B, Vainer J, Minaya D, Czaja K. Managing the microbiome: how the gut influences development and disease. Nutrients. 2021; 13: 74.

70. Weiss BL, Maltz M, Aksoy S. Obligate symbionts activate immune system development in the tsetse fly. J Immunol. 2012; 188: 3395–403.

71. Heijtz RD, Wang S, Anuar F, Qian Y, Björkholm B, Samuelsson A, et al. Normal gut microbiota modulates brain development and behavior. PNAS. 2011; 108: 3047–52.

72. Kunc M, Gabrych A, Witkowski JM. Microbiome impact on metabolism and function of sex, thyroid, growth and parathyroid hormones. Acta Biochim Pol. 2016; 63: 189–201.

73. Zheng H, Powell JE, Steele MI, Dietrich C, Moran NA. Honeybee gut microbiota promotes host weight gain via bacterial metabolism and hormonal signaling. PNAS. 2017; 114: 4775–80.

74. Weis VM, Davy SK, Hoegh-Guldberg O, Rodriguez-Lanetty M, Pringle JR. Cell biology in model systems as the key to understanding corals. Trends Ecol Evol. 2008; 23: 369–76.

75. Sharon G, Segal D, Ringo JM, Hefetz A, Zilber-Rosenberg I, Rosenberg E. Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. PNAS. 2010; 107: 20051–6.

76. Brucker RM, Bordenstein SR. The hologenomic basis of speciation: Gut bacteria cause hybrid lethality in the genus *Nasonia*. Science. 2013; 341: 667–9.

77. Kenk VC, Wilson BR. A new mussel (Bivalvia, Mytilidae) from hydrothermal vents in the Galapagos Rift zone. Malacologia. 1985; 26: 253–71.

78. Distel DL, Baco AR, Chuang E, Morrill W, Cavanaugh C, Smith CR. Do mussels take wooden steps to deep-sea vents? Nature. 2000; 403: 725–6.

79.   Duperron S, Gaudron SM, Rodrigues CF, Cunha MR, Decker C, Olu K. An overview of chemosynthetic symbioses in bivalves from the north Atlantic and Mediterranean Sea. Biogeosciences. 2013; 10: 3241–67.

80.   Van Dover CL. Community structure of mussel beds at deep-sea hydrothermal vents. Mar Ecol Prog Ser. 2002; 230: 137–58.

81.   Liu J, Liu H, Zhang H. Phylogeny and evolutionary radiation of the marine mussels (Bivalvia: Mytilidae) based on mitochondrial and nuclear genes. Mol Phylogenetics Evol. 2018; 126: 233–40.

82.   Lorion J, Kiel S, Faure B, Kawato M, Ho SYW, Marshall B, et al. Adaptive radiation of chemosymbiotic deep-sea mussels. Proc R Soc B. 2013; 280: 20131243.

83.   Génio L, Johnson S, Vrijenhoek R, Cunha M, Tyler P, Kiel S, et al. New record of *"Bathymodiolus" mauritanicus* Cosel 2002 from the Gulf of Cadiz (NE Atlantic) mud volcanoes. J Shellfish Res. 2008; 27: 53–61.

84.   Ansorge R. Strain diversity and evolution in endosymbionts of *Bathymodiolus* mussels. PhD Thesis. University of Bremen, 2019.

85.   MolluscaBase. 2021. https://www.molluscabase.org (accessed February 10, 2021).

86.   McCowin MF, Feehery C, Rouse GW. Spanning the depths or depth-restricted: three new species of *Bathymodiolus* (Bivalvia, Mytilidae) and a new record for the hydrothermal vent *Bathymodiolus thermophilus* at methane seeps along the Costa Rica margin. Deep Sea Res Part I Oceanogr. 2020; 164: 103322.

87.   Gustafson RG, Turner RD, Lutz RA, Vrijenhoek RC. A new genus and five new species of mussels (Bivalvia: Mytilidae) from deep-sea sulfide/hydrocarbon seeps in the Gulf of Mexico. Malacologia. 1998; 40: 63–112.

88.   Breusing C, Biastoch A, Drews A, Metaxas A, Jollivet D, Vrijenhoek RC, et al. Biophysical and population genetic models predict the presence of "phantom" stepping stones connecting Mid-Atlantic Ridge vent ecosystems. Curr Biol. 2016; 26: 2257–67.

89.   van der Heijden K, Petersen JM, Dubilier N, Borowski C. Genetic connectivity between north and south Mid-Atlantic Ridge chemosynthetic bivalves and their symbionts. PLOS ONE. 2012; 7: e39994.

90.   O'Mullan GD, Maas PA, Lutz RA, Vrijenhoek RC. A hybrid zone between hydrothermal vent mussels (Bivalvia: Mytilidae) from the Mid-Atlantic Ridge. Mol Ecol. 2001; 10: 2819–31.

91.   Won Y, Hallam SJ, O'Mullan GD, Vrijenhoek RC. Cytonuclear disequilibrium in a hybrid zone involving deep-sea hydrothermal vent mussels of the genus *Bathymodiolus*. Mol Ecol. 2003; 12: 3185–90.

92.   Olu-Le Roy K, Cosel R von, Hourdez S, Carney SL, Jollivet D. Amphi-Atlantic cold-seep *Bathymodiolus* species complexes across the equatorial belt. Deep Sea Res Part I Oceanogr. 2007; 54: 1890–911.

93.   Heezen BC, Bunce ET, Hersey JB, Tharp M. Chain and Romanche fracture zones. Deep-Sea Res Oceanogr Abstr. 1964; 11: 11–33.

94.   Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. Proc R Soc B. 2003; 270: 313–21.

95.   Breusing C, Vrijenhoek RC, Reusch TBH. Widespread introgression in deep-sea hydrothermal vent mussels. BMC Evol Biol. 2017; 17: 13.

96.   Arellano SM, Young CM. Spawning, development, and the duration of larval life in a deep-sea cold-seep mussel. Biol Bull. 2009; 216: 149–62.

97.   Franke M, Geier B, Hammel JU, Dubilier N, Leisch N. Becoming symbiotic – the symbiont acquisition and the early development of bathymodiolin mussels. BioRxiv. 2020: 2020.10.09.333211.

98. Hilário A, Metaxas A, Gaudron SM, Howell KL, Mercier A, Mestre NC, et al. Estimating dispersal distance in the deep sea: Challenges and applications to marine reserves. Front Mar Sci. 2015; 2: 6.

99. Young CM, Arellano SM, Hamel J-F, Mercier A. Evolutionary ecology of marine invertebrate larvae. In: Carrier T, Reitzel A, Heyland A, editors. Ecology and Evolution of Larval Dispersal in the Deep Sea, Oxford: Oxford University Press; 2017.

100. Gary SF, Fox AD, Biastoch A, Roberts JM, Cunningham SA. Larval behaviour, dispersal and population connectivity in the deep sea. Sci Rep. 2020; 10: 10675.

101. Bayne BL. Responses of *Mytilus edulis* larvae to increases in hydrostatic pressure. Nature. 1963; 198: 406–7.

102. Bayne BL. The responses of the larvae of *Mytilus edulis* L. to light and to gravity. Oikos. 1964; 15: 162–74.

103. Arellano SM, Van Gaest AL, Johnson SB, Vrijenhoek RC, Young CM. Larvae from deep-sea methane seeps disperse in surface waters. Proc Biol Sci. 2014; 281: 20133276.

104. Yahagi T, Kayama Watanabe H, Kojima S, Kano Y. Do larvae from deep-sea hydrothermal vents disperse in surface waters? Ecology. 2017; 98: 1524–34.

105. Van Gaest AL. Ecology and early life history of *Bathynerita naticoidea*: Evidence for long-distance larval dispersal of a cold seep gastropod. OIMB MS Thesis. University of Oregon, 2006.

106. DeChaine EG, Bates AE, Shank TM, Cavanaugh CM. Off-axis symbiosis found: Characterization and biogeography of bacterial symbionts of *Bathymodiolus* mussels from Lost City hydrothermal vents. Environ Microbiol. 2006; 8: 1902–12.

107. Salerno JL, Macko SA, Hallam SJ, Bright M, Won Y-J, McKiness Z, et al. Characterization of symbiont populations in life-history stages of mussels from chemosynthetic environments. Biol Bull. 2005; 208: 145–55.

108. Won Y-J, Hallam SJ, O'Mullan GD, Pan IL, Buck KR, Vrijenhoek RC. Environmental acquisition of thiotrophic endosymbionts by deep-sea mussels of the genus *Bathymodiolus*. Appl Environ Microbiol. 2003; 69: 6785–92.

109. Fontanez KM, Cavanaugh CM. Evidence for horizontal transmission from multilocus phylogeny of deep-sea mussel (Mytilidae) symbionts. Environ Microbiol. 2014; 16: 3608–21.

110. Le Pennec M, Diouris M, Herry A. Endocytosis and lysis of bacteria in gill epithelium of *Bathymodiolus thermophilus*, *Thyasira flexuosa* and *Lucinella divaricata* (bivalve, molluscs). J Shellfish Res. 1988; 7: 483–9.

111. Kádár E, Bettencourt R, Costa V, Santos RS, Lobo-da-Cunha A, Dando P. Experimentally induced endosymbiont loss and re-acquirement in the hydrothermal vent bivalve *Bathymodiolus azoricus*. J Exp Mar Biol Ecol. 2005; 318: 99–110.

112. Bright M, Bulgheresi S. A complex journey: Transmission of microbial symbionts. Nat Rev Microbiol. 2010; 8: 218–30.

113. Gosling E. Marine Bivalve Molluscs. 2nd ed. Hoboken, New Jersey: Wiley-Blackwell; 2015.

114. Distel DL, Lee HK, Cavanaugh CM. Intracellular coexistence of methano- and thioautotrophic bacteria in a hydrothermal vent mussel. PNAS. 1995; 92: 9598–602.

115. Geier B. Correlative mass spectrometry imaging of animal–microbe symbioses. PhD Thesis. University of Bremen, 2020.

116. Duperron S, Quiles A, Szafranski KM, Léger N, Shillito B. Estimating symbiont abundances and gill surface areas in specimens of the hydrothermal vent mussel *Bathymodiolus puteoserpentis* maintained in pressure vessels. Front Mar Sci. 2016; 3: 16.

117. Demopoulos AWJ, McClain-Counts JP, Bourque JR, Prouty NG, Smith BJ, Brooke S, et al. Examination of *Bathymodiolus childressi* nutritional sources, isotopic niches, and food-web linkages at two seeps in the US Atlantic margin using stable isotope analysis and mixing models. Deep Sea Res Part I Oceanogr Res Pap. 2019; 148: 53–66.

118. Duperron S, Bergin C, Zielinski F, Blazejak A, Pernthaler A, McKiness ZP, 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. 2006; 8: 1441–7.

119. Assié A, Borowski C, van der Heijden K, Raggi L, Geier B, Leisch N, et al. A specific and widespread association between deep-sea *Bathymodiolus* mussels and a novel family of Epsilonproteobacteria. Environ Microbiol Rep. 2016; 8: 805–13.

120. Zielinski FU, Pernthaler A, Duperron S, Raggi L, Giere O, Borowski C, et al. Widespread occurrence of an intranuclear bacterial parasite in vent and seep bathymodiolin mussels. Environ Microbiol. 2009; 11: 1150–67.

121. Miguel Ángel González Porras. Molecular biology and evolution of the bacterial intranuclear parasite *Ca.* Endonucleobacter. PhD Thesis. University of Bremen, 2020.

122. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. Nature. 2007; 449: 804–10.

123. Duperron S. The diversity of deep-sea mussels and their bacterial symbioses. In: Kiel S, editor. The Vent and Seep Biota, Springer Netherlands; 2010, p. 137–67.

124. Ponnudurai R, Sayavedra L, Kleiner M, Heiden SE, Thürmer A, Felbeck H, et al. Genome sequence of the sulfur-oxidizing *Bathymodiolus thermophilus* gill endosymbiont. Stand Genomic Sci. 2017; 12: 50.

125. Ponnudurai R, Kleiner M, Sayavedra L, Petersen JM, Moche M, Otto A, et al. Metabolic and physiological interdependencies in the *Bathymodiolus azoricus* symbiosis. ISME J. 2017; 11: 463–77.

126. Ansorge R, Romano S, Sayavedra L, González Porras MÁ, Kupczok A, Tegetmeyer HE, et al. Functional diversity enables multiple symbiont strains to coexist in deep-sea mussels. Nat Microbiol. 2019; 4: 2487–97.

127. Kochevar RE, Childress JJ, Fisher CR, Minnich E. The methane mussel: Roles of symbiont and host in the metabolic utilization of methane. Mar Biol. 1992; 112: 389–401.

128. Petersen JM, Dubilier N. Methanotrophic symbioses in marine invertebrates. Environ Microbiol Rep. 2009; 1: 319–35.

129. Ebert D. The epidemiology and evolution of symbionts with mixed-mode transmission. Annu Rev Ecol Evol Syst. 2013; 44: 623–43.

130. Russell SL, Corbett-Detig RB, Cavanaugh CM. Mixed transmission modes and dynamic genome evolution in an obligate animal-bacterial symbiosis. ISME J. 2017; 11: 1359–71.

131. Fisher RM, Henry LM, Cornwallis CK, Kiers ET, West SA. The evolution of host-symbiont dependence. Nat Commun. 2017; 8: 15973.

132. Sato Y, Wippler J, Wentrup C, Ansorge R, Sadowski M, Gruber-Vodicka H, et al. Fidelity varies in the symbiosis between a gutless marine worm and its microbial consortium. BioRxiv. 2021: 2021.01.30.428904.

133. de Vienne DM, Refrégier G, López-Villavicencio M, Tellier A, Hood ME, Giraud T. Cospeciation vs host-shift speciation: Methods for testing, evidence from natural associations and relation to coevolution. New Phytol. 2013; 198: 347–85.

134. Moran NA, Bennett GM. The tiniest tiny genomes. Annu Rev Microbiol. 2014; 68: 195–215.

135. Jäckle O, Seah BKB, Tietjen M, Leisch N, Liebeke M, Kleiner M, et al. Chemosynthetic symbiont with a drastically reduced genome serves as primary energy storage in the marine flatworm *Paracatenula*. PNAS. 2019; 116: 8505–14.

136. Ikuta T, Igawa K, Tame A, Kuroiwa T, Kuroiwa H, Aoki Y, et al. Surfing the vegetal pole in a small population: extracellular vertical transmission of an "intracellular" deep-sea clam symbiont. R Soc Open Sci. 2016; 3: 160130.

137. Russell SL, Chappell L, Sullivan W. Chapter Nine - A symbiont's guide to the germline. In: Lehmann R, editor. Current Topics in Developmental Biology, vol. 135, Academic Press; 2019, p. 315–51.

138. Moran NA, Wernegreen JJ. Lifestyle evolution in symbiotic bacteria: Insights from genomics. Trends Ecol Evol. 2000; 15: 321–6.

139. Nyholm SV, McFall-Ngai M. The winnowing: establishing the squid– *vibrio* symbiosis. Nat Rev Microbiol. 2004; 2: 632–42.

140. Wentrup C, Wendeberg A, Schimak M, Borowski C, Dubilier N. Forever competent: deep-sea bivalves are colonized by their chemosynthetic symbionts throughout their lifetime. Environ Microbiol. 2014; 16: 3699–713.

141. Yong E. Gut microbes keep species apart. Nature News. 2013.

142. Mayr E. Systematics and the origin of species, from the viewpoint of a zoologist. Cambridge, Massachusetts: Harvard University Press; 1942.

143. Anderson E, Hubricht L. Hybridization in *Tradescantia*. III. The evidence for introgressive hybridization. Am J Bot. 1938; 25: 396–402.

144. Martinsen GD, Whitham TG, Turek RJ, Keim P. Hybrid populations selectively filter gene introgression between species. Evolution. 2001; 55: 1325–35.

145. Arnold ML, Hodges SA. Are natural hybrids fit or unfit relative to their parents? Trends Ecol Evol. 1995; 10: 67–71.

146. Twyford AD, Ennos RA. Next-generation hybridization and introgression. Heredity. 2012; 108: 179–89.

147. Anderson E. Introgressive hybridization. John Wiley and Sons, Inc., New York, Chapman and Hall, Ltd., London; 1949.

148. Arnold ML. Natural hybridization as an evolutionary process. Annu Rev Ecol Evol Syst. 1992; 23: 237–61.

149. Gardner JPA. Hybridization in the sea. In: Blaxter J, Southward A, editors. Advances in Marine Biology, vol. 31, Academic Press; 1997, p. 1–78.

150. Harrison RG. Hybrid zones: Windows on evolutionary process. Oxford Surveys in Evolutionary Biology. 1990; 7: 69–128.

151. Bert TM, Harrison RG. Hybridization in western Atlantic stone crabs (genus *Menippe*): evolutionary history and ecological context influence species interactions. Evolution. 1988; 42: 528–44.

152. Gardner JPA. The structure and dynamics of naturally-occurring hybrid *Mytilus edulis* Linnaeus, 1758 and *Mytilus galloprovincialis* Lamarck, 1819 (Bivalvia, Mollusca) populations-review and interpretation. Arch Hydrobiol. 1994; 99: 37–71.

153. Skibinski DOF, Ahmad M, Beardmore JA. Genetic evidence for naturally occurring hybrids between *Mytilus edulis* and *Mytilus galloprovincialis*. Evolution. 1978; 32: 354–64.

154. Fisher CR, Brooks JM, Vodenichar JS, Zande JM, Childress JJ, Burke Jr. RA. The co-occurrence of methanotrophic and chemoautotrophic sulfur-oxidizing bacterial symbionts in a deep-sea mussel. Mar Ecol. 1993; 14: 277–89.

155. Fiala-Médioni A, McKiness Z, Dando P, Boulegue J, Mariotti A, Alayse-Danet A, et al. Ultrastructural, biochemical, and immunological characterization of two populations of

the mytilid mussel *Bathymodiolus azoricus* from the Mid-Atlantic Ridge: Evidence for a dual symbiosis. Mar Biol. 2002; 141: 1035–43.

156. Duperron S, Sibuet M, MacGregor BJ, Kuypers MMM, Fisher CR, Dubilier N. Diversity, relative abundance and metabolic potential of bacterial endosymbionts in three *Bathymodiolus* mussel species from cold seeps in the Gulf of Mexico. Environ Microbiol. 2007; 9: 1423–38.

157. Ikuta T, Takaki Y, Nagai Y, Shimamura S, Tsuda M, Kawagucci S, et al. Heterogeneous composition of key metabolic gene clusters in a vent mussel symbiont population. ISME J. 2016; 10: 990–1001.

158. Picazo DR, Dagan T, Ansorge R, Petersen JM, Dubilier N, Kupczok A. Horizontally transmitted symbiont populations in deep-sea mussels are genetically isolated. ISME J. 2019; 13: 2954–68.

159. Ghoul M, Mitri S. The ecology and evolution of microbial competition. Trends Microbiol. 2016; 24: 833–45.

160. Hardin G. The competitive exclusion principle. Science. 1960; 131: 1292–7.

161. Ansorge R, Romano S, Sayavedra L, Rubin-Blum M, Gruber-Vodicka HR, Scilipoti S, et al. The hidden pangenome: comparative genomics reveals pervasive diversity in symbiotic and free-living sulfur-oxidizing bacteria. BioRxiv. 2020: 2020.12.11.421487.

162. Van Rossum T, Ferretti P, Maistrenko OM, Bork P. Diversity within species: interpreting strains in microbiomes. Nat Rev Microbiol. 2020; 18: 491–506.

163. Marshall Lalish K. Elucidating the metabolic activities of SUP05, an abundant group of marine sulfur oxidizing gamma-proteobacteria. PhD Thesis. University of Washington, 2015.

164. Callbeck CM, Lavik G, Ferdelman TG, Fuchs B, Gruber-Vodicka HR, Hach PF, et al. Oxygen minimum zone cryptic sulfur cycling sustained by offshore transport of key sulfur oxidizing bacteria. Nat Commun. 2018; 9: 1729.

165. Stewart FJ, Ulloa O, DeLong EF. Microbial metatranscriptomics in a permanent marine oxygen minimum zone. Environ Microbiol. 2012; 14: 23–40.

166. Murillo AA, Ramírez-Flandes S, DeLong EF, Ulloa O. Enhanced metabolic versatility of planktonic sulfur-oxidizing γ-proteobacteria in an oxygen-deficient coastal ecosystem. Front Mar Sci. 2014; 1: 18.

167. Shah V, Morris RM. Genome sequence of "*Candidatus* Thioglobus autotrophica" strain EF1, a chemoautotroph from the SUP05 clade of marine Gammaproteobacteria. Genome Announc. 2015; 3: e01156-15.

168. Marshall KT, Morris RM. Genome sequence of "*Candidatus* Thioglobus singularis" strain PS1, a mixotroph from the SUP05 clade of marine Gammaproteobacteria. Genome Announc. 2015; 3: e01155-15.

169. Kuwahara H, Yoshida T, Takaki Y, Shimamura S, Nishi S, Harada M, et al. Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*. Curr Biol. 2007; 17: 881–6.

170. Newton ILG, Woyke T, Auchtung TA, Dilly GF, Dutton RJ, Fisher MC, et al. The *Calyptogena magnifica* chemoautotrophic symbiont genome. Science. 2007; 315: 998–1000.

171. Crépeau V, Cambon Bonavita M-A, Lesongeur F, Randrianalivelo H, Sarradin P-M, Sarrazin J, et al. Diversity and function in microbial mats from the Lucky Strike hydrothermal vent field. FEMS Microbiol Ecol. 2011; 76: 524–40.

172. Faure B, Jollivet D, Tanguy A, Bonhomme F, Bierne N. Speciation in the deep sea: multi-locus analysis of divergence and gene flow between two hybridizing species of hydrothermal vent mussels. PLOS ONE. 2009; 4: e6485.

173. Benson AK, Kelly SA, Legge R, Ma F, Low SJ, Kim J, et al. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. PNAS. 2010; 107: 18933–8.

174. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, et al. Environment dominates over host genetics in shaping human gut microbiota. Nature. 2018; 555: 210–5.

175. Davenport ER. Elucidating the role of the host genome in shaping microbiome composition. Gut Microbes. 2016; 7: 178–84.

176. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature. 2012; 486: 222–7.

177. Ho P-T, Park E, Hong SG, Kim E-H, Kim K, Jang S-J, et al. Geographical structure of endosymbiotic bacteria hosted by *Bathymodiolus* mussels at eastern Pacific hydrothermal vents. BMC Evol Biol. 2017; 17: 121.

178. Harmer TL, Rotjan RD, Nussbaumer AD, Bright M, Ng AW, DeChaine EG, et al. Free-living tube worm endosymbionts found at deep-sea vents. Appl Environ Microbiol. 2008; 74: 3895–8.

List of publications

Manuscripts included in this thesis:

1. **Deep-sea mussels from a hybrid zone on the Mid-Atlantic Ridge host genetically indistinguishable symbionts**
   Merle Ücker, Rebecca Ansorge, Yui Sato, Lizbeth Sayavedra, Corinna Breusing, Nicole Dubilier

   Manuscript published in ISME Journal (2021-05-10, available under https://doi.org/0.1038/s41396-021-00927-9)

2. **Chasing free-living stages of mussel symbionts– Metagenomic insights from the environment to the host**
   Merle Ücker, Rebecca Ansorge, Yui Sato, Ga Yan Grace Ho, Christian Borowski, Nicole Dubilier

   Manuscript in preparation

3. **Mitonuclear discordance suggests long-distance migration and mitochondrial introgression of deep-sea mussels along the Mid-Atlantic Ridge**
   Merle Ücker, Yui Sato, Rebecca Ansorge, Anne Kupczok, Maximilian Franke, Harald Gruber-Vodicka, Nicole Dubilier

   Manuscript in preparation

Contributions to manuscripts not included in this thesis:

How to survive without your symbionts? The vertical migration of *Bathymodiolus* larvae
Maximilian Franke, Manuel Kleiner, Marlene Jensen, **Merle Ücker**, Harald Gruber-Vodicka, Emilia Maggie Sogin, Nicole Dubilier, Nikolaus Leisch (author order is not fixed)

hybrid

[ˈhʌɪbrɪd] *noun*

the offspring of two plants or animals of different species or varieties

Chapter II | Symbionts in a mussel hybrid zone

Deep-sea mussels from a hybrid zone on the Mid-Atlantic Ridge host genetically indistinguishable symbionts

Merle Ücker[1,2], Rebecca Ansorge[1,3], Yui Sato[1], Lizbeth Sayavedra[1,3], Corinna Breusing[4], Nicole Dubilier[1,2,*]

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

[2]MARUM – Center for Marine Environmental Sciences of the University of Bremen, Bremen, Germany

[3]Quadram Institute Bioscience, Norwich, Norfolk, United Kingdom

[4]University of Rhode Island, Graduate School of Oceanography, Narragansett, RI, United States of America

*Corresponding author

ARTICLE

Deep-sea mussels from a hybrid zone on the Mid-Atlantic Ridge host genetically indistinguishable symbionts

Merle Ücker [1,2] · Rebecca Ansorge [1,3] · Yui Sato[1] · Lizbeth Sayavedra [1,3] · Corinna Breusing [4] · Nicole Dubilier [1,2]

Abstract

The composition and diversity of animal microbiomes is shaped by a variety of factors, many of them interacting, such as host traits, the environment, and biogeography. Hybrid zones, in which the ranges of two host species meet and hybrids are found, provide natural experiments for determining the drivers of microbiome communities, but have not been well studied in marine environments. Here, we analysed the composition of the symbiont community in two deep-sea, *Bathymodiolus* mussel species along their known distribution range at hydrothermal vents on the Mid-Atlantic Ridge, with a focus on the hybrid zone where they interbreed. In-depth metagenomic analyses of the sulphur-oxidising symbionts of 30 mussels from the hybrid zone, at a resolution of single nucleotide polymorphism analyses of ~2500 orthologous genes, revealed that parental and hybrid mussels (F2–F4 generation) have genetically indistinguishable symbionts. While host genetics does not appear to affect symbiont composition in these mussels, redundancy analyses showed that geographic location of the mussels on the Mid-Atlantic Ridge explained most of the symbiont genetic variability compared to the other factors. We hypothesise that geographic structuring of the free-living symbiont population plays a major role in driving the composition of the microbiome in these deep-sea mussels.

Introduction

The community composition of an animal's microbiome is the product of multiple interacting factors that include the environment, geography and host genetics [1–5]. To which extent host genetics affects microbiome composition is currently a topic of intense debate, in part as high-throughput sequencing is revealing the genetic makeup of host and symbiont populations in ever higher resolution

✉ Nicole Dubilier
  ndubilie@mpi-bremen.de

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

[2] MARUM—Center for Marine Environmental Sciences of the University of Bremen, Bremen, Germany

[3] Quadram Institute Bioscience, Norwich, Norfolk, UK

[4] University of Rhode Island, Graduate School of Oceanography, Narragansett, RI, USA

[6–8]. Animal hybrids are useful for assessing the effects of host genotype on microbiomes [9]. Studies of lab-reared animal hybrids, such as wasps [10], fish [11–13], and mice [14, 15] found that these hosts had different gut microbiota compositions than their parental species, based on sequencing of the microbial 16S rRNA gene. These altered gut microbiomes of hybrids affected the fitness of some hosts, suggesting that microbiomes play an important role in determining species barriers [10]. Studies on lab-reared hosts cannot, however, fully reflect the environmental conditions animals experience in their natural habitat. Hybrid zones, in which parental species interbreed and produce hybrid offspring, are excellent natural experiments for teasing apart the impact of host genotype, environment and geographic distance on microbiome composition. Yet surprisingly few studies have investigated the microbiota of hybrids from the wild, and these have yielded mixed results. For example, in a hybrid zone of the European house mouse, the composition of gut microbiota of hybrids differed from that of their parental species [15]. In contrast, in African baboons, there were no significant differences between hybrids and their parental species, and gut community composition was best

M. Ücker et al.

explained by the environment [16]. To date, all hybrid microbiome studies, whether on lab-reared animals or those from the wild, have been based on the sequencing of only a few microbial genes, with the vast majority of analyses based on the 16S rRNA gene, or a variable region of this gene. These studies were therefore limited to determining microbial community composition at the genus level or higher, and could not distinguish closely related species or strains.

Almost nothing is known about the microbial communities of hosts from marine hybrid zones, despite the pervasiveness of such zones in many regions of the oceans. Hydrothermal vents on the Mid-Atlantic Ridge (MAR), an underwater mountain range extending from the Arctic to the Southern Ocean, provide an ideal setting for investigating the microbiomes of hosts in natural hybrid zones. Many of the vents on the MAR are dominated by *Bathymodiolus* mussels that live in a nutritional symbiosis with chemosynthetic bacteria. Two mussel species colonise the northern MAR, *Bathymodiolus azoricus*, which is found at vents from 38° N to 36° N, and *Bathymodiolus puteoserpentis*, which inhabits vents further south from 23° N to 13° N. A hybrid zone between these two relatively young host species, with an estimated splitting time of 8.4 Mya [17], occurs at the Broken Spur vent field at 29° N on the MAR, where *B. puteoserpentis* co-occurs with hybrids between *B. azoricus* and *B. puteoserpentis* [18–20]. The sulphur-oxidising (SOX) symbionts of *B. azoricus* and *B. puteo-serpentis* belong to a gammaproteobacterial clade within the *Thioglobaceae*, and co-occur in the mussel gills with methane-oxidising symbionts, which belong to a gammaproteobacterial clade within the *Methylomonaceae* [21]. The relative abundance of these two symbionts in these mussels is assumed to not depend on host genetics, but rather the availability of the energy sources these symbionts use in their environment [21, 22].

The symbionts of bathymodiolin mussels are transmitted horizontally from the environment to juvenile mussels, yet each mussel species harbours a highly specific symbiont community [23–25]. This specificity suggests that the genetics of bathymodiolin mussels plays an important role in determining symbiont composition. In this study, we took advantage of the natural hybrid zone of *Bathymodiolus* mussels at the Broken Spur vent field to investigate how host genotype, geographic distance, and the vent environment affect the composition of their SOX symbionts. The recent discovery of a high diversity of SOX symbiont strains in *Bathymodiolus* from the MAR, with as many as 16 strains co-occurring in single *Bathymodiolus* mussels [22, 26, 27], made it critical to resolve genetic differences at the strain level of the SOX symbiont community (strain is defined here as suggested by Van Rossum et al. [28], as subordinate to subspecies, in our study >99% average

nucleotide identity). We achieved this resolution through multilocus phylogeny, genome-wide gene profiling, and single nucleotide polymorphism (SNP)-based population differentiation analyses of 30 *Bathymodiolus* hybrid and parental individuals collected in 1997 and 2001 at the Broken Spur vent field.

Materials and methods

A detailed description of samples (Supplementary Table S1) and methods is available in the Supplementary Information and an overview of the analyses of SOX symbionts used in this study is provided in Supplementary Table S2. Data files and scripts used for the analyses can be found in the GitHub repository (https://github.com/muecker/Symbionts_in_a_ mussel_hybrid_zone).

Broken Spur parental mussels (13 *B. puteoserpentis*) and hybrids (17 F2–F4 generation hybrids, see Supplement) were identified as described previously [20, 29] (no parental *B. azoricus* were found at Broken Spur). Briefly, mussels were genotyped based on 18 species-diagnostic markers and identified as parental or hybrid mussels using bioinformatic analyses of population structure, admixture and introgression (Supplementary Table S3). After DNA extraction and sequencing, we assembled metagenomes per mussel individual from Illumina short-read sequences. Metagenome-assembled genomes (MAGs) of the SOX symbionts from each mussel individual were binned (for statistics of symbiont MAGs, see Supplementary Table S4), representing the consensus of all SOX symbiont strains in each host individual.

To evaluate genetic differences between symbionts from the northern MAR at the level of bacterial subspecies (sensu Van Rossum et al. [28], here above 97% average nucleotide identity), we used 171 single-copy, gammaproteobacterial marker genes for phylogenomic analysis of the SOX symbiont MAGs and their closest symbiotic relatives, e.g., symbionts of *B. azoricus* from vents north of Broken Spur and *B. puteoserpentis* mussels from vents south of Broken Spur, and free-living relatives (see Supplementary Table S5). To understand which factors affect symbiont composition on the strain level at the northern MAR, we assessed the influence of geographic distance, host species, vent type (basaltic versus ultramafic rock) and depth on SOX symbiont allele frequencies using redundancy analysis (RDA). We analysed Broken Spur symbiont MAGs at the genome-wide level by comparing their average nucleotide identities (ANI) to resolve differences on the subspecies level. To resolve strain-level differences between SOX symbionts from Broken Spur, we analysed pairwise F_{ST} values based on SNPs in 2496 orthologous genes from Broken Spur SOX symbiont MAGs. To identify genes that

differed between hybrid and parental symbiont populations, we analysed the presence/absence and differential abundance of these orthologues, and further investigated pairwise F_{ST} values of all 2496 orthologous genes.

Results and discussion

Phylogenomic analysis of 171 single-copy genes revealed the presence of two SOX symbiont subspecies, one specific to *B. azoricus* from the more northern vents Menez Gwen, Lucky Strike and Rainbow, and one specific to *B. puteoserpentis* from the vents further south, Logatchev and Semenov (Fig. 1A, C).

This substantiates previous analyses based on sequencing of the 16S rRNA gene and internal transcribed spacer that these two *Bathymodiolus* species harbour different SOX symbiont subspecies of the same bacterial species [21, 25, 30]. Our phylogenomic analyses revealed that all *Bathymodiolus* individuals from Broken Spur harboured a third SOX symbiont subspecies (Fig. 1A, C). This new subspecies is most closely related to the *B. puteoserpentis* SOX symbiont subspecies from mussels collected south of Broken Spur. These two symbiont subspecies form a sister group to the SOX symbiont subspecies of *B. azoricus* collected at vents north of Broken Spur.

To evaluate if the SOX symbionts of Broken Spur parental and hybrid *Bathymodiolus* differed, we compared their ANI and estimated genomic differentiation (F_{ST}) based on ~2500 orthologous genes (for more information on the host, see Supplementary section 2 and Supplementary Table S3). Symbiont ANI values ranged from 96.7 to 99.9% with a median of 99.7%. We found no correlation between symbiont differentiation and the sampling year or genetic differentiation of the mussels (Mantel test of symbiont ANI and F_{ST} versus sampling year and host pairwise genetic distances based on 18 SNP markers, Fig. 2). Our analyses of SNPs per individual gene revealed that not even one of the ~2500 orthologous genes had significantly differing F_{ST} values (Mann-Whitney U test of F_{ST} per gene between versus within symbionts of hybrids and parental mussels). Similarly, there was also no significant difference between hybrids and parental individuals in the abundance of symbiont genes (based on a general linear model and Kruskal-Wallace test in ALDEx2 using Benjamini-Hochberg cor-rected p value < 0.05) or their presence/absence. These results indicate that the composition and gene repertoire of SOX symbionts in Broken Spur mussels was highly similar or identical in hybrids and parental *B. puteoserpentis*. A study of SOX symbionts in hybrids of *Bathymodiolus thermophilus* and *Bathymodiolus ant-arcticus* at 23° S in the eastern Pacific also found that

these could not be distinguished from parental mussels, based on PCR analyses of seven bacterial marker genes in five parental and three hybrid individuals [31].

Our results raise the question at what level of genetic divergence between two host species differences in their symbiont communities evolve. *Bathymodiolus brooksi* and *Bathymodiolus heckerae*, which regularly co-occur in the Gulf of Mexico, harbour different symbiont species that are only distantly related to each other (Fig. 1A, B). These two mussel species have an estimated splitting time of 15.4 Mya [17], and are not known to hybridise. More closely related hosts, such as *B. thermophilus* and *B. antarcticus* (estimated splitting time of 2.5–5.3 Mya [32]), and *B. azoricus* and *B. puteoserpentis* (estimated splitting time of 8.4 Mya [17]), produce fertile hybrids [19, 33], and have genetically indistinguishable symbionts in zones where they hybridise. This suggests that specificity at the symbiont species level in these horizontally transmitted symbioses evolves only after extended divergence times of tens of millions of years, during which these hosts become genetically dissimilar enough to evolve specific symbiont selection mechanisms.

While *Bathymodiolus* mussels on the northern MAR host the same SOX symbiont species, our phylogenomic analyses revealed clear genetic differentiation in three SOX symbiont subspecies: *B. azoricus*, *B. puteoserpentis* and Broken Spur subspecies (Fig. 1). To better understand the factors that drive this symbiont differentiation, we tested which influence host species, geographic distance, vent type (basaltic versus ultramafic rock) and depth have on symbiont allele frequencies. All variables were highly collinear. For example, the water depth of the vents studied here increases with geographic distance, from 800 m at 37.8° N, to 3050 m at 14.7° N (only the southernmost vent at 13.5° N and 2320 m depth interrupted this pattern). When the four variables were considered individually, geographic distance explained 13% of symbiont differentiation, while the three other variables water depth, host species and vent type each explained 0.2%, 0.0%, and 0.2%, respectively. When interaction effects between the four variables were considered, geographic distance and interactions involving this variable explained 45% of symbiont differentiation, while the three other variables water depth, host species, vent type and the interactions with these explained 14%, 12%, and 9%, respectively (p value < 0.001, Fig. 3).

There are at least three explanations for why geographic distance has such a large effect on the SOX symbiont composition of *Bathymodiolus* mussels from the northern MAR. The first is that with increasing geographic distance, environmental differences between vents become larger and these environmental differences affect symbiont composition (genetic isolation by environment versus distance

M. Ücker et al.

Fig. 1 Phylogenetic relationships of *Bathymodiolus* SOX symbionts and their mussel hosts. **A** Overview tree based on 171 single-copy marker genes. The tree was reconstructed based on a 36,949 bp alignment using the LG + F + R6 amino acid model and 1000 samples for ultrafast bootstrap with IQ-TREE. The *Bathymodiolus* SOX symbionts from the northern Mid-Atlantic Ridge (blue, yellow and pink) form a clade within the gammaproteobacterial SUP05 clade. *Thiomicrospira* spp. and *Ca.* T. singularis PS1 were used as outgroups. MAG accessions are listed in Supplementary Table S5. **B** Host phylogeny based on published mitochondrial cytochrome oxidase subunit I (COI) sequences. "*B.*" *childressi* was used as an outgroup. Sequence accessions are listed in the Supplement ("1.3 Reconstruction of *Bathymodiolus* phylogeny"). **C** Zoom in of sequences shown in box in (**A**): Phylogeny of *Bathymodiolus* SOX symbionts from vents on the northern Mid-Atlantic Ridge. Black and white circles indicate the vent type (basaltic or ultramafic rock), colours correspond to vent sites shown in the map. Hybrid individuals from Broken Spur are marked with a black star. *Bathymodiolus* SOX symbionts from the vent sites Clueless (5° S) and Lilliput (9° S) were used as outgroups. B. *Bathymodiolus*, MAR Mid-Atlantic Ridge, GoM Gulf of Mexico, EPR East Pacific Rise, J Japan, NP North Pacific (colour figure online).
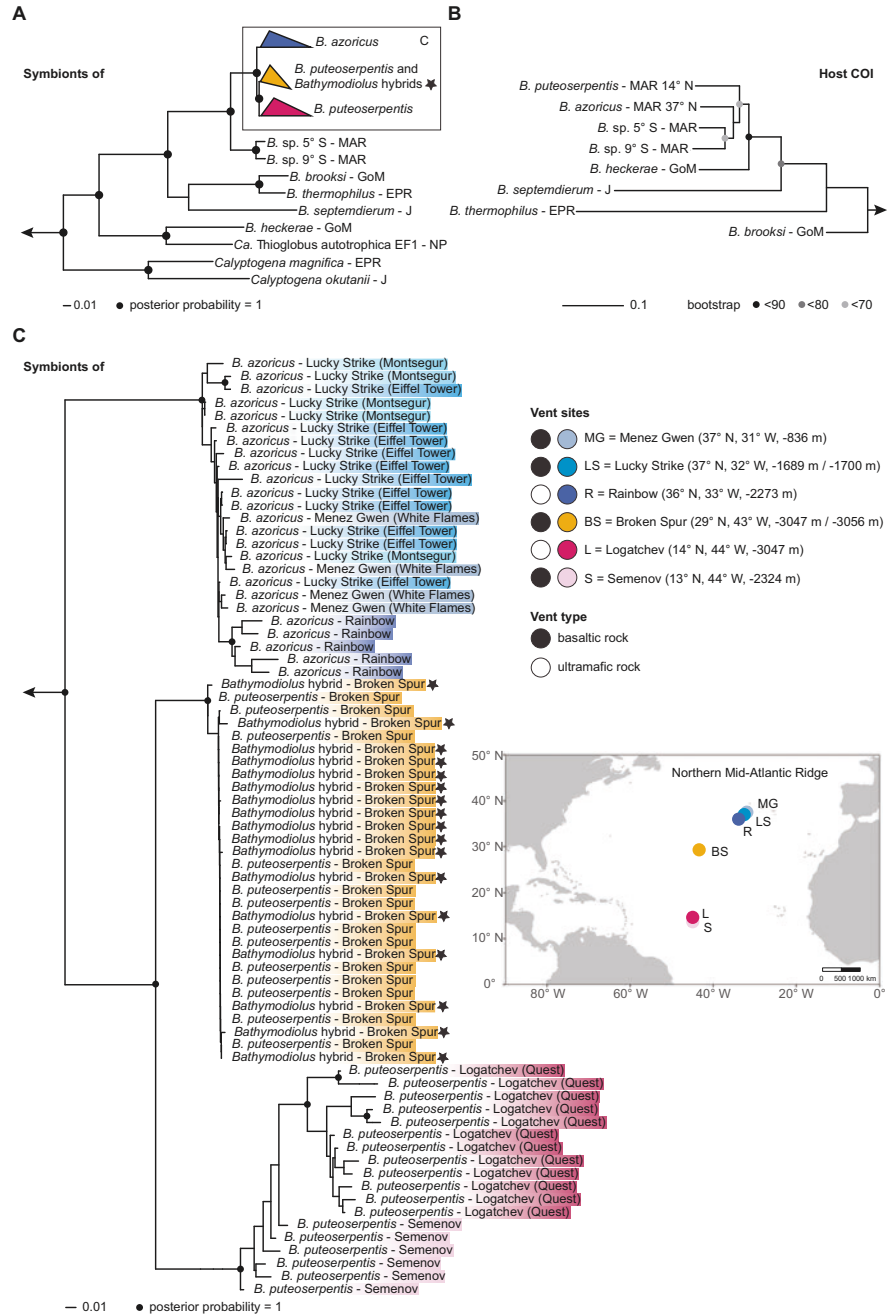
[34]). *Bathymodiolus* mussels acquire their symbionts horizontally from the environment, presumably when the larvae settle on the seafloor [25, 35], and it would be advantageous for the mussels if they selected symbionts that are best adapted to local environmental conditions. We tested the effect of vent type based on one of the key environmental variables at hydrothermal vents, basaltic and ultramafic rock. These two rock types have major effects on the biogeochemistry of vent fluids, including the relative concentrations of the symbiotic energy sources sulphide, methane and hydrogen [36]. However, vent type alone explained only 0.2% of symbiont genetic differentiation, similar to another environmental variable water depth, which also only explained 0.2% of symbiont

Deep-sea mussels from a hybrid zone on the Mid-Atlantic Ridge host genetically indistinguishable...
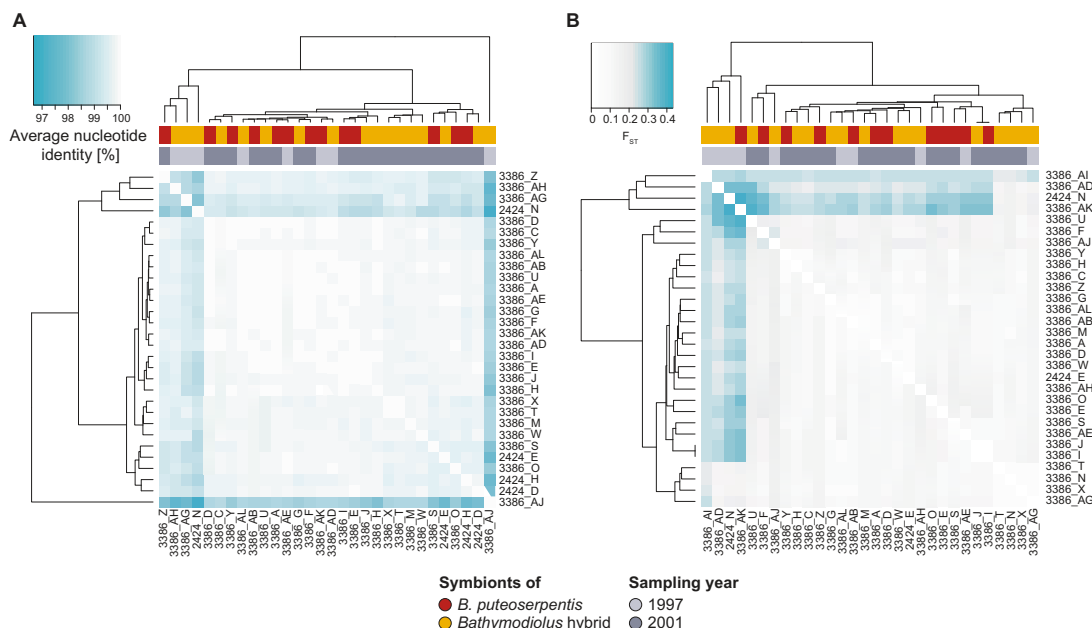
Fig. 2 Genome-wide differentiation of *Bathymodiolus* SOX symbionts at Broken Spur. A Differentiation based on pairwise average nucleotide identity. **B** Differentiation based on pairwise average F_{ST} in 2496 orthologous genes. Colour bars represent host genotypes (red: *B. puteoserpentis*, yellow: hybrids) and the sampling year (light grey: 1997, dark grey: 2001). Turquoise indicates a higher differentiation or more dissimilar genomes. Based on a Mantel test, neither clustering based on ANI (**A**), nor F_{ST} (**B**) correlated with host pairwise genetic distances based on 18 SNP markers (**A** $r = 0.054$, $p = 0.222$, **B** $r = 0.006$, $p = 0.435$) or sampling year (**A** $r = -0.191$, $p = 0.949$, **B** $r = 0.105$, $p = 0.150$) (colour figure online).
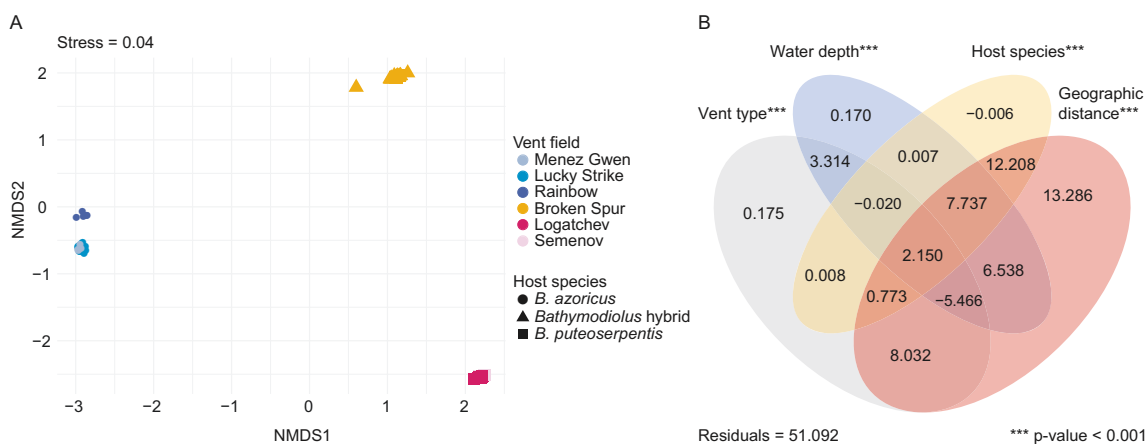
Fig. 3 Differentiation of *Bathymodiolus* SOX symbionts at the northern Mid-Atlantic Ridge and the influence of geographic distance, host species and environmental parameters (vent type and water depth). A NMDS plot of SOX symbiont allele frequencies show clear separation of the three symbiont subspecies at the northern MAR: *B. azoricus*, *B. puteoserpentis*, and Broken Spur symbiont subspecies. Symbionts from Broken Spur cluster together regardless of the species affiliation of their host (hybrid versus *B. puteoserpentis*). Colours correspond to vent fields, shapes to host species. **B** Variation partitioning of explanatory variables used in the RDA (Supplementary Fig. S3). Values are shown in percent. The variables vent type, water depth, host species, and geographic distance and their interaction effects explained 49% of the total variation, with 13% of the variation solely explained by geographic distance. Negative adjusted R values can occur for several reasons, e.g., negative eigenvalues in the model underlying the varpart function in R. P values are based on permutation tests with 1000 repetitions.

differentiation. It is therefore unlikely that environmental differences between vents underlie the symbiont population structures we observed in this study.

The second explanation for why geographic distance has such a large effect on the SOX symbiont composition of *Bathymodiolus* mussels is that genetic differences between the hosts increased with geographic distance. However, population genetic analyses of *B. azoricus* and *B. puteoserpentis* from the same vents as in our study indicated no genetic structuring within each of these host species [29]. This indicates that host genetics did not play a major role in structuring the SOX symbiont composition. The third, and most likely explanation is that the free-living pool of SOX symbionts is geographically structured. At Broken Spur, hybrids and *B. puteoserpentis* host genetically indistinguishable symbionts, and these differed from the symbionts of *B. azoricus* and *B. puteoserpentis* from vents to the north and south of Broken Spur. This indicates that in these two closely related host species, geographic location but not host genetics drives the composition of their SOX symbiont communities. Furthermore, the environment, based on vent type, cannot explain why the mussels at Broken Spur had symbionts that are genetically distinct from other vent sites. Broken Spur is basalt-hosted, while the vents to the north are both basalt- (Menez Gwen and Lucky Strike) and ultramafic-hosted (Rainbow). Yet the symbionts from the vents to the north of Broken Spur are more closely related to each other than to the symbionts of Broken Spur mussels (Fig. 1C).

The validity of these three explanations could be tested in future studies by sampling the free-living microbial populations at hydrothermal vents on the MAR. This is, however, not as simple as it sounds because of multiple challenges including obtaining representative samples from the immediate environment of *Bathymodiolus* mussels, collecting environmental data at scales relevant to the microbial population, and characterising the free-living symbiont population at sufficiently high resolution.

Understanding the biogeography of the free-living stages of microbial symbionts and other as yet uncultured microorganisms is currently one of the biggest challenges in microbial ecology. While there is evidence that 'everything is everywhere, but the environment selects' [37, 38], there is also increasing data showing that dispersal limitation shapes the biogeography of marine microorganisms [39, 40]. Almost nothing is known about the biogeography of uncultivable marine microorganisms at the subspecies or strain level, as most species are rarely abundant enough to allow phylogenetic analyses at such high resolution. Advances in high-throughput short-read, and particularly long-read sequencing, coupled with bioinformatic methods for revealing genetic structuring of microbial populations, are now providing us with the tools for resolving the

intraspecific diversity of environmental microorganisms. Our study highlights the importance of gaining a better understanding of the free-living community of microbial symbionts to disentangle the genetic, environmental, and geographic factors that contribute to the ecological and evolutionary success of animal–microbe associations in which the symbionts are acquired from the environment.

Data availability

Sequence data (metagenomes and symbiont MAGs) are available in the European Nucleotide Archive (ENA) at EMBL-EBI under project accession number PRJEB36976. The data, together with their metadata, were deposited using the data brokerage service of the German Federation for Biological Data (GFBio [41]), with the standard information on sequence data provided as recommended [42].

Code availability

Additional data files and scripts used in this study are available in the GitHub repository (https://github.com/muecker/Symbionts_in_a_mussel_hybrid_zone).

Compliance with ethical standards

Deep-sea mussels from a hybrid zone on the Mid-Atlantic Ridge host genetically indistinguishable…

References

1. Benson AK, Kelly SA, Legge R, Ma F, Low SJ, Kim J, et al. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. PNAS. 2010;107:18933–8.

2. Davenport ER. Elucidating the role of the host genome in shaping microbiome composition. Gut Microbes. 2016;7:178–84.

3. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, et al. Environment dominates over host genetics in shaping human gut microbiota. Nature. 2018;555:210–5.

4. Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. Nat Rev Microbiol. 2011;9:279–90.

5. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature. 2012;486:222–7.

6. Di Bella JM, Bao Y, Gloor GB, Burton JP, Reid G. High throughput sequencing methods and analysis for microbiome research. J Microbiol Methods. 2013;95:401–14.

7. Ellegren H. Genome sequencing and population genomics in non-model organisms. Trends Ecol Evol. 2014;29:51–63.

8. Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet. 2003;4:981–94.

9. Lim SJ, Bordenstein SR. An introduction to phylosymbiosis. Proc R Soc B. 2020;287:20192900.

10. Brucker RM, Bordenstein SR. The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus *Nasonia*. Science. 2013;341:667–9.

11. Li W, Liu J, Tan H, Yang C, Ren L, Liu Q, et al. Genetic effects on the gut microbiota assemblages of hybrid fish from parents with different feeding habits. Front. Microbiol. 2018;9:2972.

12. Rennison DJ, Rudman SM, Schluter D. Parallel changes in gut microbiome composition and function during colonization, local adaptation and ecological speciation. Proc R Soc B. 2019;286:20191911.

13. Sevellec M, Laporte M, Bernatchez A, Derome N, Bernatchez L. Evidence for host effect on the intestinal microbiota of whitefish (*Coregonus* sp.) species pairs and their hybrids. Ecol Evol. 2019;9:11762–74.

14. Korach-Rechtman H, Freilich S, Gerassy-Vainberg S, Buhnik-Rosenblau K, Danin-Poleg Y, Bar H, et al. Murine genetic background has a stronger impact on the composition of the gut microbiota than maternal inoculation or exposure to unlike exogenous microbiota. Appl Environ Microbiol. 2019;85:e00826–19.

15. Wang J, Kalyan S, Steck N, Turner LM, Harr B, Künzel S, et al. Analysis of intestinal microbiota in hybrid house mice reveals evolutionary divergence in a vertebrate hologenome. Nat Commun. 2015;6:1–10.

16. Grieneisen LE, Charpentier MJE, Alberts SC, Blekhman R, Bradburd G, Tung J, et al. Genes, geology and germs: gut microbiota across a primate hybrid zone are explained by site soil properties, not host species. Proc R Soc B. 2019;286:20190431.

17. Faure B, Schaeffer SW, Fisher CR. Species distribution and population connectivity of deep-sea mussels at hydrocarbon seeps in the Gulf of Mexico. PLoS ONE. 2015;10:e0118460.

18. Won Y, Hallam SJ, O'Mullan GD, Vrijenhoek RC. Cytonuclear disequilibrium in a hybrid zone involving deep-sea hydrothermal vent mussels of the genus *Bathymodiolus*. Mol Ecol. 2003;12:3185–90.

19. O'Mullan GD, Maas PA, Lutz RA, Vrijenhoek RC. A hybrid zone between hydrothermal vent mussels (Bivalvia: Mytilidae) from the Mid-Atlantic Ridge. Mol Ecol. 2001;10:2819–31.

20. Breusing C, Vrijenhoek RC, Reusch TBH. Widespread introgression in deep-sea hydrothermal vent mussels. BMC Evol Biol. 2017;17:13.

21. Duperron S, Bergin C, Zielinski F, Blazejak A, Pernthaler A, McKiness ZP, 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. 2006;8:1441–7.

22. Ansorge R, Romano S, Sayavedra L, González Porras MÁ, Kupczok A, Tegetmeyer HE, et al. Functional diversity enables multiple symbiont strains to coexist in deep-sea mussels. Nat Microbiol. 2019;4:2487–97.

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

24. Van Dover CL, German CR, Speer KG, Parson LM, Vrijenhoek RC. Evolution and biogeography of deep-sea vent and seep invertebrates. Science. 2002;295:1253–7.

25. Won Y-J, Hallam SJ, O'Mullan GD, Pan IL, Buck KR, Vrijenhoek RC. Environmental acquisition of thiotrophic endosymbionts by deep-sea mussels of the genus *Bathymodiolus*. Appl Environ Microbiol. 2003;69:6785–92.

26. Ikuta T, Takaki Y, Nagai Y, Shimamura S, Tsuda M, Kawagucci S, et al. Heterogeneous composition of key metabolic gene clusters in a vent mussel symbiont population. ISME J. 2016;10:990–1001.

27. Picazo DR, Dagan T, Ansorge R, Petersen JM, Dubilier N, Kupczok A. Horizontally transmitted symbiont populations in deep-sea mussels are genetically isolated. ISME J. 2019;13:2954–68.

28. Van Rossum T, Ferretti P, Maistrenko OM, Bork P. Diversity within species: interpreting strains in microbiomes. Nat Rev Microbiol. 2020;18:491–506.

29. Breusing C, Biastoch A, Drews A, Metaxas A, Jollivet D, Vrijenhoek RC, et al. Biophysical and population genetic models predict the presence of "phantom" stepping stones connecting Mid-Atlantic Ridge vent ecosystems. Curr Biol. 2016;26:2257–67.

30. DeChaine EG, Bates AE, Shank TM, Cavanaugh CM. Off-axis symbiosis found: characterization and biogeography of bacterial symbionts of *Bathymodiolus* mussels from Lost City hydrothermal vents. Environ Microbiol. 2006;8:1902–12.

31. Ho P-T, Park E, Hong SG, Kim E-H, Kim K, Jang S-J, et al. Geographical structure of endosymbiotic bacteria hosted by *Bathymodiolus* mussels at eastern Pacific hydrothermal vents. BMC Evol Biol. 2017;17:121.

32. Won Y-J, Young CR, Lutz RA, Vrijenhoek RC. Dispersal barriers and isolation among deep-sea mussel populations (Mytilidae: *Bathymodiolus*) from eastern Pacific hydrothermal vents. Mol Ecol. 2003;12:169–84.

33. Johnson SB, Won Y-J, Harvey JB, Vrijenhoek RC. A hybrid zone between *Bathymodiolus* mussel lineages from eastern Pacific hydrothermal vents. BMC Evol Biol. 2013;13:21.

34. Sexton JP, Hangartner SB, Hoffmann AA. Genetic isolation by environment or distance: which pattern of gene flow is most common? Evolution. 2014;68:1–15.

35. Laming SR, Duperron S, Cunha MR, Gaudron SM. Settled, symbiotic, then sexually mature: adaptive developmental anatomy in the deep-sea, chemosymbiotic mussel *Idas modiolaeformis*. Mar Biol. 2014;161:1319–33.

36. Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R, et al. Hydrogen is an energy source for hydrothermal vent symbioses. Nature. 2011;476:176–80.

37. Baas Becking LGM. Geobiologie of inleiding tot de milieukunde. Den Haag: W.P. Van Stockum & Zoon; 1934.

38. Wit RD, Bouvier T. 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? Environ Microbiol. 2006;8:755–8.

39. Dick GJ. The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. Nat Rev Microbiol. 2019;17:271–83.

40. Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, et al. Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol. 2006;4:102–12.

41. Diepenbroek M, Glöckner FO, Grobe P, Güntsch A, Huber R, König-Ries B, et al. Towards an integrated biodiversity and ecological research data management and archiving platform: the German federation for the curation of biological data (GFBio). In: Plödereder E, Grunske L, Schneider E, Ull D, editors. Informatik 2014. Bonn: Gesellschaft für Informatik e.V.; 2014, p. 1711–24.

42. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, et al. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat Biotechnol. 2011;29:415–20.

Supplementary Information

(Formatting deviates from the published version.)

Deep-sea mussels from a hybrid zone on the Mid-Atlantic Ridge host genetically indistinguishable symbionts

Merle Ücker[1,2], Rebecca Ansorge[1,3], Yui Sato[1], Lizbeth Sayavedra[1,3], Corinna Breusing[4], Nicole Dubilier[1,2,*]

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

[2]MARUM – Center for Marine Environmental Sciences of the University of Bremen, Bremen, Germany

[3]Quadram Institute Bioscience, Norwich, Norfolk, United Kingdom

[4]University of Rhode Island, Graduate School of Oceanography, Narragansett, RI, United States of America

*Corresponding author

1. Supplementary materials & methods

1.1 Sampling, DNA extraction and metagenomic sequencing

Mussels were sampled at Broken Spur (29°10.0'N, 43°10.0'W at 3045 – 3056 m water depth) with the deep-sea submersible Alvin during the Atlantis cruises AT-03/03 (1997) and AT-05/03 (2001). Upon recovery on board, small mussels (< 25 mm) were frozen whole at -70°C, while larger mussels were dissected before freezing at -70°C [1]. An overview map of sites analysed in this study was plotted with RStudio v1.3.959 using R v3.6.3 and the packages rnaturalearth v0.1.0, legendMap v1.0 and ggplot2 v3.3.0 [2–5].

Genomic DNA was extracted from gill (metagenomic libraries 3386-A~AL) or a combination of gill, mantel and digestive tissue (called mixed tissue hereafter, metagenomic libraries 2424-A~O) depending on sample availability. DNA extractions were performed with either the AllPrep DNA/RNA/Protein MiniKit (Qiagen, Hilden, Germany) or DNAeasy Blood & Tissue kit (Qiagen, Hilden, Germany) according to the manufacturer's protocols with the following modifications: Prior to extraction, frozen sample pieces (5-10 mm) were homogenised by bead beating in MP Biomedicals Lysing Matrix B using an MP Biomedicals

FastPrep-24 (Thermo Fisher Scientific, Waltham, USA) for 30 s at 6.5 m/s. In the elution step, samples were incubated for 10 min at room temperature before centrifugation. Volumes of eluent were halved, and elution was repeated with the first eluate to maximise DNA yields. Metagenomic libraries were generated with the Nextera DNA Flex Library Prep Kit (Illumina, San Diego, CA, USA) and the Illumina TruSeq DNA Samples Prep Kit (BioLABS, Frankfurt, Germany). Library preparation and sequencing of 150 bp paired-end metagenomic reads were performed by the Max Planck-Genome-centre Cologne, Germany (https://mpgc.mpipz.mpg.de/home/) on HiSeq 2500 or 3000 machines. Details of sampling, DNA extraction and sequencing of all samples used in this study are summarised in Supplementary Table S 1.

1.2   Identification of hybrid host individuals

Mussels were genotyped based on 18 species-diagnostic single-nucleotide polymorphism (SNP) markers and identified as hybrid or parental species with subsequent bioinformatic analyses using 1) STRUCTURE v2.3.4 with strauto v1.0 and CLUMPAK (http://clumpak.tau.ac.il/), 2) introgress v1.22 in RStudio, and 3) NEWHYBRIDS v1.1 [6–13]. The analysis is based on the method developed in [14,15]. Results of all three programmes are shown in Supplementary Table S 3. Classification based on introgress was used as the basis for analyses of mussel symbionts, as it had low misidentification rates for hybrid and parental species as reported in [15]. Since not all programmes supported the classification of backcrosses and their identification was less reliable than for hybrid and parental species in [15], these samples were excluded from further analyses.

1.3   Reconstruction of *Bathymodiolus* phylogeny

Sequences of the mitochondrial marker gene cytochrome c oxidase subunit I or full mitochondrial genomes were downloaded from NCBI (database accessed 2020-02-11) for *B. azoricus* (LN833437), *B. brooksi* (KU597634), *B. heckerae* (KU659139), *B. puteoserpentis* (KU597632), *B.* sp. Lilliput (LN833440), *B.* sp. Clueless (LT674164), *B. septemdierum* (AP014562), *"B." childressi* (ANY30357) and *B. thermophilus* (MK721544) [16]. We aligned the sequences with MUSCLE v3.8.31 [17,18], and reconstructed a phylogenetic tree using IQTREE v1.6.9 with 1000 samples for ultrafast bootstrap and the mtZoa model, which was selected as the best model by Model Finder based on the Bayesian Information Criterion [19–22]. The tree was visualised with iTol v5.5 [23] and edited with Adobe Illustrator 2020 [24].

1.4   Metagenome assembly and symbiont binning

Metagenomic reads were adapter-trimmed and quality-filtered to a PHRED score of 2 with BBDuk, merged with BBMerge, and error-corrected and normalized to 80x average coverage with BBNorm from BBTools v37.28 [25]. Merged and unmerged reads were assembled with Megahit v1.0.3 using a maximum k-mer size of 127 [26,27]. Initial metagenome-assembled genomes (MAGs) were obtained with Metabat2 v2.10.2 automated binning, after mapping with BBMap, and sorting of bam files with samtools v1.9 [28,29]. MAGs were identified based on small subunit ribosomal RNA gene sequences (SSUs, detected by barrnap v0.6 and classified with vsearch v2.6.2 against the SILVA SSU database v132) and other taxonomic marker genes (detected by Amphora2), through visualisation with gbtools v2.6.0 in RStudio [30–35]. This metagenomic workflow was initially performed separately for gill and mixed tissue. If both sample types were available for the same mussel individual, we confirmed that their symbiont MAGs were identical based on GC content, coverage, and average nucleotide identity (ANI) before pooling the reads and repeating the workflow to increase symbiont coverage. Similarly, if Metabat2 split the sulphur-oxidising (SOX) symbiont sequences into multiple bins, these were pooled after assessment of their GC content, coverage and ANI. Completeness, contamination and strain heterogeneity of MAGs were estimated throughout the binning process with CheckM v1.0.17 based on 280 gammaproteobacterial marker genes [36–39]. Due to the presence of multiple strains in the symbiont population [40], we expected duplicates of marker genes with highly similar sequences, such as the ones reported by CheckM as strain heterogeneity. We therefore corrected contamination rates and calculated the contamination that could not be attributed to strain heterogeneity.

For MAGs with a completeness below 90 %, an additional step of assembly and binning was performed to yield MAGs with higher completeness. In such cases, the incomplete bin was used as a reference for mapping to recruit symbiont reads for assembly with SPAdes v3.12.0 [41]. These draft genome assemblies were manually binned in Bandage v0.8.1 based on sequence nodes connected in the assembly graph [42].

Additional statistics of symbiont MAGs were calculated using the stats.sh script of BBTools (Supplementary Table S 4). Read coverage of symbiont MAGs was estimated with samtools after mapping raw reads against MAGs with BBMap. High-quality symbiont MAGs with a completeness above 90% and a contamination below 5% (after correction for strain heterogeneity) were used for further analysis with one exception: The MAG of library 3386_H was 89 % complete, only <1 % less than the cutoff, but had no contamination.

1.5   Analyses of SOX symbionts based on symbiont MAGs

All analyses of SOX symbionts conducted in this study, their input data and the level of resolution are summarised in Supplementary Table S 2.

1.5.1   Phylogenomic analysis of SOX symbionts

To investigate the phylogeny of SOX symbionts from mussels in Broken Spur, we constructed a phylogenomic tree with 171 gammaproteobacterial, single-copy marker genes from the SOX MAGs as well as closely-related symbiotic and free-living bacteria. Accession numbers and publications for all reference MAGs/genomes used in the analysis are listed in Supplementary Table S 5.

A protein alignment of the 171 marker genes (Data file: Phylogenomics_NMAR.fasta, available at https://github.com/muecker/Symbionts_in_a_mussel_hybrid_zone) was obtained with GToTree v1.4.11 [17,18,38,39,43,44]. The alignment was visually inspected in Geneious v11.1.5 [45]. One gene sequence (of sample 1586K) had a high proportion of mismachtes to all other sequences and was identified as contamination by blasting against the NCBI database. This sequence was removed from the alignment. Using the LG+F+R6 amino acid model ([46], best model according to Model Finder), and 1000 samples for ultrafast bootstrap, we reconstructed a phylogenomic tree of SOX symbionts and their closest relatives with IQ-TREE v1.6.7.1, and edited it with iTol and Adobe Illustrator.

Correlation between symbiont F_{ST} and geographic distance was tested using the Mantel test. The multiple sequence alignment was imported into R with the read.alignment function of R package seqinr v3.4-5 and transformed into a genind object using the alignment2genind function of adegenet v2.1.1 [47,48]. We caluclated the pairwise F_{ST} of symbionts between all locations on the northern MAR using pairwise.fst of R package hierfstat v0.04-33 [49]. Geographic distances between vent sites were calculated based on the coordinates using https://www.movable-type.co.uk/scripts/latlong.html, and transformed into a dist object using the R function dist. To account for the geographic subdivision of the host species that can also cause patterns similar to isolation-by-distance [50], we performed a stratified Mantel test using the mantel function of R package vegan v2.5-5, and the host species groups as strata (mantel(FST, geo_distances,strata=groups_NMAR) [51]. After a statistical significant result, the F_{ST} values between symbionts from different sites were plotted against their geographic distances for visual inspection.

1.5.2   Average nucleotide identity of SOX symbionts

To analyse how similar symbiont MAGs from Broken Spur were to each other, we analysed their pairwise ANI values of the aligned fraction (0.48 – 0.99%) with fastANI v1.1 [52] (Data file: Average_nucleotide_identity_SOX_symbionts_Broken_Spur.csv). Samples were clustered and represented based on their average ANI values in a heatmap with dendrogram, generated in RStudio using the packages gplots v3.0.1.1 and maditr v0.6.2 [53,54].

Correlation of SOX symbiont ANI values and sampling year was tested with a Mantel test (mantel function of R package vegan, 5039 permutations) using the Spearman's rank correlation coefficient after transforming sampling year information into euclidean distances. To test the correlation between SOX symbiont ANI values and host genetics, pairwise genetic distances between host individuals were calculated based on 18 species-diagnostic SNP markers (see "1.2. Identification of hybrid host individuals"). Host SNP markers were imported to RStudio and converted into a dataframe using the read.structure and genind2df functions of the package adegenet. We subsequently calculated pairwise genetic distances between individuals with the dist.gene function of R package ape v5.3 [55]. Correlation was tested as described above for ANI values vs. sampling year.

1.6   Analyses of symbiont population based on single-nucleotide polymorphisms

Recent advances in whole-(meta)genome approaches have increased resolution and sensitivity of analyses, and advanced our knowledge on strain diversity in deep-sea mussel symbiont populations [40,56,57]. We therefore performed genome-wide SNP analyses of SOX symbionts based on 2496 orthologous genes to investigate symbiont population differentiation between different mussel individuals from Broken Spur.

We used a gene catalogue of 3204 orthologues (see below) as a reference for SNP identification. The catalogue was annotated using prokka v1.1, resulting in 2496 genes with annotation (including hypothetical proteins) [58]. Genes without any annotation, mostly short (<300 bp), probably fragmented genes, were excluded from the analysis. SNP calling was performed as described in [40] using scripts available at https://github.com/rbcan/MARsym_paper with a few adjustments to newer software versions. In summary, raw reads were adapter trimmed and quality filtered to a PHRED score of 20 and subsequently mapped to the reference with a minimum identitiy of 95 % using BBMap. We realigned reads around indels and downsampled to an average coverage of 70x. Samples that did not meet the coverage threshold were excluded from the analysis. The steps above

were performed with samtools, Picard tools v1.1.02 and the Genome Analysis Toolkit (GATK) v3.7-0 [29,59,60]. SNPs were called with GATK HaplotypeCaller, and unreliable SNPs were filtered with GATK VariantFiltration (settings: QD < 2; FS > 60; MQ < 40, MQRankSum < -20, ReadPosRankSum < -8). Spearman's rank correlation coefficient was calculated in RStudio using the R function cor to test for correlation of SNP density (#SNPs/kb) with shell size. The fixation index F_{ST} was calculated for each gene with a script (https://github.com/deropi/BathyBrooksiSymbionts/tree/master/Population_structure_analyse s) previously used in [57], and averaged per host individual (Data file: Pairwise_mean_FST_SOX_symbionts_Broken_Spur.csv). Mean pairwise F_{ST} values were plotted in a heatmap, and correlation between F_{ST} and sampling year and F_{ST} and host genotype was tested as described above for ANI values.

1.7 Analysis of differences in gene repertoire between symbionts from hybrids and parental species

1.7.1 Gene presence/absence and abundance analyses

To examine whether there are differences in the gene repertoire of the symbiont populations between hybrid and parental mussels, we analysed the presence/absence of genes specific to either group of mussels and their relative abundances. We annotated all MAGs with prokka and clustered orthologues with GET_HOMOLOGUES v3.2.3 using the OrthoMCL algorithm [39,61–69], resulting in a gene catalogue of 3204 orthologues (Data file: Orthologue_gene_catalogue_OMCL_SOX_symbionts_Broken_Spur.fasta; also used in SNP-identification above). Using the parse_pangenome_matrix.pl script of GET_HOMOLOGUES, we tested for genes that were present in at least 90 % of symbiont genomes from *B. puteoserpentis* and absent in at least 90 % of symbiont genomes from hybrids, and vice versa.

To further analyse gene abundances, raw reads of all libraries were mapped to the orthologous gene catalogue with BBMap and downsampled to 70x coverage with samtools. The fasta sequences were extracted from downsampled bam files using samtools and pseudoaligned to the catalogue using kallisto v0.46.0 [25,70]. The gene coverage was estimated using the abundance_estimates_to_matrix.pl script of Trinity v2.5.1 [71,72] (Data file: Gene_counts_kallisto_SOX_symbionts_Broken_Spur.matrix).

To account for the compositionality of the data, the gene abundances were statistically evaluated using ALDEx2 v1.16.0 and data.table v1.12.2 in RStudio [2,73–76]. We used the

aldex.clr module to prepare the data using host categories (hybrid or *B. puteoserpentis*) as condition. With the aldex.kw command, we ran a general linear model and a Kruskal Wallace test for one way ANOVA.

1.7.2   Analysis of gene differentiation between symbionts from hybrids and parentals

We analysed population differentiation (F_{ST}) based on SNP frequencies in 2496 orthologue genes to find genes that have higher differentiation 'between' symbionts of hybrids and parental species than variations 'within' symbionts of the same host category (hybrids or *B. puteoserpentis*, Supplementary Figure S 1). F_{ST} values were acquired as described above (1.6 "Analyses of symbiont population based on single-nucleotide polymorphisms") and reformatted for 'between' versus 'within' statistical comparisons (Data file: Per_gene_FST_SOX_symbionts_Broken_Spur.zip). We used a Mann–Whitney U test in RStudio to test the null hypothesis that there is no significant difference between F_{ST} of genes in the 'between' (hybrids versus parental species) and the 'within' (hybrids versus hybrids, parental versus parental species) categories. To reduce false discovery rates, the test was repeated with a dataset of random F_{ST} values. More genes with p-value <0.05 were detected in the random dataset than in the actual data, and no p-values of the real dataset were below those from the random dataset.

This indicates that all genes detected as significant (p<0.05) for the real data can be attributed to type I error, and that there was no gene more differentiated between symbionts of hybrid and parental mussels than within the same host category.
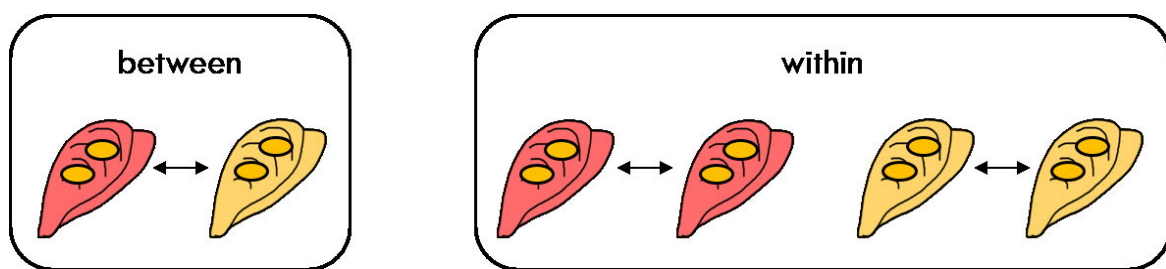
Supplementary Figure S 1 | Categories for per gene F_{ST} analysis. Between: Comparison of SOX symbionts from *B. puteoserpentis* (red) and hybrid mussels (yellow). Within: Comparison among SOX symbionts from *B. puteoserpentis* and among SOX symbionts from hybrid mussels.

1.7.3 Redundancy analysis of SOX symbiont allele frequencies from Bathymodiolus mussels along the northern Mid-Atlantic Ridge

Redundancy analysis (RDA) allows to eliminate redundant information in genetic data and its associations with environmental variables, and to assess the proportion of variation explained by these environmental variables [77]. We performed a RDA in RStudio using the vegan package to test how much of the variation in symbiont allele frequencies can be explained by geographic distance, vent type (basaltic versus ultramafic rock), the associated host species and depth. In hydrothermal systems, rock type plays a central in role in determining biogeochemical conditions, including pH and the energy sources available for chemosynthetic microorganisms [78], which is why we chose this environmental parameter, which we called vent type. Depth and vent type were retrieved from the InterRidge Vents Database v3.4 (https://vents-data.interridge.org/, accessed 2020-06-15). As a reference for SNP analysis, we constructed a gene catalogue based on all SOX symbiont MAGs from the northern MAR (Data file: Orthologue_gene_catalogue_OMCL_SOX_symbionts_NMAR.fasta) using the workflow described above (1.7.1 "Gene presence/absence and abundance analyses"). To obtain allele frequencies, we performed a SNP analysis based on the gene catalogue of all SOX symbionts from the northern MAR as described above (1.6 "Analyses of symbiont population based on single-nucleotide polymorphisms"). We extracted the AD (read depth per allele) and DP (read depth) field from VCF files using GATK's VariantToTable tool, divided AD by DP to obtain allele frequencies per sample and merged the individual tables using join on the Linux command line (Data file: Allele_frequencies_SOX_symbionts_NMAR.csv). For the RDA, site coordinates were scaled and computed as orthogonal polynomials with R package stats v3.6.3 (function poly) as suggested by [77,79,80]. We performed a forward selection on the polynomials with the ordistep function of R package vegan, ran the RDA with all variables and calculated an adjusted R^2. We assessed the significance of the RDA, the individual axes and the explanatory variables with the vegan package function anova.cca using 1000 permutations. To explore how much variation could be explained by each explanatory variable, we performed a variation partitioning using the varpart function of vegan and plotted it with R base function plot. The RDA triplot was plotted using the ggord package v1.1.4 [81]. For an overview visualisation of the allele frequencies, we calculated a NMDS using the metaMDS function of vegan and plotted it with ggplot2. All figures were modified with Adobe Illustrator.

2 Supplementary results & discussion

2.1 *B. puteoserpentis* and hybrid individuals identified in Broken Spur

We genotyped mussels from Broken Spur and identified *B. puteoserpentis* and hybrid individuals. *B. azoricus* mussels were not detected in all methods used, except for one mussel (3676-15/3386_N) that was identified as *B. azoricus* by NEWHYBRIDS. However, this result was not supported by the two other programmes, suggesting that the mussel is more likely a hybrid.

The absence of *B. azoricus* in Broken Spur could be due to bathymetric limitation, as *B. azoricus* usually occurs at shallower depths. Another possible explanation is that the actual hybrid zone might be further north as suggested by [15]. Lastly, it cannot be ruled out that *B. azoricus* mussels were not found during sampling as the number of mussels collected was limited and their distribution quite patchy at Broken Spur.

All hybrids identified by INTROGRESS were in the F2 to F4 generation indicating that the hybrids are fertile. Although the exact status of backcrosses, especially which generations of backcrosses are actually present, was uncertain, multiple mussels were identified as backcrosses by NEWHYBRIDS and INTROGRESS. Together with the admixture values reported by STRUCTURE, this suggests that there is still gene flow between hybrids and the populations of parental species.

2.2 Future studies on *Bathymodiolus* hybrids

Hybrids from Broken Spur were clearly able to successfully reproduce given the presence of F2 – F4 hybrids, but we have no information on hybrid performance and fitness. In lab-held organisms, fitness of hybrids can be assessed by various measurements, e.g. comparisons of offspring survival rates or developmental times [82]. To study these parameters in *Bathymodiolus* mussels, the mussels would have to be maintained in aquaria until they are ready to spawn, which may occur only once a year in January in *B. azoricus* [83]. After spawning, eggs could be collected for counting and genotype determination. However, embryo development in aquaria-held mussels has not yet succeeded, as previous attempts failed because development was abnormal or stopped at the 4-cell stage [83,84]. An alternative option, equally challenging, would be to collect *Bathymodiolus* mussels from a hybrid zone, e.g. Broken Spur on the MAR or the vents at 23°S on the East Pacific Rise [85], and determine their genotype prior to cultivation (e.g. by removing hemolymph from their adductor muscles). To ensure reproducible results, enough replicates would be needed, which

is generally challenging for most deep-sea species. In model organisms that can be easily cultured in the laboratory, such studies can be performed more easily. Studies in *Nasonia* or *Drosophila* pointed towards poor hybrid performance, i.e. high lethality and sterility of hybrid individuals [86–89].

2.3 No difference in gene abundances between symbionts from hybrids and parental species

We analysed 3204 orthologous genes to detect genes that are exclusive to either symbionts of hybrid or symbionts of *B. puteoserpentis*. GET_HOMOLOGUES detected none of such genes, even with lower stringency (presence in >90 % in one and <90 % in the other group).

When comparing gene abundances between symbionts from hybrid and *B. puteoserpentis* mussels using the statistical analysis with ALDEx2, no genes were significantly different in their abundances (Benjamini-Hochberg corrected p-value < 0.05). Functional variation has previously been shown to occur among symbiont populations from different vents along the MAR [40]. However, the gene repertoire of SOX symbiont populations within Broken Spur did not vary according to host genotype, indicating that hybrids and parental species do not select their symbionts based on different functions.

2.4 No correlation of SNPs/kb with mussel shell size

Picazo et al. (2019) previously detected lower strain diversity in large (146–241 mm) compared to medium-sized (72–141 mm) mussels in the Gulf of Mexico that might be explained by self-infection and slower symbiont uptake in older mussels [57]. We did not detect any correlation of SNPs/kb with shell size, which might be due to the relatively limited size range of the analysed mussels (24–133 mm).

2.5 Isolation-by-distance of *Bathymodiolus* SOX symbiont subspecies at the northern MAR

Phylogenomic analysis of 171 gammaproteobacterial marker genes revealed that symbiont genetic variation (F_{ST} based on the amino acid alignment) was positively correlated with geographic distance (r = 0.7471, p = 0.035). However, a gradual genetic change along a geographic gradient as would be expected under an evolutionary isolation-by-distance (IBD) model [90] could not be observed (Supplementary Figure S 2). Mantel tests are often used to test for isolation-by-distance which is why we included this analysis here. However, its use has been discouraged [91]. We therefore used a redundancy analysis in this study (Figure 3, Supplementary Figure S 3, Supplement 1.8 "Redundancy analysis of SOX symbiont allele

frequencies from *Bathymodiolus* mussels along the northern Mid-Atlantic Ridge" and main text).
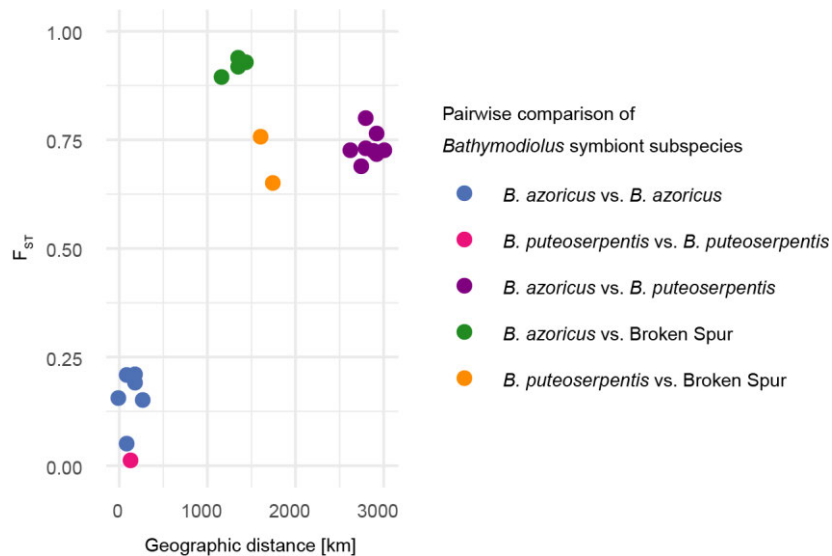
Supplementary Figure S 2 | Relation of F_{ST} and geographic distance between *Bathymodiolus* **SOX symbiont populations from different vent fields along the northern MAR.** Displayed F_{ST} values are averaged pairwise F_{ST} between symbionts from all mussels at a vent field. Each dot represents one pairwise comparison between two sites, comparisons of a site with itself are not shown. Colours correspond to the comparisons of the different symbiont subspecies (*B. azoricus* type, present at Menez Gwen – White Flames, Lucky Strike – Montsegur, Lucky Strike – Eiffel Tower and Rainbow; *B. puteoserpentis* type, present at Logatchev Quest and Semenov; Broken Spur type, present at Broken Spur).
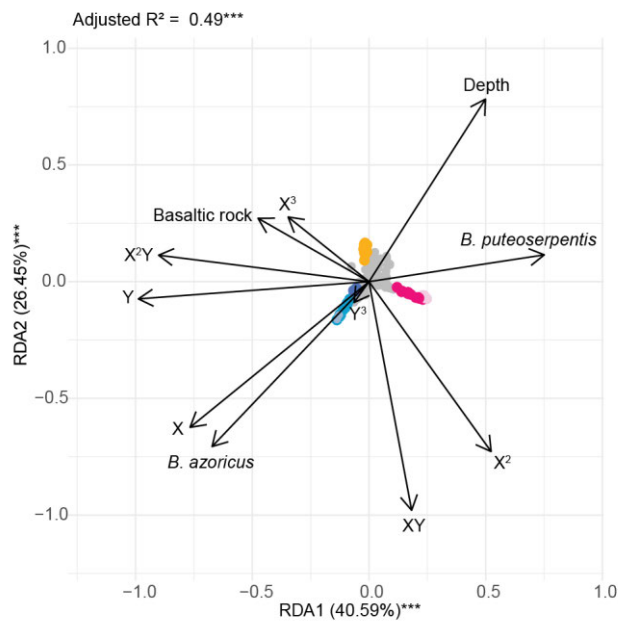
Supplementary Figure S 3 | Influence of geographic distance, host species and environmental parameters on differentiation of *Bathymodiolus* SOX symbionts at the northern MAR. Redundancy analysis triplot (scaling 2, wa scores) showing the influence of geographic distance (forward selected variables X, Y, X^2, X^3, X^2Y, Y^3 represent orthogonal polynomials of latitude and longitude), host species (*B. azoricus* and *B. puteoserpentis*), the vent type (only basaltic rock displayed) and water depth on symbiont allele frequencies. *** p-value < 0.001. P-values are based on permutation tests with 1000 repetitions.

Supplementary Table S 1 | Sample overview. ID consists of dive number and mussel ID. Shell length (size) is listed when data were available according to cruise material in http://dlacruisedata.whoi.edu/AT/AT003L03/ (accessed 2019-05-29). AllPrep: AllPrep DNA/RNA/ Protein MiniKit (Qiagen); DNAeasy: DNAeasy Blood & Tissue kit (Qiagen), Nex: Nextera DNA Flex Library Prep Kit (Illumina); TruSeq: Illumina TruSeq DNA Samples Prep Kit (BioLABS).

Cruise	ID	Latitude	Longitude	Depth [m]	Sampling date	Material	DNA extraction	Libprep	Sequencing	Library	Size [mm]
AT_05/03	3676-1	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_A	119.8
AT_05/03	3676-2	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_B	113.8
AT_05/03	3676-3	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_C	112.2
AT_05/03	3676-4	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_D	116.3
AT_05/03	3676-5	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_E	119.1
AT_05/03	3676-6	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_F	114.6
AT_05/03	3676-7	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_K	107.6
AT_05/03	3676-8	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_G	103.6
AT_05/03	3676-9	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_H	75
AT_05/03	3676-10	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_I	122.9
AT_05/03	3676-11	29.1672	-43.1742	3045	July 19, 2001	mixed & gill	DNAeasy & AllPrep	TruSeq, Nex	HiSeq 2500 & 3000	3386_J, 2424_A	101.9
AT_05/03	3676-12	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_L	114.8
AT_05/03	3676-14	29.1672	-43.1742	3045	July 19, 2001	mixed & gill	DNAeasy & AllPrep	TruSeq & Nex	HiSeq 2500 & 3000	3386_M, 2424_B	81.4
AT_05/03	3676-15	29.1672	-43.1742	3045	July 19, 2001	mixed & gill	DNAeasy & AllPrep	TruSeq & Nex	HiSeq 2500 & 3000	3386_N, 2424_C	92.9
AT_05/03	3676-16	29.1672	-43.1742	3045	July 19, 2001	mixed & gill	DNAeasy & AllPrep	TruSeq & Nex	HiSeq 2500 & 3000	3386_O, 2424_F	12.2
AT_05/03	3676-17	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_P	110.2
AT_05/03	3676-18	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_Q	89.9

Cruise	ID	Latitude	Longitude	Depth [m]	Sampling date	Material	DNA extraction	Libprep	Sequencing	Library	Size [mm]
AT_05/03	3676-19	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_R	64.7
AT_05/03	3676-20	29.1672	-43.1742	3045	July 19, 2001	mixed & gill	DNAeasy & AllPrep	TruSeq & Nex	HiSeq 2500 & 3000	3386_S, 2424_G	93.6
AT_05/03	3676-21	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_T	93.2
AT_05/03	3676-22	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_U	81.5
AT_05/03	3676-26	29.1672	-43.1742	3045	July 19, 2001	mixed	DNAeasy	TruSeq	HiSeq 2500	2424_D	31.7
AT_05/03	3676-27	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_V	65.4
AT_05/03	3676-28	29.1672	-43.1742	3045	July 19, 2001	mixed	DNAeasy	TruSeq	HiSeq 2500	2424_E	66.7
AT_05/03	3676-29	29.1672	-43.1742	3045	July 19, 2001	mixed	DNAeasy	TruSeq	HiSeq 2500	2424_H	51.8
AT_05/03	3676-30	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_W	94.4
AT_05/03	3676-31	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_X	64.5
AT_05/03	3676-32	29.1672	-43.1742	3045	July 19, 2001	mixed & gill	DNAeasy & AllPrep	TruSeq & Nex	HiSeq 2500 & 3000	3386_Y, 2424_I	88.8
AT_05/03	3676-33	29.1672	-43.1742	3045	July 19, 2001	mixed & gill	DNAeasy & AllPrep	TruSeq & Nex	HiSeq 2500 & 3000	3386_Z, 2424_J	100.5
AT_05/03	3676-37	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_AA	111.4
AT_05/03	3676-38	29.1672	-43.1742	3045	July 19, 2001	gill	AllPrep	Nex	HiSeq 3000	3386_AB	104.7
AT_03/03	3125-1	29.1667	-43.1733	3056	July 17, 1997	gill	AllPrep	Nex	HiSeq 3000	3386_AC	
AT_03/03	3125-2	29.1667	-43.1733	3056	July 17, 1997	gill	AllPrep	Nex	HiSeq 3000	3386_AD	
AT_03/03	3125-3	29.1667	-43.1733	3056	July 17, 1997	gill	AllPrep	Nex	HiSeq 3000	3386_AE	

Cruise	ID	Latitude	Longitude	Depth [m]	Sampling date	Material	DNA extraction	Libprep	Sequencing	Library	Size [mm]
AT_03/03	3125-4	29.1667	-43.1733	3056	July 17, 1997	gill	AllPrep	Nex	HiSeq 3000	3386_AF	
AT_03/03	3125-5	29.1667	-43.1733	3056	July 17, 1997	mixed & gill	DNAeasy & AllPrep	TruSeq & Nex	HiSeq 2500 & 3000	3386_AG, 2424_K	
AT_03/03	3125-6	29.1667	-43.1733	3056	July 17, 1997	mixed & gill	DNAeasy & AllPrep	TruSeq & Nex	HiSeq 2500 & 3000	3386_AH, 2424_L	
AT_03/03	3125-7	29.1667	-43.1733	3056	July 17, 1997	mixed & gill	DNAeasy & AllPrep	TruSeq & Nex	HiSeq 2500 & 3000	3386_AI, 2424_M	
AT_03/03	3125-8	29.1667	-43.1733	3056	July 17, 1997	mixed	DNAeasy	TruSeq	HiSeq 2500	2424_N	
AT_03/03	3125-9	29.1667	-43.1733	3056	July 17, 1997	mixed & gill	DNAeasy & AllPrep	TruSeq & Nex	HiSeq 2500 & 3000	3386_AJ, 2424_O	
AT_03/03	3125-10	29.1667	-43.1733	3056	July 17, 1997	gill	AllPrep	Nex	HiSeq 3000	3386_AK	
AT_03/03	3125-11	29.1667	-43.1733	3056	July 17, 1997	gill	AllPrep	Nex	HiSeq 3000	3386_AL	

Supplementary Table S 2 | Overview of analyses of *Bathymodiolus* SOX symbionts.

Analysis	Input data/reference	Level of resolution
Phylogenomics	171 marker genes extracted from NMAR MAGs	symbiont subspecies
Average nucleotide identity	Broken Spur MAGs	symbiont subspecies
Gene abundances	Mussel metagenomes from Broken Spur mapped against Broken Spur orthologous gene catalogue	symbiont strain
SNP analyses	Mussel metagenomes from Broken Spur mapped against Broken Spur orthologous gene catalogue	symbiont strain
Redundancy analysis	Mussel metagenomes from NMAR sites mapped against NMAR orthologous gene catalogue	symbiont strain

Supplementary Table S 3 | Genotyping results of 42 mussel individuals from Broken Spur based on analyses with NEWHYBRIDS, INTROGRESS and STRUCTURE. ID: Dive number–mussel ID; Lib: Metagenomic library name; Bazo: *B. azoricus*; BC azo: Backcross between *B. azoricus* and hybrid; FX: Hybrid in generation X; BC put: Backcross between *B. puteoserpentis* and hybrid; Bput: *B. puteoserpentis*; Genotype: NEWHYBRIDS' genotype category with highest probability.

ID	Lib	NEWHYBRIDS Bazo	BC1 azo	BC2 azo	BC3 azo	BC4 azo	F1	F2-4	BC1 put	BC2 put	BC3 put	BC4 put	Bput	Genotype	INTROGRESS	STRUCTURE Bazo	Bput
3676-1	3386_A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.09	0.84	Bput	*B. puteoserpentis*	0.00	1.00
3676-2	3386_B	0.00	0.00	0.00	0.00	0.00	0.00	0.23	0.13	0.36	0.27	0.01	0.00	BC put	BC *puteoserpentis*	0.19	0.81
3676-3	3386_C	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	F2-4	F2-4	0.32	0.68
3676-4	3386_D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.07	0.89	Bput	*B. puteoserpentis*	0.00	1.00
3676-5	3386_E	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.08	0.84	Bput	*B. puteoserpentis*	0.00	1.00
3676-6	3386_F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.28	0.16	0.54	BC put	BC *puteoserpentis*	0.01	0.99
3676-7	3386_K	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.50	0.17	0.20	BC put	BC *puteoserpentis*	0.04	0.96
3676-8	3386_G	0.00	0.01	0.00	0.00	0.00	0.60	0.36	0.04	0.00	0.00	0.00	0.00	F1	F2-4	0.37	0.63
3676-9	3386_H	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	F2-4	F2-4	0.28	0.72
3676-10	3386_I	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.09	0.84	Bput	*B. puteoserpentis*	0.00	1.00
3676-11	3386_J	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	F2-4	F2-4	0.37	0.63
3676-12	3386_L	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.10	0.41	0.45	0.03	0.00	BC put	BC *puteoserpentis*	0.13	0.87
3676-14	3386_M	0.03	0.63	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	BC azo	F2-4	0.57	0.43
3676-15	3386_N	0.86	0.12	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	Bazo	F2-4	0.69	0.31
3676-16	3386_O	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.10	0.79	Bput	*B. puteoserpentis*	0.00	1.00
3676-17	3386_P	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.40	0.50	0.04	0.00	BC put	BC *puteoserpentis*	0.13	0.87
3676-18	3386_Q	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.28	0.14	0.53	Bput	BC *puteoserpentis*	0.01	0.99

ID	Lib	Bazo	BC1 azo	BC2 azo	BC3 azo	BC4 azo	F1	F2-4	BC1 put	BC2 put	BC3 put	BC4 put	Bput	Genotype	INTROGRESS	Bazo	Bput
		NEWHYBRIDS														STRUCTURE	
3676-19	3386_R	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.38	0.52	0.06	0.00	BC put	BC *puteoserpentis*	0.12	0.88
3676-20	3386_S	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.21	0.15	0.63	Bput	*B. puteoserpentis*	0.00	1.00
3676-21	3386_T	0.00	0.02	0.00	0.00	0.00	0.31	0.66	0.01	0.00	0.00	0.00	0.00	F2-4		0.40	0.60
3676-22	3386_U	0.00	0.00	0.00	0.00	0.00	0.20	0.59	0.19	0.03	0.00	0.00	0.00	F2-4	F2-4	0.29	0.71
3676-26	2339_D	0.00	0.07	0.00	0.00	0.00	0.24	0.69	0.00	0.00	0.00	0.00	0.00	F2-4	F2-4	0.43	0.57
3676-27	3386_V	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.50	0.18	0.19	BC put	BC *puteoserpentis*	0.02	0.98
3676-28	2424_E	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	F2-4	F2-4	0.36	0.64
3676-29	2424_H	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.19	0.13	0.66	Bput	*B. puteoserpentis*	0.00	1.00
3676-30	3386_W	0.01	0.56	0.00	0.00	0.00	0.00	0.43	0.00	0.00	0.00	0.00	0.00	BC azo	F2-4	0.54	0.46
3676-31	3386_X	0.15	0.79	0.01	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	BC azo	F2-4	0.63	0.37
3676-32	3386_Y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.08	0.84	Bput	*B. puteoserpentis*	0.00	1.00
3676-33	3386_Z	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.07	0.89	Bput	*B. puteoserpentis*	0.00	1.00
3676-37	3386_AA	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.17	0.43	0.36	0.01	0.00	BC put	BC *puteoserpentis*	0.17	0.83
3676-38	3386_AB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.12	0.10	0.78	Bput	*B. puteoserpentis*	0.00	1.00
3125-1	3386_AC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.32	0.14	0.49	Bput	*B. puteoserpentis*	0.01	0.99
3125-2	3386_AD	0.02	0.82	0.00	0.00	0.00	0.01	0.15	0.00	0.00	0.00	0.00	0.00	BC azo	F2-4	0.57	0.43
3125-3	3386_AE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.07	0.89	Bput	*B. puteoserpentis*	0.00	1.00

ID	Lib	Bazo	BC1 azo	BC2 azo	BC3 azo	BC4 azo	F1	F2-4	BC1 put	BC2 put	BC3 put	BC4 put	Bput	Genotype	INTROGRESS	Bazo	Bput
						NEWHYBRIDS										STRUCTURE	
3125-4	3386_AF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.42	0.18	0.33	BC put	BC *puteoserpentis*	0.02	0.98
3125-5	3386_AG	0.00	0.04	0.00	0.00	0.00	0.06	0.90	0.00	0.00	0.00	0.00	0.00	F2-4	F2-4	0.43	0.57
3125-6	3386_AH	0.00	0.32	0.00	0.00	0.00	0.00	0.68	0.00	0.00	0.00	0.00	0.00	F2-4	F2-4	0.51	0.49
3125-7	3386_AI	0.05	0.52	0.00	0.00	0.00	0.00	0.42	0.00	0.00	0.00	0.00	0.00	BC azo	F2-4	0.57	0.43
3125-8	2424_N	0.00	0.03	0.00	0.00	0.00	0.00	0.96	0.00	0.00	0.00	0.00	0.00	F2-4	F2-4	0.45	0.55
3125-9	3386_AJ	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	F2-4	F2-4	0.37	0.63
3125-10	3386_AK	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.11	0.10	0.78	Bput	B. *puteoserpentis*	0.00	1.00
3125-11	3386_AL	0.00	0.09	0.00	0.00	0.00	0.01	0.91	0.00	0.00	0.00	0.00	0.00	F2-4	F2-4	0.45	0.55

Supplementary Table S 4 | Statistics of SOX symbiont MAGs. Complete: completeness based on gammaproteobacterial marker genes in CheckM; Contam (Strain): Contamination and percentage of this contamination which can be explained by strain heterogeneity according to CheckM; Contam Corr: Contamination after correction for strain heterogeneity (i.e. strain variants were not considered as contamination).

MAG ID	Complete [%]	Contam (Strain) [%]	Contam Corr [%]	# Contigs	GC [%]	Genome size [Mb]	Read coverage [x]
SOX_BS_2424_D	94.25	0 (0)	0.00	914	0.37	2.40	22
SOX_BS_2424_E	93.69	0.16 (33.33)	0.11	3206	0.37	2.89	100
SOX_BS_2424_H	92.28	0 (0)	0.00	772	0.37	2.13	17
SOX_BS_2424_N	92.28	6.87 (95.45)	0.31	4717	0.36	3.57	1001
SOX_BS_3386_A	94.44	1.11 (66.67)	0.37	3998	0.37	2.76	241
SOX_BS_3386_AB	94.25	1.64 (100)	0.00	4095	0.37	2.80	372
SOX_BS_3386_AD	94.25	0.14 (100)	0.00	3928	0.37	2.69	472
SOX_BS_3386_AE	94.25	0.28 (100)	0.00	3899	0.37	2.73	105
SOX_BS_3386_AG	95.09	7.24 (84)	1.16	4767	0.36	3.46	1744
SOX_BS_3386_AH	94.25	3.5 (12.5)	3.06	4113	0.36	3.44	542
SOX_BS_3386_AJ	92	8.81 (86.11)	1.22	5570	0.36	4.04	989
SOX_BS_3386_AK	94.63	0.45 (40)	0.27	3648	0.37	2.70	301
SOX_BS_3386_AL	94.25	3.04 (58.33)	1.27	3415	0.37	2.62	387
SOX_BS_3386_C	94.25	3.51 (8.33)	3.22	3038	0.37	2.50	502
SOX_BS_3386_D	94.25	2.09 (25)	1.57	4408	0.37	2.92	432
SOX_BS_3386_E	90.18	0 (0)	0.00	143	0.38	1.43	488
SOX_BS_3386_F	94.25	2.2 (100)	0.00	4006	0.37	2.70	382
SOX_BS_3386_G	94.25	2.57 (90.91)	0.23	3656	0.37	2.63	601
SOX_BS_3386_H	89.05	0 (0)	0.00	145	0.38	1.37	527
SOX_BS_3386_I	92.43	0 (0)	0.00	157	0.38	1.48	365
SOX_BS_3386_J	91.3	0 (0)	0.00	163	0.38	1.48	2588
SOX_BS_3386_M	94.25	4.45 (33.33)	2.97	3425	0.37	2.70	2704
SOX_BS_3386_O	94.81	3.79 (36.36)	2.41	4500	0.37	3.07	515
SOX_BS_3386_S	94.25	4.03 (69.23)	1.24	3607	0.37	2.90	489
SOX_BS_3386_T	93.97	4.35 (63.64)	1.58	2934	0.37	2.60	298

MAG ID	Complete [%]	Contam (Strain) [%]	Contam Corr [%]	# Contigs	GC [%]	Genome size [Mb]	Read coverage [x]
SOX_BS_3386_U	94.25	0.14 (100)	0.00	4081	0.37	2.81	249
SOX_BS_3386_W	94.25	2.2 (87.5)	0.28	3531	0.37	2.72	398
SOX_BS_3386_X	94.25	4.57 (28.57)	3.21	3161	0.37	2.79	419
SOX_BS_3386_Y	94.25	0.14 (100)	0.00	3032	0.37	2.66	431
SOX_BS_3386_Z	94.53	3.31 (61.54)	1.27	4674	0.36	3.48	951

Supplementary Table S 5 | List of external data. BioProject and dataset accessions for the European Nucleotide Archive and the respective publication were listed if available. Lat: Latitude; Lon: Longitude; Complete: completeness based on gammaproteobacterial marker genes in CheckM; Contam: Contamination; Strain: Percentage of this contamination which can be explained by strain heterogeneity according to CheckM; Pub: Publication.

MAG	Site	Lat	Lon	Cruise	Host species	Complete	Contam	Strain	Pub	BioProject	Accession
1048F	Lucky Strike (Montsegur)	37.288	-32.276	Biobaz (2013)	*B. azoricus*	94.53	0	0	[92]	PRJEB36091	GCA_903813355
1048G	Lucky Strike (Montsegur)	37.288	-32.276	Biobaz (2013)	*B. azoricus*	94.53	0	0	[92]	PRJEB36091	GCA_903813365
1048H	Lucky Strike (Montsegur)	37.288	-32.276	Biobaz (2013)	*B. azoricus*	95.09	0	0	[92]	PRJEB36091	GCA_903819415
1048I	Lucky Strike (Eiffel Tower)	37.283	-32.276	Biobaz (2013)	*B. azoricus*	93.97	0.03	0	[92]	PRJEB36091	GCA_903819345
1048J	Lucky Strike (Eiffel Tower)	38.283	-32.276	Biobaz (2013)	*B. azoricus*	94.53	0.56	100	[92]	PRJEB36091	GCA_903819365
1586B	Lucky Strike (Eiffel Tower)	37.289	-32.275	Biobaz (2013)	*B. azoricus*	93.41	0.84	100	[92]	PRJEB36091	GCA_903813405
1586C	Lucky Strike (Eiffel Tower)	37.289	-32.275	Biobaz (2013)	*B. azoricus*	92.85	0.84	50	[92]	PRJEB36091	GCA_903819425
1586D	Lucky Strike (Eiffel Tower)	37.289	-32.275	Biobaz (2013)	*B. azoricus*	93.97	0	0	[92]	PRJEB36091	GCA_903813415
1586E	Lucky Strike (Eiffel Tower)	37.289	-32.275	Biobaz (2013)	*B. azoricus*	92.85	0	0	[92]	PRJEB36091	GCA_903813425
1586F	Lucky Strike (Eiffel Tower)	37.283	-32.276	Biobaz (2013)	*B. azoricus*	92.85	1.4	33.3	[92]	PRJEB36091	GCA_903813665

MAG	Site	Lat	Lon	Cruise	Host species	Complete	Contam	Strain	Pub	BioProject	Accession
1586G	Lucky Strike (Eiffel Tower)	37.283	-32.276	Biobaz (2013)	*B. azoricus*	93.97	0	0	[92]	PRJEB36091	GCA_903813675
1586I	Lucky Strike (Eiffel Tower)	37.283	-32.276	Biobaz (2013)	*B. azoricus*	94.53	0	0	[93]		SAMEA6822959
1586J	Lucky Strike (Eiffel Tower)	37.283	-32.276	Biobaz (2013)	*B. azoricus*	93.97	0.56	0	[92]	PRJEB36091	GCA_903799955
1586K	Lucky Strike (Montsegur)	37.288	-32.276	Biobaz (2013)	*B. azoricus*	92.28	0.56	0	[92]	PRJEB36091	GCA_903813645
1586N	Lucky Strike (Montsegur)	37.288	-32.276	Biobaz (2013)	*B. azoricus*	94.53	0	0	[92]	PRJEB36091	GCA_903813615
1586O	Lucky Strike (Montsegur)	37.288	-32.276	Biobaz (2013)	*B. azoricus*	93.6	1.12	50	[92]	PRJEB36091	GCA_903813655
1586P	Menez Gwen (White Flames)	37.844	-31.519	Biobaz (2013)	*B. azoricus*	91.72	0	0	[92]	PRJEB36091	GCA_903813625
1586Q	Menez Gwen (White Flames)	37.844	-31.519	Biobaz (2013)	*B. azoricus*	92.28	0	0	[92]	PRJEB36091	GCA_903813695
1586R	Menez Gwen (White Flames)	37.844	-31.519	Biobaz (2013)	*B. azoricus*	93.97	0	0	[93]		SAMEA6822960
1586S	Menez Gwen (White Flames)	37.844	-31.519	Biobaz (2013)	*B. azoricus*	93.97	0	0	[92]	PRJEB36091	GCA_903813685
1600F	Rainbow	36.229	-33.902	Biobaz (2013)	*B. azoricus*	93.97	0	0	[92]	PRJEB36091	GCA_903813635

MAG	Site	Lat	Lon	Cruise	Host species	Complete	Contam	Strain	Pub	BioProject	Accession
1600G	Rainbow	36.229	-33.902	Biobaz (2013)	*B. azoricus*	93.97	0	0	[92]	PRJEB36091	GCA_903813705
1600H	Rainbow	36.229	-33.902	Biobaz (2013)	*B. azoricus*	93.93	0	0	[92]	PRJEB36091	GCA_903819385
1600I	Rainbow	36.229	-33.902	Biobaz (2013)	*B. azoricus*	91.72	0	0	[92]	PRJEB36091	GCA_903813715
1600J	Rainbow	36.229	-33.902	Biobaz (2013)	*B. azoricus*	93.97	0.14	100	[92]	PRJEB36091	GCA_903813725
BBROOKSOX	Chapopote (Mexico)	21.90002	-93.4353	M114-2	*B. brooksi*	93.27	0.56	0	[93]	PRJEB17996	GCA_900128405.1
BHECKSOX	Chapopote (Mexico)	21.90005	-93.4354	M114-2	*B. heckerae*	92.66	4.31	63.6	[93]	PRJEB17996	GCA_900128515.1
1115A	Semenov	13.513	-44.963	Odemar (2014)	*B. puteoserpentis*	94.53	3.09	75	[92]	PRJEB36091	GCA_903813375
1115B	Semenov	13.513	-44.963	Odemar (2014)	*B. puteoserpentis*	93.41	3.09	85.7	[92]	PRJEB36091	GCA_903813385
1115C	Semenov	13.513	-44.963	Odemar (2014)	*B. puteoserpentis*	94.53	0.56	100	[92]	PRJEB36091	GCA_903813395
2065A	Logatchev Quest	14.753	-44.979	M64-2 Logatchev (2005)	*B. puteoserpentis*	94.72	7.3	100	[92]	PRJEB36091	GCA_903813905
2065B	Logatchev Quest	14.753	-44.979	M64-2 Logatchev (2005)	*B. puteoserpentis*	94.72	1.15	100	[92]	PRJEB36091	GCA_903813925
2487A	Logatchev Quest	14.753	-44.980	M126 (2016)	*B. puteoserpentis*	94.16	1.4	57.1	[92]	PRJEB36091	GCA_903819405
2487B	Logatchev Quest	14.753	-44.980	M126 (2016)	*B. puteoserpentis*	94.72	1.4	57.1	[92]	PRJEB36091	GCA_903819245

MAG	Site	Lat	Lon	Cruise	Host species	Complete	Contam	Strain	Pub	BioProject	Accession
2487C	Logatchev Quest	14.753	-44.980	M126 (2016)	*B. puteoserpentis*	94.72	1.4	83.3	[92]	PRJEB36091	GCA_903813875
2487D	Semenov	13.514	-44.963	M126 (2016)	*B. puteoserpentis*	94.53	1.12	66.7	[92]	PRJEB36091	GCA_903813855
2487E	Semenov	13.514	-44.963	M126 (2016)	*B. puteoserpentis*	94.72	1.12	66.7	[92]	PRJEB36091	GCA_903819375
2487F	Semenov	13.514	-44.963	M126 (2016)	*B. puteoserpentis*	94.53	1.97	83.3	[92]	PRJEB36091	GCA_903813935
3722CJ	Logatchev Quest	14.753	-44.981	MSM10-03 Hydrom ar VII	*B. puteoserpentis*	94.72	5.15	50	[92]	PRJEB36091	GCA_903814125
3722CK	Logatchev Quest	14.753	-44.981	MSM10-03 Hydrom ar VII	*B. puteoserpentis*	94.72	3.65	78.6	[92]	PRJEB36091	GCA_903819395
3722CL	Logatchev Quest	14.753	-44.981	MSM10-03 Hydrom ar VII	*B. puteoserpentis*	94.72	3.93	92.9	[92]	PRJEB36091	GCA_903814105
3722CM	Logatchev Quest	14.753	-44.981	MSM10-03 Hydrom ar VII	*B. puteoserpentis*	94.72	3.23	100	[92]	PRJEB36091	GCA_903814155
3722CN	Logatchev Quest	14.753	-44.981	MSM10-03 Hydrom ar VII	*B. puteoserpentis*	93.6	2.39	76.9	[92]	PRJEB36091	GCA_903814145
3722CO	Logatchev Quest	14.753	-44.981	MSM10-03 Hydrom ar VII	*B. puteoserpentis*	93.03	2.81	92.3	[92]	PRJEB36091	GCA_903814115

MAG	Site	Lat	Lon	Cruise	Host species	Complete	Contam	Strain	Pub	BioProject	Accession
3722CP	Logatchev Quest	14.753	-44.980	M126 (2016)	*B. puteoserpentis*	94.72	8.8	90.9	[92]	PRJEB36091	GCA_903814135
Endosymbiont of *Bathymodiolus septemdierum* str. Myojin knoll DNA, complete genome	Izu-Bonin Arc, Myojin knoll (Japan)	32.104	139.219		*B. septemdierum*	94.83	0.56	100	[56]	PRJDB949	NZ_AP013042.1
C112	Clueless	-4.803	-12.372	M78-2 (2009)	*B.* sp. Clueless	94.53	1.31	75	[92]	PRJEB36091	GCA_903814195
C113	Clueless	-4.803	-12.372	M78-2 (2009)	*B.* sp. Clueless	94.53	1.4	66.7	[92]	PRJEB36091	GCA_903814185
C114	Clueless	-4.803	-12.372	M78-2 (2009)	*B.* sp. Clueless	94.53	1.4	66.7	[92]	PRJEB36091	GCA_903814175
L102	Lilliput	-9.547	-13.210	M78-2 (2009)	*B.* sp. Lilliput	93.33	1.12	100	[92]	PRJEB36091	GCA_903813445
L51	Lilliput	-9.547	-13.210	M78-2 (2009)	*B.* sp. Lilliput	94.08	1.69	100	[92]	PRJEB36091	GCA_903813485
L54	Lilliput	-9.547	-13.210	M78-2 (2009)	*B.* sp. Lilliput	94.36	0.84	100	[92]	PRJEB36091	GCA_903813455
Bathymodiolus thermophilus thioautotrophic gill symbiont strain:BAT/CrabSpa'14	East Pacific Rise (EPR) 9°N	9.839833	-104.292	R/V *Atlantis* cruise AT26–10	*B. thermophilus*	96.98	11.32	81.4	[94]	PRJNA339702	GCA_001875585
Ca. Ruthia magnifica str. Cm	9° East Pacific Rise vent field	9.830	-104.290		*C. magnifica*	86.67	0	0	[95]	PRJNA16841	CP000488

MAG	Site	Lat	Lon	Cruise	Host species	Complete	Contam	Strain	Pub	BioProject	Accession
Ca. Vesicomyosocius okutanii HA	Sagami Bay	35.117	139.383		*C. okutanii*	85.69	0	0	[96]	PRJDA18267	AP009247
Ca. Thioglobus autotrophicus strain EF1	Effingham Inlet (estimated coordinates)	49.029	-125.154			94.64	0	0	[97]	PRJNA224116	NZ_CP010552
Ca. Thioglobus singularis PS1	Puget Sound	47.600	-122.450			94.36	0	0	[98]	PRJNA229178	CP006911
Thiomicrospira crunogena XCL-2 (*Hydrogenovibrio crunogenus* XCL-2)						99.72	0.19	0	[99]	PRJNA13018	NC_007520.2

3   References

1.   Won Y-J, Hallam SJ, O'Mullan GD, Pan IL, Buck KR, Vrijenhoek RC. Environmental acquisition of thiotrophic endosymbionts by deep-sea mussels of the genus *Bathymodiolus*. Appl Environ Microbiol. 2003; 69: 6785–92.
2.   RStudio Team. RStudio: integrated development for R. Boston, MA: RStudio, Inc.; 2015. http://www.rstudio.com/ (accessed December 19, 2016).
3.   South A. rnaturalearth: World map data from Natural Earth. 2017. https://CRAN.R-project.org/package=rnaturalearth (accessed June 23, 2020).
4.   Gallic E. legendMap: north arrow and scale bar for ggplot2 graphics. 2016. https://rdrr.io/github/3wen/legendMap/man/legendMap-package.html (accessed June 23, 2020).
5.   Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, et al. ggplot2: create elegant data visualisations using the grammar of graphics. 2019. https://CRAN.R-project.org/package=ggplot2 (accessed November 25, 2019).
6.   Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 2003; 164: 1567–87.
7.   Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155: 945–59.
8.   Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour. 2009; 9: 1322–32.
9.   Chhatre VE, Emerson KJ. StrAuto: Automation and parallelization of STRUCTURE analysis. BMC Bioinformatics. 2017; 18: 192.
10.   CLUMPAK server. n.d. http://clumpak.tau.ac.il/ (accessed November 25, 2019).
11.   Anderson EC, Thompson EA. A model-based method for identifying species hybrids using multilocus genetic data. Genetics. 2002; 160: 1217–29.
12.   Gompert Z, Buerkle CA. introgress: a software package for mapping components of isolation in hybrids. Mol Ecol Resour. 2010; 10: 378–84.
13.   Gompert Z, Buerkle CA. A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. Mol Ecol. 2009; 18: 1207–24.
14.   Breusing C, Biastoch A, Drews A, Metaxas A, Jollivet D, Vrijenhoek RC, et al. Biophysical and population genetic models predict the presence of "phantom" stepping stones connecting Mid-Atlantic Ridge vent ecosystems. Curr Biol. 2016; 26: 2257–67.
15.   Breusing C, Vrijenhoek RC, Reusch TBH. Widespread introgression in deep-sea hydrothermal vent mussels. BMC Evol Biol. 2017; 17: 13.
16.   NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2016; 44: D7–19.
17.   Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004; 5: 113.
18.   Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32: 1792–7.
19.   Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015; 32: 268–74.
20.   Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017; 14: 587–

21. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013; 30: 1188–95.

22. Rota-Stabelli O, Yang Z, Telford MJ. MtZoa: a general mitochondrial amino acid substitutions model for animal evolutionary studies. Mol Phylogenet Evol. 2009; 52: 268–72.

23. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019; 47: W256–9.

24. Adobe. Adobe Illustrator. 2020. https://www.adobe.com/de/products/illustrator.html (accessed February 3, 2020).

25. Bushnell B. BBMap. 2014. http://sourceforge.net/projects/bbmap/ (accessed October 21, 2016).

26. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015; 31: 1674–6.

27. Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, et al. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods. 2016; 102: 3–11.

28. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 2015; 3: e1165.

29. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009; 25: 2078–9.

30. Seemann T. barrnap. 2014. http://www.vicbioinformatics.com/software.barrnap.shtml (accessed October 21, 2016).

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

32. Edgar RC. Usearch. 2010. http://www.drive5.com/usearch/ (accessed October 21, 2016).

33. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013; 41: D590–6.

34. Wu M, Scott AJ. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinformatics. 2012; 28: 1033–4.

35. Seah BKB, Gruber-Vodicka HR. gbtools: Interactive visualization of metagenome bins in R. Front Microbiol. 2015; 6: 1451.

36. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015; 25: 1043–55.

37. Matsen FA, Kodner RB, Armbrust EV. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics. 2010; 11: 538.

38. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010; 11: 119.

39. HMMER. n.d. http://hmmer.org/ (accessed November 25, 2019).

40. Ansorge R, Romano S, Sayavedra L, González Porras MÁ, Kupczok A, Tegetmeyer HE, et al. Functional diversity enables multiple symbiont strains to coexist in deep-sea mussels. Nat Microbiol. 2019; 4: 2487–97.

41.   Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012; 19: 455–77.

42.   Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of *de novo* genome assemblies. Bioinformatics. 2015; 31: 3350–2.

43.   Lee MD. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics. 2019; 35: 4162–4.

44.   Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009; 25: 1972–3.

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

46.   Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol. 2008; 25: 1307–20.

47.   Jombart T, Kamvar ZN, Collins C, Lustrik R, Beugin M-P, Knaus BJ, et al. adegenet: exploratory analysis of genetic and genomic data. 2020. https://CRAN.R-project.org/package=adegenet (accessed February 6, 2020).

48.   Charif D, Clerc O, Frank C, Lobry JR, Necşulea A, Palmeira L, et al. seqinr: biological sequences retrieval and analysis. 2019. https://CRAN.R-project.org/package=seqinr (accessed February 20, 2020).

49.   Goudet J, Jombart T. hierfstat: estimation and tests of hierarchical F-statistics. 2015. https://CRAN.R-project.org/package=hierfstat (accessed February 20, 2020).

50.   Meirmans PG. The trouble with isolation by distance. Mol Ecol. 2012; 21: 2839–46.

51.   Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community ecology package. 2019. https://CRAN.R-project.org/package=vegan (accessed November 25, 2019).

52.   Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018; 9: 5114.

53.   Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, et al. gplots: various R programming tools for plotting data. 2019. https://CRAN.R-project.org/package=gplots (accessed November 25, 2019).

54.   Demin G. maditr: fast data aggregation, modification, and filtering with pipes and "data.table." 2019. https://CRAN.R-project.org/package=maditr (accessed November 25, 2019).

55.   Paradis E, Blomberg S, Bolker B, Brown J, Claude J, Cuong HS, et al. ape: Analyses of phylogenetics and evolution. 2019. https://CRAN.R-project.org/package=ape (accessed February 6, 2020).

56.   Ikuta T, Takaki Y, Nagai Y, Shimamura S, Tsuda M, Kawagucci S, et al. Heterogeneous composition of key metabolic gene clusters in a vent mussel symbiont population. ISME J. 2016; 10: 990–1001.

57.   Picazo DR, Dagan T, Ansorge R, Petersen JM, Dubilier N, Kupczok A. Horizontally transmitted symbiont populations in deep-sea mussels are genetically isolated. ISME J. 2019; 13: 2954–68.

58.   Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014; 30: 2068–9.

59.   Broad Institute. Picard Tools. 2013. https://github.com/broadinstitute/picard (accessed January 17, 2017).

60. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20: 1297–303. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2928508/ (accessed January 17, 2017).

61. Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol. 2013; 79: 7696–701.

62. Vinuesa P, Contreras-Moreira B. Robust identification of orthologues and paralogues for microbial pan-genomics using GET_HOMOLOGUES: a case study of pIncA/C plasmids. In: Mengoni A, Galardini M, Fondi M, editors. Bact. Pangenomics, vol. 1231, New York, NY: Springer; 2015, p. 203–32.

63. Li L, Stoeckert CJ, Roos DS. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. 2003; 13: 2178–89.

64. Kristensen DM, Kannan L, Coleman MK, Wolf YI, Sorokin A, Koonin EV, et al. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. Bioinformatics. 2010; 26: 1481–7.

65. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25: 3389–402.

66. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, et al. The Bioperl Toolkit: Perl modules for the life sciences. Genome Res. 2002; 12: 1611–8.

67. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015; 12: 59–60.

68. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016; 44: D279–85.

69. Brown NP, Leroy C, Sander C. MView: a web-compatible database search or multiple alignment viewer. Bioinformatics. 1998; 14: 380–1.

70. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016; 34: 525–7.

71. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol. 2011; 29: 644–52.

72. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013; 8: 1494–512.

73. Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB. ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-Seq. PLOS ONE. 2013; 8: e67019.

74. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-Seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. Microbiome. 2014; 2: 15.

75. Gloor GB, Macklaim JM, Fernandes AD. Displaying variation in large datasets: plotting a visual summary of effect sizes. J Comput Graph Stat. 2016; 25: 971–9.

76. Dowle M, Srinivasan A, Gorecki J, Chirico M, Stetsenko P, Short T, et al. data.table: extension of "data.frame." 2019. https://CRAN.R-project.org/package=data.table (accessed November 25, 2019).

77. Borcard D, Gillet F, Legendre P. Numerical ecology with R. Springer; 2018.

78. Amend JP, McCollom TM, Hentscher M, Bach W. Catabolic and anabolic energy for chemolithoautotrophs in deep-sea hydrothermal systems hosted in different rock types. Geochim Cosmochim Acta. 2011; 75: 5736–48.

79. Ter Braak CJF. The analysis of vegetation-environment relationships by canonical correspondence analysis. Vegetatio. 1987; 69: 69–77.

80. Legendre P. Spatial autocorrelation: Trouble or new paradigm? Ecology. 1993; 74: 1659–73.

81. Beck MW. ggord: ordination plots with ggplot2. 2019. https://rdrr.io/github/fawda123/ggord/man/ggord.html (accessed June 23, 2020).

82. Arnold ML, Hodges SA. Are natural hybrids fit or unfit relative to their parents? Trends Ecol Evol. 1995; 10: 67–71.

83. Colaço A, Martins I, Laranjo M, Pires L, Leal C, Prieto C, et al. Annual spawning of the hydrothermal vent mussel, *Bathymodiolus azoricus*, under controlled aquarium conditions at atmospheric pressure. J Exp Mar Biol Ecol. 2006; 333: 166–71.

84. Miyake H, Kitada M, Itoh T, Nemoto S, Okuyama Y, Watanabe H, et al. Larvae of deep-sea chemosynthetic ecosystem animals in captivity. Cah Biol Mar. 2010; 51: 441–50.

85. Johnson SB, Won Y-J, Harvey JB, Vrijenhoek RC. A hybrid zone between *Bathymodiolus* mussel lineages from eastern Pacific hydrothermal vents. BMC Evol Biol. 2013; 13: 21.

86. Bordenstein SR, O'Hara FP, Werren JH. *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. Nature. 2001; 409: 707–10.

87. Breeuwer JAJ, Werren JH. Hybrid breakdown between two haplodiploid species: the role of nuclear and cytoplasmic genes. Evolution. 1995; 49: 705–17.

88. Coyne JA. Genetics and speciation. Nature. 1992; 355: 511–5.

89. Wu CI, Davis AW. Evolution of postmating reproductive isolation: the composite nature of Haldane's rule and its genetic bases. Am Nat. 1993; 142: 187–212.

90. Wright S. Isolation by distance. Genetics. 1943; 28: 114–38.

91. Meirmans PG. Seven common mistakes in population genetics and how to avoid them. Mol Ecol. 2015; 24: 3223–31.

92. Ansorge R, Romano S, Sayavedra L, Rubin-Blum M, Gruber-Vodicka HR, Scilipoti S, et al. The hidden pangenome: comparative genomics reveals pervasive diversity in symbiotic and free-living sulfur-oxidizing bacteria. BioRxiv. 2020: 2020.12.11.421487.

93. Sayavedra L, Ansorge R, Rubin-Blum M, Leisch N, Dubilier N, Petersen JM. Horizontal acquisition followed by expansion and diversification of toxin-related genes in deep-sea bivalve symbionts. BioRxiv. 2019: 605386.

94. Ponnudurai R, Sayavedra L, Kleiner M, Heiden SE, Thürmer A, Felbeck H, et al. Genome sequence of the sulfur-oxidizing *Bathymodiolus thermophilus* gill endosymbiont. Stand Genomic Sci. 2017; 12: 50.

95. Newton ILG, Woyke T, Auchtung TA, Dilly GF, Dutton RJ, Fisher MC, et al. The *Calyptogena magnifica* chemoautotrophic symbiont genome. Science. 2007; 315: 998–1000.

96. Kuwahara H, Yoshida T, Takaki Y, Shimamura S, Nishi S, Harada M, et al. Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*. Curr Biol. 2007; 17: 881–6.

97. Shah V, Morris RM. Genome sequence of "*Candidatus* Thioglobus autotrophica" strain EF1, a chemoautotroph from the SUP05 clade of marine Gammaproteobacteria. Genome Announc. 2015; 3: e01156-15.

98.   Marshall KT, Morris RM. Genome sequence of "*Candidatus* Thioglobus singularis" strain PS1, a mixotroph from the SUP05 clade of marine Gammaproteobacteria. Genome Announc. 2015; 3: e01155-15.

99.   Scott KM, Sievert SM, Abril FN, Ball LA, Barrett CJ, Blake RA, et al. The genome of deep-sea vent chemolithoautotroph *Thiomicrospira crunogena* XCL-2. PLoS Biol. 2006; 4: e383.

water

[ˈwɔːtə] *noun*

a colourless, transparent, odourless liquid that forms the seas, lakes, rivers, and rain and is the basis of the fluids of living organisms

Chapter III | Symbionts in the water column

Chasing free-living stages of mussel symbionts– Metagenomic insights from the environment to the host

Merle Ücker[1,2], Rebecca Ansorge[1,3], Yui Sato[1], Ga Yan Grace Ho[1], Christian Borowski[1], Nicole Dubilier[1,2,*]

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

[2]MARUM – Center for Marine Environmental Sciences of the University of Bremen, Bremen, Germany

[3]Quadram Institute Bioscience, Norwich, Norfolk, United Kingdom

[*]Corresponding author

This manuscript is in preparation and has not been reviewed by all authors.

Abstract

In marine animal–microbe symbioses, horizontal transmission is a common way of symbiont transfer between host generations. The horizontally transmitted sulphur-oxidising symbionts of the deep-sea mussel *Bathymodiolus* enable their host to be among the most successful fauna at hydrothermal vents worldwide. Horizontal transmission can have different forms and only by investigating the free-living stage, we can learn more about the interaction of the partners, potential uptake mechanisms, and the specificity of the association. Despite the importance of their free-living stage, only few studies investigated the symbionts in the surrounding seawater. Thus, their geographic distribution is unresolved. In our metagenomics study, we analysed the free-living community from seven vent sites along the northern Mid-Atlantic Ridge that are inhabited by *Bathymodiolus* mussels. We screened for marker genes, including the small ribosomal subunit, to resolve the diversity of free-living bacteria surrounding the mussels and the geographic distribution of symbiont subspecies in the water column. Despite the low diversity community within the mussel gills, we found a high diversity of bacteria closely related to *Bathymodiolus* symbionts in the water, highlighting the specificity of the host-symbiont association. Our results revealed that symbiont DNA was present in the water column but always co-occurred with mussel DNA, indicating that symbionts are mostly present associated to host tissue, e.g. in released bacteriocytes. We established a workflow to identify marker genes distinguishing symbiont subspecies and found that one subspecies was dominant per site. Our study provides further insights into the low abundant free-living stage of *Bathymodiolus* symbionts, discusses a likely dominant mode of transmission via host particles, and provides suggestions for future sampling campaigns.

Introduction

Animal–microbe symbioses are ubiquitous and play a vital role for animal development, speciation, and the evolution of eukaryotic life [1–4]. The transmission mode of symbionts between host generations has important implications for the stability of the association and the genetic diversity of the symbiont population [5,6]. There are two main modes of symbiont transmission: vertical and horizontal transmission. During vertical transmission, symbionts are directly passed on from one generation to the next, often from the female to the offspring [5,7]. Vertical transmission is associated with a high congruence of host and symbiont phylogeny, strong host fidelity, decreased genetic diversity, and genome reduction of the symbiont [3,6,8]. In contrast, during horizontal transmission symbionts are acquired from the

environment or neighbouring hosts in each generation. This allows for more genetically diverse symbiont communities due to comparatively high rates of genetic exchange with the free-living pool of bacteria. Horizontal transmission often displays little to no phylogenetic congruence between host and symbionts [5,6]. In many symbioses, transmission does not strictly follow either of the two modes but is rather a mixed-mode transmission, e.g. vertical transmission with events of horizontal transmission.

Throughout the world's oceans, horizontal transmission is the most commonly found mode of symbiont transmission [9]. One of the few well-studied symbiosis with horizontal transmission is the association of deep-sea *Bathymodiolus* mussels with their chemosynthetic symbionts. *Bathymodiolus* mussels live in an environment deprived of light and photosynthesis-derived carbon, and are among the most successful fauna at hydrothermal vents and cold seeps [10,11]. Two mussel species occur at vent sites along the northern Mid-Atlantic Ridge (NMAR). While mussels of the species *B. azoricus* populate the northern sites close to the Azores (White Flames and Woody in the Menez Gwen vent field, Eiffel Tower and Montsegur in the Lucky Strike field, and Rainbow; Figure 1), mussels of the species *B. puteoserpentis* populate the southern sites (Logatchev and Semenov; Figure 1).

Symbionts of *Bathymodiolus* mussels are hosted in the gills. Using reduced chemical compounds as energy sources to produce biomass from inorganic carbon, they provide nutrition to their hosts [11–13]. Two types of symbionts are found in mussels at the NMAR: Sulphur-, on which this study will focus, and methane-oxidising (SOX and MOX) symbionts [14]. SOX symbionts of both *Bathymodiolus* species from the NMAR belong to the order of Thiomicrospirales. More specifically, they are part of the *Thioglobaceae* family and fall into the SUP05/Arctic96BD-19 clade, a widely distributed clade of marine sulphur-oxidising bacteria with diverse metabolisms, and both symbiotic and free-living lifestyles [15]. Bacteria within the SUP05/Arctic96BD-19 clade share 94 % 16S rRNA nucleotide identity indicating that they are closely related [15]. The SOX symbionts of *B. azoricus* and *B. puteoserpentis* can be considered one bacterial species (above 95 % average nucleotide identity (ANI)[16]) but belong to different subspecies (as defined by Van Rossum et al. (2020), above 97 % ANI [18,19]). Horizontal transmission was first described for *Bathymodiolus* SOX symbionts in 1988, when Le Pennec and colleagues suggested the endocytosis of bacteria in the gills of the eastern Pacific mussel *B. thermophilus* [20]. Molecular studies later found a lack of congruence between host and symbiont phylogeny, which is in line with symbiont acquisition from the environment or via mixed mode transmission [19,21–23].

Horizontally transmitted symbionts experience a free-living stage in which they are exposed to and potentially influenced by the environment. Energy metabolism of microbial communities at hydrothermal plumes across the Eastern Lau Spreading Centre seem to be 'dictated' by plume chemistry [24]. Microbial communities and metabolisms in the north-eastern Pacific Ocean have been shown to be shaped by the geochemistry and physical structure of the investigated vents [25]. Recent studies of *Bathymodiolus* symbionts found that the bacterial population structure within mussels is highly correlated with the biogeography, but independent of their host's genotype. The geographical structuring of symbionts was hypothesised to reflect a geographically structured population of symbionts in the free-living stage [18,26]. Alternatively, if the free-living pool of bacteria and host genetics are assumed identical, the environmental differences between sites might define which symbionts are taken up.

Horizontal acquisition of symbionts is associated with high risks for the host, as a free-living symbiont population that can be recognised by the mussels must be available [5]. Once the free-living symbionts are taken up into their host, they contribute to genetic and functional diversity of the symbiont community. The uptake of genetically diverse symbionts potentially benefits their host by being adaptable to environmental changes. Although the free-living stage of symbionts is important to understand the interaction between partners, potential uptake mechanisms, and the specificity in horizontally transmitted symbioses, knowledge about this stage is scarce. A few studies identified *Bathymodiolus* symbiont-related genes (e.g. 16S rRNA or *cbbL*) in seawater samples or microbial mats around mussel beds in the Atlantic and the Pacific [22,27–29]. These studies investigated symbionts at only one site each, even though multi-site studies are pivotal to resolve phylogeographic patterns of the free-living stage of *Bathymodiolus* symbionts. Yet, the distribution of different symbiont subspecies, and whether the free-living population is indeed geographically structured, remains unclear.

Here, we analyse seawater metagenomes from seven vent sites in five vent fields along the northern Mid-Atlantic Ridge to investigate 1) the diversity of free-living marine bacteria including *Bathymodiolus* symbionts, 2) the presence of symbionts in the free-living stage, and 3) the geographic distribution of different symbiont subspecies.
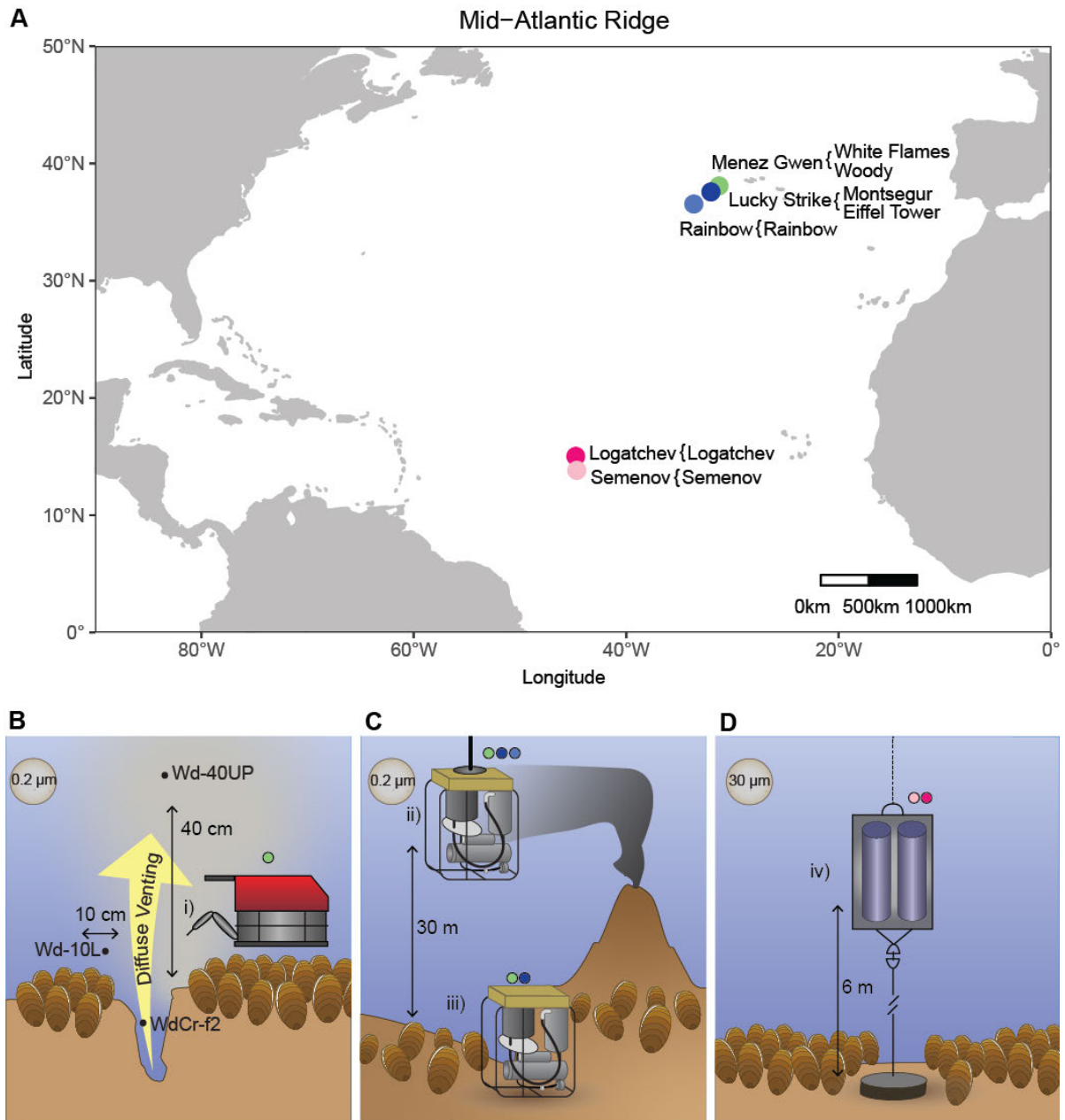
Figure 1 | Overview of sampling along the northern Mid-Atlantic Ridge. A: Vent fields and vent sites located within each field where water was sampled. B: Research cruise M82-3 (2010). Black points show sample locations relative to the crack with diffuse venting at site Woody in the Menez Gwen vent field. i) Kiel In Situ Pumping System (KIPS) mounted to ROV Quest. Modified after Meier et al. (2016). C: Research cruise BioBaz (2013). ii) Wire-mounted in-situ pump (WISP); iii) in-situ pump deposited on the seafloor by ROV Victor 6000 (EISP). D: Research cruise M126 BigMAR (2016). iv) In-situ pump mounted to a mooring. Coloured circles represent the location corresponding to the colour scheme of the map in panel A. Circles in top left corners state the filter pore sizes used.

Materials and methods

Collection of seawater filters

Bacterial populations from bottom water were sampled at five hydrothermal vent fields at the Mid-Atlantic Ridge (Supplementary Table S 1). An overview of sampling sites was plotted with RStudio v1.2.1335 using R v3.6.2 and the packages mapplots v1.5.1, plotrix v3.7-6, and mapproj v1.2.6 (Figure 1 A). During the French research cruise BioBaz in August 2013 with RV Pourquoi Pas?, sampling was conducted at the fields Menez Gwen, Lucky Strike, and Rainbow (Figure 1 C). During the German research cruise M126 (BigMAR) with RV METEOR in May 2016, sampling was conducted at the vent fields Logatchev and Semenov (Figure 1 D). Seawater volumes ranging from 550 to 24,000 L per site were filtered *in situ* adjacent to hydrothermal venting.

During the BioBaz cruise, bacterial populations were collected with McLane WTS-LV04 in-situ pumps (ISP). ISPs were either deployed on a wire from the research vessel or positioned on the seafloor by the remotely operated vehicle (ROV) Victor 6000. Wire deployments of in-situ pumps (WISP) were positioned in 20-30 m height above seafloor and as close as ship navigation allowed to the focused discharges of the White Flames vent in the Menez Gwen hydrothermal vent field, Eiffel Tower at Lucky Strike, and between edifices Iris 9 and France 5 at Rainbow. For ROV-positioned deployments (EISP), in-situ pumps were shuttled by an instrument carrier to the seafloor and placed by the ROV within 3 m distance to the Woody vent site in Menez Gwen (Supplementary Figure S 1 A) and the Montsegur edifice in Lucky Strike (Supplementary Figure S 1 C). Water was filtered through 142 mm diameter cellulose acetate filters with 0.22 µm pore size. Filters were frozen at -80°C upon recovery on board.

During the M126 cruise, a McLane WTS-LV30 in-situ pump was moored free falling near the Michelangelo vent in the Semenov 2 field and the Irina 2 vent at Logatchev. Water collection height was 6 m above ground. Water was filtered over a 30 µm plankton mesh that was fixed upon recovery on board for fluorescence in-situ hybridisation (2 % paraformaldehyde in 1x PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 2 mM KH2PO4) solution) and stored at -20°C in a 60 % Ethanol/PBS solution. No mussel larvae were present on any of the filters (M. Franke, personal communication).

DNA extraction and metagenomic sequencing of seawater samples

DNA was extracted from filters from the Biobaz cruise using the ZR Duet DNA/RNA MiniPrep Plus Kit (Zymo Research, Freiberg, Germany). DNA extraction was performed in duplicates, i.e. for each sample, one slice of 1-2 cm of the filter was cut with disinfected

scissors into four pieces of which two were used per duplicate. Samples were prepared according to the Tough-to-Lyse Protocol. Briefly, 400 µl DNA/RNA Shield (Zymo Research, Freiberg, Germany) and silica beads were added. Samples were treated for 2 min at 20/s with the TissueRuptor II (QIAGEN, Hilden, Germany) and centrifuged for 3 min at 4000 rpm. 400 µl of DNA/RNA Lysis Buffer (Zymo Research, Freiberg, Germany) was added to the mix and the steps with the TissueRuptor II (QIAGEN, Hilden, Germany) and centrifugation was repeated. The supernatant was transferred onto the spin column for DNA extraction following the manufacturer's protocol with the following modifications: All centrifugation steps were performed for 60 instead of 30 s. In the washing steps, the centrifugation was repeated to remove all liquid (Step 4 and 5). Instead of 100 µl of DNase/RNase-Free Water (Zymo Research, Freiberg, Germany), 50 µl were used for the elution.

DNA was extracted from 1-2 cm slices of plankton gaze from the BigMAR cruise (fixed with paraformaldehyde). Slices were divided into smaller pieces for better detachment of the biological material before adding 450 µl Buffer PKD (QIAGEN, Hilden, Germany) and 30 µl proteinase K (QIAGEN, Hilden, Germany) and incubating at 37°C and 400 rpm-shaking for 4 hours in a heat bath. Another 60 µl of proteinase K were added in two steps, once after 8 hours and again after 24 hours. DNA was extracted with the Mo Bio PowerSoil DNA Isolation Kit (QIAGEN, Hilden, Germany) according to the manufacturer's protocol with one modification: Instead of 100 µl of Solution C6, 2 x 20 µl were used for elution (step 20 of the manufacturer's protocol).

Sequencing libraries were prepared with the Nextera DNA Flex Library Prep Kit (Illumina, San Diego, USA), and a total of 10 million 150 bp paired-end reads per sample were sequenced on HiSeq3000 machines. Samples, for which DNA extraction duplicates were prepared, were only sequenced with 5 million read pairs per duplicate library to result in a total 10 million per original sample. Library preparation and metagenomic sequencing was performed by the Max Planck-Genome-centre Cologne, Germany. Metagenomic data of duplicate sequencing libraries were pooled per sample after evaluation of the taxonomic composition (see Supplementary Figure S 2).

Data from public databases used in bioinformatic analysis

In addition to water samples sequenced for this study, we used three metagenomes from Meier et al. (2016). A detailed description of sampling is available in the publication [30]. In short, seawater was sampled at different spots at the Woody site in the Menez Gwen hydrothermal vent field (Figure 1 B): The diffuse fluid outflow was collected directly at the

subsurface within the crack, i.e. in the middle of the flow stream (WdCr-f1, Supplementary Figure S 1 B), 10 cm left to the crack opening, i.e. more peripheral to the flow stream (Wd-10L), and 40 cm above the crack surface (WD-40UP) by in-situ filtration on filters with 0.2 µm pore size using a stainless steel pressure filter holder mounted on a Kiel In Situ Pumping System (KIPS) [31]. Metagenomic data were retrieved from ENA project PRJEB11362 (https://www.ebi.ac.uk/ena/data/view/PRJEB11362).

Assessment of taxonomic composition

To investigate the diversity of the bacterial deep-sea communities, we assessed the taxonomic composition of metagenomes from seawater filters. We reconstructed the small ribosomal subunits (16S and 18S rRNA) by running phyloFlash v3.3b1 [32] in single-cell mode using the SILVA non-redundant database v132 for classification [33]. Taxonomic compositions were highly similar among all sequencing runs from each filter sample (Supplementary Figure S 2) and we therefore pooled all sequenced reads per sample. The analysis was performed before and after pooling the data. Running phyloFlash_compare.pl (implemented in phyloFlash), we plotted the taxonomic composition of all samples.

Phylogeny of Thiomicrospirales and evolutionary placement of 16S rRNA genes from water filters

We identified the taxonomic affiliation of 16S rRNA genes detected with phyloFlash using an evolutionary placement algorithm (EPA). First, we prepared a set of 16S rRNA sequences of *Bathymodiolus* symbionts and phylogenetically related bacteria that were analysed previously (Figure 2, figure supplement 1, Chapter II, [34]). To retrieve similar sequences, we ran the SINA aligner v1.2.11 [35] using a minimum identity of 0.95, 10 neighbours per query sequence, and the "Add to neighbours tree" option. A RAxML tree was reconstructed and downloaded [36]. All sequences in the resulting tree, including our initial set and the sequences recruited with SINA, were clustered with usearch v9.1.13 [37] at 99.9 % identity to reduce redundancy (148 out of 196 sequences were kept). Secondly, we aligned all reference sequences with MAFFT v.7.310 in G-INSI mode (--global pair --maxiterate 1000) [38,39]. Thirdly, we reconstructed the reference tree using IQ-TREE v1.6.7.1 with the TN+F+R3 model (best model according to ModelFinder) [40–42]. Lastly, assembled 16S rRNA sequences detected in our libraries and classified as Thiomicrospirales by phyloFlash were placed in the reference tree using the EPA [43]. In detail, 16S rRNA sequences were extracted from the phyloFlash output files using bash commands and aligned with MAFFT as described above. RAxML-NG v0.9.0 ModelFinder was used to determine the best model of

sequence evolution for the EPA run based on the reference alignment and reference tree (raxml-ng --evaluate --msa mafft_ref.fasta --tree mafft_ref_fasta.contree --prefix info --model GTR+G+F) [36,44]. The Thiomicrospirales 16S rRNA sequences were placed on the reference tree using epa-ng v0.3.6 [43] with the model determined by ModelFinder keeping only the best match (epa-ng --ref-msa mafft_ref.fasta --tree mafft_ref.fasta.contree --query mafft_query.fasta --model info.raxml.bestModel -- filter-max 1). The final tree was visualized with iTol v5 [45] and edited in Adobe Illustrator [46].

Screening for symbiont 16S rRNA and mussel 18S rRNA and COI gene sequences

To assess the presence and abundance of *Bathymodiolus* symbiont- and host-related genes with an assembly-independent approach, we mapped metagenomic reads against references for symbiont 16S rRNA and host 18S rRNA and cytochrome c oxidase I (COI) genes. We downloaded 16S rRNA gene sequences of *B. azoricus*/*B. puteoserpentis* symbionts (AY235676.1.1425, AY235677.1.1425, AY951931.1.1495, CDSC02000433.4475.6000, CVUD02000265.1.829, DQ321711.1.1208, DQ321712.1.1208, FR670517.1.468, FR670518.1.468, LN871183.1.1341) and their hosts (AF221640.1.1740 and AY649822.1.1751) from the SILVA database (https://www.arb-silva.de/, accessed 2021-01-18 [33]). Host COI gene sequences longer than 1 000 bp (KU597643.1, KU597642.1, KU597641.1, LN833437.1, LN833436.1, LN833435.1, LN833434.1, LN833433.1, KU597631.1, KU597630.1, KU597629.1, KU597632.1) were downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov/, accessed 2021-01-18 [47]). Reads from water metagenomes were mapped against the 16S rRNA, 18S rRNA and COI references with a minimum identity of 99 %, while tossing ambiguous reads using BBSplit from BBTools v37.28 [48]. Results were plotted with R v3.6.3 in RStudio v1.3.959 using the packages ggplot2 v3.3.0, tidyverse v1.3.0, dpylr v0.8.5, and readr v1.3.1 [49–52], and modified in Abode Illustrator [46].

Screening for diagnostic marker genes for symbiont subspecies differentiation and application to mussel and water metagenomes

The resolution of 16S rRNA gene sequences is limited when it comes to differentiating closely related organisms, such as the SOX symbiont subspecies associated with *B. azoricus* and *B. puteoserpentis*. We therefore identified marker genes suitable for differentiation of symbiont subspecies and screened for them in mussel and water metagenomes to compare the distribution of subspecies-specific gene sequences.

The GToTree marker gene set for Gammaproteobacteria (comprising 171 markers) was used as a starting set of diagnostic marker genes [53]. We determined which of these marker genes could be used as diagnostic genes. Therefore, we extracted marker genes from reference MAGs from symbionts of *B. azoricus* (n = 25), *B. puteoserpentis* (n = 27), and other bacteria from the *Thioglobaceae* family (Supplementary Table S 2) and generated pairwise comparisons of identity using blastp. Briefly, the GToTree marker genes extracted from reference MAGs were converted into a custom blast database with makeblastdb v2.9.0+, [54–56]. The custom database was used as a query to search against itself in a series of blastp commands v2.9.0+, [54–56], with parsing of the table-format outputs to make a pairwise identity matrix. The output matrix was plotted as heatmap with R v3.6.3 in RStudio v1.3.959 using the R package pheatmap v1.0.12 [49,57]. Heatmaps were visually inspected to find suitable diagnostic marker genes as described in the supplementary information ("Three diagnostic markers can discriminate *B. azoricus* and *B. puteoserpentis* symbiont subspecies", see https://github.com/gracegyho/markergenescreen for workflow and script). The suitability was determined based on specificity to one symbiont subspecies (high similarity within but low similarity between the two symbiont subspecies).

To quantify the presence of these diagnostic marker genes within the water column and mussel metagenomes, a read mapping-based approach was employed. Metagenome reads (Supplementary Table S 3) were mapped to the subspecies-specific marker genes using BBSplit from the BBToolsv37.28 suite [48]. BBSplit assigns an input read based on the similarity to the specified reference sequence (here, *B. azoricus* symbiont or *B. puteoserpentis* symbiont), and marks it as ambiguous in case of a tie in similarity, or if it falls below the specified minimum identity threshold (here stringent minimum identity of 99 %). Results were plotted with R v3.6.3 in RStudio v1.3.959 using the packages ggplot2 v3.3.0, tidyverse v1.3.0, dpylr v0.8.5, and readr v1.3.1 [49–52], and modified in Adobe Illustrator [46].

Data availability

Sequence data (metagenomic reads and marker gene sequences) and the metadata (recommended standard information on sequence data [58]) are available in the European Nucleotide Archive (ENA) at EMBL-EBI under project accession number PRJEB41358 (https://www.ebi.ac.uk/ena/data/view/PRJEB41358).

Results

Taxonomic composition of the deep-sea community in filtered seawater samples

To investigate the diversity in water filter metagenomes from seven sites in five vent fields colonised by *Bathymodiolus* mussels, we analysed their taxonomic composition. The communities of all samples were of comparable composition as identified with phyloFlash based on reconstruction of 16S/18S rRNA gene sequences, however, the proportions varied (Figure 2). Proteobacteria dominated the communities, especially Campylobacterales (Epsilonproteobacteria/Epsilonbacteraeota), SAR11 clade (Alphaproteobacteria), Thiomicrospirales, and Alteromonadales (both Gammaproteobacteria). Besides the Proteobacteria, bacteria from the Chloroflexi, Flavobacteriales and Marinimicrobia, and archaea belonging to the Nitrosopumilales and the Marine Group II were among the most abundant members of the communities. Up to a third of each community could be attributed to Eukaryota (Retaria, Protalveolata, Discicristata, and Metazoa).
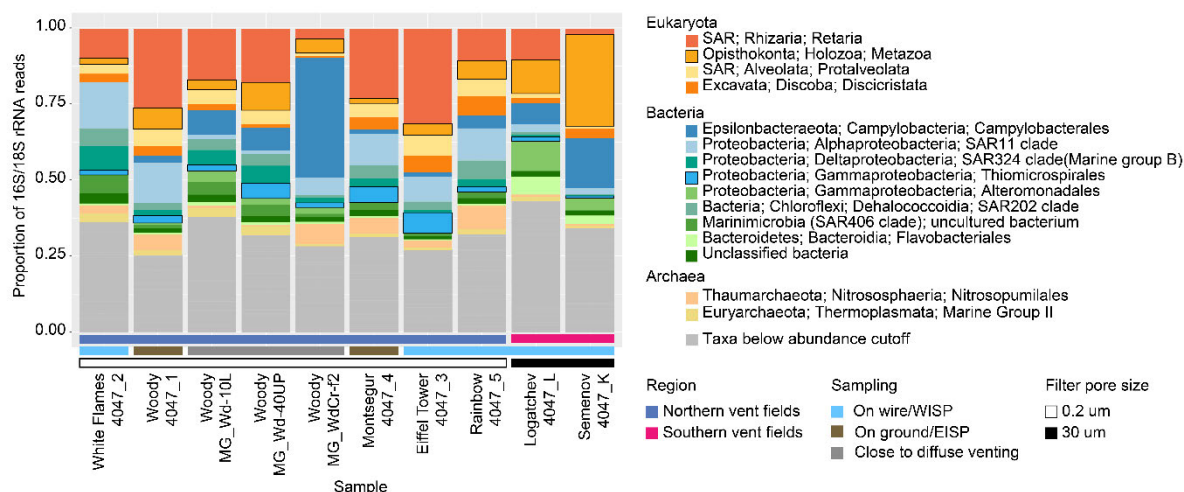
Figure 2 | Taxonomic composition of the deep-sea community in water metagenomes. We analysed the composition with phyloFlash to detect and classify reads matching 16S and 18S rRNA gene sequences against the SILVA database. Reads from *Bathymodiolus* mussels fall into the Metazoa fraction while the Thiomicrospirales fraction includes *Bathymodiolus* SOX symbionts (both highlighted with black rim). Colour bar indicates the region of sampling: Northern vent fields (blue) or southern vent fields (pink). Colour coded squares represent the sampling procedure and the filter pore size: On wire/WISP (blue), on ground/EISP (brown), close to diffuse venting (grey), 0.2 µm pore size (white), 30 µm pore size (black). Labels indicate sampling locations and sample names.

Among the samples, we observed individual differences: samples from vent fields Logatchev and Semenov (4047_K and 4047_L) had a higher proportion of Flavobacteria and

Alteromonadales compared to water samples from the Azores region (samples 4047_1 to 4047_5). The Azores samples from our study were similar to one another but they differed from the metagenomes from [30] (samples starting with MG-Wd) that were sampled in the same region or in some cases at the same vent field. Our Azores samples (4047_1 to 4047_5) showed higher proportions of bacteria from the SAR11 clade/Pelagibacterales and archaea belonging to the Nitrosopumilales, whereas Campylobacterales were more prominent in the metagenomes from [30] sampled at Woody in Menez Gwen. This difference might be due to different sampling spots, either directly in the diffuse fluid flow [30] or a few meters away from focused or diffuse fluid discharge near the seafloor or even tens of meters away in the water column (our study). The relative proportion of Metazoa-associated reads, the taxon to which *Bathymodiolus* mussels belong, varied randomly between the water metagenomes (1.55 to 30.07 % of all 16S and 18S rRNA sequences) and was highest in Semenov (4047_K).

We were particularly interested in whether we could find symbionts of *Bathymodiolus* mussels in the water samples from seven vent sites along the northern Mid-Atlantic Ridge. Methane-oxidising (MOX) symbionts that belong to the Methylococcales were only detected at a single location and in low abundance (Semenov, sample 4047_K, 0.3 % read mapping coverage). We therefore focused exclusively on the sulphur-oxidising (SOX) symbionts for the rest of the study. Across all samples, the order Thiomicrospirales, to which *Bathymodiolus* SOX symbionts belong, only made up a small proportion of all reads (0 to 7 %). This indicates that the Thiomicrospirales are only low abundant in the high diversity microbial community living in the water column. In comparison to the high microbial diversity in the surrounding water, the diversity of taxonomic groups within mussel gill metagenomes was much lower (see Supplement "Low taxonomic diversity of mussel gill metagenomes"). As a result, the relative proportion of Thiomicrospirales in the overall taxonomic composition was at least six times higher in the gill compared to water-derived metagenomes (Supplementary Figure S 3).

Evolutionary placement of 16S rRNA sequences from water samples on a phylogeny of Thiomicrospirales

After initial screening of 16S rRNA sequence reconstructions (17 full and partial sequences), we identified their phylogenetic affiliation using an evolutionary placement analysis. Our analysis revealed a clear separation between two main clades, a symbiotic clade consisting of mostly mussel symbionts (>97 % sequence identity), and a mixed clade that included free-

living marine bacteria and clam symbionts (>95 % sequence identity, Figure 3). The placement of water sample 16S rRNA genes showed that different bacterial species from the Thiomicrospirales were present at vent sites where mussels occur. Most of the detected sequences fell on branches of free-living marine bacteria from different ocean regions or clam symbionts. In detail, 16S rRNA genes from water samples were similar to free-living bacteria sampled from vent environments (Tonga Arc, Lau Basin, and Mid-Okinawa Trough), oxygen minimum zones (Byfjorden in Norway and Saanich Inlet in Canada), and scattered locations in the Pacific and Atlantic Ocean, mostly from coastal areas. One sequence was most similar to *Ca.* Thioglobus singularis, a cultured representative of the *Thioglobaceae* family isolated from surface waters (5 m) in Puget Sound. Only few of our detected sequences were similar to symbiotic relatives. Out of the 17 16S rRNA sequences, one was similar to symbionts of *Adipicola* and tubeworms (in the symbiotic clade) and two fell into the subclade of clam symbionts (in the mixed clade).
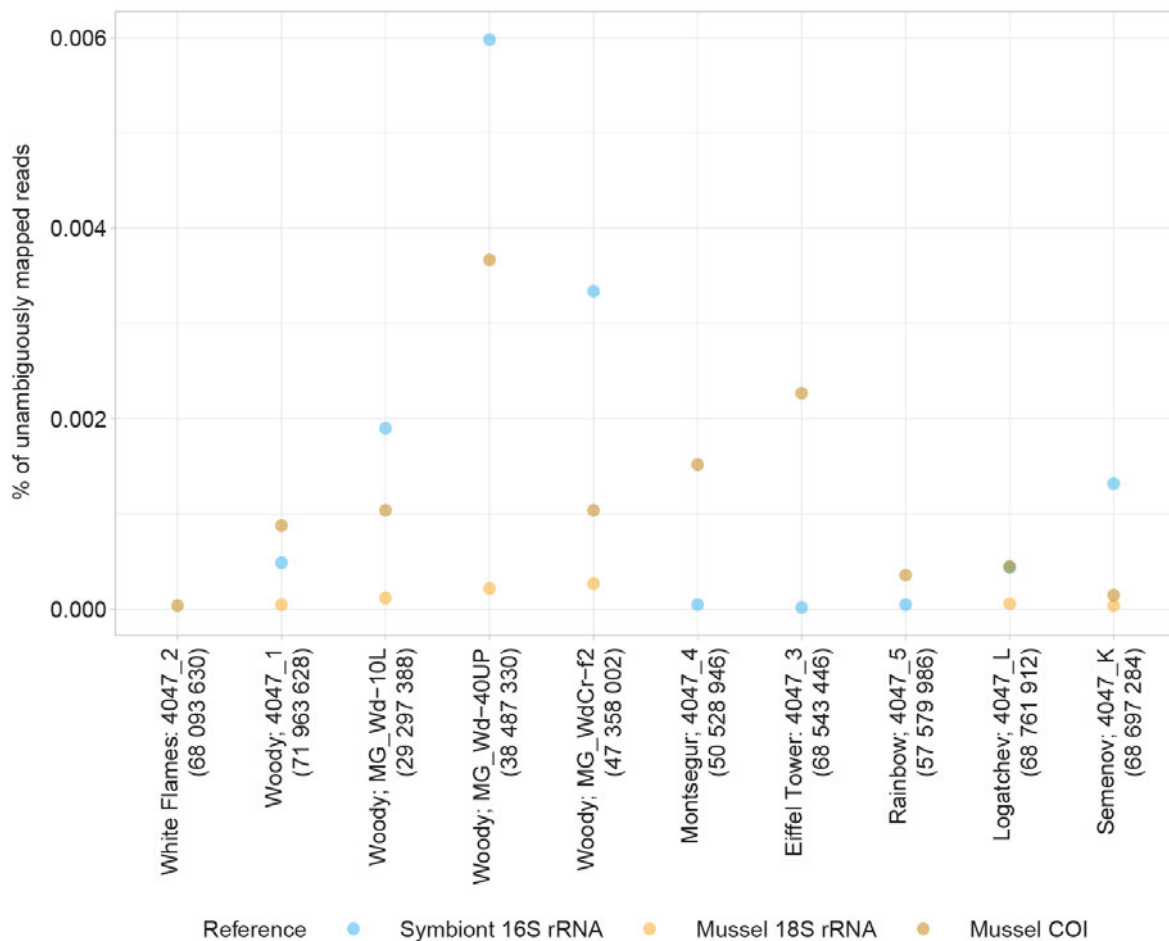
Presence of symbiont 16S rRNA and mussel 18S rRNA and COI genes in water filter metagenomes

To increase our sensitivity for detecting free-living *Bathymodiolus* symbionts in water-derived metagenomes, we performed an assembly-independent read-recruitment using 16S/18S rRNA and COI genes as references for alignment. To assess whether hits to symbionts could be attributed to free-living or host-associated stages, we also screened for co-occurring host material in the same sample. To identify such co-occurrence we mapped the reads simultaneously against 18S rRNA (for initial screening) and cytochrome c oxidase subunit 1 (COI) gene sequences of the hosts (for analysis of species affiliation).

16S rRNA sequences of *Bathymodiolus* symbionts were present (at least one read mapped) in six out of seven samples in low abundances (Figure 4). In White Flames in the Menez Gwen vent field (sample 4047_2), we did not detect any symbiont 16S rRNA sequences. In Montsegur and Eiffel Tower in the Lucky Strike vent field (4047_4 and 4047_3) and in Rainbow (4047_5), only few reads matching the symbiont 16S rRNA sequences and none matching the mussel 18S rRNA sequences were detected. We also analysed sequence reads in the metagenomes that mapped to COI references of *B. puteoserpentis* or *B. azoricus* and detected hits in all samples. The observed COI gene sequences were 99.8-100 % identical to the expected *Bathymodiolus* host species at each site.
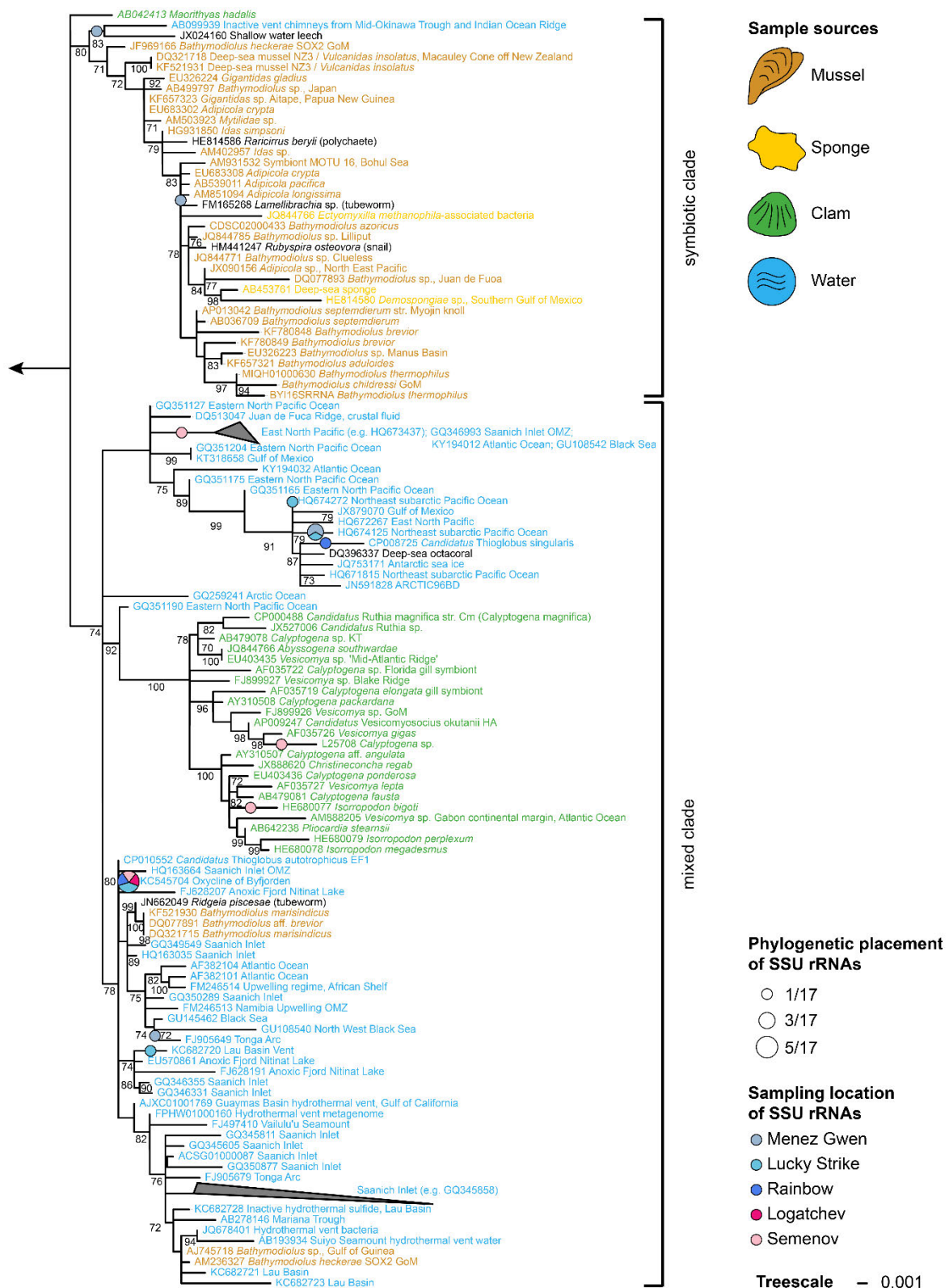
Figure 3 | Evolutionary placement of 16S rRNA sequences on a phylogenetic reconstruction of Thiomicrospirales encompassing *Bathymodiolus* symbionts, and closely related bacteria with symbiotic and free-living lifestyles. Label colours represent the sampling sources: mussel (brown), sponge (yellow), clam (green), or water (blue). The phylogeny was reconstructed with IQ-TREE (TN+F+R3 model) using sequences from the SILVA database. Circles indicate the phylogenetic placement of 16S rRNA sequences from water samples by the Evolutionary Placement Algorithm (epa-ng), their colours represent the sampling location. When several 16S rRNA sequences from different sample locations were placed on the same branch/clade, the proportions of sample locations are indicated in pie charts.

Figure 4 | Presence and abundance of 16S rRNA, 18S rRNA and COI gene sequences of symbionts and their host in water metagenomes. Blue symbols represent reads mapping to the 16S rRNA from *B. azoricus/B. puteoserpentis* symbionts (minimum identity 99 %). Yellow symbols represent reads mapping to the 18S rRNA of *B. azoricus/B. puteoserpentis*. *Bathymodiolus* mussels share large parts of their 18S rRNA gene sequences with other mussel species from the same family, which is why COI sequences were additionally analysed and plotted in brown. Reads matching *Bathymodiolus* symbiont 16S rRNA sequences were only present in samples from which reads also mapped to 18S rRNA and COI gene sequences of their host. X-axis labels indicate location and name of each sample, numbers in brackets represent the total number of reads per sample.

All metagenomes with reads matching symbiont 16S rRNA also contained reads mapping to host 18S rRNA, COI sequences, or both. This co-occurrence indicates that symbiont 16S rRNA gene sequences on the filters were associated with host-tissue, e.g. mussel tissue that is floating around the mussel bed. Thus, we could not distinguish whether symbiont sequences originated from the host-associated or the free-living stage.

Diagnostic markers in mussel and water metagenomes

It has been hypothesised that the free-living population of *Bathymodiolus* symbionts is geographically structured [18]. To assess whether the symbiont subspecies follow the same geographical distribution as their host species (*B. azoricus* occurs at 38°N to 36°N, *B. puteoserpentis* at 23°N to 13°N), we established a workflow to identify genetic markers able to distinguish different subspecies within one symbiont species. Out of 171 gammaproteobacterial marker genes tested, three showed clear differences between the symbionts of *B. azoricus* and *B. puteoserpentis* (see Supplement "Three diagnostic markers can discriminate *B. azoricus* and *B. puteoserpentis* symbiont subspecies", and Supplementary Figure S 4 and Supplementary Figure S 5). Gene 23 was annotated as a 3-dehydroquinate synthase, gene 39 as a hypothetical protein with unknown function, and gene 117 as a 30S ribosomal protein S16.

We used stringent read mapping (minimum identity of 99 %) to study the presence of marker gene sequences associated with either *B. azoricus* or *B. puteoserpentis* symbionts (gene sequences will be referred to as Bazo and Bput, respectively). We analysed metagenomes from the filtered water samples and compared them to metagenomes from mussel hosts collected at the same vent sites. In the mussels, the marker gene reads reached proportions of up to 0.08 % of all metagenomic reads. They clearly reflected the expected marker distributions, i.e. metagenomes recovered from Menez Gwen, Lucky Strike, and Rainbow were dominated by marker sequences unambiguously mapping to the Bazo type, and mussel metagenomes recovered from Logatchev and Semenov where dominated by the Bput type (Figure 5). However, in both host species, additional small proportions of symbiotic marker gene reads mapped to the unexpected type, i.e. to Bput in metagenomes of *B. azoricus* and to Bazo in metagenomes of *B. puteoserpentis*. In the water metagenomes, reads mapping to the marker genes occurred only in some of our samples, and if so, their relative abundance was 1000-fold less than in the mussel metagenomes. However, the distribution of the marker types Bazo and Bput at the various locations corresponded to the expectations based on the local distributions of host species.
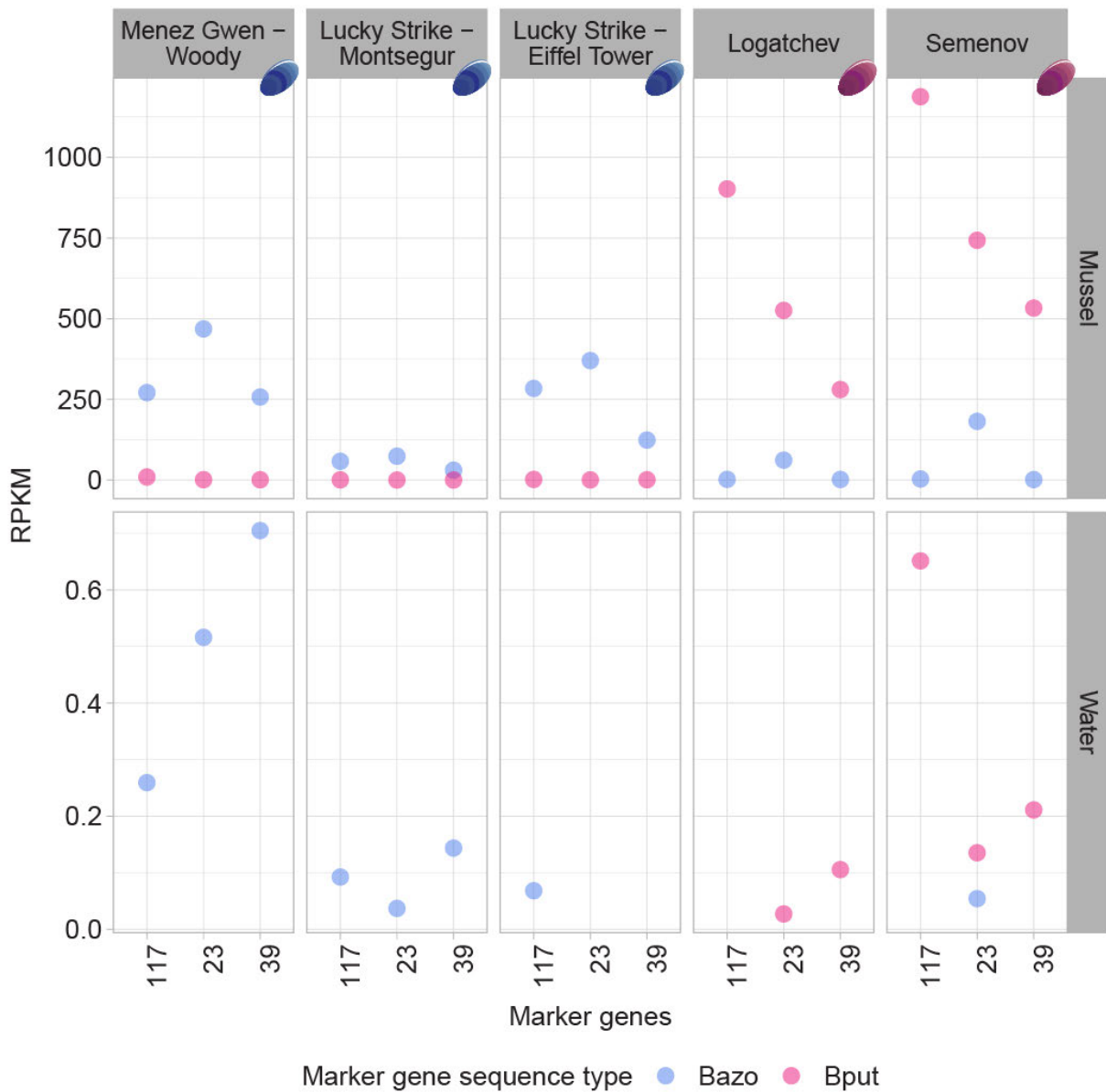
Figure 5 | Presence and abundance of three diagnostic markers for *Bathymodiolus* symbionts in mussel and water metagenomes from different vent sites in RPKM. PRKM normalises for sequencing depth and sequence length. Reads from mussel metagenomes used in the analysis are listed in Supplementary Table S 3. For assessment of water metagenomes, only reads sequenced in this study were used. Mapped reads were obtained with BBSplit and a minimum identity of 99 %. Colours represent whether the gene sequence was derived from *B. azoricus* (Bazo, blue) or *B. puteoserpentis* (Bput, pink) symbiont MAGs. Mussel icons represent the host species at each site: *B. azoricus* (blue) or *B. puteoserpentis* (pink). Dots are absent for a sample if no reads were detected. Annotations of marker genes are: Gene 117 = 30S ribosomal protein 16S, gene 23 = 3-dehydroquinate synthase, gene 39 = hypothetical protein.

Discussion

Low abundance of *Bathymodiolus* symbionts in the water column

We investigated the taxonomic composition of seven water metagenomes from vent sites at the Mid-Atlantic Ridge and found that the bacterial communities comprised mostly Proteobacteria. This is in line with previous findings that, despite the high diversity of rare taxa in deep-sea and diffuse fluid environments, only few very abundant taxa from the Proteobacteria dominate the bacterial community in these environments [30,59,60]. The core community of low-temperature diffuse fluids, around which mussel beds occur, has been described to consist of mostly Epsilon- (mainly *Caminibacter*, *Sulfurimonas*, *Sulfurovum* and Campylobacterales) and Gammaproteobacteria (mainly *Marinobacter*, *Alcanivorax* and *Thiomicrospira* [59]) which is comparable to the community compositions in our study.

Recently, Meier and colleagues showed that metagenomes from different sampling spots could be clustered based on the similarity of their microbial communities [30]. Epsilonproteobacteria dominated samples that were collected directly from diffuse effluent in seafloor fissures, while Gamma- and Alphaproteobacteria dominated the bacterial community of samples obtained in the vicinity of the diffuse venting orifices where mixing with seawater had already occurred (horizontal distance up to 1 m or vertical distance around 40 cm). The composition of samples derived from rising plumes and the deep water column differed the most, as these were dominated by Alpha- and Deltaproteobacteria. The categorisation of bacterial communities based on sampling spots as established by Meier and his colleagues also applies to the data shown here. For example, 4047_2 from the vent site White Flames was sampled in the water column around 30 m above the seafloor. Alpha- and Deltaproteobacteria dominated its taxonomic composition, as would be expected for a sample from a rising plume or the water column. Another example is 4047_1 from the vent site Woody, which was sampled by in-situ filtration on the ground and is comprised of mostly Alphaproteobacteria. The taxonomic composition corresponds best to a sample from the vicinity of diffuse venting, which is in line with the sampling device being deployed max. 2 m away from diffuse venting (Supplementary Figure S 1 A). Altogether, the taxonomic composition of metagenomes presented in this study is in agreement with the range of taxa reported for hydrothermal vent environments in the literature.

16S rRNA sequences belonging to Thiomicrospirales, the order that includes the *Bathymodiolus* SOX symbionts, covered 7 % or less of the taxa present in the analysed water metagenomes indicating that Thiomicrospirales represent only a small proportion in the

whole community. Our evolutionary placement analysis showed that the 16S rRNA sequences of Thiomicrospirales at hydrothermal vents were most similar to sequences of various free-living marine bacteria and chemosynthetic symbionts from vent and seep clams and mussels. This indicates that Thiomicrospirales closely related to *Bathymodiolus* SOX symbionts are indeed present around mussel beds.

Mussels constantly filter water through their gills, and with it, the microbial community that is present in the water column [61]. The water flow and the mussels' filtration activity represent the source for symbionts colonising the gills. However, Thiomicrospirales bacteria other than *Bathymodiolus* symbiont species have not been found to occur within the tissue and tissue surfaces of mussel hosts suggesting that the mussels very specifically acquire one symbiont phylotype. The specificity of symbiont selection is remarkable - even more in the light of the very low abundance of highly diverse Thiomicrospirales in the water column described here. This means that the host can overcome two challenges: Detecting the symbiont although it is low abundant and only acquiring the suited symbiont from a great mixture of very similar bacteria. The mechanisms underlying this specificity are to date unclear. One strategy followed by the host could involve selection of specific bacteria, e.g. based on intracellular recognition receptors such as the receptor BpLRR-1 that was suggested to play a role in the recognition of methane-oxidising symbionts in *B. platifrons* [62], via cell surface components of *Bathymodiolus* symbionts [63] or via yet unknown mechanisms. The symbionts may in turn have evolved strategies for successful host colonisation, e.g. by escaping host immune defence [64,65]. To fully understand the mechanisms leading to such a high host-symbiont specificity, further studies on cellular processes within host and symbionts are needed.

Do *Bathymodiolus* symbionts live as a free-living population in the water column?
Based on metagenomic evidence, it is likely that *Bathymodiolus* symbionts from gill tissue have the genetic potential to survive outside their hosts. In a recent comparative genomics study, Ansorge and her colleagues compared bacterial MAGs and genomes from the *Thioglobaceae* family, including many *Bathymodiolus* symbionts and free-living relatives. They did not detect arrays of genes that are exclusively specific to one lifestyle, i.e. symbiotic or free-living, indicating that *Bathymodiolus* SOX symbionts are probably able to survive in a free-living stage [16]. To what extent the free-living symbionts are able to thrive in the environment, i.e. whether they are proliferating or in a dormant stage, is unclear.

We used symbiont 16S rRNA sequences in water filter metagenomes as a proxy to study the presence of free-living *Bathymodiolus* SOX symbionts. We detected reads mapping to the symbiont 16S rRNA gene, however, their abundance was only in the range of 10^{-5} % of all water metagenome reads. To clarify whether detected symbiont reads came from a free-living or host-associated stage, we screened for mussel DNA and found it to be present in most of the samples. Two possible reasons to detect the host DNA on the water filters are the presence of host tissue, or at least fragments of it, in the water column or the occurrence of environmental DNA (eDNA). While the presence of host-derived eDNA might not necessarily influence the SOX symbionts, a potential presence of host material in the water column could indicate that symbionts are still associated to host tissue and potentially also transmitted via host cells floating in the water column. Based on our finding that symbiont reads always co-occurred with mussel reads, it is inconclusive whether *Bathymodiolus* symbionts live as a free-living population in the water column or whether the symbionts are only transferred and taken up through host particles.

Previous studies have reported on *Bathymodiolus* symbiont genes in the water column and microbial mats, however, some of them provide contradictory findings as they detected functional genes or symbiont transcripts but no corresponding 16S rDNA or rRNA, or vice versa (Table 1). A study on bacterial seawater communities next to diffuse outflow in the Lilliput vent field on the southern MAR targeted specific bacterial genes using PCR (Perner et al., 2007). They detected *cbbL* gene sequences 95.6 % similar to those of northern MAR *Bathymodiolus* symbionts, while they did not find 16S rDNA sequences resembling mussel symbiont phylotypes. Another study investigated symbiont ribosomal and other marker genes in DNA and RNA samples from microbial mats at the northern MAR using PCR [27]. They detected ribosomal sequences that were 93.7 % and 100 % similar to published *B. azoricus* symbionts from the Lucky Strike vent field but did not find transcripts of SOX symbiont-specific *soxB* and *aprA* genes. A third study focused on water samples from seamounts around Japan, investigating PCR-amplified symbiont marker genes. Variation in DNA fragments of a hydrogenase and a nitrate reductase in water samples resembled the variation detected in the mussel symbionts, and thus, could be an indication for the presence of these symbionts in the water [28]. While none of these studies screened for host DNA, this was done only in one study where biofilms and water samples from the Lau Basin were investigated with PCR, qPCR and southern blots [22]. In biofilms, up to 0.25 % of total bacterial ribosomal copy numbers represented the symbionts whereas host DNA was absent.

Water samples, that were filtered through 1 μm and subsequent 0.2 μm filters, also contained symbiont-related genes but they were not screened for host DNA.

Table 1 | Summary of studies targeting the free-living stage of *Bathymodiolus* symbionts. Pub: Publication.

Material	Location	Method	Genes	Result	Host-check	Pub
Seawater filters	Lilliput (SMAR)	PCR	16S rRNA, *cbbL*, *cbbM*, *aclB*	*cbbL* 95.6 % similar to NMAR *Bathymodiolus* symbionts, no 16S rDNA detected	no	[29]
Microbial mats	Lucky Strike (NMAR)	PCR	16S rRNA, *aprA*, *soxB*, *pmoA*, *cbbM*, *cbbL*, *aclB*	ribosomal genes 93.7 and 100 % similar to *B. azoricus*, no symbiont specific *soxB* and *aprA* genes	no	[27]
Seawater filters	Myojin Knoll & Suiyo Seamount (Japan)	PCR	*hup*, *nar*	variation of DNA fragments in water similar to variation in mussels	no	[28]
Seawater filters	Lau Basin	qPCR	16S rRNA	symbiont-related genes detected	no, but filtered through 1 μm and 0.2 μm filters	[22]
Biofilm	Lau Basin	qPCR	16S rRNA	0.25 % of bacterial ribosomal reads belonged to symbionts	yes	[22]

We found that symbiont and mussel DNA was always present together in the analysed metagenomes. At this point, it remains unclear whether a truly free-living *Bathymodiolus* SOX symbiont population exists. To resolve this question in future studies, our results highlight the need of screening for host DNA in water samples when detecting symbiont reads. Symbiont-related genes in our study were low abundant and represented only a small proportion of the overall bacterial community at hydrothermal vents. If the free-living pool of *Bathymodiolus* SOX symbionts was highly abundant, we assume that it would have been detected more abundant in our and previous marine metagenomic studies. However, we acknowledge that the detection of low abundant organisms is limited in metagenome-based studies, as it requires deep sequencing unlike other amplification-based methods. Thus, our findings point towards the existence of a small seed population of *Bathymodiolus* SOX

symbionts in the water column surrounding mussel beds that might still be associated with host tissue.

Are all symbionts everywhere?

The question of what governs the distribution of microorganisms has haunted microbial ecologist for decades. With recent improvements in sequencing technologies, an increasing number of studies confirms that the concept of Baas Becking and Beijerinck that *everything is everywhere, but, the environment selects* seems to be true in the marine environment [66,67]. The concept implies that dispersal limitation does not seem to be the main driver for shaping microbial communities but that environmental selection has a great influence on which taxa become abundant [68,69]. We examined if this is also the case for the free-living stage of *Bathymodiolus* symbionts and used three symbiont subspecies-diagnostic markers to compare their distribution in the water column to that within mussel gills.

We found that marker gene type sequences expected for the local mussel species were indeed dominant in the gill and also water metagenomes of the corresponding sites. However, sequences of the unexpected subspecies were also detectable in low abundance. The results of our marker analysis point towards a scenario in which one symbiont subspecies is dominant per mussel species and site while other closely-related symbiont subspecies might be present in low abundance. The similar distributions of marker gene types in mussel and water metagenomes suggest that mussels pick up the symbionts available in the environment without strict selection between highly similar but yet distinct SOX symbiont subspecies. However, the predictive function of our analysis is limited by the overall low abundance of symbiont-derived genetic material in the water samples, and should be repeated with more samples and deeper sequencing.

How *Bathymodiolus* mussels might acquire their symbionts

In horizontally transmitted symbioses, the symbionts must be acquired from the environment or neighbouring hosts with each host generation. There are different potential ways of symbiont acquisition (Figure 6). Besides the uptake of symbionts from a free-living stage in the water column, symbionts could be acquired through biofilms surrounding the mussel bed [22,27]. Biofilms on basaltic rock, hydrothermal deposits, and mussel assemblages have been previously shown to contain *Bathymodiolus* symbionts, and were suggested to enable the maintenance of large mussel assemblages and their symbiosis by providing a favourable environment for both mussel larvae and potential symbionts [22,27]. Another possibility would be the acquisition of symbionts that are released from their dying host [70,71]. For

hydrothermal vent species other than *Bathymodiolus*, it has been proposed that symbionts leave their dying hosts, a phenomenon that might enhance transmission in the proximity of animals. Experiments with the tubeworm *Riftia* demonstrated that up to 7×10^5 symbionts escape from a dead tubeworm within half a day under cold deep-sea conditions [70]. Such dispersal events have not been studied for the *Bathymodiolus* mussels but it cannot be excluded that symbionts are released to the environment after death of their mussel host. Within the gills, newly formed cells are colonised only after formation and it has been proposed that this colonisation is mainly driven by self-infection with symbionts released from neighbouring filaments [72]. Another alternative for symbiont acquisition might be through symbionts associated with host tissue.

In our study, symbiont DNA co-occurred with host DNA in all analysed samples. Therefore, we propose that a transmission of *Bathymodiolus* symbionts via host material between neighbouring hosts could play an important role in established mussel populations. Free-floating host tissue pieces could be released to the water column either through water currents, predation of the mussels by other vent fauna or during the process of "adaptive feeding". Geier et al. recently proposed the concept of "adaptive feeding", which describes an additional mode of symbiont release by the mussel host. They show that the mussels shed symbiont-containing bacteriocytes as part of their epithelial turnover, followed by an incomplete digestion and potential release of living symbionts back to the environment [73]. Unlike the occasional death of mussels, mussel faeces filled with symbionts could be a constant source of co-occurring host and symbiont DNA as detected across samples in our study.

A transmission via host-derived particles would indicate that the abundance of symbionts available for transmission at a vent site correlates with the distance to the mussel bed and that symbiont densities would be highest close to the mussels. We therefore hypothesise that proximity of the mussel host is key for an effective symbiont transmission. Results from our analyses of symbiont 16S rRNA gene sequences support the requirement of mussel proximity for symbiont transmission. Read numbers mapping to *Bathymodiolus* symbionts were highest in samples from [30], where water was sampled close to diffuse venting and the mussels surrounding the venting. In samples from our study, we detected most symbiont-related genes in samples that were obtained with in-situ pumps deployed on the sea floor at Woody and Montsegur or with ISP Mooring 6 m above the seafloor at Semenov and Logatchev. In contrast, fewer reads for symbiont-related genes were observed in samples obtained with an

in-situ pump hanging on a wire ~30 m above the seafloor at Rainbow, Eiffel Tower, and White Flames. This corroborates with our hypothesis that the proportion of symbionts available for transmission might decrease drastically with increasing distance from active venting and the mussel bed. Evidence of within-mussel symbiont differentiation further supports our hypothesis. Recent studies showed that there is considerable diversity of strains within the same symbiotic phylotype in a host population and even within single host individuals [63,74]. Intriguingly, the genetic diversity between symbiont populations of single host individuals correlated with spatial distance, i.e. mussels sitting in the same mussel patch revealed more similar strain communities based on the population differentiation index F_{ST} than mussels collected from separate and more distant patches [63].

Bathymodiolus mussels take up their symbionts after larvae settlement [75] and therefore, a transmission exclusively via host associated symbionts is unlikely. Mussel larvae that are colonising a new vent site without a local mussel population must be able to acquire symbionts and form a new population at this site. Our finding that different symbiont subspecies are present but only one becomes more abundant, e.g. through selection by the environment or the mussel host, is a further indication that transmission might happen through a mix of mechanisms. Even if one dominant symbiont subspecies was transmitted from one host to another via host material, the non-dominant subspecies might still be taken up directly from the environment. We presented evidence that symbionts made up only a small proportion of the bacterial community in the water column around hydrothermal vents. However, in a symbiotic association with high host-symbiont specificity, i.e. only suited symbionts are recognised and taken up, a few bacterial cells can be enough to inoculate the host. Mussels constantly filter seawater through their gills, which are also their symbiotic organ so that water-borne symbionts could accumulate if there was a specific selection mechanism. In other symbiotic systems such as the squid–*vibro* symbiosis, the inoculation with only a single symbiotic cell has been demonstrated [76]. To investigate the uptake of free-living bacteria in *Bathymodiolus*, further studies are needed. These could include exposing aquaria mussels to seawater with different bacterial communities, e.g. seawater from their sampling site and a controlled microbial community in sterile seawater.
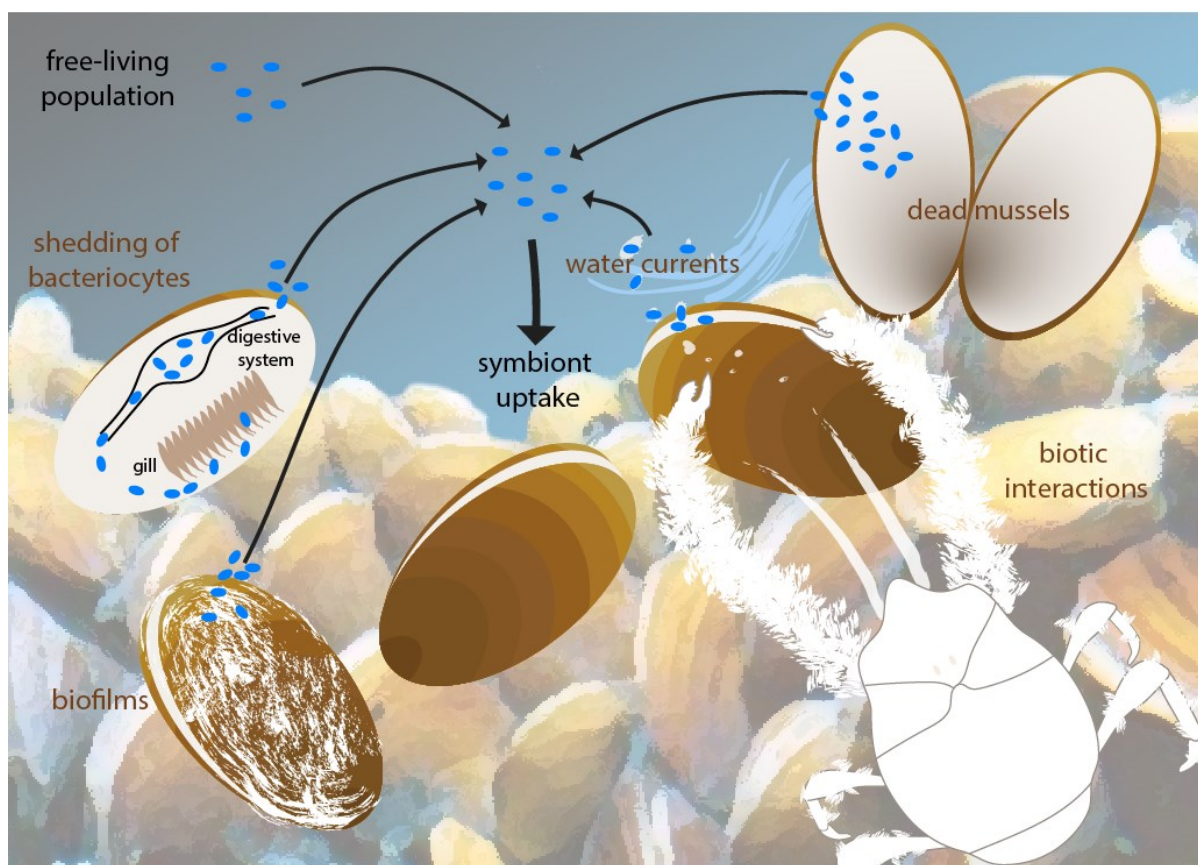
Figure 6 | Potential ways of symbiont transmission in *Bathymodiolus* mussels. For symbiont sources written in brown, symbionts might be associated with host tissue. Host tissue particles may play a major role for symbiont transmission. Host cells can enter the water column after shedding of bacteriocytes ("adaptive feeding" as described by [73]), due to disturbances by water currents or biotic interactions, e.g. predation, with other fauna. Symbionts might be released from dead mussels or biofilms grown on the mussel shells and enter the environmentally available pool of symbiont cells for transmission. For most depicted ways of symbiont transmission, the symbiont density is likely highest close to the mussel bed, suggesting that the proximity between mussels is a key factor for an effective transmission. Symbionts are first acquired after settlement of the mussel larvae but also taken up throughout the mussel's lifetime.

Directions for future studies of the free-living symbiont population

Studying environmental populations of bacteria in the proximity of symbiotic animals can provide important insights into the functioning of horizontally-transmitted symbioses [22,77]. However, sampling can come with great challenges depending on the sampled environment. When conducting research in the deep sea, each sample is tied to high costs, extensive planning, great technological requirements and long expeditions. Therefore, we condensed experiences gained from our and similar studies into a sampling concept for future research. In the light of our results and different transmission scenarios presented above, an ideal sampling approach would be holistic and should cover mussels, seawater, biofilms on shells and rocks, and sediments from the same site. Accompanying metadata – such as coordinates,

distances to mussel beds, rocks and other remarkable features, in-situ temperature and pH, chemical composition of hosting rocks and fluids, images of sampling devices, sampling spots, the samples themselves and surroundings, or any other data that can be collected at the sampling site – is essential for the interpretation of the sequence-based data and should be collected as thoroughly as possible.

To examine whether symbiont density in the water column decreases with increasing distance to the mussel bed, water samples should cover a range of distances, e.g. just above the mussel bed, close to the mussel bed, and a few meters away. In addition, seawater from the plume might give helpful insights into whether the upward movement of the plume transports symbionts further in the water column. Our analysis revealed that the sample MG-Wd40UP, which was obtained 40 cm above the orifice, had higher numbers of reads mapped to symbiont genes than MG-Wd10L, which was sampled 10 cm left of the orifice, which is why we think that the upward movement of the plume might play a role in the transport of symbionts.

In our study, we did not observe any evidence that in-situ filtration of large volumes of water increases the chances to detect symbiont-derived genetic material. In contrast, highest numbers of reads mapping to symbiont genes were detected in the metagenomes from [30] where smaller volumes of seawater were filtered. Thus, we suggest filtering 20 L of seawater with the Kiel In Situ Pumping System. To improve future sampling campaigns of free-living symbiont stages, we suggest sampling seawater using sequential filters with decreasing pore size. This could help to pinpoint whether symbiont and host DNA always occur together or whether a symbiont fraction can be detected after larger particles, i.e. host cells, were filtered out. Previous experiments in the laboratory showed that some SOX symbiont cells might pass through filters with 0.2 µm pore size depending on their orientation. Therefore, we suggest analysing the flow through or to add an additional filtration size of 0.01 µm that is usually used for viruses. Laboratory experiments with seawater containing homogenised mussel tissue are needed to determine the optimal filter sizes, e.g. 3 µm, 1 µm and 0.2 µm, prior to an expedition. Ideally, filters should be divided and prepared for different analyses such as metagenomics, proteomics, and imaging. When targeting the symbiont fraction on the filters, screening for host DNA should always be included. To get a better resolution of rare bacteria in the metagenomic analyses that might represent the seed population of symbionts, we suggest sequencing at least a few samples at high depth.

Conclusion

Research of horizontally transmitted symbioses should not stop at the inside of a symbiotic animal but include studying the external environment to investigate the complexity of these associations further. Using a metagenomics approach analysing in-situ filtered seawater samples from the deep sea, we showed that *Bathymodiolus* mussels are exposed to a diverse range of bacteria from the Thiomicrospirales that are closely related to their symbionts. However, these bacteria are not taken up into the host tissue, stressing the high specificity of the *Bathymodiolus* symbiosis. We showed the presence of low abundant symbiont-related genes in the water column, as others have done before, but our analysis of host-derived reads revealed a co-occurrence of host and symbiont DNA. This indicates that a free-living symbiont community might only play a minor role as a pool for symbiont acquisition in established mussel populations. Instead, we propose a symbiont transmission that is facilitated by free-floating host cells and biofilms in which close proximity of the hosts is key for an effective transmission. With our marker gene analysis, we present first evidence that different symbiont subspecies might be present at a given vent site but only the suited candidate grows to larger abundances, possibly through selection by the environment and the mussel host occurrence. These concepts enhance our understanding of the *Bathymodiolus* symbiosis and should inspire further research and conceptual discussions about every aspect of horizontally transmitted animal–microbe associations.

Author Contributions

MÜ, RA, YS, CB & ND conceived the study. RA, CB & ND conducted the sampling. MÜ performed laboratory work, analysed the data, prepared figures & tables, and wrote the manuscript. RA & YS contributed to data analyses, figure design, conceptual ideas, and data interpretation of the study. GYGH developed the workflow to identify symbiont markers, published the code, and prepared Figure 1 B, C & D. RA, YS, GYGH & CB revised the manuscript.

Acknowledgements

References

1. Archibald JM. Endosymbiosis and eukaryotic cell evolution. Curr Biol. 2015; 25: R911–21.
2. Brucker RM, Bordenstein SR. The hologenomic basis of speciation: Gut bacteria cause hybrid lethality in the genus *Nasonia*. Science. 2013; 341: 667–9.
3. Fisher RM, Henry LM, Cornwallis CK, Kiers ET, West SA. The evolution of host-symbiont dependence. Nat Commun. 2017; 8: 15973.
4. McFall-Ngai MJ. Unseen forces: the influence of bacteria on animal development. Dev Biol. 2002; 242: 1–14.
5. Bright M, Bulgheresi S. A complex journey: Transmission of microbial symbionts. Nat Rev Microbiol. 2010; 8: 218–30.
6. Vrijenhoek RC. Genetics and evolution of deep-sea chemosynthetic bacteria and their invertebrate hosts. In: Kiel S, editor. The Vent and Seep Biota, Springer Netherlands; 2010, p. 15–49.
7. Russell SL, Chappell L, Sullivan W. Chapter Nine - A symbiont's guide to the germline. In: Lehmann R, editor. Current Topics in Developmental Biology, vol. 135, Academic Press; 2019, p. 315–51.
8. Hosokawa T, Kikuchi Y, Nikoh N, Shimada M, Fukatsu T. Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. PLoS Biol. 2006; 4: e337.
9. Russell SL. Transmission mode is associated with environment type and taxa across bacteria-eukaryote symbioses: a systematic review and meta-analysis. FEMS Microbiol Lett. 2019; 366: fnz013.
10. Jones WJ, Won Y-J, Maas PAY, Smith PJ, Lutz RA, Vrijenhoek RC. Evolution of habitat use by deep-sea mussels. Mar Biol. 2006; 148: 841–51.
11. Van Dover CL. The Ecology of Deep-sea Hydrothermal Vents. Princeton University Press; 2000.
12. Dubilier N, Bergin C, Lott C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. Nat Rev Microbiol. 2008; 6: 725–40.
13. Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R, et al. Hydrogen is an energy source for hydrothermal vent symbioses. Nature. 2011; 476: 176–80.
14. Duperron S, Bergin C, Zielinski F, Blazejak A, Pernthaler A, McKiness ZP, 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. 2006; 8: 1441–7.
15. Marshall Lalish K. Elucidating the metabolic activities of SUP05, an abundant group of marine sulfur oxidizing gamma-proteobacteria. PhD Thesis. 2015.
16. Ansorge R, Romano S, Sayavedra L, Rubin-Blum M, Gruber-Vodicka HR, Scilipoti S, et al. The hidden pangenome: comparative genomics reveals pervasive diversity in symbiotic and free-living sulfur-oxidizing bacteria. BioRxiv. 2020: 2020.12.11.421487.
17. Van Rossum T, Ferretti P, Maistrenko OM, Bork P. Diversity within species: interpreting strains in microbiomes. Nat Rev Microbiol. 2020; 18: 491–506.
18. Ücker M, Ansorge R, Sato Y, Sayavedra L, Breusing C, Dubilier N. Deep-sea mussels from a hybrid zone on the Mid-Atlantic Ridge host genetically indistinguishable symbionts. ISME J. 2021; 1–8.
19. Won Y-J, Hallam SJ, O'Mullan GD, Pan IL, Buck KR, Vrijenhoek RC. Environmental acquisition of thiotrophic endosymbionts by deep-sea mussels of the genus *Bathymodiolus*. Appl Environ Microbiol. 2003; 69: 6785–92.

20. Le Pennec M, Diouris M, Herry A. Endocytosis and lysis of bacteria in gill epithelium of *Bathymodiolus thermophilus*, *Thyasira flexuosa* and *Lucinella divaricata* (bivalve, molluscs). J Shellfish Res. 1988; 7: 483–9.

21. DeChaine EG, Bates AE, Shank TM, Cavanaugh CM. Off-axis symbiosis found: Characterization and biogeography of bacterial symbionts of *Bathymodiolus* mussels from Lost City hydrothermal vents. Environ Microbiol. 2006; 8: 1902–12.

22. Fontanez KM, Cavanaugh CM. Evidence for horizontal transmission from multilocus phylogeny of deep-sea mussel (Mytilidae) symbionts. Environ Microbiol. 2014; 16: 3608–21.

23. Won Y-J, Jones WJ, Vrijenhoek RC. Absence of cospeciation between deep-sea mytilids and their thiotrophic endosymbionts. J Shellfish Res. 2008; 27: 129–38.

24. Anantharaman K, Breier JA, Dick GJ. Metagenomic resolution of microbial functions in deep-sea hydrothermal plumes across the Eastern Lau Spreading Center. ISME J. 2016; 10: 225–39.

25. Fortunato CS, Larson B, Butterfield DA, Huber JA. Spatially distinct, temporally stable microbial populations mediate biogeochemical cycling at and below the seafloor in hydrothermal vent fluids. Environ Microbiol. 2018; 20: 769–84.

26. Ho P-T, Park E, Hong SG, Kim E-H, Kim K, Jang S-J, et al. Geographical structure of endosymbiotic bacteria hosted by *Bathymodiolus* mussels at eastern Pacific hydrothermal vents. BMC Evol Biol. 2017; 17: 121.

27. Crépeau V, Cambon Bonavita M-A, Lesongeur F, Randrianalivelo H, Sarradin P-M, Sarrazin J, et al. Diversity and function in microbial mats from the Lucky Strike hydrothermal vent field. FEMS Microbiol Ecol. 2011; 76: 524–40.

28. Ikuta T, Takaki Y, Nagai Y, Shimamura S, Tsuda M, Kawagucci S, et al. Heterogeneous composition of key metabolic gene clusters in a vent mussel symbiont population. ISME J. 2016; 10: 990–1001.

29. Perner M, Seifert R, Weber S, Koschinsky A, Schmidt K, Strauss H, et al. Microbial CO_2 fixation and sulfur cycling associated with low-temperature emissions at the Lilliput hydrothermal field, southern Mid-Atlantic Ridge (9°S). Environ Microbiol. 2007; 9: 1186–201.

30. Meier DV, Bach W, Girguis PR, Gruber-Vodicka HR, Reeves EP, Richter M, et al. Heterotrophic Proteobacteria in the vicinity of diffuse hydrothermal venting. Environ Microbiol. 2016; 18: 4348–68.

31. Schmidt K, Koschinsky A, Garbe-Schönberg D, Carvalho L, Seifert R. Geochemistry of hydrothermal fluids from the ultramafic-hosted Logatchev field, 15°N MAR. Chem Geol. 2007; 242: 1–21.

32. Gruber-Vodicka HR, Seah BKB, Pruesse E. phyloFlash: rapid small-subunit rRNA profiling and targeted assembly from metagenomes. MSystems. 2020; 5.

33. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013; 41: D590–6.

34. Sayavedra L. Host-symbiont interactions and metabolism of chemosynthetic symbiosis in deep-sea *Bathymodiolus* mussels. PhD Thesis. University of Bremen, 2016.

35. Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012; 28: 1823–9.

36. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30: 1312–3.

37. Edgar RC. Usearch. 2010. http://www.drive5.com/usearch/ (accessed October 21, 2016).

38. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002; 30: 3059–66.

39. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013; 30: 772–80.

40. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017; 14: 587–9.

41. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015; 32: 268–74.

42. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993; 10: 512–26.

43. Berger SA, Krompass D, Stamatakis A. Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. Syst Biol. 2011; 60: 291–302.

44. Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. Lectures on Mathematics in the Life Sciences. 1986; 17: 57–86.

45. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019; 47: W256–9.

46. Adobe. Adobe Illustrator. 2020. https://www.adobe.com/de/products/illustrator.html (accessed February 3, 2020).

47. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2016; 44: D7–19.

48. Bushnell B. BBMap. 2014. http://sourceforge.net/projects/bbmap/ (accessed October 21, 2016).

49. RStudio Team. RStudio: integrated development for R. Boston, MA: RStudio, Inc.; 2015. http://www.rstudio.com/ (accessed December 19, 2016).

50. Wickham H. ggplot2: elegant graphics for data analysis. 1st ed. 2009. Corr. 3rd printing 2010 edition. Springer; 2016. https://ggplot2.tidyverse.org/.

51. Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, et al. ggplot2: create elegant data visualisations using the grammar of graphics. 2019. https://CRAN.R-project.org/package=ggplot2 (accessed November 25, 2019).

52. Wickham H, RStudio. tidyverse: easily install and load the "Tidyverse." 2019. https://CRAN.R-project.org/package=tidyverse (accessed July 27, 2020).

53. Lee MD. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics. 2019; 35: 4162–4.

54. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215: 403–10.

55. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25: 3389–402.

56. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009; 10: 421.

57. Kolde R. pheatmap: pretty heatmaps. 2019. https://CRAN.R-project.org/package=pheatmap (accessed July 27, 2020).

58. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, et al. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat Biotech. 2011; 29: 415–20.

59. Perner M, Gonnella G, Hourdez S, Böhnke S, Kurtz S, Girguis P. *In situ* chemistry and microbial community compositions in five deep-sea hydrothermal fluid samples from Irina II in the Logatchev field. Environ Microbiol. 2013; 15: 1551–60.

60. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored "rare biosphere." PNAS. 2006; 103: 12115–20.

61. Page HM, Fisher CR, Childress JJ. Role of filter-feeding in the nutritional biology of a deep-sea mussel with methanotrophic symbionts. Mar Biol. 1990; 104: 251–7.

62. Chen H, Wang M, Zhang H, Wang H, Lv Z, Zhou L, et al. An LRR-domain containing protein identified in *Bathymodiolus platifrons* serves as intracellular recognition receptor for the endosymbiotic methane-oxidation bacteria. Fish Shellfish Immunol. 2019; 93: 354–60.

63. Ansorge R, Romano S, Sayavedra L, González Porras MÁ, Kupczok A, Tegetmeyer HE, et al. Functional diversity enables multiple symbiont strains to coexist in deep-sea mussels. Nat Microbiol. 2019; 4: 2487–97.

64. Bettencourt R, Roch P, Stefanni S, Rosa D, Colaço A, Serrão Santos R. Deep sea immunity: unveiling immune constituents from the hydrothermal vent mussel *Bathymodiolus azoricus*. Mar Environ Res. 2007; 64: 108–27.

65. Bettencourt R, Dando P, Collins P, Costa V, Allam B, Serrão Santos R. Innate immunity in the deep sea hydrothermal vent mussel *Bathymodiolus azoricus*. Comp Biochem Physiol Part A Mol Integr Physiol. 2009; 152: 278–89.

66. Baas Becking LGM. Geobiologie of inleiding tot de milieukunde. Den Haag: W.P. Van Stockum & Zoon; 1934.

67. Wit RD, Bouvier T. 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? Environ Microbiol. 2006; 8: 755–8.

68. Dick GJ. The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. Nat Rev Microbiol. 2019; 17: 271–83.

69. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Microbiol. 2012; 10: 497–506.

70. Klose J, Polz MF, Wagner M, Schimak MP, Gollner S, Bright M. Endosymbionts escape dead hydrothermal vent tubeworms to enrich the free-living population. PNAS. 2015; 112: 11300–5.

71. Polz MF, Cavanaugh CM. Dominance of one bacterial phylotype at a Mid-Atlantic Ridge hydrothermal vent site. PNAS. 1995; 92: 7232–6.

72. Wentrup C, Wendeberg A, Schimak M, Borowski C, Dubilier N. Forever competent: deep-sea bivalves are colonized by their chemosynthetic symbionts throughout their lifetime. Environ Microbiol. 2014; 16: 3699–713.

73. Geier B. Correlative mass spectrometry imaging of animal–microbe symbioses. PhD Thesis. University of Bremen, 2020.

74. Picazo DR, Dagan T, Ansorge R, Petersen JM, Dubilier N, Kupczok A. Horizontally transmitted symbiont populations in deep-sea mussels are genetically isolated. ISME J. 2019; 13: 2954–68.

75. Franke M, Geier B, Hammel JU, Dubilier N, Leisch N. Becoming symbiotic – the symbiont acquisition and the early development of bathymodiolin mussels. BioRxiv. 2020: 2020.10.09.333211.

76. Wollenberg MS, Ruby EG. Population structure of *Vibrio fischeri* within the light organs of *Euprymna scolopes* squid from Two Oahu (Hawaii) populations. Appl Environ Microbiol. 2009; 75: 193–202.

77. Polzin J, Arevalo P, Nussbaumer T, Polz MF, Bright M. Polyclonal symbiont populations in hydrothermal vent tubeworms and the environment. Proc R Soc B. 2019; 286: 20181281.

78. Shah V, Morris RM. Genome sequence of "*Candidatus* Thioglobus autotrophica" strain EF1, a chemoautotroph from the SUP05 clade of marine Gammaproteobacteria. Genome Announc. 2015; 3: e01156-15.

79. Marshall KT, Morris RM. Genome sequence of "*Candidatus* Thioglobus singularis" strain PS1, a mixotroph from the SUP05 clade of marine Gammaproteobacteria. Genome Announc. 2015; 3: e01155-15.

80. Sayavedra L, Ansorge R, Rubin-Blum M, Leisch N, Dubilier N, Petersen JM. Horizontal acquisition followed by expansion and diversification of toxin-related genes in deep-sea bivalve symbionts. BioRxiv. 2019: 605386.

81. Kuwahara H, Yoshida T, Takaki Y, Shimamura S, Nishi S, Harada M, et al. Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*. Curr Biol. 2007; 17: 881–6.

82. Ponnudurai R, Sayavedra L, Kleiner M, Heiden SE, Thürmer A, Felbeck H, et al. Genome sequence of the sulfur-oxidizing *Bathymodiolus thermophilus* gill endosymbiont. Stand Genomic Sci. 2017; 12: 50.

83. Newton ILG, Woyke T, Auchtung TA, Dilly GF, Dutton RJ, Fisher MC, et al. The *Calyptogena magnifica* chemoautotrophic symbiont genome. Science. 2007; 315: 998–1000.

Supplementary Information

Chasing free-living stages of mussel symbionts– Metagenomic insights from the environment to the host

Merle Ücker[1,2], Rebecca Ansorge[1,3], Yui Sato[1], Ga Yan Grace Ho[1], Christian Borowski[1], Nicole Dubilier[1,2,*]

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

[2]MARUM – Center for Marine Environmental Sciences of the University of Bremen, Bremen, Germany

[3]Quadram Institute Bioscience, Norwich, Norfolk, United Kingdom

Includes

Supplementary Results and Discussion

- Low taxonomic diversity of mussel gill metagenomes
- Abundance of symbionts in the water column
- Three diagnostic markers can discriminate *B. azoricus* and *B. puteoserpentis* symbiont subspecies

Supplementary Figures

Supplementary Tables

Supplementary Results and Discussion

Low taxonomic diversity of mussel gill metagenomes

To compare the taxonomic diversity observed in water metagenomes from hydrothermal vents to the diversity within mussel gill metagenomes, we assessed the taxonomic composition of three mussel metagenomes per site along the MAR with phyloFlash. In mussel gill metagenomes, most of the community could be attributed to the symbiotic partners: mussels (represented in the taxonomic group of Metazoa), SOX symbionts (represented in the taxonomic group of Thiomicrospirales) and methane-oxidising symbionts (represented in the taxonomic group of Methylococcales, Supplementary Figure S 2). Only a few percent of the community consisted of other, mostly bacterial, taxa. Compared to the high diversity of different bacterial taxa observed in water metagenomes (Figure 2), the mussel gill metagenomes represented a distinct low diversity system.

Abundance of symbionts in the water column

The percentage of reads mapping to symbiont 16S rRNA genes in our study was close to 0 %. These numbers seem low, especially compared to Fontanez and Cavanaugh (2014) who quantified the abundance of *B. breviour* symbionts in seawater to be up to 1 % of total bacteria. One technical explanation for the low abundance of symbiont reads in our samples might be the sampling scheme, in which water samples were taken 6-20 m away from the mussel bed. Fontanez & Cavanaugh (2014) investigated water samples that were taken among, near (< 12 m) and away (> 20 m) from mussel beds and found that symbiont density decreased with increasing distance from the mussels. Thus, the symbiont density that can be sampled at 6-20 m in the water column might be rather low.

Three diagnostic markers can discriminate *B. azoricus* and *B. puteoserpentis* symbiont subspecies

We developed a simple workflow to find suitable marker genes that distinguish symbiont subspecies of *B. azoricus* and *B. puteoserpentis* mussels. Such marker genes are crucial to assess the distribution of closely related bacteria, especially when markers that are usually used in these cases such as the 16S rRNA gene, are identical between the subspecies. During our assessment of potential marker genes, we found markers with

(1) sequences that were (almost) identical between symbionts of *B. azoricus* and *B. puteoserpentis*,

(2) sequences that showed no pattern with regard to phylogeny,

(3) at least two copies of the marker with different sequences in each bin, and

(4) sequences that were highly similar within but clearly different between symbiont subspecies (Supplementary Figure S 4 D, Supplementary Figure S 5).

The individual genes that we assessed showed a variety of patterns that does not reflect the overall phylogeny based on a larger set of marker genes. Similarly, Fontanez and Cavanaugh (2014) reported that the seven functional genes that they analysed showed varying patterns in their phylogenies and none of the functional phylogenies was congruent with 16S rRNA phylogeny or any other functional gene phylogeny. For our analysis, we focused on markers that displayed a pattern as described in scenario (4) above and investigated their similarity within and between the symbiont subspecies, leading us to three markers that were chosen for further analysis (Supplementary Figure S 5).

We propose this approach can be employed to identify marker genes for differentiating closely related microbial species by adapting the input marker gene sets and reference genomes to the desired species of interest. The scripts, written in bash commands and R, are available on Github (https://github.com/gracegyho/markergenescreen).

Supplementary Figures

Supplementary Figure S 1 | Images of sampling. A: Sampling with EISP at site Woody in the Menez Gwen vent field during Biobaz cruise (2013). B: Sampling with KIPS inside the Woody crack at the Menez Gwen vent field (Meier et al., 2016). C: Sampling with EISP at site Montsegur in the Lucky Strike vent field during Biobaz cruise (2013).
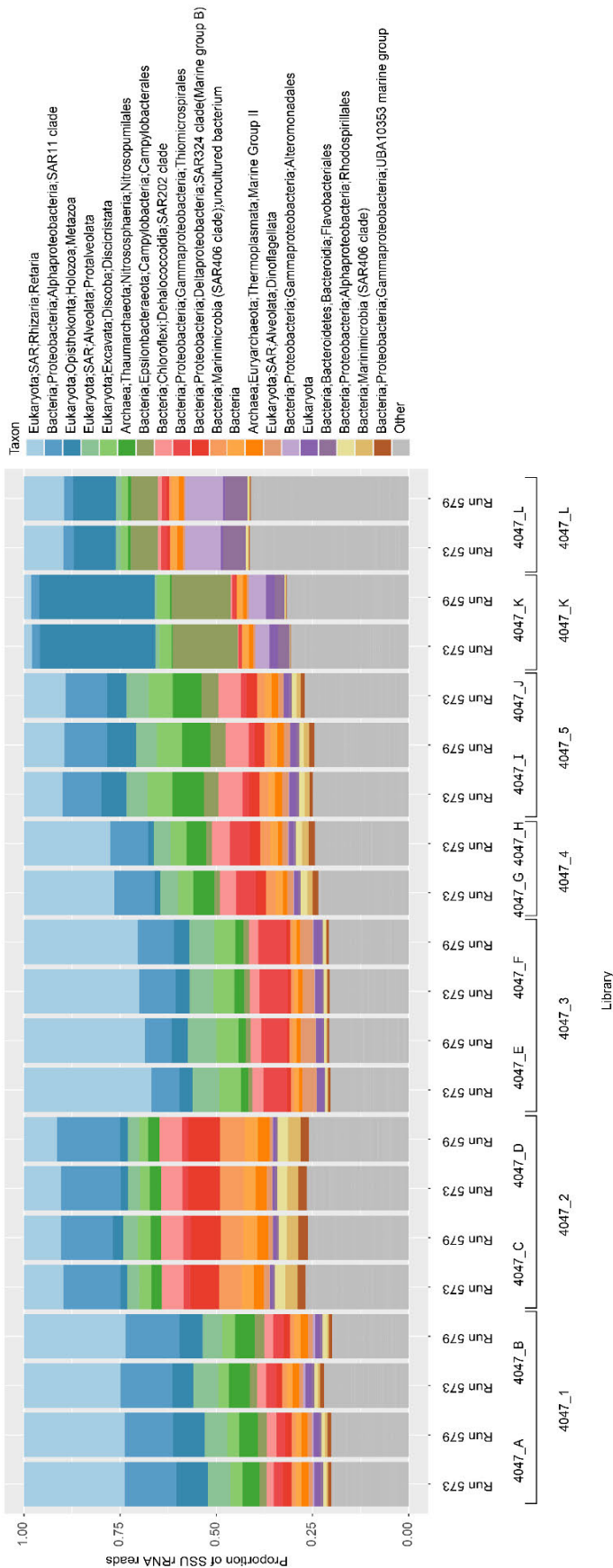
Supplementary Figure S 2 | Taxonomic composition of water samples. Metagenomes from duplicate (e.g., 4047_A and 4047_B) and re-sequenced libraries (e.g., run 573 and run 579) were assessed separately to ensure that they are highly similar before being pooled. The composition was analysed with phyloFlash by classification of detected 16S/18S rRNA reads against the SILVA database.
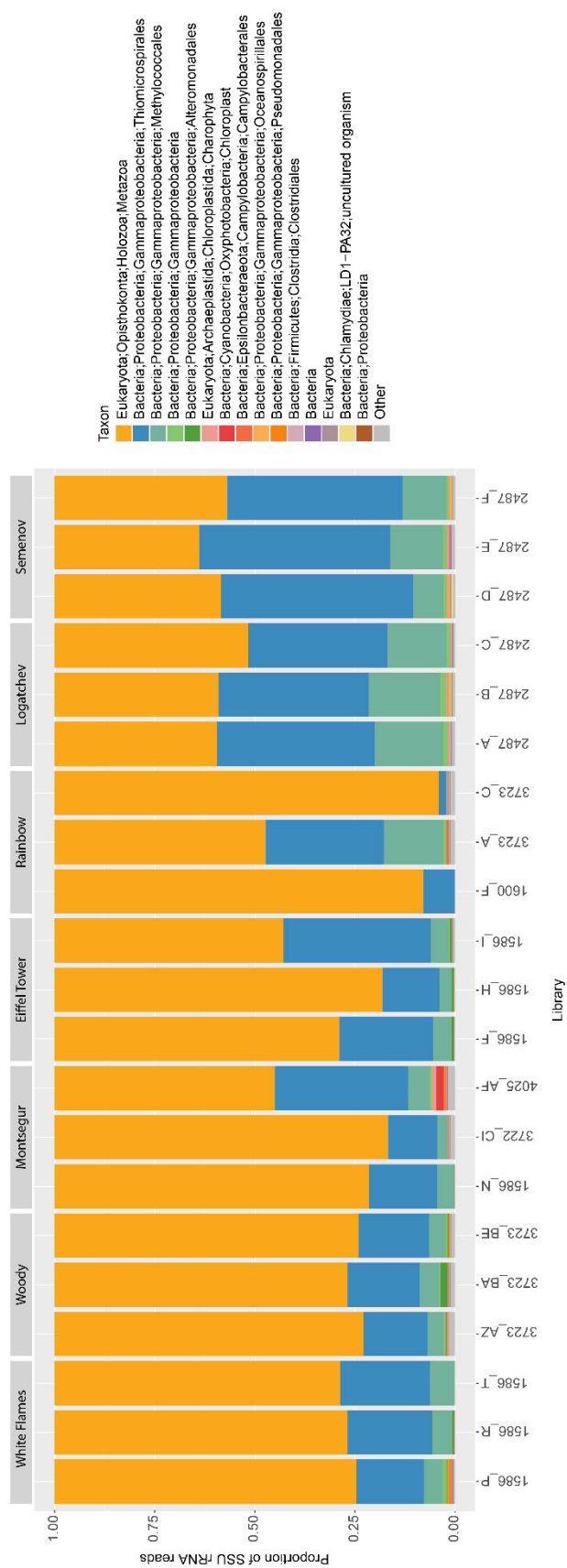
Supplementary Figure S 3 | Taxonomic composition of mussel gill metagenomes from *Bathymodiolus* along the northern MAR. The partners of this symbiosis (mussel host, SOX symbiont and methane-oxidising symbionts) belong to the taxonomic groups depicted in yellow, blue and mint green. The composition was analysed with phyloFlash by classification of reads matching 16S/18S rRNA genes against the SILVA database.
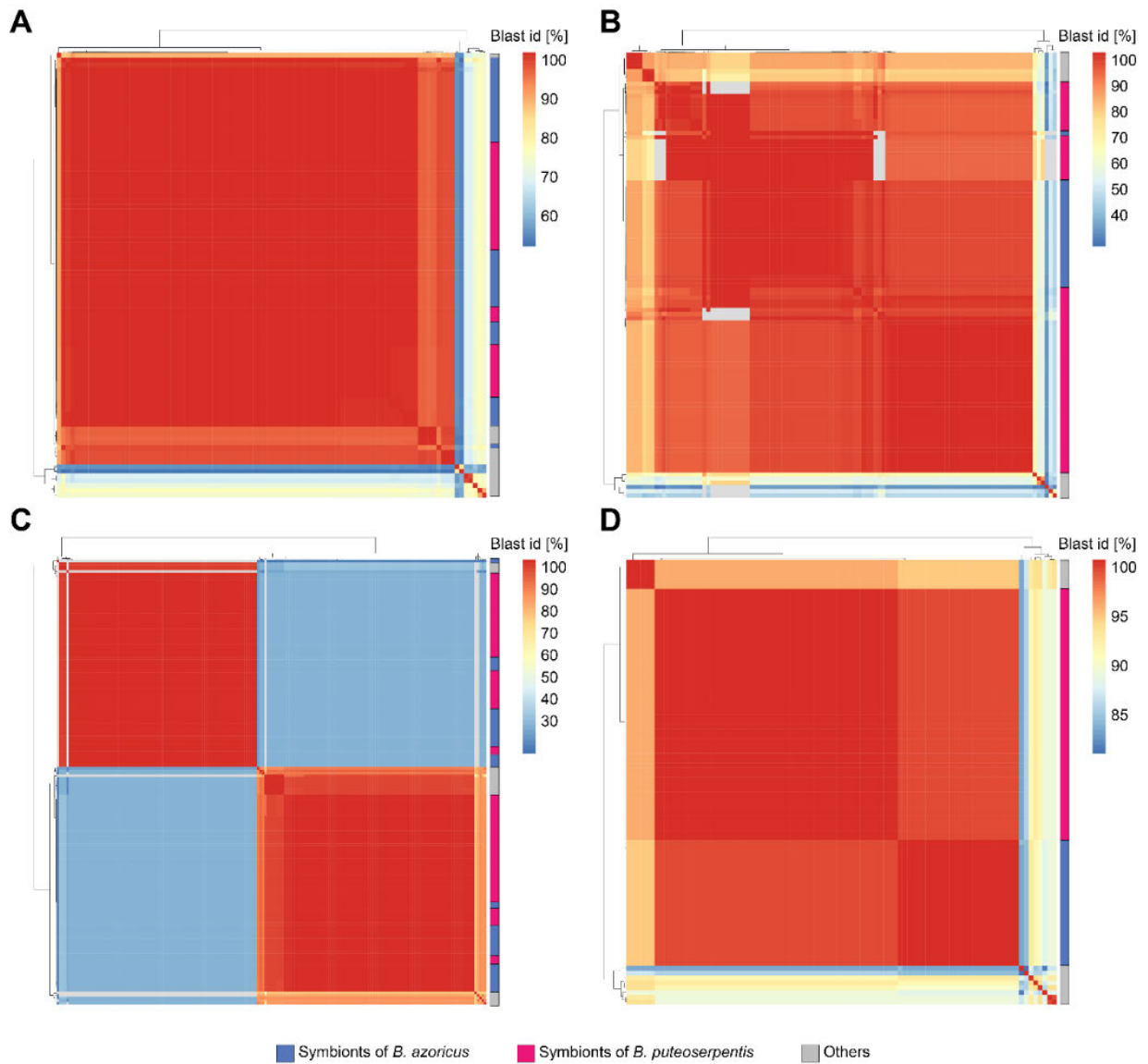
Supplementary Figure S 4 | Heatmaps of pairwise blastp similarities of gammaproteobacterial markers in *Bathymodiolus* symbionts and other SUP05 MAGs and genomes. Heatmap colour corresponds to high (red) or low (blue) similarity, grey indicates that no hits could be found in the pairwise blast. Colour bars next to the heatmap correspond to the MAG affiliation: Symbionts of *B. azoricus* (blue), symbionts of *B. puteoserpentis* (pink) and others (grey) including *Bathymodiolus* and other symbionts as well as two free-living *Thiologlobus* ssp. Four different patterns of similarity were observed. A: Protein is (almost) identical between symbionts of *B. azoricus* and *B. puteoserpentis*; B: Protein had no similarity pattern related to phylogeny; C: Protein had two different copies in most MAGs; D: Protein was highly similar within a symbiont subspecies but less similar between subspecies. Markers showing a pattern as depicted in D were used for further analyses.
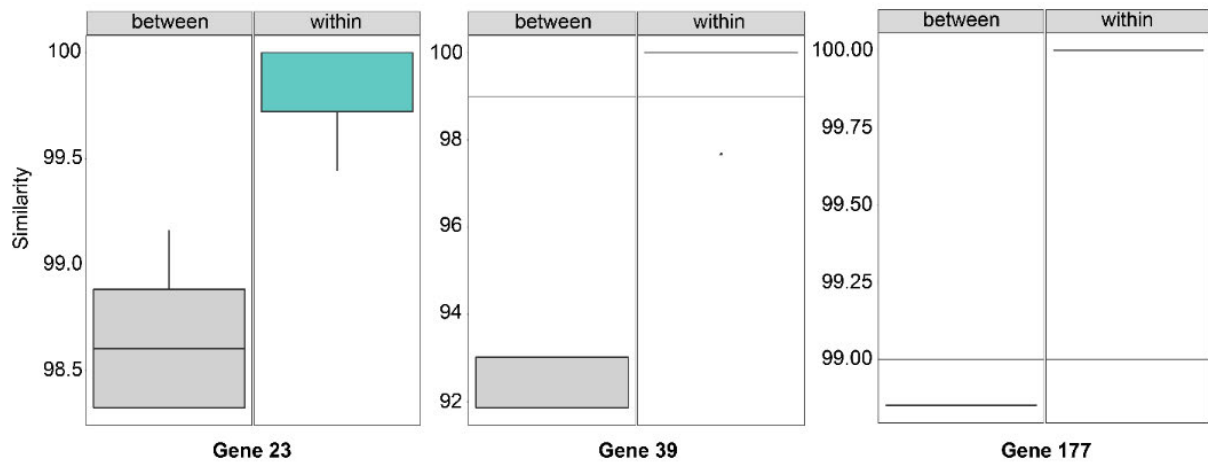
Supplementary Figure S 5 | Boxplots representing pairwise similarity within (lightseagreen) and between (grey) symbionts of *B. azoricus* and *B. puteoserpentis* for the three diagnostic markers. Horizontal line indicates a similarity of 99 %. All three markers were used for subsequent analyses because the similarity within the lineages was much higher than between.

Supplementary Tables

Supplementary Table S 1 | Overview of water filter samples. Lat = Latitude, Lon = Longitude, Metag = Metagenome, Sampling = sampling device/setup, CA = cellulose acetate, PG = plankton gaze, Volume = total volume of filtered water, Dist = distance to seafloor.

Site	Cruise	Date	Lat	Lon	Depth	DNA library	Metag.	Sampling	Filter	Pore size	Volume	Dist	Fixation
White Flames (Menez Gwen vent field)	Biobaz	2013-08-04	37.928	-31.519	-780	4047_C, D	4047_2	WISP1, in-situ pump Roger on Wire	CA	0.22 μm	550 L	30 m	-80°C
Woody (Menez Gwen vent field)	Biobaz	2013-08-18	37.845	-31.519	-815	4047_A, B	4047_1	EISP3, in-situ pump Norbert deployed with ROV	CA	0.22 μm	1570.30 L	0 m	-80°C
Eiffel Tower (Lucky Strike vent field)	Biobaz	2013-08-16	37.289	-32.276	-1688	4047_E,F	4047_3	WISP5, in-situ pump Norbert on Wire	CA	0.22 μm	1755.56 L	30 m	-80°C
Montsegur (Lucky Strike vent field)	Biobaz	2013-08-08	37.288	-32.276	-1698	4047_G, H	4047_4	EISP2, in-situ pump Norbert deployed with ROV	CA	0.22 μm	683 L	0 m	-80°C
Rainbow	Biobaz	2013-08-13	36.229	-33.902	-2276	4047_I, J	4047_5	WISP6, in-situ pump Norbert on Wire	CA	0.22 μm	1270.8 L	30 m	-80°C
Logatchev	BigMAR (M126)	2016-05-15	14.753	-44.979	-2970	4047_L	4047_L	Mooring 5	PG	30 μm	24062.56 L	6 m	0,5xPBS/ 60%EtOH
Semenov-2	BigMAR (M126)	2016-05-11	13.513	-44.963	-2237	4047_K	4047_K	Mooring 3	PG	30 μm	22324.82 L	6 m	0,5xPBS/ 60%EtOH

Supplementary Table S 2 | *Bathymodiolus* **SOX symbiont bins and genomes of close relatives from public databases.** The bins were screened for marker genes that discriminate *B. azoricus* and *B. puteoserpentis* symbionts. MAG = Metagenome-assembled genome, Lat = Latitude, Lon = Longitude, Pub = Publication, BC = backcross, Hybrid = Hybrid between *B. azoricus* and *B. puteoserpentis*.

MAG	Site	Lat	Lon	Cruise	Host species	Pub	Acession
Ca. Thioglobus autotrophicus strain EF1	Effingham Inlet	49.029	-125.15			[78]	PRJNA224116
Ca. Thioglobus singularis PS1	Puget Sound	47.6	-122.45			[79]	PRJNA229178
1048J	Lucky Strike (Eiffel Tower)	38.283	-32.276	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1586P	Menez Gwen (White Flames)	37.844	-31.519	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1586Q	Menez Gwen (White Flames)	37.844	-31.519	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1586R	Menez Gwen (White Flames)	37.844	-31.519	Biobaz (2013)	*B. azoricus*	[80]	SAMEA6822960
1586S	Menez Gwen (White Flames)	37.844	-31.519	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1586B	Lucky Strike (Eiffel Tower)	37.289	-32.275	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1586C	Lucky Strike (Eiffel Tower)	37.289	-32.275	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1586D	Lucky Strike (Eiffel Tower)	37.289	-32.275	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1586E	Lucky Strike (Eiffel Tower)	37.289	-32.275	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1048F	Lucky Strike (Montsegur)	37.288	-32.276	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1048G	Lucky Strike (Montsegur)	37.288	-32.276	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1048H	Lucky Strike (Montsegur)	37.288	-32.276	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1586K	Lucky Strike (Montsegur)	37.288	-32.276	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1586N	Lucky Strike (Montsegur)	37.288	-32.276	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1586O	Lucky Strike (Montsegur)	37.288	-32.276	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091

MAG	Site	Lat	Lon	Cruise	Host species	Pub	Acession
1048I	Lucky Strike (Eiffel Tower)	37.283	-32.276	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1586F	Lucky Strike (Eiffel Tower)	37.283	-32.276	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1586G	Lucky Strike (Eiffel Tower)	37.283	-32.276	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1586I	Lucky Strike (Eiffel Tower)	37.283	-32.276	Biobaz (2013)	*B. azoricus*	[80]	SAMEA6822959
1586J	Lucky Strike (Eiffel Tower)	37.283	-32.276	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1600F	Rainbow	36.229	-33.902	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1600G	Rainbow	36.229	-33.902	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1600H	Rainbow	36.229	-33.902	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1600I	Rainbow	36.229	-33.902	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
1600J	Rainbow	36.229	-33.902	Biobaz (2013)	*B. azoricus*	[16]	PRJEB36091
Ca. Vesicomyosocius okutanii HA	Sagami Bay	35.117	139.383		*C. okutanii*	[81]	PRJDA18267
Endosymbiont of *Bathymodiolus septemdierum* str. Myojin knoll DNA, complete genome	Izu-Bonin Arc, Myojin knoll (Japan)	32.104	139.219		*B. septemdierum*	[28]	PRJDB949
3386_A	Broken Spur	29.17	-43.17	AT_05/03	*B. puteoserpentis*	[18]	PRJEB36976
3386_B	Broken Spur	29.17	-43.17	AT_05/03	BC *puteoserpentis*	[18]	PRJEB36976
3386_C	Broken Spur	29.17	-43.17	AT_05/03	Hybrid	[18]	PRJEB36976
3386_D	Broken Spur	29.17	-43.17	AT_05/03	*B. puteoserpentis*	[18]	PRJEB36976
3386_E	Broken Spur	29.17	-43.17	AT_05/03	*B. puteoserpentis*	[18]	PRJEB36976
3386_F	Broken Spur	29.17	-43.17	AT_05/03	*B. puteoserpentis*	[18]	PRJEB36976
3386_K	Broken Spur	29.17	-43.17	AT_05/03	BC *puteoserpentis*	[18]	PRJEB36976
3386_G	Broken Spur	29.17	-43.17	AT_05/03	Hybrid	[18]	PRJEB36976
3386_H	Broken Spur	29.17	-43.17	AT_05/03	Hybrid	[18]	PRJEB36976
3386_I	Broken Spur	29.17	-43.17	AT_05/03	*B. puteoserpentis*	[18]	PRJEB36976
3386_L	Broken Spur	29.17	-43.17	AT_05/03	BC *puteoserpentis*	[18]	PRJEB36976
3386_P	Broken Spur	29.17	-43.17	AT_05/03	BC *puteoserpentis*	[18]	PRJEB36976
3386_Q	Broken Spur	29.17	-43.17	AT_05/03	BC *puteoserpentis*	[18]	PRJEB36976
3386_R	Broken Spur	29.17	-43.17	AT_05/03	BC *puteoserpentis*	[18]	PRJEB36976
3386_T	Broken Spur	29.17	-43.17	AT_05/03	Hybrid	[18]	PRJEB36976
3386_U	Broken Spur	29.17	-43.17	AT_05/03	Hybrid	[18]	PRJEB36976
3386_V	Broken Spur	29.17	-43.17	AT_05/03	BC *puteoserpentis*	[18]	PRJEB36976
3386_W	Broken Spur	29.17	-43.17	AT_05/03	Hybrid	[18]	PRJEB36976
3386_X	Broken Spur	29.17	-43.17	AT_05/03	Hybrid	[18]	PRJEB36976
3386_AA	Broken Spur	29.17	-43.17	AT_05/03	BC *puteoserpentis*	[18]	PRJEB36976
3386_AB	Broken Spur	29.17	-43.17	AT_05/03	*B. puteoserpentis*	[18]	PRJEB36976

MAG	Site	Lat	Lon	Cruise	Host species	Pub	Acession
3386_AC	Broken Spur	29.17	-43.17	AT_03/03	*B. puteoserpentis*	[18]	PRJEB36976
3386_AD	Broken Spur	29.17	-43.17	AT_03/03	Hybrid	[18]	PRJEB36976
3386_AE	Broken Spur	29.17	-43.17	AT_03/03	*B. puteoserpentis*	[18]	PRJEB36976
3386_AF	Broken Spur	29.17	-43.17	AT_03/03	BC *puteoserpentis*	[18]	PRJEB36976
3386_AK	Broken Spur	29.17	-43.17	AT_03/03	*B. puteoserpentis*	[18]	PRJEB36976
3386_AL	Broken Spur	29.17	-43.17	AT_03/03	Hybrid	[18]	PRJEB36976
BHECKSOX	Chapopote (Mexico)	21.9001	-93.435	M114-2	*B. heckerae*	[80]	PRJEB17996
BBROOKSOX	Chapopote (Mexico)	21.9	-93.435	M114-2	*B. brooksi*	[80]	PRJEB17996
2065A	Logatchev Quest	14.753	-44.979	M64-2 Logatchev (2005)	*B. puteoserpentis*	[16]	PRJEB36091
2065B	Logatchev Quest	14.753	-44.979	M64-2 Logatchev (2005)	*B. puteoserpentis*	[16]	PRJEB36091
2487A	Logatchev Quest	14.753	-44.98	M126 (2016)	*B. puteoserpentis*	[16]	PRJEB36091
2487B	Logatchev Quest	14.753	-44.98	M126 (2016)	*B. puteoserpentis*	[16]	PRJEB36091
2487C	Logatchev Quest	14.753	-44.98	M126 (2016)	*B. puteoserpentis*	[16]	PRJEB36091
3722CJ	Logatchev Quest	14.753	-44.981	MSM10-03 Hydromar VII	*B. puteoserpentis*	[16]	PRJEB36091
3722CK	Logatchev Quest	14.753	-44.981	MSM10-03 Hydromar VII	*B. puteoserpentis*	[16]	PRJEB36091
3722CL	Logatchev Quest	14.753	-44.981	MSM10-03 Hydromar VII	*B. puteoserpentis*	[16]	PRJEB36091
3722CM	Logatchev Quest	14.753	-44.981	MSM10-03 Hydromar VII	*B. puteoserpentis*	[16]	PRJEB36091
3722CN	Logatchev Quest	14.753	-44.981	MSM10-03 Hydromar VII	*B. puteoserpentis*	[16]	PRJEB36091
3722CO	Logatchev Quest	14.753	-44.981	MSM10-03 Hydromar VII	*B. puteoserpentis*	[16]	PRJEB36091
3722CP	Logatchev Quest	14.753	-44.98	M126 (2016)	*B. puteoserpentis*	[16]	PRJEB36091
2487D	Semenov	13.514	-44.963	M126 (2016)	*B. puteoserpentis*	[16]	PRJEB36091
2487E	Semenov	13.514	-44.963	M126 (2016)	*B. puteoserpentis*	[16]	PRJEB36091
2487F	Semenov	13.514	-44.963	M126 (2016)	*B. puteoserpentis*	[16]	PRJEB36091
1115A	Semenov	13.513	-44.963	Odemar (2014)	*B. puteoserpentis*	[16]	PRJEB36091
1115B	Semenov	13.513	-44.963	Odemar (2014)	*B. puteoserpentis*	[16]	PRJEB36091

MAG	Site	Lat	Lon	Cruise	Host species	Pub	Acession
1115C	Semenov	13.513	-44.963	Odemar (2014)	*B. puteoserpentis*	[16]	PRJEB36091
Bathymodiolus thermophilus thioautotrophic gill symbiont strain:BAT/CrabSpa'14	East Pacific Rise (EPR) 9°N	9.83983	-104.29	R/V *Atlantis* cruise AT26–10	*B. thermophilus*	[82]	PRJNA339702
Ca. Ruthia magnifica str. Cm	9° East Pacific Rise vent field	9.83	-104.29		*C. magnifica*	[83]	PRJNA16841
C112	Clueless	-4.803	-12.372	M78-2 (2009)	*B.* sp. Clueless	[16]	PRJEB36091
C113	Clueless	-4.803	-12.372	M78-2 (2009)	*B.* sp. Clueless	[16]	PRJEB36091
C114	Clueless	-4.803	-12.372	M78-2 (2009)	*B.* sp. Clueless	[16]	PRJEB36091
L102	Lilliput	-9.547	-13.21	M78-2 (2009)	*B.* sp. Lilliput	[16]	PRJEB36091
L51	Lilliput	-9.547	-13.21	M78-2 (2009)	*B.* sp. Lilliput	[16]	PRJEB36091
L54	Lilliput	-9.547	-13.21	M78-2 (2009)	*B.* sp. Lilliput	[16]	PRJEB36091

Supplementary Table S 3 | Overview of external metagenomic reads. Data were used for mapping against marker genes and investigation of taxonomic diversity within mussel gills. Lat = Latitude, Lon = Longitude.

Metagenome	Cruise	Date	Lat	Lon	Depth	Site	Accession
3723_AZ	Biobaz	2013-08-06	37.84	-31.52	-828	Menez Gwen - Woody	
3723_BA	Biobaz	2013-08-06	37.84	-31.52	-828	Menez Gwen - Woody	PRJEB42345
3723_BE	Biobaz	2013-08-06	37.84	-31.52	-828	Menez Gwen - Woody	PRJEB42345
3723_BF	Biobaz	2013-08-06	37.84	-31.52	-828	Menez Gwen - Woody	PRJEB42345
3723_BG	Biobaz	2013-08-06	37.84	-31.52	-828	Menez Gwen - Woody	
3723_BI	Biobaz	2013-08-06	37.84	-31.52	-828	Menez Gwen - Woody	PRJEB42345
3723_BJ	Biobaz	2013-08-06	37.84	-31.52	-828	Menez Gwen - Woody	PRJEB42345
3723_BL	Biobaz	2013-08-06	37.84	-31.52	-828	Menez Gwen - Woody	PRJEB42345
3723_BM	Biobaz	2013-08-06	37.84	-31.52	-828	Menez Gwen - Woody	PRJEB42345
3723_BN	Biobaz	2013-08-06	37.84	-31.52	-828	Menez Gwen - Woody	PRJEB42345
1586_P	M82-3	2021-09-13	37.45	-31.52	-836	Menez Gwen - White Flames	PRJEB36091
1586_R	M82-3	2021-09-13	37.45	-31.52	-836	Menez Gwen - White Flames	
1586_T	M82-3	2021-09-13	37.45	-31.52	-836	Menez Gwen - White Flames	
1586_N	Biobaz	2013-08-09	37.29	-32.28	-1700	Lucky Strike - Montsegur	PRJEB36091
3722_CI	Biobaz	2013-08-08	37.29	-32.28	-1700	Lucky Strike - Montsegur	PRJEB42345
4025_AF	Biobaz	2013-08-08	37.29	-32.28	-1700	Lucky Strike - Montsegur	PRJEB42345
1586_F	Biobaz	2013-08-17	37.28	-32.28	-1689	Lucky Strike - Eiffel Tower	PRJEB36091
1586_H	Biobaz	2013-08-17	37.28	-32.28	-1689	Lucky Strike - Eiffel Tower	
1586_I	Biobaz	2013-08-17	37.28	-32.28	-1689	Lucky Strike - Eiffel Tower	
3722_AA	Biobaz	2013-08-17	37.28	-32.28	-1689	Lucky Strike - Eiffel Tower	
3722_W	Biobaz	2013-08-17	37.28	-32.28	-1689	Lucky Strike - Eiffel Tower	PRJEB42345
3722_Y	Biobaz	2013-08-17	37.28	-32.28	-1689	Lucky Strike - Eiffel Tower	PRJEB42345
1600_F	Biobaz	2013-08-13	36.23	-33.90	-2273	Rainbow	PRJEB36091
3723_A	Biobaz	2013-08-13	36.23	-33.90	-2273	Rainbow	
3723_C	Biobaz	2013-08-13	36.23	-33.90	-2273	Rainbow	PRJEB42345
2487_A	M126	2016-04-28	14.75	-44.98	-3047	Logatchev	PRJEB36091
2487_B	M126	2016-04-28	14.75	-44.98	-3047	Logatchev	PRJEB36091
2487_C	M126	2016-04-28	14.75	-44.98	-3047	Logatchev	PRJEB36091
3722_CQ	M126	2016-04-28	14.75	-44.98	-3047	Logatchev	PRJEB42345
3723_CH	M126	2016-04-28	14.75	-44.98	-3047	Logatchev	
2065_A	M64-2 Logatchev 2005	2005-05-30	14.75	-44.98	-3045	Logatchev	PRJEB36091
2065_B	M64-2 Logatchev 2005	2005-05-30	14.75	-44.98	-3045	Logatchev	PRJEB36091

Metagenome	Cruise	Date	Lat	Lon	Depth	Site	Accession
2065_C	M64-2 Logatchev 2005	2005-05-30	14.75	-44.98	-3045	Logatchev	PRJEB36091
2487_D	M126	2016-05-02	13.51	-44.96	-2447	Semenov	PRJEB36091
2487_E	M126	2016-05-02	13.51	-44.96	-2447	Semenov	PRJEB36091
2487_F	M126	2016-05-02	13.51	-44.96	-2447	Semenov	PRJEB36091
3723_P	M126	2016-05-02	13.51	-44.96	-2447	Semenov	
3723_R	M126	2016-05-02	13.51	-44.96	-2447	Semenov	PRJEB42345
3723_S	M126	2016-05-02	13.51	-44.96	-2447	Semenov	PRJEB42345
3723_T	M126	2016-05-02	13.51	-44.96	-2447	Semenov	PRJEB42345
Sem_B	Odemar	2014-11-22	13.51	-44.96	-2432	Semenov	PRJEB36091

population
[pɒpjʊˈleɪʃ(ə)n] *noun*

a community of animals, plants, or humans among whose members interbreeding occurs

Chapter IV | Hosts from the Mid-Atlantic Ridge

Mitonuclear discordance suggests long-distance migration and mitochondrial introgression of deep-sea mussels along the Mid-Atlantic Ridge

Merle Ücker[1,2], Yui Sato[1], Rebecca Ansorge[1,3], Anne Kupczok[1,4], Maximilian Franke[1], Harald Gruber-Vodicka[1], Nicole Dubilier[1,2,*]

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

[2]MARUM – Center for Marine Environmental Sciences of the University of Bremen, Bremen, Germany

[3]Quadram Institute Bioscience, Norwich, Norfolk, United Kingdom

[4]Wageningen University & Research, Wageningen, Netherlands

[*]Corresponding author

This manuscript is in preparation and has not been reviewed by all authors.

Abstract

Populations of hydrothermal vent fauna at different sites are separated by large geographic distances and potential dispersal barriers, however, genetic connectivity is much higher than anticipated. Population genomic studies of deep-sea vent fauna are limited by the accessibility of their environment and insufficient resolution of markers such as mitochondrial COI genes. Here, we investigated 175 individuals of *Bathymodiolus* mussels, one of the most successful fauna at hydrothermal vents, from 22 vent sites along the Mid-Atlantic Ridge. We applied metagenome sequencing and exome-wide SNP profiling to assess how mitochondrial and nuclear genomes relate, and to analyse contemporary population genomic patterns. Our analysis revealed that mitochondrial clades were specific to the region of sampling but not always congruent with chromosomal genetic differentiation. We observed this mitonuclear discordance for 18 out of 175 individuals. While there are different processes that can lead to the observed pattern, we consider long-distance migration of dispersing larvae followed by mitochondrial introgression the most likely explanation. This is supported by the lack of clear subpopulation structure between conspecific mussels from different vent sites suggesting regularly mixing of populations enabled through exchange of larvae between these sites. Our study sheds light on the geographic ranges of *Bathymodiolus* species at the Mid-Atlantic Ridge and shows the potential of using available metagenomic data for high-resolution analyses to enhance our knowledge about deep-sea *Bathymodiolus* mussels.

Introduction

Animal populations at deep-sea hydrothermal vents are naturally separated by distinct features of their environment [1]. Depth differences and large geographic distances separate vent sites. Topological features, such as transform faults and high walls of the rift valley, cross-axis currents, and the influence of changing volcanic and tectonic activity make hydrothermal vents unstable and heterogeneous environments. All these factors were expected to enhance subdivision and potentially speciation between animal populations at different vent sites [2–4]. However, some vent fauna, such as deep-sea mussels and clams, have been shown to be genetically connected across thousands of kilometres despite interrupting geological features [5–8]. Most vent fauna are sessile organisms that rely on larval dispersal to escape changing environments and colonise new locations [9,10]. Animal species from hydrothermal vents are endemic to this particular environment. Vent fields are far apart from each other (up to 4000 km between known vent sites), especially in the

Atlantic Ocean where the ridge system is slow spreading. Thus, migration is crucial for genetic connectivity, gene flow, and introgression among vent populations [11].

Bathymodiolus mussels are among the most successful fauna at hydrothermal vents throughout the world's ocean. Adult mussels obtain their nutrition through chemosynthetic endosymbionts that oxidise reduced chemical compounds from the environment to obtain energy for biomass production [12,13]. Mussels acquire their symbionts from the environment throughout their lifetime[14–18], and it has been hypothesised that the symbionts are geographically structured [19,20]. Four mussel lineages have been described at the Mid-Atlantic Ridge (MAR), each specific to a geographic region [7,21]: On the northern MAR, *B. azoricus* occurs at vent fields between 38°N and 36°N and *B. puteoserpentis* at vent fields between 29°N and 13°N. The intermediate field Broken Spur at 29°N has been previously described as hybrid zone and is inhabited by *B. puteoserpentis* and hybrids of *B. azoricus* and *B. puteoserpentis* [20,22,23]. On the southern MAR, two lineages are present at vent fields, one at 5°S and one at 9°S. *Bathymodiolus* mussels have a life cycle in which adult mussels are sessile and only the planktonic larvae can disperse over larger distances. A mean larval dispersal distance of max. 150 km has been predicted for the MAR species based on biophysical modelling [5] whereas vertical migration and transport of larvae over distances > 150 km with surface water currents has been proposed for *"B." childressi* [24]. Larvae are likely filter feeders that switch their mode of nutrition once they settle [13].

Studying organisms of ecological relevance, such as *Bathymodiolus* mussels, can help us understand evolutionary history and the population dynamics of these species in their natural context. However, population genomic studies of these mussels come with two main challenges. First, such studies usually require a sufficient number of replicates per population or site, e.g. a minimum of 3-8 individuals has been suggested for the harlequin lady beetle [25]. Obtaining such replicates of deep-sea mussels from hydrothermal vents across long geographic distances (e.g. a total range of 5 600 km in this study) comes with high costs and immense time requirements. Cruises for deep-sea research and exploration usually consist of interdisciplinary teams, limiting the time and sample capacity for each of the research disciplines. The acquisition of mussel samples requires technologies such as remotely operated vehicles and relies on the previous (re-)detection of hydrothermal vent sites, which can be challenging in itself, depending on weather, water currents, technical issues, and other factors.

Secondly, population genomic studies of non-model organisms such as *Bathymodiolus* were so far often limited to few loci or mitochondrial DNA [5,7,26–31] that might not reflect the diversity and differentiation of the whole genome [32]. Species determination for these mussels has mostly been done based on the mitochondrial cytochrome c-oxidase (COI) gene, the most common marker for species assessment in animals [33]. While mitochondria are maternally inherited in most animals, their inheritance is complex or unresolved in many bivalve species, e.g. double-uniparental inheritance in marine clams [34,35] or sex-biased inheritance in *Mytilus* spp. [36,37]. This can complicate the interpretation of mitochondria-derived population information. The four mussel lineages on the MAR were confirmed by Breusing et al. (2016) based on 100 molecular markers [5]. Their analysis further revealed a lack of genetic differentiation of conspecific mussels from different vent fields in the same region (e.g., *B. azoricus* in Menez Gwen, Lucky Strike, and Rainbow covering ~280 km between 38°N and 36°N). However, markers were designed to discriminate between lineages and might not capture contemporary subpopulation structure within a lineage.

For the first time, we used profiling of exome-wide single-nucleotide polymorphisms (SNPs) to assess the genetic differentiation of 175 *Bathymodiolus* mussels sampled at 22 sites in 10 hydrothermal vent fields along the MAR. By investigating population genomics of these deep-sea mussels on genome-wide level, we assessed how mitochondrial and nuclear genomes relate, and whether there is subpopulation structure in conspecific mussels from different vent fields.

Methods

Sampling, DNA extraction, and metagenomic sequencing of deep-sea mussels

175 mussels of the genus *Bathymodiolus* were sampled at 10 vent fields covering 22 vent sites along the Mid-Atlantic Ridge between 1997 and 2016 (Figure 1 A). A summary of vent fields included in this study is available in Table 1. Mussels were recovered with either a scoop net, a metal claw of the remotely operated vehicle, with a scratch shovel or on rock samples. The samples were dissected on board, and the gills (or mixed tissue in case of 2424D, E, H, N from Broken Spur) were either fixed in RNA*later* according to the manufacturer's protocol (Sigma-Aldrich, St. Louis, MO, USA), frozen at -20°C or -80°C or fixed in 2 % paraformaldehyde/PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 2 mM KH2PO4) solution and stored in 0.5x PBS/60 % Ethanol. A complete sample list is available in Supplementary Table S 1. Our dataset includes newly sequenced metagenomes and

metagenomic reads that were previously sequenced. Previously sequenced data was merged with newer data from the same biological sample. Samples sequenced in project 3722 and 3723 were extracted from 3-4 mm sized pieces with the DNAeasy 96 Blood & Tissue Kit (QIAGEN, Hilden, Germany) according to the manufacturer's protocol with the following modifications: Before elution, samples were incubated for 10 min at room temperature and subsequently centrifuged. Eluent volume was halved and first eluate was re-used for the second elution to maximise DNA yields. Samples sequenced in project 4025 and 4132 were re-sequenced from previously extracted DNA.

For sequencing projects 3722 and 3723, libraries were prepared with the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs, Frankfurt, Germany). For project 4025, libraries were prepared with the Nextera DNA Flex Library Prep Kit (Illumina, San Diego, CA, USA). Samples in project 4132 were re-sequenced from the previously prepared libraries. All samples were sequenced as 150 bp paired end reads on Illumina HiSeq3000 machines. Library preparation and sequencing was performed at the Max Planck-Genome-centre Cologne, Germany (https://mpgc.mpipz.mpg.de/home/). The DNA extraction and sequencing procedures of previously sequenced metagenomes are available in the respective publications [20,38,39].

Analysis of mitochondrial phylogeny

To investigate mitochondrial clades of *Bathymodiolus* mussels, we assembled mitochondrial sequences from 175 mussel gill metagenomes from 10 vent sites on the MAR (Figure 1 A). Metagenomic reads were cleaned by removing PCR duplicates with FastUniq v1.1 [40], trimming sequencing adapters, first 10 base pairs and bases with a PHRED score <2 with Trimmomatic v0.36 [41], and interleaving the reads with the script interleave-fastqgz-MITOBIM.py from MITObim v1.9.1 [42]. We ran MITObim v1.9.1 using a published COI sequence of *B. azoricus* (accession FJ766947 [26]) for mitochondrial baiting and iterative mapping. With BBMap from the BBTools Suite v38.34 [43], we mapped cleaned reads to the MITObim output with 90 % minimum identity (bbmap.sh in1 in2 ref=mitobim_out.fasta nodisk ow minid=0.90 killbadpair pairedonly mappedonly maxindel=1000 pairlen=1000 out1 out2) and assembled the mapped reads with SPAdes v3.14.1 [44] using kmers 21, 33, 55, 77, 99, and 127, and the plasmid option. Resulting assemblies were visualised in Bandage v0.8.0 [45] based on their assembly graph. Only assemblies that represented a closed circular genome or that consisted of one long contig were kept for further analysis. If ends of the long contig were connected by repeats, these were removed.

Table 1 | Summary of vent sites and samples per site in this study. Lat: Latitude, Lon: Longitude, NMAR: Northern Mid-Atlantic Ridge, SMAR: Southern Mid-Atlantic Ridge. Depth in m.

Region	Vent field	Vent site	Lat site	Lon site	Depth	# of samples
38°N-36°N	Menez Gwen	White Flames	37.84	-31.52	-836	1
		Woody	37.84	-31.52	-828	11
		Station 719-10	37.84	-31.52	-839	7
		Station 729-4	37.84	-31.52	-840	2
		Station 736-3	37.84	-31.52	-836	1
	Bubbylon	Bubbylon	37.80	-31.54	-1002	1
	Lucky Strike	Montsegur	37.29	-32.28	-1700	6
		Eiffel Tower	37.28	-32.28	-1689	13
	Rainbow	Rainbow	36.23	-33.90	-2273	10
29°N-13°N	Broken Spur	Broken Spur	29.17	-43.17	-3056	27
	Logatchev-1	Quest	14.75	-44.98	-3047	13
		Irina II	14.75	-44.98	-3037	8
	Semenov-2	Ash Lighthouse	13.51	-44.96	-2447	7
		Semenov-2	13.51	-44.96	-2324	3
	Irinovskoe	Irinovskoe	13.33	-44.91	-2791	8
5°S	Comfortless Cove	Desperate	-4.80	-12.37	-2988	11
		Wideawake	-4.81	-12.37	-2958	6
		Golden Valley	-4.80	-12.37	-2986	6
		Foggy Corner	-4.80	-12.37	-2987	11
		Clueless	-4.80	-12.37	-2995	8
9°S	Lilliput	Lilliput	-9.55	-13.21	-1494	15

To assess the mitochondrial phylogeny, we reconstructed a mitochondrial tree including *B. thermophilus* (MK721544) and *B. septemdierum* (AP014562) as outgroups. First, we aligned the sequences with MARS [46], a cyclic aligner. The resulting multifasta file was split into single fasta files to be annotated with prokka v1.13 [47]. Genbank files of the annotated mitochondrial sequences were visualised in Geneious v11.1.5 [48] for manual determination of the sequence direction. Sequences were aligned with MAFFT v7.407 [49,50] on the command line. The resulting alignment covered 38 601 bp and was used for tree reconstruction with IQ-TREE v1.6.10 [51] with 1000 replicates for both SH-aLRT test [52] and ultrafast bootstrap [53]. We used the TN+F+R9 model [54], which was determined as best model by ModelFinder [55], for tree building. The tree was visualised in iTol v5.7 [56] and edited with Adobe Illustrator [57].

Not all metagenomes yielded mitochondrial assemblies of sufficient quality for further analysis which is why we compared the resulting mitochondrial clades to the affiliation of the cytochrome c oxidase I (COI) gene when analysed with BathyFlash/GENEFlash v1.0, and inferred a haplotype network. BathyFlash is a modified version of PhyloFlash [58] and enables high throughout phylogenetic screening using *Bathymodiolus* COI gene abundances. The COI haplotype network was inferred with PopART v1.7.2 [59] using the median-joining network inference method with Epsilon of 0 [60]. We compared the mitochondrial clades to vent fields where mussels were sampled, using the packages tidyverse v1.3.0 [61], dplyr v1.0.2 [62], htmlwidgets v1.5.2 [63], and the function sankeyNetwork from networkD3 v0.4 [64] in RStudio v1.3959 [65] using R v3.6.3. All figures were modified with Adobe Illustrator [57].

Curation of transcriptomic references

To reduce redundancy in our transcriptome references, we applied a multistep workflow of quality filtering the contigs following transcriptomic best practice guidelines by De Wit, Pespeni, and Palumbi (2015). We used transcriptome assemblies from three mussel species as references that were provided by C. Breusing. Raw RNA reads were downloaded from NCBI [67] from the following accessions: *B. puteoserpentis* - SRR3714556, *B. azoricus* - SRR3714583, and *B.* sp. 5°S - SRR3714574 [5]. Completeness of the transcriptomes was estimated throughout the curation process based on the metazoan single copy orthologous gene database (odb10) using BUSCO v3.1.0 [68]. First, we filtered contigs based on their coverage (4x) using a custom perl script and the filterbyname.sh script from BBTools after mapping with BBMap v38.34 [43]. Contigs without open reading frames were discarded after analysis with the LongOrfs and Predict function of TransDecoder v5.5.0 [69]. Using the diamond v0.8.36.98 [70] blastx command, we blasted the contigs against the NCBI-nr protein database (accessed 2019-05) [67] using the --more-senstive and -- max-target-seqs 50 options. Subsequently, we ran MEGAN v6.15.2 [71] to remove contamination and only keep contigs of the taxonomic group of Protostomia. With bowtie2 v2.2.8 [72] for mapping and subsequent analysis with Corset v1.09 [73], we clustered contigs based on shared reads and expression to remove redundant contigs. In a final step, we selected the longest isoform with a custom perl script.

Exome-wide SNP profiling

To investigate the mussel populations from the MAR based on their nuclear genome, we identified exome-wide SNPs from low-coverage sequences based on a workflow described

by Therkildsen and Palumbi (2017). In summary, we removed PCR duplicates and adapter sequences from raw reads, quality filtered (PHRED score above 2 for window of 4 bases) and merged overlapping paired-end reads. We removed potential contamination from mitochondria, *Bathymodiolus* symbionts as well as ribosomal, human, bacterial, and viral genomes. Symbiont genomes used for the filtering were previously published under project accessions PRJEB36976 at the European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/view/PRJEB36976). Mitochondrial genomes used were assembled in this study (see Methods section "Analysis of mitochondrial phylogeny"). Reads were normalised to 1.5 Gb of host reads per sample (Supplementary Figure S 1 A). This should approximate 1x coverage of the mussel genome assuming that genome size is similar to the published *B. platifrons* genome (1.64 Gb [75]). We used reformat.sh from the BBTools suite v38.34 [43] for normalisation. Normalised reads were mapped to the reference transcriptome and samples were filtered based on mapping-quality and horizontal coverage of the host transcriptome (Supplementary Figure S 1 B). We estimated genotype Bayesian posterior probabilities with ANGSD v0.911 [76] for all sites that were present in at least 50 % of the individuals, with a SNP site p-value cutoff of < 0.001, a global minor allele frequency of > 1 %, and assuming a uniform prior.

We tested different parameters to assess their influence on the resulting genotype posterior probabilities. SNP sites were filtered based on the following parameters: (1) minimum number of individuals in which the SNP site must be present (40, 50, and 60 %), (2) minor allele frequency (> 1 %, > 10 %, and > 20 %), (3) departure from Hardy-Weinberg equilibrium (HWE) (either no filtering (-1) or filtering of sites that violate HWE with a p-value less than 0.001). In addition, we assessed the influence of two different priors during genotyping (1 - estimate the posterior genotype probability based on the allele frequency as a prior and 2 - estimate the posterior genotype probability assuming a uniform prior) and two different thresholds for per base quality filtering during the data filtering (Q2 and Q30). To make sure that the observed structuring is not biased by the reference transcriptome, we ran the pipeline for all data against three different references: *B. azoricus*, *B. puteoserpentis*, and *B.* sp. 5°S.

Analysis of population structure

We used a maximum individual coverage depth of 4 in ANGSD to reduce a potential source of genotyping errors by excluding SNPs with high coverages that were probably identified from genetic repeat regions. Mapping depth retrieved with ANGSD was plotted throughout

for visual quality control with plotQC.R from ngsTools [77]. The mapping depth range was adjusted for each run with custom code to capture the main peak of the read depth distribution (see Supplementary Text 1, Supplementary Figure S 2).

We estimated pairwise genetic distances with ngsDist v1.0.2 [78] and used the script getMDS.R from M. Sikora to calculate a multidimensional scaling matrix for visualisation using the packages lattice v0.20-41 [79], methods v3.6.3 [80], optparse v1.6.6 [81], and ggplot2 v3.3.2 [82] in RStudio v1.3959. 3D plots of the MDS matrix were obtained using the packages devtools v2.3.2 [83] and rgl v0.100.54 [84]. We analysed the dataset in four subsequent steps: First, we analysed all samples from the MAR for an overview of potential clusters at species level. Secondly, we estimated the individual admixture proportions based on the genotype posterior probabilities assuming 3, 4, and 5 clusters (K = 3, 4, or 5) using NGSadmix v32 [85] and plotted them in RStudio v1.3959 with the ggplot2 package v3.3.2. [82]. We determined the affiliation of each sample to a nucDNA cluster based on the NGSadmix results using awk commands (Supplementary Text 2). Thirdly, we investigated each detected nucDNA cluster separately, regardless of the sampling location. Clustering of samples in each nucDNA cluster was evaluated running a permutational multivariate analysis of variance (PERMANOVA) to test whether samples from different mitochondrial clades/locations were also different based on their nuclear pairwise genetic distances. We performed an overall PERMANOVA using the function adonis from package vegan v2.5-7 [86] in RStudio v1.3959 with post-hoc pairwise PERMANOVA with the package pairwiseAdonis v0.0.1 [87]. P-values were corrected for multiple testing using the Bonferroni method. Lastly, we analysed all samples that fell into the same nucDNA cluster and were sampled in the same geographic region, e.g. all samples from cluster 1 that were sampled at 38°N-36°N. Performing statistical evaluation with PERMANOVA as described above, we tested whether samples from different vent fields and sites were significantly different based on their nuclear pairwise genetic differences.

We assessed whether the frequency of combinations of mitochondrial clades and nucDNA clusters correlated to geographic distance or the direction of the combination. As an example, direction south to north would be a mussel sampled at the Azores, having mitochondrial clade A (38°N-36°N) but nucDNA cluster 3 (5°S). For this analysis, we excluded samples from Broken Spur as two mitochondrial clades were present at the same site. In detail, we plotted distances between regions and the frequency of the corresponding combination of mitochondrial and nuclear genome. Additionally, we ran a pairwise comparison of

combinations with "north to south" against "south to north" direction using the Wilcoxon-test for pairwise nonparametric testing (wilcox_test from R package rstatix v0.6.0 [88]). Statistical analysis and visualisation was performed in RStudio v1.3959 using the packages ggplot2 v3.3.2 [82], tidyverse v1.3.0 [61], dplyr v1.0.2 [62], readr v1.4.0, ggpubr v0.4.0 [89], coin v1.4-0 [90], and psych v2.0.9 [91]. Plots were modified in Adobe Illustrator [57].

Symbiont binning and analysis

Raw reads were processed using BBDuk from BBTools v38.34 (options qin=33 minlen=15 qtrim=r trimq=2 ktrim=r k=23 mink=11 hdist=1) and merged using BBMerge [92]. Only if at least 10 % of the reads got merged, the merged reads were taken into account for the assembly, otherwise only the paired reads were used. Assembly was performed with metaSpades v3.12 (option -k 21,33,55,77 for read length 100, -k 21,33,55,77,99 for read length 150, -k 21,33,55,77,99,127 for read length >=250) [93]. Coverages were obtained by mapping the error corrected reads against the contigs with bwa mem v0.7.17-r1188 [94] and generating a depth file with jgi_summarize_bam_contig_depths [95]. Binning was performed with DAS Tool v1.1.2 based on phyloFlash v3.3b2, MetaBAT 2 v2.15(option -m 1500), and Maxbin v2.2.7 (option -min_contig_length 1500) [58,95–97]. CheckM v1.1.2 with the lineage_wf workflow was used for the quality assessment of all bins [98]. Only bins with at least 90 % completeness and less than 3 % contamination were retained. Taxonomic affiliation was assigned based on comparisons of average nucleotide identity (ANI) with known *Bathymodiolus* symbiont MAGs. If the SOX bin did not meet these quality criteria and was of very high coverage, we downsampled the read data to 20 %, 30 %, or 50 % of the original reads using reformat.sh from the BBTools package v38.34 and repeated the assembly and binning procedure. In some cases, this resulted in SOX bins with the required quality. For the remaining samples, we pursued co-abundance binning by including samples from the same site. Those samples were mapped against the contigs and binning was performed with DAS Tool v1.1.2 [97] based on CONCOCT v1.1.0 (option -l 1500) [99], MetaBAT2 v2.15 (option -m 1500) [95], and Maxbin v2.2.7 (option -min_contig_length 1500) [96]. Subsequent quality assessment with CheckM v1.1.2 confirmed that all samples resulted in bins with at least 90 % completeness and less than 3 % contamination.

ANI was estimated using fastANI v1.1 [100] and visualised in a heatmap with dendrogram using the packages ComplexHeatmap v2.0.0 [101] and maditr v0.7.4 [102] in RStudio v1.3959.

Data availability

Sequence data (metagenomic reads and curated transcriptome assemblies) and the metadata (recommended standard information on sequence data [103]) are available in the European Nucleotide Archive (ENA) at EMBL-EBI under project accession number PRJEB42345 (https://www.ebi.ac.uk/ena/data/view/PRJEB42345).

Results

Mitochondrial clades of mussels along the Mid-Atlantic Ridge

To investigate mitochondrial clades of *Bathymodiolus* mussels along the Mid-Atlantic Ridge (MAR, Figure 1 A), we assembled mitochondrial genomes from metagenomes. We obtained assemblies of complete mitochondrial genomes (>17 500 bp) for 95 out of 175 metagenomes and reconstructed their phylogeny. The phylogeny revealed four different clades: Mussels from the northern MAR (NMAR) grouped into Clade A and B while mussels from the southern MAR (SMAR) fell in Clade C and D (Supplementary Figure S 3). Despite their distant sampling locations, NMAR clade B and SMAR clade C grouped together with SMAR clade D as sister clade. Within the clades, there were sub-clades, especially in clade C and D, however, there was no pattern linked to the different sampling sites.

COI gene sequences are commonly used to identify animal species. Assembled COI genes, that are found on the mitochondrial genome, corresponded to the detected mitochondrial clade of the full mitochondrial assemblies in all cases (Supplementary Figure S 4). The topology, i.e. the branching structure, of the COI haplotype network and the mitochondrial phylogeny was not congruent. The clades in which the samples grouped were the same. Therefore, we used the COI gene as a proxy for mitochondrial clades as it could be retrieved for all 175 individuals in this study. Our analysis revealed a clear congruency of the geographic region and the associated mitochondrial clade, e.g. all mussels sampled at the vent fields Menez Gwen, Bubbylon, Lucky Strike, and Rainbow between 38°N and 36°N fell into mitochondrial clade A (Figure 1 B). Only exception was the hybrid zone Broken Spur where 11 out 27 mussels fell into clade A and the other 16 into clade B. Mussels from Broken Spur that fell into mitochondrial clade A were previously genotyped as hybrids between the mussels from sites north and south of Broken Spur based on 18 species-diagnostic SNP markers [20,23]. *Bathymodiolus* mussels harbour chemosynthetic symbionts that are acquired from the environment and have been shown to be geographically structured [14–17,19,20].

Similar to the mitochondrial clades, also the symbionts investigated in our study were location-specific (Supplementary Figure S 5).
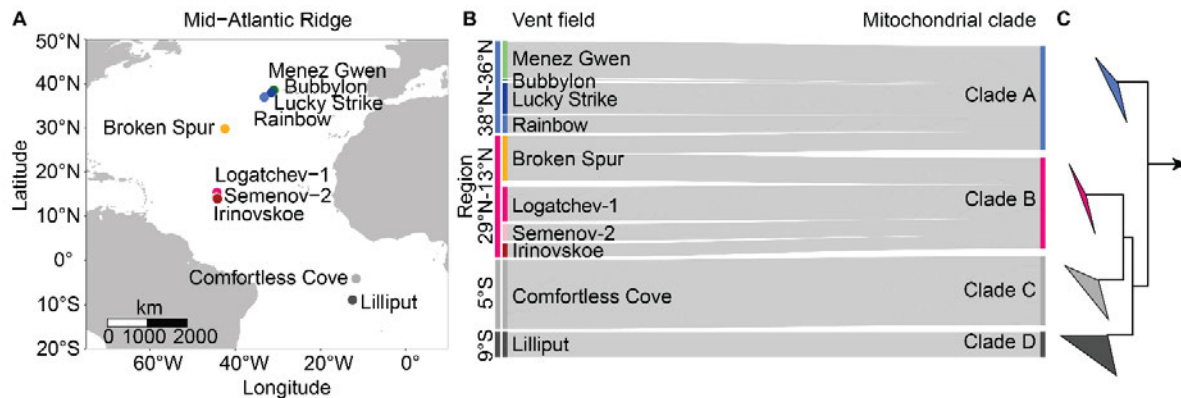
Figure 1 | Mitochondrial clades of mussels at 10 vent fields along the Mid-Atlantic Ridge. A: Map of vent fields where mussels were sampled. B: Sankey diagram depicts the relation of the mitochondrial clades detected in each mussel gill metagenome to the vent fields. Vent fields correspond to the 10 locations depicted in the map. C: Cladogram of mitochondrial clades based complete mitochondrial genomes of 95 mussels. Arrow indicates outgroup consisting of *B. septemdierum* and *B. thermophilus*. For full phylogeny, please see Supplementary Figure S 3**.** Mitochondrial clades were congruent with their region of sampling. Colours indicate the sampled regions: Blue – 38°N to 36°N, pink – 29°N to 13°N, light grey – 5°S, and dark grey – 9°S.

NucDNA clusters of mussels along the Mid-Atlantic Ridge

To ensure confident SNP profiling, we cleaned transcriptome assemblies from contamination and redundancy prior to SNP analysis, while keeping overall high completeness. As an example, the transcriptome of *B. azoricus* had a completeness of 98.5 % and duplication rate of 21.8 % before curation and a completeness of 92.9 % with a duplication rate of 0.4 % after curation (Supplementary Table S 2). The file size was reduced drastically (e.g., for *B. azoricus* from 132 MB to 28 MB) which decreased analysis runtime.

To investigate the mussel populations from the MAR based on their nuclear genome, we performed an exome-wide SNP profiling. We tested transcriptome assemblies of the different mussel species and found that the different transcriptome references had no influence on the clustering (Supplementary Figure S 6) and we chose the most complete transcriptome, *B. puteoserpentis*, for the analysis of all samples (Supplementary Table S 2). We assessed different parameters to find settings that allowed for enough resolution to see differences between individual samples but were stringent enough to observe clustering. Including SNP sites that were present in at least 50 % of the individuals and had a minimum minor allele

frequency of 1 % (Supplementary Figure S 7) yielded the best balance between resolution and stringency.

In total, 635 948 SNP sites passed all filters and were used to find similarities between 175 mussels from the MAR. Multidimensional scaling (MDS) of genetic distances obtained from the SNP profiling revealed four clusters of mussel genotypes with few samples from Broken Spur grouping between two of the clusters (Figure 2).

We ran an NGSadmix analysis based on the exome-wide SNP profiling and classified the individuals into nucDNA clusters based on the results assuming four populations (K = 4, Supplementary Figure S 8). When assuming K = 3, mussels from 5°S/Comfortless Cove were assigned to be a mix with high proportion of 9°S and a smaller proportion of the 29°N-13°N population. When assuming K = 5, the mussels from the region at 38°N-36°N were split into two populations. Thus, K = 4 best represented the clusters observed in the MDS. These clusters represent groups of mussels that had admixture proportions of at least 90 % for one of the four assumed populations and will be called "nucDNA clusters" in the following. Samples that could not be assigned to a cluster were called "mixed". In total, 12 mussels from the Broken Spur hybrid zone and one from Logatchev-1 were assigned to the mixed clusters indicating that they are hybrids.

While mitochondrial clades were congruent with their geographic region of sampling (Figure 1 B), nucDNA clusters were congruent with the regions in many but not all cases. In 18 out of 175 mussels, the assigned nucDNA cluster was not in line with the sampling location and the corresponding location-specific mitochondrial clade (Figure 3). As an example, most mussels sampled at Lilliput at 9°S on the SMAR had the location-specific mitochondrial clade D and were observed in nucDNA cluster 4. In contrast, a few mussels from 9°S also had mitochondrial clade D but were observed in nucDNA cluster 1 that mostly consisted of samples from vent fields between 38°N and 36°N.
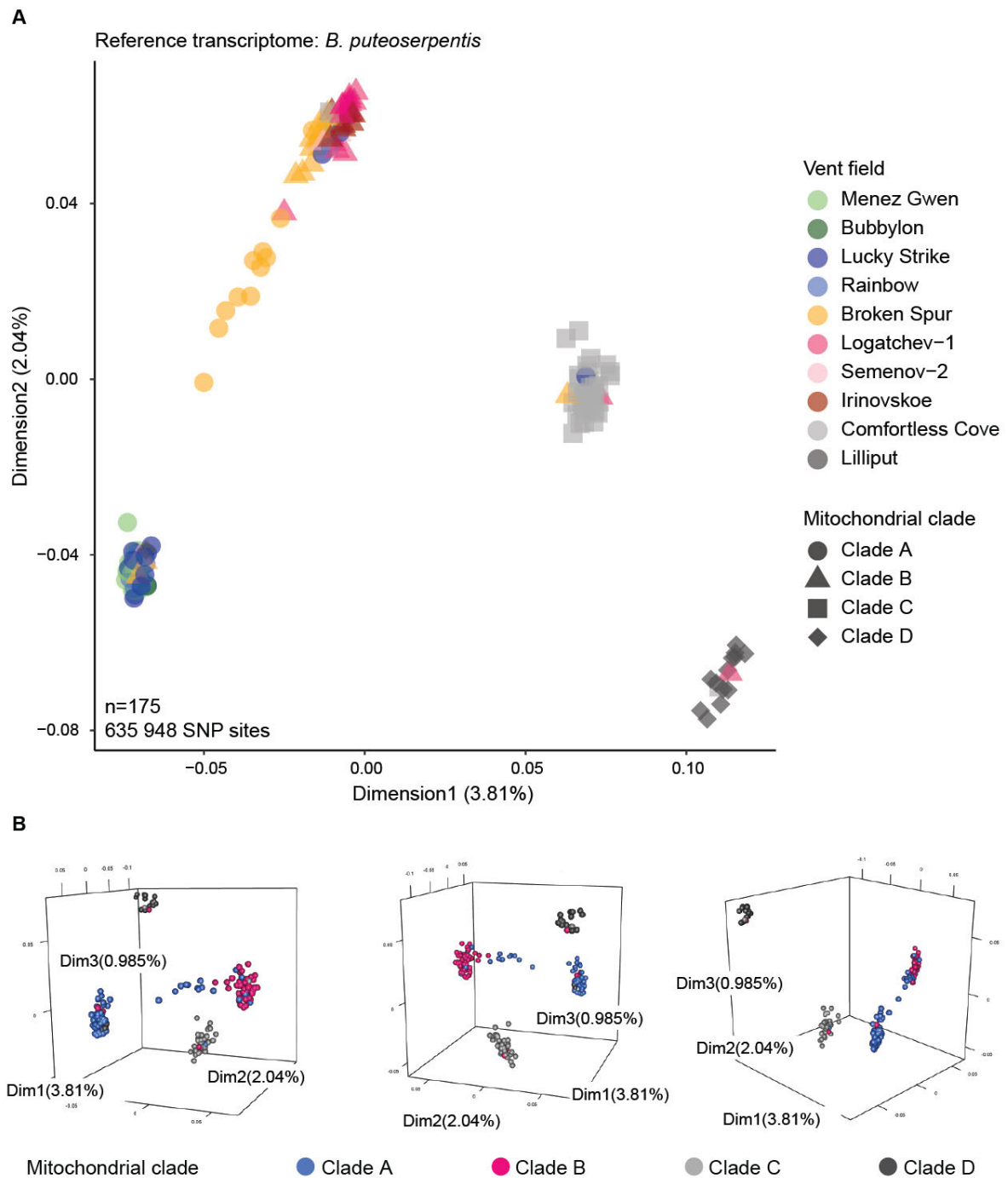
Figure 2 | nucDNA clusters of mussels from the Mid-Atlantic Ridge. A: Multidimensional scaling of genotype likelihoods based on exome-wide SNPs (635 948 SNP sites). Colours correspond to vent fields, shapes to the mitochondrial clade detected for each metagenome. Mussels from Broken Spur falling into mitochondrial clade A were previously genotyped as hybrids between the mussel species from north and south of Broken Spur based on 18 species-diagnostic SNP markers [20,23]. B: 3D MDS of genotype likelihoods based on exome-wide SNPs viewed from three different angles. Colours represent the mitochondrial clade of each sample.

We compared the frequency of combinations of mitochondrial clades and nucDNA cluster to geographic distance and did not find any pattern (Supplementary Table S 3, Supplementary Figure S 9 A). Statistical pairwise comparison of the combinations was performed considering the direction of nucDNA-mitochondria incongruences, i.e. a mussel from the Azores with clade A mitochondria (38°N-36°N) but nucDNA cluster 3 (5°S) would be south to north direction. The analysis showed no significant differences (Supplementary Figure S 9 B). Altogether, the results of our exome-wide SNP profiling revealed a discordance between the nucDNA clusters and the mitochondrial clades while mitochondrial clades and symbionts were always location specific.
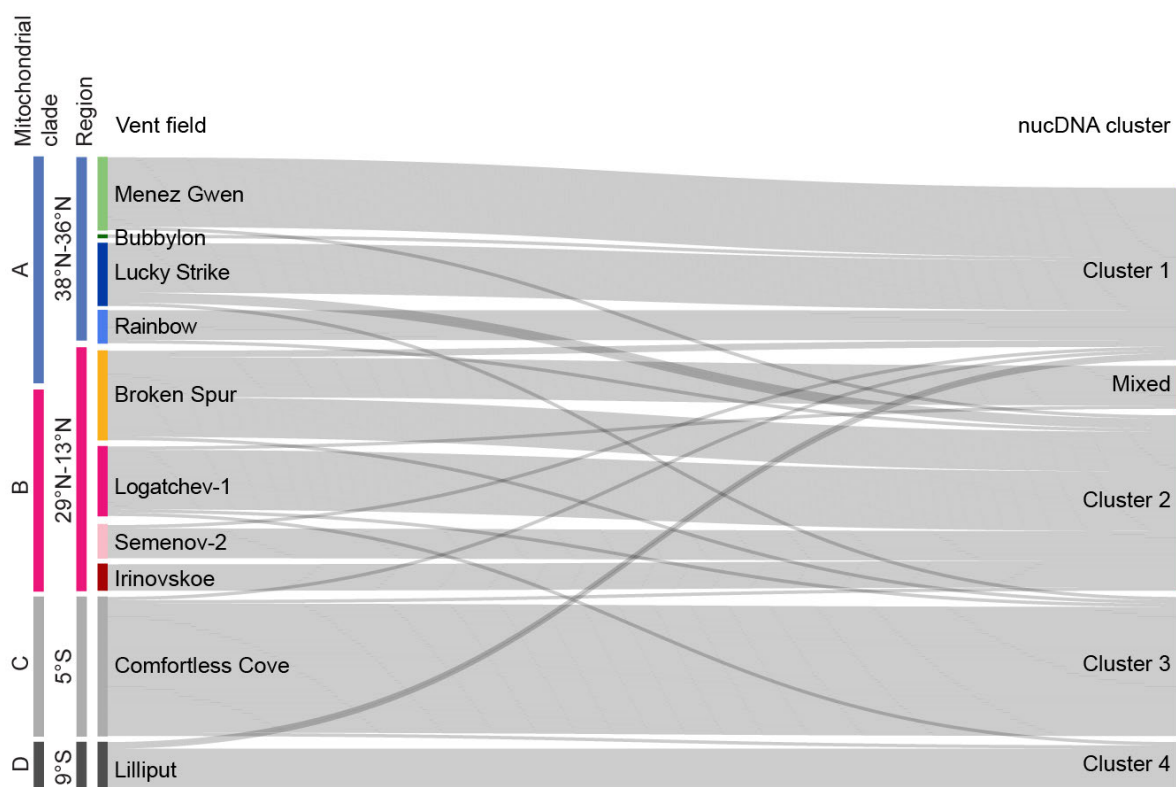
Figure 3 | nucDNA clusters of mussels along the Mid-Atlantic Ridge compared to mitochondrial clade, region and vent field. Clusters were obtained with NGSadmix assuming four clusters (K4) based on exome-wide SNP analysis and correspond to the clusters displayed in Figure 2. NGSadmix results for K3-5 are shown in Supplementary Figure S 8. Clusters were not congruent with their sampling region. Colours indicate the mitochondrial clade and sampled regions: Blue – clade A, region 38°N to 36°N, pink – clade B, region 29°N to 13°N, light grey – clade C, region 5°S, and dark grey – clade D, region 9°S.

Genetic differentiation within each nucDNA cluster

To investigate the genetic differentiation within each of the four nucDNA clusters, we performed the exome-wide SNP profiling separately for each cluster using the transcriptome reference that corresponded to the majority of samples in this cluster, e.g. *B. azoricus* transcriptome for nucDNA cluster 1. Our analysis revealed that within the nucDNA clusters, samples clustered based on their mitochondrial clade and the corresponding region (Figure 4, Supplementary Figure S 10). As an example, from all samples that were classified into nucDNA cluster 1, those that fell into mitochondrial clade A clustered tightly together based on their exome-wide SNPs while those from mitochondrial clade B, C, or D clustered on their own or with other samples from the respective mitochondrial clade and geographic region (Figure 4 A, Supplementary Figure S 10 A, B). To statistically test the observed pattern that samples from the same mitochondrial clade were also more similar on a nuclear level when compared within their assigned nucDNA cluster, we performed a PERMANOVA with pairwise post-hoc tests for all significant results (p-value < 0.05, Supplementary Table S 4 Mitochondrial clade). Samples from nucDNA cluster 1 were significantly different for mussels from mitochondrial clade A versus clade D and for clade A versus clade B (Bonferroni-corrected p-value < 0.05, Supplementary Table S 5 Mitochondrial clade). For nucDNA cluster 2, samples from mitochondrial clade A versus clade B were significantly different (Bonferroni-corrected p-value =0.003). For nucDNA clusters 3 and 4, none of the pairwise comparisons indicated significant differences. For comparisons with insignificant results, the number of samples in at least one of the groups was very low, e.g. only one mussel from mitochondrial clade C in nucDNA cluster 1. The analysis was conducted testing for mitochondrial clades, however, results were the same when tested for geographical regions as these are directly linked, except for 29°N / Broken Spur.
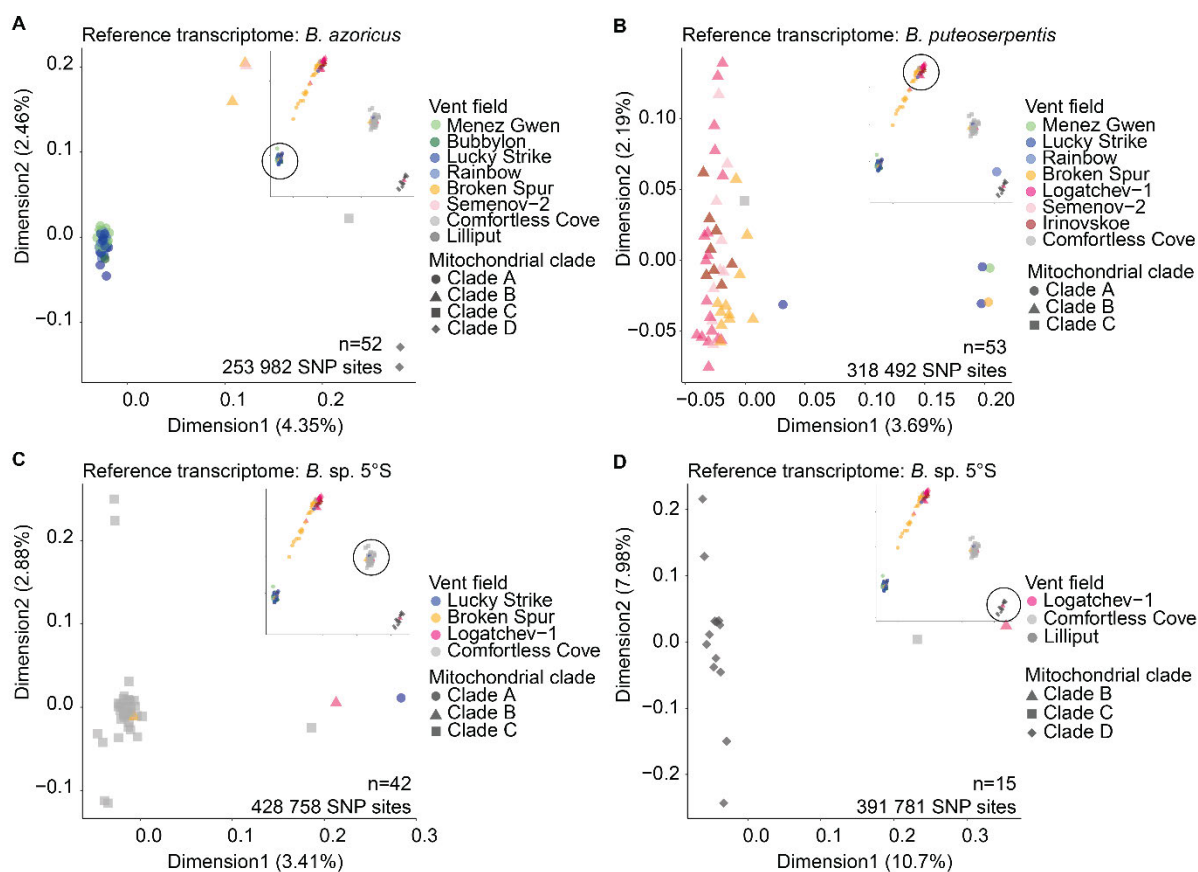
Figure 4 | Genetic differentiation within nucDNA clusters. A: nucDNA cluster 1. B: nucDNA cluster 2. C: nucDNA cluster 3. D: nucDNA cluster 4. Colours represent vent fields, shapes correspond to mitochondrial clade. Number of samples, number of SNP sites, and the respective transcriptome references are noted in the plot. Inlets represent the origin of each nucDNA cluster in the analysis of all 175 mussels from the MAR (Figure 2). Samples within each nucDNA cluster group according to their mitochondrial clade. 3D visualisations are available in Supplementary Figure S 10.

To assess whether mussels are genetically structured based on finer geographic scales, i.e. the vent field or site where they were sampled, we performed separate SNP profiling for each subset of samples that were assigned to the same nucDNA cluster and sampled in the same region (Figure 5, Supplementary Figure S 11). For example, all samples from nucDNA cluster 1 that were sampled at 38°N to 36°N were assessed based on a *B. azoricus* reference transcriptome (Figure 5 A, Supplementary Figure S 11 A). Our analyses did not show clearly separated clusters, which would have indicated a strict genetic isolation between vent fields. However, samples from some vent fields, e.g. Broken Spur in nucDNA cluster 2 (Figure 5 B, Supplementary Figure S 11 B), clustered more tightly together while samples from the other vent fields, here Logatchev-1, Semenov-2, and Irinovskoe, were rather spread out. To test whether mussels from geographically more distant vent fields have more divergent genomes,

we performed a PERMANOVA of genetic distances of groups from different vent fields but within the same nucDNA cluster with pairwise post-hoc tests for all significant results (p-value < 0.05, Supplementary Table S 4 Vent field).
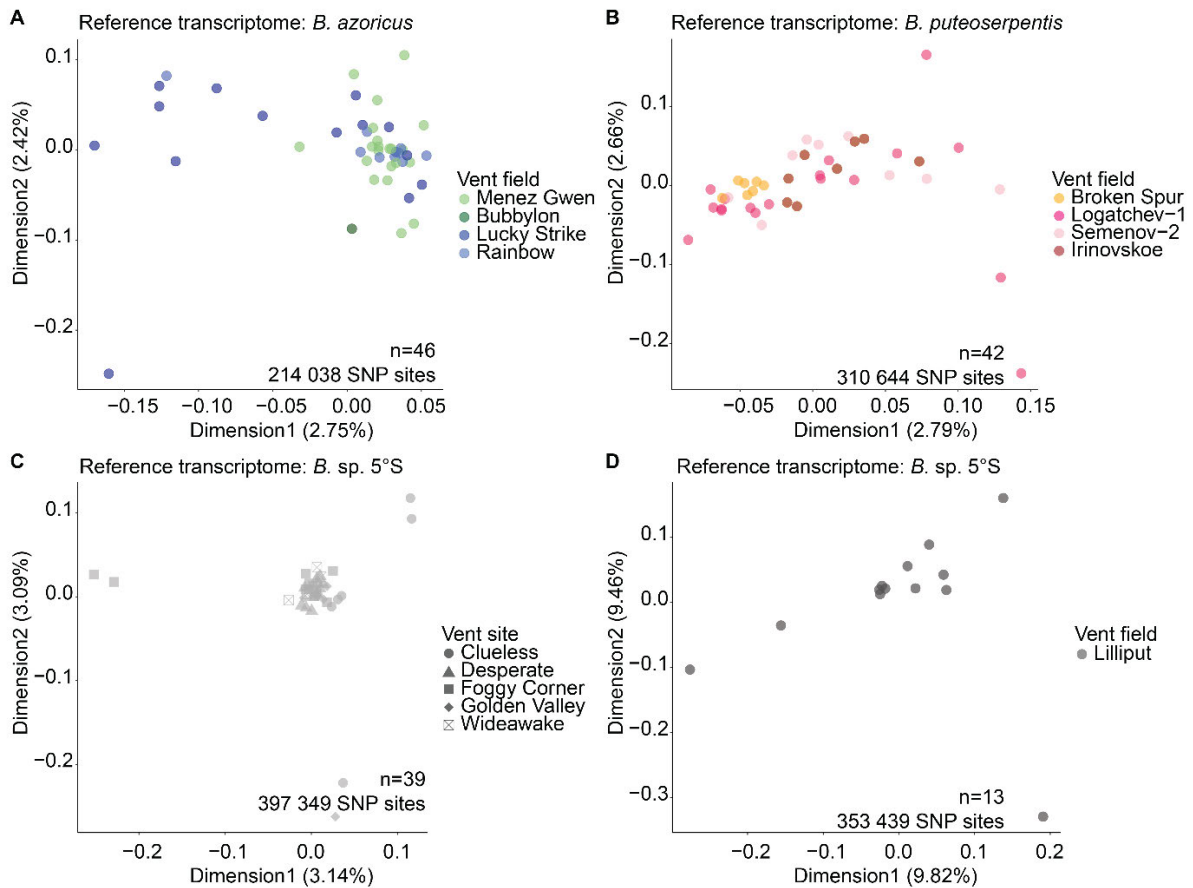
Figure 5 | Genetic differentiation within nucDNA clusters of mussels from the same geographic region. A: nucDNA cluster 1. B: nucDNA cluster 2. C: nucDNA cluster 3. D: nucDNA cluster 4. Colours represent vent fields or vent sites within a field (when all samples come from the same field, here Comfortless Cove for cluster 3). 3D visualisations are available in Supplementary Figure S 11.

Our statistical analysis only revealed significant differences based on vent field for the comparison of Broken Spur versus Irinovskoe from nucDNA cluster 2 (Bonferroni-corrected p-value = 0.006, Supplementary Table S 5 Vent field). None were detected in nucDNA cluster 1. We further tested for differences between samples from different vent sites within the same vent field, the finest geographical scale available in our dataset, and found no overall significant difference for nucDNA cluster 1 (vent fields Menez Gwen and Lucky

Strike, Supplementary Table S 5 Vent site). For nucDNA cluster 2, only mussels within vent field Semenov-2 were significantly different based on their sampling site (Bonferroni-corrected p-value = 0.036). Within nucDNA cluster 3 (vent field Comfortless Cove), comparisons of sites Clueless versus Foggy Corner, Desperate, and Wideawake showed significant differences (Bonferroni-corrected p-value < 0.05). All other comparisons between sites were not significant (Supplementary Table S 4, Supplementary Table S 5 Vent site). Overall, our results indicate that there is little to no subpopulation structure for most mussels within nucDNA clusters from the same region based on vent field and vent site.

Discussion

In our study, we investigated the population structure in 175 mussel individuals from four geographical regions at the Mid-Atlantic Ridge (MAR) covering 10 vent fields, one of which has been described as a hybrid zone. Examining more than 600 000 exome-wide SNPs sites, we detected four nuclear DNA (nucDNA) clusters that corresponded in most, but not all cases, to the geographic region of sampling. In 18 out of 175 investigated mussel individuals, the genetic affiliation did not correspond to sampling location nor to the location-specific mitochondrial clades (Supplementary Table S 6). Our finding is in line with the four previously described mussels species on the MAR but highlights the how scalability of metagenomics increases our chances to find the unexpected exceptions. Approaches such as *in silico* exome capture further expands our resolution to investigate the underlying evolutionary history of these lineages.

Comparison of species affiliation based on nucDNA clusters to the vent fields of origin revealed a surprising pattern for ~10 % of the investigated individuals: Mussels that were sampled at geographically distant vent fields, up to 5600 km apart, were observed in the same nucDNA clusters, indicating high similarity of the nuclear genome based on 600 000 exome-wide SNPs. The most plausible explanation to observe mussels from SMAR vent sites such as Lilliput in the nucDNA cluster 1 that mostly consists of mussels sampled at the Azores (38°N-36°N), is migration of mussels over very large distances.

We optimised and applied our high-resolution analysis to investigate potential subpopulation structures in conspecific mussels from different vent fields in the same geographic region and did not find a clear link between vent field and clustering based on the nuclear genome for most pairwise comparisons. In only a few cases, significant differences in the nuclear genomes of conspecific mussels from different vent fields were observed, e.g. between

mussels from Broken Spur and Irinovskoe in nucDNA cluster 2. Different factors could contribute to genetic isolation in these populations, e.g. ocean currents, other geographical barriers such as fracture zones, or the lack of 'phantom' stepping stones in between known vent fields which were suggested to enable genetic connectivity between mussel populations at the MAR [5]. Which of these factors drive the genetic differentiation between mussel populations at the MAR needs to be assessed in the future. The lack of clear subpopulation structure between vent fields and sites at a resolution of thousands of SNPs suggests a regular migration of larvae between vent fields in the same region which is in line with immigration rates modelled by Breusing and colleagues where populations mostly consisted of self-recruits from the same region [5].

Our results point towards a scenario in which mussels from the same geographic region are often highly similar based on their nuclear genome indicating a regular genetic exchange between vent fields. Few cases of mitonuclear discordance further suggest that long-distance migration is, at least occasionally, happening at the MAR (Figure 6). Given the sessile lifestyle of the adults, such migration has to happen via long distance dispersal during the planktonic larval stage. The larval dispersal distance of *Bathymodiolus* mussels at the MAR was modelled by Breusing et al. (2016) [5]. They concluded that, while the maximum median dispersal distance was around 150 km, episodic dispersal over distances of 200-400 km might happen. Even episodic events of dispersal over 400 km could not completely explain the pattern observed in our study, however, larval behaviour such as vertical migration, that was not included in their simulations, could drastically increase the potential dispersal distances [104–106]. Such long-distance dispersal could explain the genetic connectivity between NMAR and SMAR as observed by van der Heijden et al. (2012) and might be enabled through larvae migrating in the surface water as has been suggested for deep-sea limpet and snail larvae [107,108].

Mitochondrial clades corresponded to the geographic regions where mussels were sampled for all analysed individuals indicating a geographical structuring regardless of nucDNA clusters. The phylogenetic difference between mitochondrial and nuclear genomes, that we observed for 10 % of the investigated mussels, is referred to as mitonuclear (or cytonuclear) discordance and has been described in few other marine molluscs, e.g. blue mussels [109] or nudibranchs [110]. Several biological reasons but also human mistake during analysis can lead to mismatches between the nuclear and mitochondrial genome information. In our study, sampling location, mitochondrial information, and the associated sulphur-oxidising

symbionts, that were previously described to be location-specific [20], were congruent for all mussel individuals. Thus, the observed mitonuclear discordance cannot be simply explained by mistakes in the metadata. Biologically, these patterns can be caused by processes such as introgression, incomplete lineage sorting, local adaptation of distinct mtDNA lineages, and androgenesis, among others [111–113].
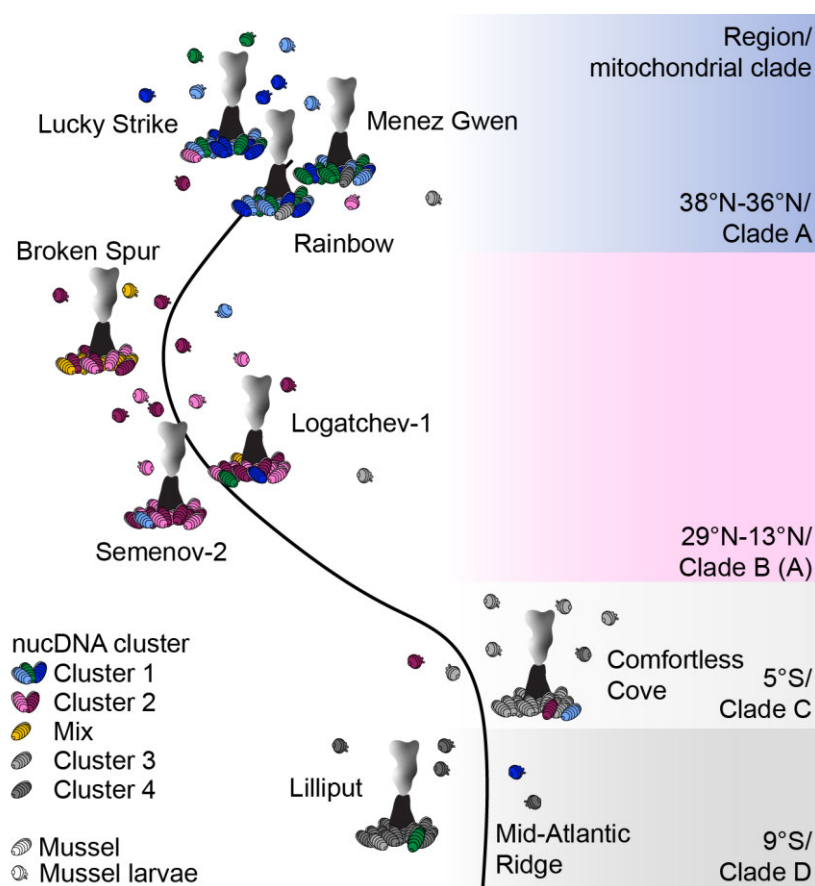
Figure 6 | Scenario of larval dispersal between geographic regions at the Mid-Atlantic Ridge. Within a region, mussels from the same nucDNA cluster (e.g. nucDNA cluster 1) that are very similar but not identical (dark blue, light blue, and green) dominate the population at this site. Larvae mostly disperse within the region of their origin, however, some disperse to other regions, e.g. blue larva at Lilliput/9°S. The region of sampling is congruent with mitochondrial clade except for Broken Spur, where not only clade B but also clade A occurred in hybrid individuals.

Based on the observed pattern alone, it is difficult to distinguish introgression, where genes or here mitochondria from one species are mixed in the population of another species through ongoing backcrossing, from incomplete lineage sorting, a process where ancestral polymorphisms are retained so that closely-related species not necessarily have the most-closely related polymorphisms of a gene [114]. Moreover, the two processes are nonexclusive causes for mitonuclear discordance. Still, mitochondrial introgression is the more likely explanation compared to incomplete lineage sorting for several reasons. First, hybridisation between the NMAR species *B. azoricus* and *B. puteoserpentis* has been described for the Broken Spur vent field [22] and all individuals identified as hybrids belonged to the mitochondrial clade of their genetically (nucDNA) less similar parent. This already indicates that asymmetric mitochondrial introgression is happening in these species. Second,

introgression has also been observed on the nuclear level in *Bathymodiolus* with backcrosses detected at many sites at the NMAR [22]. Lastly, mitochondria clearly followed biogeographic patterns as the geographic range, in which a certain mitochondrial clade was observed, was limited. Under a scenario of incomplete lineage sorting, a clear dominance of one mitochondrial clade between the different geographic ranges is unlikely [111].

We did not observe mussel individuals where both mitochondrial and nuclear genomes were different from the ones dominating at a given vent field, which would be expected for first-generation migrants from distant vent fields. One the one hand, this could by explained by step-wise long-distance migration over multiple generations. On the other hand, we would still expect the first-generation migrants in this case but at sites closer to the original location. Our results raise the question how the local mitochondria are obtained in the first place. There are at least four possible explanations for the observed pattern. One explanation are fitness disadvantages of the foreign, non-adapted mitochondrial clade at a given vent field. Mussels that fall within the non-adapted mitochondrial clade might survive long enough to proliferate and pass their nuclear genome information on to the next generation but do not persist at this vent field long-term. The following steps could have happened to lead to the observed pattern of local mitochondria but foreign nuclear genome: The descendants of migrants, i.e. first generation hybrids, regularly backcrossed with other migrants. This could have resulted in the retention of the foreign nuclear genome in a small part of the population. Local mitochondria were under positive selection at this location and introgressed asymmetrically. Therefore, also mussels with foreign nuclear genomes could have ended up having the adapted mitochondrial genome. The first-generation migrants and hybrids with non-adapted mitochondria became rare or extinct.

The second explanation comes with the fact that *Bathymodiolus* mussels harbour chemosynthetic symbionts whose presence may influence their host's mitochondrial genomes through selective sweeps or incompatibilities [115–117]. Environmentally acquired *Bathymodiolus* SOX symbionts were previously described to be geographically structured [19,20] and matched their region of sampling and the corresponding mitochondrial clade in all mussel individuals analysed in this study. If incompatibilities between the SOX symbionts and their host's mitochondria existed, this would explain breakdown of mussels with foreign mitochondrial clades, independent of their (potentially foreign) nuclear DNA. However, since the symbionts correspond to geographic regions or even vent fields, any effect related to the symbionts could not be teased apart from the impact that the geological setting or location-

dependant mitochondrial adaptation (as described above) could have. One example where interactions between bacterial partners and the mitochondria have been described to occur are *Wolbachia* infected *Drosophila* flies in which the spread of *Wolbachia* is associated with selective sweeps of the mitochondria [115]. Ivanov and colleagues investigated mitonuclear discordance in symbiotic wolf spider species and suggested that endosymbionts may be important in promoting the fixation of mitochondrial genomes in this system [118]. However, nothing is known about such processes in marine chemosynthetic symbioses and further studies are needed to specifically target the potential interaction between symbionts and processes shaping the mitochondrial genome.

The third explanation for not observing mussels with mitochondria and nuclear genomes from a foreign vent field would be unusual reproductive processes such as androgenesis. During this androgenesis, the male is the only source of nuclear genetic material while the female only provides mitochondrial DNA without nucDNA contribution [119]. Thus, this process can lead to mitonuclear discordance within one generation. Androgenesis can either happen through elimination of the maternal nuclear information by the paternal genome after the fusion of egg and sperm [120] or when sperm fertilises an egg that is lacking a nucleus [121]. Androgenesis has been demonstrated to occur in freshwater clams [122,123] and potentially carp [124], and might be associated with inter-species hybridisation [119]. To detect androgenesis in animals, the parentage has to be identified genetically or via cytological studies [121]. In addition, androgenesis was suggested based on observations such as low genetic diversity, polyploidy, and biflagellate sperm in *Corbicula* clams, the latter being proposed as a potential biomarker of androgenesis in these clams [125]. Microscopic evidence on the spermatogenesis of *B. azoricus* suggests that the spermatozoa only have one flagellum [126]. Still, to clarify whether androgenesis occurs in these species, further studies identifying the parentage of individual mussels are needed. Identification of parentage in wild *Bathymodiolus* from the deep sea is probably not feasible as it is impossible to know the parents of a given mussel to compare the genetic similarity. However, spawning experiments of mussels maintained in aquaria have been conducted in species from shallower vents or seeps such as *B. azoricus* or *"B." childressi* [24,127]. While such experiments might not be possible with the here analysed species from vents at greater water depths (*B. puteoserpentis* and mussels from 5°S and 9°S), thorough investigation of fertilisation events in *B. azoricus* might provide further insight into the reproductive processes in these mussels.

Lastly, there is some uncertainty about how representative the samples analysed in this study are for the whole population at each site. Considering that we analysed 1 to 27 mussel individuals per site, these numbers are low compared to the thousands of mussels that often occur in one mussel bed. At this point, we do not know how widespread the mitonuclear discordance is within the whole population and more importantly, we cannot exclude the presence of first-generation migrant mussels just because they were absent from our dataset. More extensive sampling and analyses of genetic differentiation of mussels might shed light on the genetic heterogeneity of individuals within one mussel bed in the future.

We observed that mitochondrial phylogeographic patterns do not necessarily match those of nuclear DNA. Our results highlight the need to treat the COI marker gene or mtDNA cautiously if used for species determination of animals, especially when the mitochondrial inheritance is complex or unresolved, or mitonuclear discordance is known for related species. In these cases, information derived from the nuclear genome might provide a higher resolution to better understand the population structure of these species.

Conclusion

Our study revealed mitonuclear discordance and genetic clustering unrelated to sampling regions in *Bathymodiolus* mussels on the MAR. The detection of mussels from the 'Azores' cluster at the southern MAR challenged our understanding of how far mussel larvae might be able to disperse in the water column, and highlighted how little we still know about reproduction and dispersal and the related evolutionary history of even the well-studied deep-sea vent populations. Intriguingly, we detected no 'first-generation migrants', i.e. mussels with both mitochondrial and nuclear genomes being incongruent with the sampling location. This opens up a plethora of exciting questions about mitochondrial inheritance, interactions between chemosynthetic endosymbionts and mitochondria, and processes underlying the mitonuclear discordance in these species that are awaiting future research.

Author Contributions

MÜ, YS, HGV, and ND designed the study. MÜ compiled the dataset, performed exome-wide SNP analysis, prepared figures and tables, submitted the sequencing data and wrote the initial draft. YS set up the exome-wide SNP pipeline, provided workflow (with input from HGV) and scripts for improvement of transcriptome references, and assisted with parameter optimisation. YS and RA supported the bioinformatic analyses and figure design. AK curated

the dataset and binned the symbionts. MF dissected Bubbylon specimen prior to DNA extraction. All authors interpreted the results. YS and RA revised the manuscript.

Acknowledgements

References

1. Vrijenhoek RC. Gene flow and genetic diversity in naturally fragmented metapopulations of deep-sea hydrothermal vent animals. J Hered. 1997; 88: 285–93.
2. Vrijenhoek RC. Genetic diversity and connectivity of deep-sea hydrothermal vent metapopulations. Mol Ecol. 2010; 19: 4391–411.
3. Juniper SK, Tunnicliffe V. Crustal accretion and the hot vent ecosystem. Phil Trans R Soc A. 1997; 355: 459–74.
4. Van Dover CL, German CR, Speer KG, Parson LM, Vrijenhoek RC. Evolution and biogeography of deep-sea vent and seep invertebrates. Science. 2002; 295: 1253–7.
5. Breusing C, Biastoch A, Drews A, Metaxas A, Jollivet D, Vrijenhoek RC, et al. Biophysical and population genetic models predict the presence of "phantom" stepping stones connecting Mid-Atlantic Ridge vent ecosystems. Curr Biol. 2016; 26: 2257–67.
6. Breusing C, Johnson SB, Tunnicliffe V, Vrijenhoek RC. Population structure and connectivity in Indo-Pacific deep-sea mussels of the *Bathymodiolus septemdierum* complex. Conserv Genet. 2015; 16: 1415–30.
7. van der Heijden K, Petersen JM, Dubilier N, Borowski C. Genetic connectivity between north and south Mid-Atlantic Ridge chemosynthetic bivalves and their symbionts. PLOS ONE. 2012; 7: e39994.
8. Craddock C, Hoeh WR, Lutz RA, Vrijenhoek RC. Extensive gene flow among mytilid (*Bathymodiolus thermophilus*) populations from hydrothermal vents of the eastern Pacific. Mar Biol. 1995; 124: 137–46.
9. Cowen RK, Sponaugle S. Larval dispersal and marine population connectivity. Ann Rev Mar Sci. 2009; 1: 443–66.
10. Mullineaux LS, Adams DK, Mills SW, Beaulieu SE. Larvae from afar colonize deep-sea hydrothermal vents after a catastrophic eruption. PNAS. 2010; 107: 7829–34.
11. Mullineaux LS, Metaxas A, Beaulieu SE, Bright M, Gollner S, Grupe BM, et al. Exploring the ecology of deep-sea hydrothermal vents in a metacommunity framework. Front Mar Sci. 2018; 5: 49.
12. Dubilier N, Bergin C, Lott C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. Nat Rev Microbiol. 2008; 6: 725–40.
13. Franke M, Geier B, Hammel JU, Dubilier N, Leisch N. Becoming symbiotic – the symbiont acquisition and the early development of bathymodiolin mussels. BioRxiv. 2020: 2020.10.09.333211.
14. DeChaine EG, Cavanaugh CM. Symbioses of methanotrophs and deep-sea mussels (Mytilidae: Bathymodiolinae). Prog Mol Subcell Biol. 2006; 41: 227–49.
15. Won Y-J, Hallam SJ, O'Mullan GD, Pan IL, Buck KR, Vrijenhoek RC. Environmental acquisition of thiotrophic endosymbionts by deep-sea mussels of the genus *Bathymodiolus*. Appl Environ Microbiol. 2003; 69: 6785–92.
16. Le Pennec M, Beninger PG. Ultrastructural characteristics of spermatogenesis in three species of deep-sea hydrothermal vent mytilids. Canadian Journal of Zoology. 1997; 75: 308–16. http://www.nrcresearchpress.com/doi/abs/10.1139/z97-039 (accessed February 17, 2017).
17. Fontanez KM, Cavanaugh CM. Evidence for horizontal transmission from multilocus phylogeny of deep-sea mussel (Mytilidae) symbionts. Environ Microbiol. 2014; 16: 3608–21.
18. Wentrup C, Wendeberg A, Schimak M, Borowski C, Dubilier N. Forever competent: deep-sea bivalves are colonized by their chemosynthetic symbionts throughout their lifetime. Environ Microbiol. 2014; 16: 3699–713.

19.   Ho P-T, Park E, Hong SG, Kim E-H, Kim K, Jang S-J, et al. Geographical structure of endosymbiotic bacteria hosted by *Bathymodiolus* mussels at eastern Pacific hydrothermal vents. BMC Evol Biol. 2017; 17: 121.

20.   Ücker M, Ansorge R, Sato Y, Sayavedra L, Breusing C, Dubilier N. Deep-sea mussels from a hybrid zone on the Mid-Atlantic Ridge host genetically indistinguishable symbionts. ISME J. 2021; 1–8.

21.   Von Cosel R, Comtet T, Krylova EM. *Bathymodiolus* (Bivalvia: Mytilidae) from hydrothermal vents on the Azores Triple Junction and the Logatchev hydrothermal field, Mid-Atlantic Ridge. Veliger. 1999; 42: 218–48.

22.   O'Mullan GD, Maas PA, Lutz RA, Vrijenhoek RC. A hybrid zone between hydrothermal vent mussels (Bivalvia: Mytilidae) from the Mid-Atlantic Ridge. Mol Ecol. 2001; 10: 2819–31.

23.   Breusing C, Vrijenhoek RC, Reusch TBH. Widespread introgression in deep-sea hydrothermal vent mussels. BMC Evol Biol. 2017; 17: 13.

24.   Arellano SM, Young CM. Spawning, development, and the duration of larval life in a deep-sea cold-seep mussel. Biol Bull. 2009; 216: 149–62.

25.   Li H, Qu W, Obrycki JJ, Meng L, Zhou X, Chu D, et al. Optimizing sample size for population genomic study in a global invasive lady beetle, *Harmonia Axyridis*. Insects. 2020; 11: 290.

26.   Faure B, Jollivet D, Tanguy A, Bonhomme F, Bierne N. Speciation in the deep sea: multi-locus analysis of divergence and gene flow between two hybridizing species of hydrothermal vent mussels. PLOS ONE. 2009; 4: e6485.

27.   Carney SL, Formica MI, Divatia H, Nelson K, Fisher CR, Schaeffer SW. Population structure of the mussel *"Bathymodiolus" childressi* from Gulf of Mexico hydrocarbon seeps. Deep Sea Res Part I Oceanogr Res Pap. 2006; 53: 1061–72.

28.   Kyuno A, Shintaku M, Fujita Y, Matsumoto H, Utsumi M, Watanabe H, et al. Dispersal and differentiation of deep-sea mussels of the genus *Bathymodiolus* (Mytilidae, Bathymodiolinae). J Mar Biol. 2009; 2009: e625672.

29.   Miyazaki J-I, Shintaku M, Kyuno A, Fujiwara Y, Hashimoto J, Iwasaki H. Phylogenetic relationships of deep-sea mussels of the genus *Bathymodiolus* (Bivalvia: Mytilidae). Mar Biol. 2004; 144: 527–35.

30.   Smith PJ, McVeagh SM, Won Y, Vrijenhoek RC. Genetic heterogeneity among New Zealand species of hydrothermal vent mussels (Mytilidae: *Bathymodiolus*). Mar Biol. 2004; 144: 537–45.

31.   Won Y-J, Young CR, Lutz RA, Vrijenhoek RC. Dispersal barriers and isolation among deep-sea mussel populations (Mytilidae: *Bathymodiolus*) from eastern Pacific hydrothermal vents. Mol Ecol. 2003; 12: 169–84.

32.   Morin PA, Luikart G, Wayne RK, the SNP workshop group. SNPs in ecology, evolution and conservation. Trends Ecol Evol. 2004; 19: 208–16.

33.   Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. Proc R Soc B. 2003; 270: 313–21.

34.   Skibinski DOF, Gallagher C, Beynon CM. Mitochondrial DNA inheritance. Nature. 1994; 368: 817–8.

35.   Zouros E, Ball AO, Saavedra C, Freeman KR. Mitochondrial DNA inheritance. Nature. 1994; 368: 818.

36.   Skibinski DOF, Gallagher C, Beynon CM. Sex-limited mitochondrial DNA transmission in the marine mussel *Mytilus edulis*. Genetics. 1994; 138: 801–9.

37.   Zouros E, Ball AO, Saavedra C, Freeman KR. An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. PNAS. 1994; 91: 7463–7.

38. Ansorge R, Romano S, Sayavedra L, Rubin-Blum M, Gruber-Vodicka HR, Scilipoti S, et al. The hidden pangenome: comparative genomics reveals pervasive diversity in symbiotic and free-living sulfur-oxidizing bacteria. BioRxiv. 2020: 2020.12.11.421487.

39. Sayavedra L. Host-symbiont interactions and metabolism of chemosynthetic symbiosis in deep-sea *Bathymodiolus* mussels. PhD Thesis. University of Bremen, 2016.

40. Xu H, Luo X, Qian J, Pang X, Song J, Qian G, et al. FastUniq: a fast *de novo* duplicates removal tool for paired short reads. PLOS ONE. 2012; 7: e52249.

41. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30: 2114–20.

42. Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads - a baiting and iterative mapping approach. Nucleic Acids Res. 2013; 41: e129.

43. Bushnell B. BBMap. 2014. http://sourceforge.net/projects/bbmap/ (accessed October 21, 2016).

44. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012; 19: 455–77.

45. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of *de novo* genome assemblies. Bioinformatics. 2015; 31: 3350–2.

46. Ayad LAK, Pissis SP. MARS: improving multiple circular sequence alignment using refined sequences. BMC Genomics. 2017; 18: 86.

47. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014; 30: 2068–9.

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

49. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013; 30: 772–80.

50. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002; 30: 3059–66.

51. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015; 32: 268–74.

52. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003; 52: 696–704.

53. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013; 30: 1188–95.

54. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993; 10: 512–26.

55. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017; 14: 587–9.

56. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019; 47: W256–9.

57. Adobe. Adobe Illustrator. 2020. https://www.adobe.com/de/products/illustrator.html (accessed February 3, 2020).

58. Gruber-Vodicka HR, Seah BKB, Pruesse E. phyloFlash: rapid small-subunit rRNA profiling and targeted assembly from metagenomes. MSystems. 2020; 5: e00920-20.

59. Leigh JW, Bryant D. popart: full-feature software for haplotype network construction. Methods Ecol Evol. 2015; 6: 1110–6.

60. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999; 16: 37–48.

61. Wickham H, RStudio. tidyverse: easily install and load the "Tidyverse." 2019. https://CRAN.R-project.org/package=tidyverse (accessed July 27, 2020).

62. Wickham H, François R, Henry L, Müller K, RStudio. dplyr: a grammar of data manipulation. 2020. https://CRAN.R-project.org/package=dplyr (accessed December 10, 2020).

63. Vaidyanathan R, Xie Y, Allaire JJ, Cheng J, Sievert C, Russell K, et al. htmlwidgets: HTML widgets for R. 2020. https://CRAN.R-project.org/package=htmlwidgets (accessed December 10, 2020).

64. Allaire JJ, Ellis P, Gandrud C, Kuo K, Lewis BW, Owen J, et al. networkD3: D3 JavaScript network graphs from R. 2017. https://CRAN.R-project.org/package=networkD3 (accessed December 10, 2020).

65. RStudio Team. RStudio: integrated development for R. Boston, MA: RStudio, Inc.; 2015. http://www.rstudio.com/ (accessed December 19, 2016).

66. De Wit P, Pespeni MH, Palumbi SR. SNP genotyping and population genomics from expressed sequences – current advances and future possibilities. Mol Ecol. 2015; 24: 2310–23.

67. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2016; 44: D7–19.

68. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015; 31: 3210–2.

69. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013; 8: 1494–512.

70. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Meth. 2015; 12: 59–60.

71. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, et al. MEGAN Community Edition - interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Comput Biol. 2016; 12: e1004957.

72. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9: 357–9.

73. Davidson NM, Oshlack A. Corset: enabling differential gene expression analysis for *de novo* assembled transcriptomes. Genome Biol. 2014; 15: 410.

74. Therkildsen NO, Palumbi SR. Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. Mol Ecol Resour. 2017; 17: 194–208.

75. Sun J, Zhang Y, Xu T, Zhang Y, Mu H, Zhang Y, et al. Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. Nature Ecology & Evolution. 2017; 1: 1–7.

76. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of next generation sequencing data. BMC Bioinformatics. 2014; 15: 356.

77. Fumagalli M, Vieira FG, Linderoth T, Nielsen R. ngsTools: methods for population genetics analyses from next-generation sequencing data. Bioinformatics. 2014; 30: 1486–7.

78.   Vieira FG, Lassalle F, Korneliussen TS, Fumagalli M. Improving the estimation of genetic distances from next-generation sequencing data. Biol J Linn Soc Lond. 2016; 117: 139–49.

79.   Sarkar D. Lattice: multivariate data visualization with R. New York: Springer; 2008.

80.   R Team. methods: formal methods and classes. n.d. https://www.rdocumentation.org/packages/methods/versions/3.6.2 (accessed December 10, 2020).

81.   Davis TL, Day A, Python Software Foundation, Lianoglou S, Nikelski J, Müller K, et al. optparse: Command line option parser. 2020. https://CRAN.R-project.org/package=optparse (accessed December 10, 2020).

82.   Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, et al. ggplot2: create elegant data visualisations using the grammar of graphics. 2019. https://CRAN.R-project.org/package=ggplot2 (accessed November 25, 2019).

83.   Wickham H, Hester J, Chang W, RStudio, R Core team. devtools: tools to make developing R packages easier. 2020. https://CRAN.R-project.org/package=devtools (accessed December 10, 2020).

84.   Adler D, Murdoch D. rgl: 3D visualization using OpenGL. 2020. https://CRAN.R-project.org/package=rgl (accessed December 10, 2020).

85.   Skotte L, Korneliussen TS, Albrechtsen A. Estimating individual admixture proportions from next generation sequencing data. Genetics. 2013; 195: 693–702.

86.   Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community ecology package. 2019. https://CRAN.R-project.org/package=vegan (accessed November 25, 2019).

87.   Arbizu PM. pairwiseAdonis. 2020. https://github.com/pmartinezarbizu/pairwiseAdonis (accessed December 10, 2020).

88.   Kassambara A. rstatix: pipe-friendly framework for basic statistical tests. 2020. https://CRAN.R-project.org/package=rstatix (accessed January 22, 2021).

89.   Kassambara A. ggpubr: "ggplot2" based publication ready plots. 2020. https://CRAN.R-project.org/package=ggpubr (accessed January 22, 2021).

90.   Hothorn T, Winell H, Hornik K, van de Wiel MA, Zeileis A. coin: conditional inference procedures in a permutation test framework. 2021. https://CRAN.R-project.org/package=coin (accessed January 22, 2021).

91.   Revelle W. psych: Procedures for psychological, psychometric, and personality research. 2020. https://CRAN.R-project.org/package=psych (accessed January 22, 2021).

92.   Bushnell B, Rood J, Singer E. BBMerge – accurate paired shotgun read merging via overlap. PLOS ONE. 2017; 12: e0185056.

93.   Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017; 27: 824–34.

94.   Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv:13033997 [q-Bio]. 2013. http://arxiv.org/abs/1303.3997 (accessed January 8, 2021).

95.   Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019; 7: e7359.

96.   Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome. 2014; 2: 26.

97. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol. 2018; 3: 836–43.

98. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015; 25: 1043–55.

99. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nat Meth. 2014; 11: 1144–6.

100. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018; 9: 5114.

101. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016; 32: 2847–9.

102. Demin G. maditr: fast data aggregation, modification, and filtering with pipes and "data.table." 2019. https://CRAN.R-project.org/package=maditr (accessed November 25, 2019).

103. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, et al. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat Biotech. 2011; 29: 415–20.

104. Hilário A, Metaxas A, Gaudron SM, Howell KL, Mercier A, Mestre NC, et al. Estimating dispersal distance in the deep sea: Challenges and applications to marine reserves. Front Mar Sci. 2015; 2: 6.

105. Metaxas A, Saunders M. Quantifying the "bio-"components in biophysical models of larval transport in marine benthic invertebrates: Advances and pitfalls. Biol Bull. 2009; 216: 257–72.

106. Cowen RK, Paris CB, Srinivasan A. Scaling of connectivity in marine populations. Science. 2006; 311: 522–7.

107. Yahagi T, Kayama Watanabe H, Kojima S, Kano Y. Do larvae from deep-sea hydrothermal vents disperse in surface waters? Ecology. 2017; 98: 1524–34.

108. Van Gaest AL. Ecology and early life history of *Bathynerita naticoidea*: Evidence for long-distance larval dispersal of a cold seep gastropod. Thesis. University of Oregon, 2006.

109. Quesada H, Wenne R, Skibinski DO. Interspecies transfer of female mitochondrial DNA is coupled with role-reversals and departure from neutrality in the mussel *Mytilus trossulus*. Mol Biol Evol. 1999; 16: 655–65.

110. Layton KKS, Carvajal JI, Wilson NG. Mimicry and mitonuclear discordance in nudibranchs: new insights from exon capture phylogenomics. Ecol Evol. 2020; 10: 11966–82.

111. Toews DPL, Brelsford A. The biogeography of mitochondrial and nuclear discordance in animals. Molecular Ecology. 2012; 21: 3907–30.

112. Pavlova A, Amos JN, Joseph L, Loynes K, Austin JJ, Keogh JS, et al. Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an australian bird. Evolution. 2013; 67: 3412–28.

113. Hedtke SM, Hillis DM. The potential role of androgenesis in cytoplasmic–nuclear phylogenetic discordance. Syst Biol. 2011; 60: 87–96.

114. Twyford AD, Ennos RA. Next-generation hybridization and introgression. Heredity. 2012; 108: 179–89.

115. Jiggins FM. Male-killing *Wolbachia* and mitochondrial DNA: selective sweeps, hybrid introgression and parasite population dynamics. Genetics. 2003; 164: 5–12.

116. Rousset F, Solignac M. Evolution of single and double *Wolbachia* symbioses during speciation in the *Drosophila simulans* complex. PNAS. 1995; 92: 6389–93.
117. Montchamp-Moreau C, Ferveur JF, Jacques M. Geographic distribution and inheritance of three cytoplasmic incompatibility types in *Drosophila simulans*. Genetics. 1991; 129: 399–407.
118. Ivanov V, Lee KM, Mutanen M. Mitonuclear discordance in wolf spiders: genomic evidence for species integrity and introgression. Mol Ecol. 2018; 27: 1681–95.
119. Schwander T, Oldroyd BP. Androgenesis: where males hijack eggs to clone themselves. Philos Trans R Soc Lond B Biol Sci. 2016; 371: 20150534.
120. Rieger R, Michaelis A, Green MM. A glossary of genetics and cytogenetics: classical and molecular. 3rd ed. Berlin Heidelberg: Springer-Verlag; 1968.
121. Pigneur L-M, Hedtke SM, Etoundi E, Van Doninck K. Androgenesis: a review through the study of the selfish shellfish *Corbicula* spp. Heredity. 2012; 108: 581–91.
122. Komaru A, Kawagishi T, Konishi K. Cytological evidence of spontaneous androgenesis in the freshwater clam *Corbicula leana* Prime. Dev Gene Evol. 1998; 208: 46–50.
123. Ishibashi R, Ookubo K, Aoki M, Utaki M, Komaru A, Kawamura K. Androgenetic reproduction in a freshwater diploid clam *Corbicula fluminea* (Bivalvia: Corbiculidae). Zoolog Sci. 2003; 20: 727–32.
124. Schmidt DJ, Bond NR, Adams M, Hughes JM. Cytonuclear evidence for hybridogenetic reproduction in natural populations of the Australian carp gudgeon (Hypseleotris: Eleotridae). Mol Ecol. 2011; 20: 3367–80.
125. Korniushin AV. A revision of some Asian and African freshwater clams assigned to *Corbicula fluminalis* (Müller, 1774) (Mollusca: Bivalvia: Corbiculidae), with a review of anatomical characters and reproductive features based on museum collections. Hydrobiologia. 2004; 529: 251–70.
126. Kádár E, Lobo-da-Cunha A, Santos RS, Dando P. Spermatogenesis of *Bathymodiolus azoricus* in captivity matching reproductive behaviour at deep-sea hydrothermal vents. J Exp Mar Biol Ecol. 2006; 335: 19–26.
127. Colaço A, Martins I, Laranjo M, Pires L, Leal C, Prieto C, et al. Annual spawning of the hydrothermal vent mussel, *Bathymodiolus azoricus*, under controlled aquarium conditions at atmospheric pressure. J Exp Mar Biol Ecol. 2006; 333: 166–71.

Supplementary Information

Mitonuclear discordance suggests long-distance migration and mitochondrial introgression of deep-sea mussels along the Mid-Atlantic Ridge

Merle Ücker[1,2], Yui Sato[1], Rebecca Ansorge[1,3], Anne Kupczok[1,4], Maximilian Franke[1], Harald Gruber-Vodicka[1], Nicole Dubilier[1,2]

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

[2]MARUM – Center for Marine Environmental Sciences of the University of Bremen, Bremen, Germany

[3]Quadram Institute Bioscience, Norwich, Norfolk, United Kingdom

[4]Wageningen University & Research, Wageningen, Netherlands

Includes

Supplementary Text

1. Code to determine read depth distribution
2. Code to classify individuals based on NGSadmix proportions

Supplementary Figures

Supplementary Tables

Supplementary Text

1. Code to determine read depth distribution

```
echo "calculating mindepth and maxdepth"

#Get the min and max depth values from the globalDepth files

depthfile=$bamdir/$reffile/$(cat $sampleinfo | wc -
l)Samples_to_${reffile}_MinQ20_qc.depthGlobal

fmt -1 $depthfile > t_$(basename $depthfile)

head -n -1 t_$(basename $depthfile) > depthGlobal

max=$(sort -rn depthGlobal | head -n 1)

line_number=$(grep -n $max depthGlobal | perl -pe 's/\:\d+//')

min_counts=$(awk "NR >= 0 && NR <= $line_number" depthGlobal | grep -v -Fw
'0' | sort -n | head -n 1)

max_counts=$(awk "NR >= $line_number && NR <= $(wc depthGlobal| perl -pe
's/(\d+)\s.+/$1/')" depthGlobal | awk -v min="$min_counts" '$1<min' | sort
-rn | head -n 1)

mindepth=$(grep -n $min_counts depthGlobal | perl -pe 's/\:\d+//')

maxdepth=$(grep -n $max_counts depthGlobal | perl -pe 's/\:\d+//')

echo "done with mindepth and maxdepth"

echo " "
```

2. Code to classify individuals based on NGSadmix proportions

For K = 3

```
awk '{if ($1 > 0.9)
print "Pop1"
else if ($2 > 0.9)
print "Pop2"
else if ($3 > 0.9)
print "Pop3"
else
print "Hybrid"}' 175Samples_to_Bput_clean_MinQ20_SNPpval1e-3_Prior2_K3.qopt
```

For K = 4

```
awk '{if ($1 > 0.9)
print "Pop1"
else if ($2 > 0.9)
print "Pop2"
else if ($3 > 0.9)
print "Pop3"
else if ($4 > 0.9)
```

```
print "Pop4"
else
print "Hybrid"}' 175Samples_to_Bput_clean_MinQ20_SNPpval1e-3_Prior2_K4.qopt
```

For K = 5

```
awk '{if ($1 > 0.9)
print "Pop1"
else if ($2 > 0.9)
print "Pop2"
else if ($3 > 0.9)
print "Pop3"
else if ($4 > 0.9)
print "Pop4"
else if ($5 > 0.9)
print "Pop5"
else
print "Hybrid"}' 175Samples_to_Bput_clean_MinQ20_SNPpval1e-3_Prior2_K5.qopt
```

For K = 6

```
awk '{if ($1 > 0.9)
print "Pop1"
else if ($2 > 0.9)
print "Pop2"
else if ($3 > 0.9)
print "Pop3"
else if ($4 > 0.9)
print "Pop4"
else if ($5 > 0.9)
print "Pop5"
else if ($6 > 0.9)
print "Pop6"
else
print "Hybrid"}' 175Samples_to_Bput_clean_MinQ20_SNPpval1e-3_Prior2_K6.qopt
```

Supplementary Figures
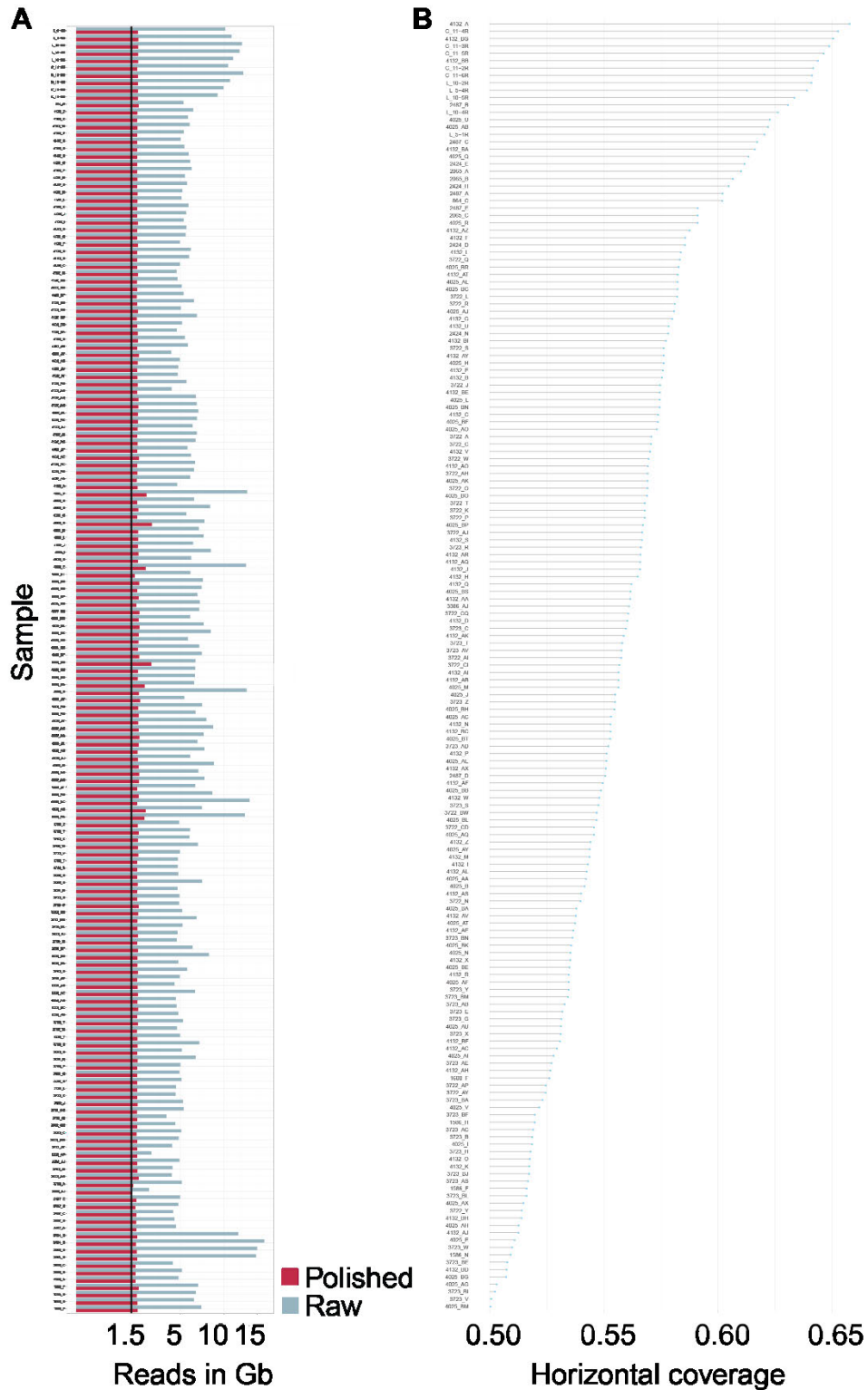
Supplementary Figure S 1 | Quality control of samples included in the dataset.
A: Total host coverage of sequencing reads. Target depth of 1.5 Gb is shown with a black line. Blue: Raw reads prior to data curation, Red: "Polished" reads after data curation. B: Horizontal coverage across the mussel transcriptome. Samples with a horizontal coverage above 50 % were found to deliver reproducible results (data not shown), thus this was chosen as a cutoff.
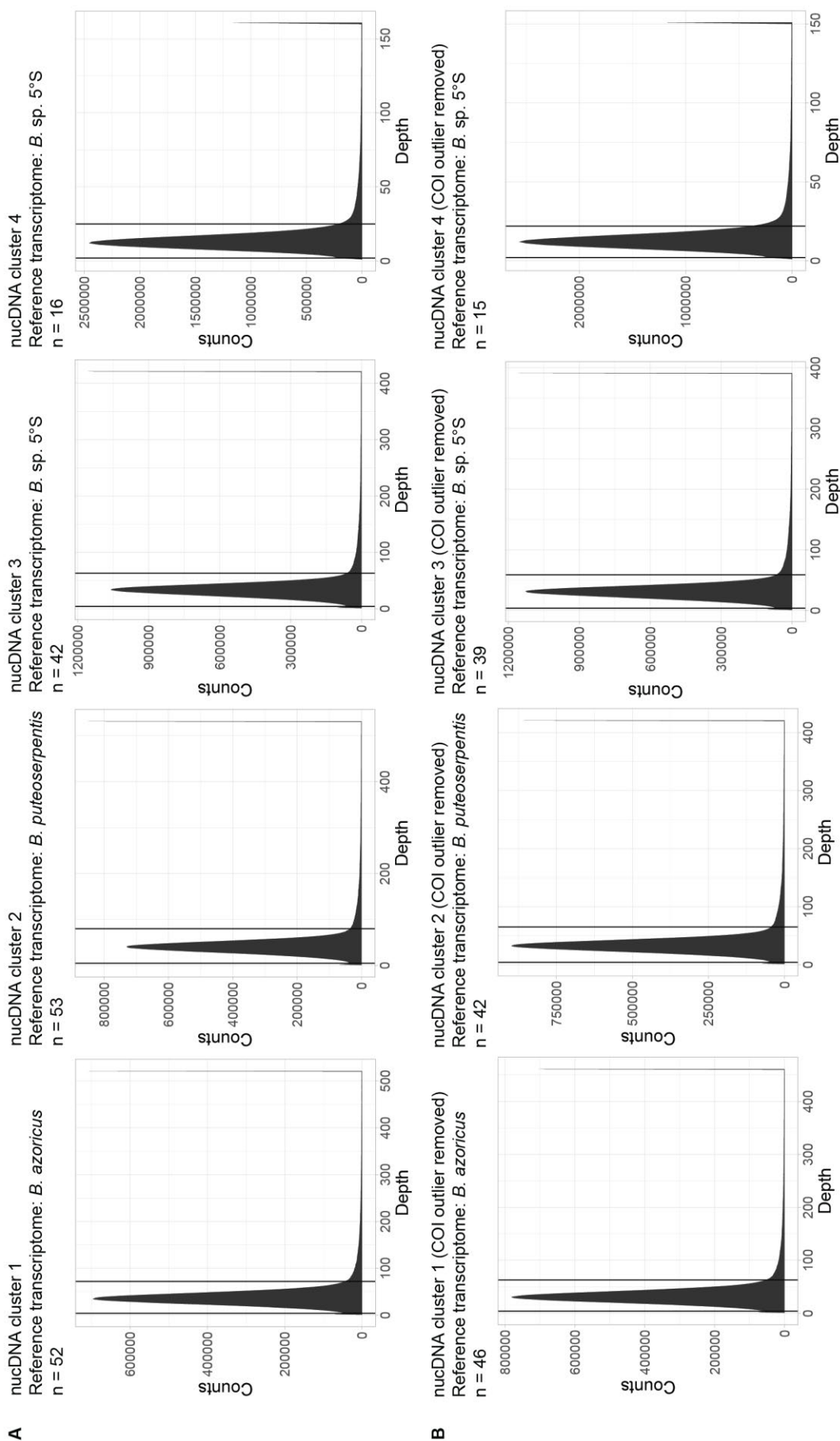
Supplementary Figure S 2 | Read depth distribution. A: Read depth distributions for nucDNA clusters 1 to 4 against their respective reference transcriptome. **B:** Read depth distributions for nucDNA clusters 1 to 4 from the same geographic region (COI/mitochondrial clade outliers were removed) against their respective reference transcriptome. Black lines illustrate the distribution peak that was automatically chosen with custom code (Supplementary Text 1).
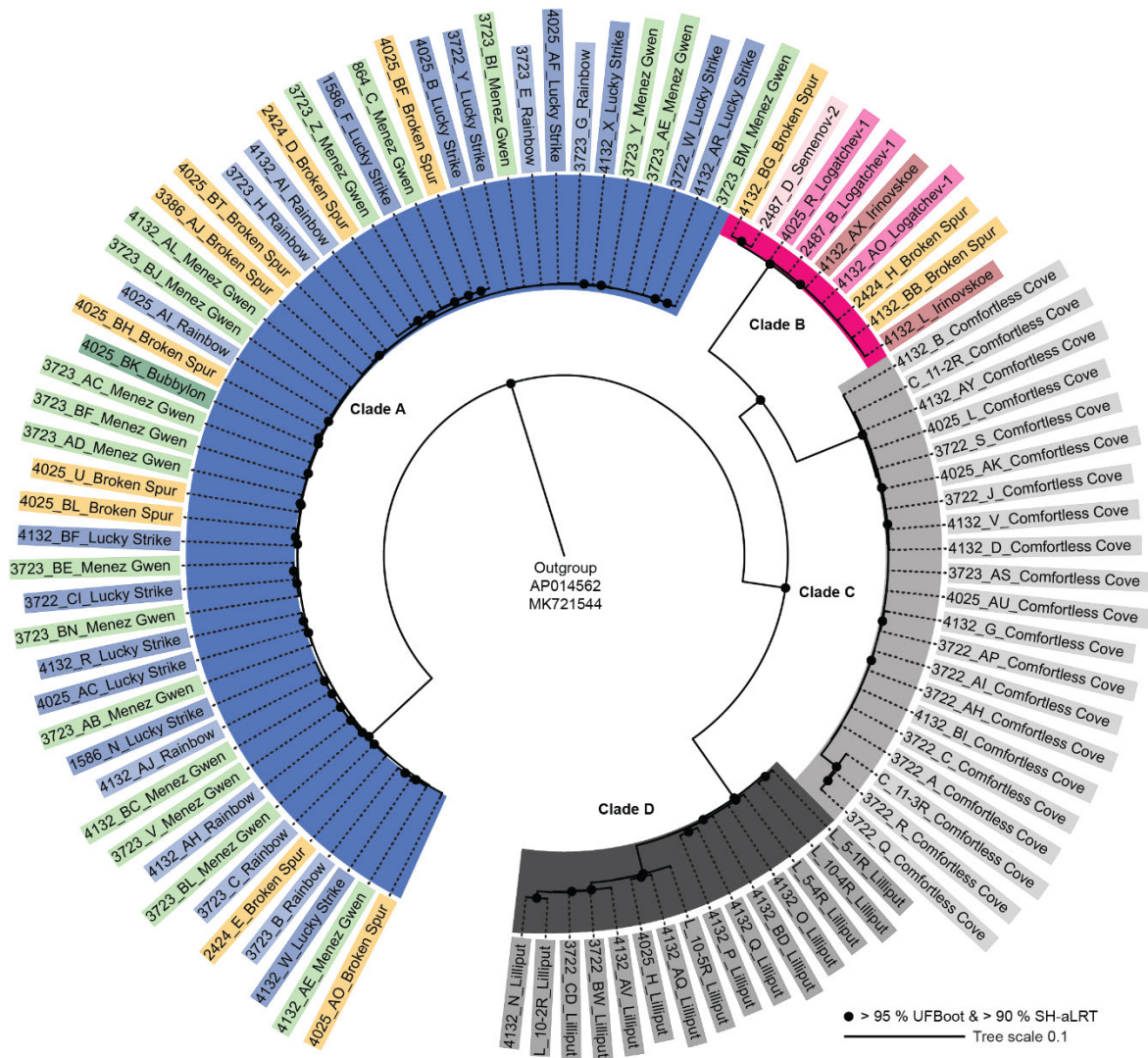
Supplementary Figure S 3 | Mitochondrial phylogeny of mussels from the Mid-Atlantic Ridge. Mitochondria were assembled from metagenomic reads and only full mitochondria (> 17 500 bp) were kept for further analysis. The tree was reconstructed based on a 38 601 bp alignment with IQ-TREE using the substitution model TN+F+R9 determined by ModelFinder. Mitochondrial genomes of *B. septemdierum* (AP014562) and *B. thermophilus* (MK721544) from public databases were used as outgroups. Mitochondria fall into four clades that correspond to the four geographic regions at the MAR: Clade A (blue) – 38°N-36°N, clade B (pink) – 29/15°N-13°N, clade C (light grey) – 5°S, and clade D (dark grey) – 9°S.

Supplementary Figure S 4 | Median-joining haplotype network of the mitochondrial cytochrome c-oxidase gene of mussels from the Mid-Atlantic Ridge. The network was reconstructed with PopART. Marks on the edges represent the number of mutations, sizes of circles show the number of samples. COI haplotypes fall into four groups that correspond to the four clades in the mitochondrial phylogeny (Supplementary Figure S 3) and the respective geographic regions at the MAR: Clade A (blue) – 38°N-36°N, clade B (pink) – 29°N-13°N, clade C (light grey) – 5°S, and clade D (dark grey) – 9°S.
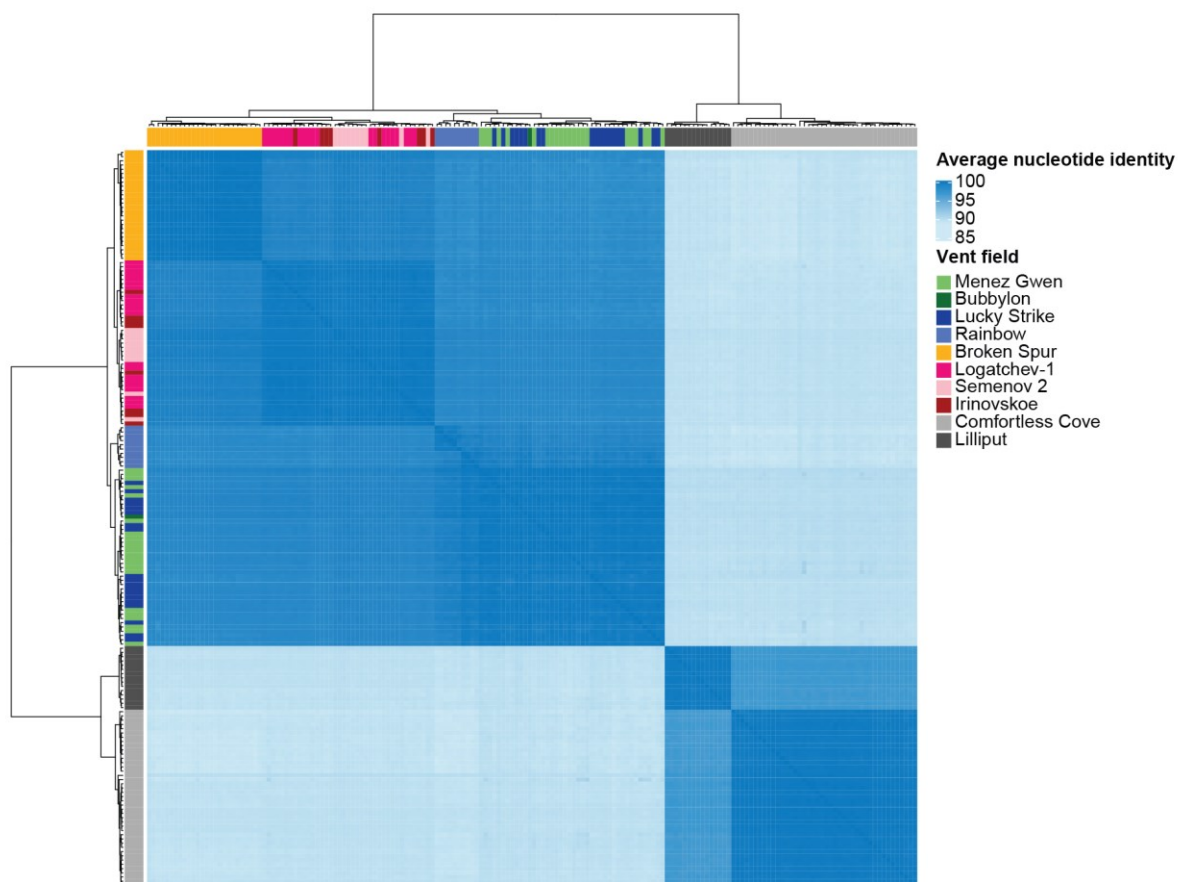
Supplementary Figure S 5 | Average nucleotide identity of sulphur-oxidising symbionts of *Bathymodiolus* mussels along the Mid-Atlantic Ridge. Metagenome-assembled genomes of symbionts were analysed with fastANI. Dark blue: high genome similarity; light blue: lower genome similarity. Colour stripes correspond to vent fields where mussel hosts were sampled. Symbionts cluster in the dendrogram according to geographic region: yellow/pinks/red - 29°N-13°N, blue/green - 38°N-36°N, dark grey - 9°S, and light grey - 5°S.
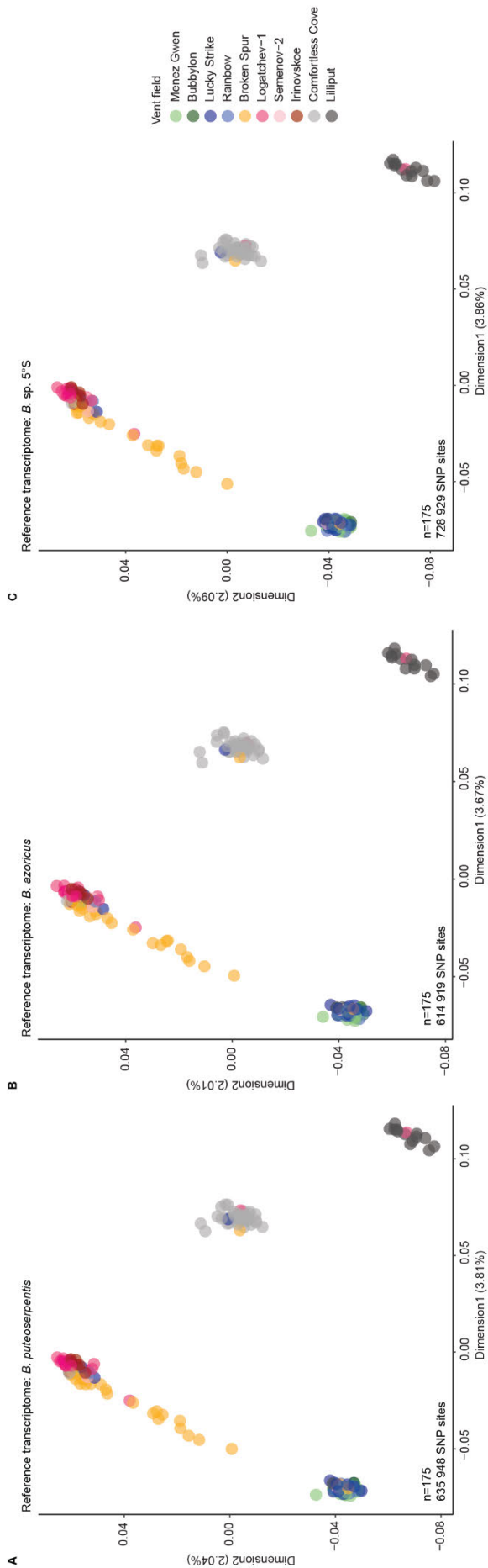
Supplementary Figure S 6 | Comparison of transcriptome references and their effect on nucDNA clusters. A: *B. puteoserpentis*. B: *B. azoricus*. C: *B.* sp. 5°S. No effect of transcriptome reference on the resulting nucDNA clusters was observed.
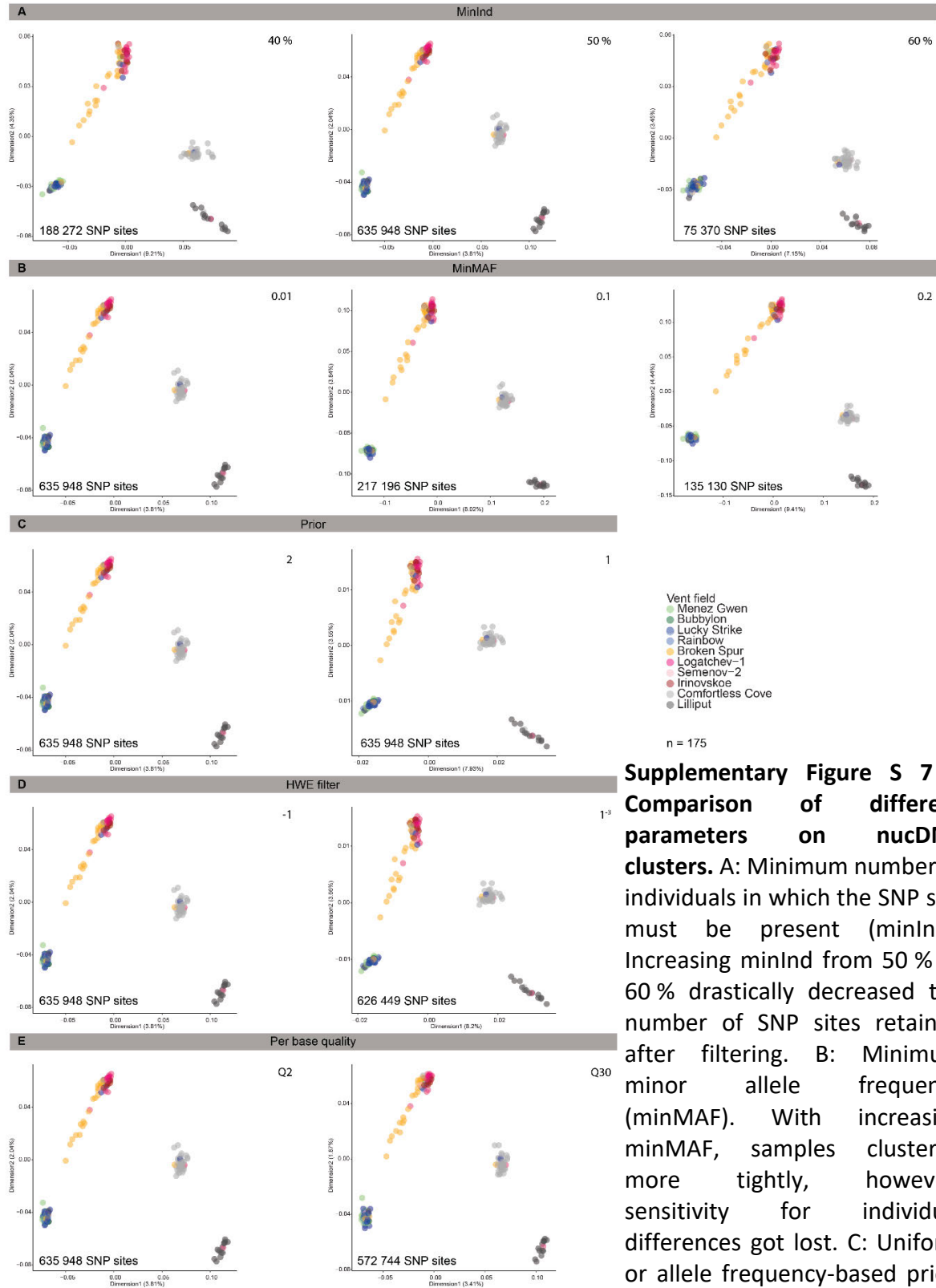
Supplementary Figure S 7 | Comparison of different parameters on nucDNA clusters. A: Minimum number of individuals in which the SNP site must be present (minInd). Increasing minInd from 50 % to 60 % drastically decreased the number of SNP sites retained after filtering. B: Minimum minor allele frequency (minMAF). With increasing minMAF, samples clustered more tightly, however, sensitivity for individual differences got lost. C: Uniform or allele frequency-based prior. D: Filter for violation of Hardy-Weinberg-equilibrium (HWE). E: Per base quality score of sequencing reads. Settings shown in C, D and E only had minimal influence on nucDNA clusters.
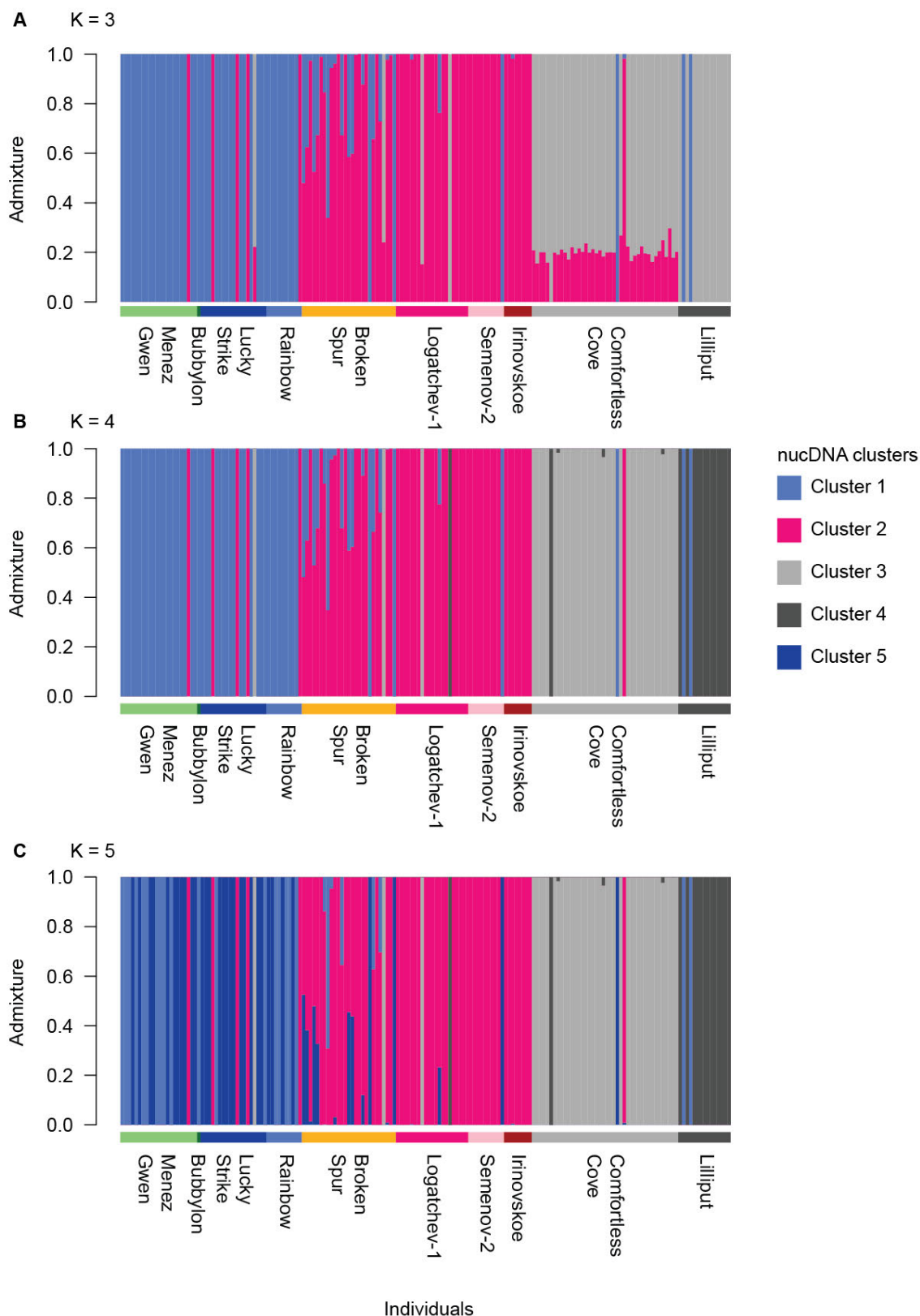
Supplementary Figure S 8 | Admixture proportions based on posterior genotype probabilities assuming A: 3, B: 4 or C: 5 clusters (K) using NGSadmix. Assuming K = 3, mussels from Comfortless Cove were mixed between Cluster 2 and 3. Assuming K = 4, mussels fell into four geographically constrained cluster that corresponded to mitochondrial lineages for most samples. Assuming K=5, Cluster 1 from K = 4 got split into two clusters.

A
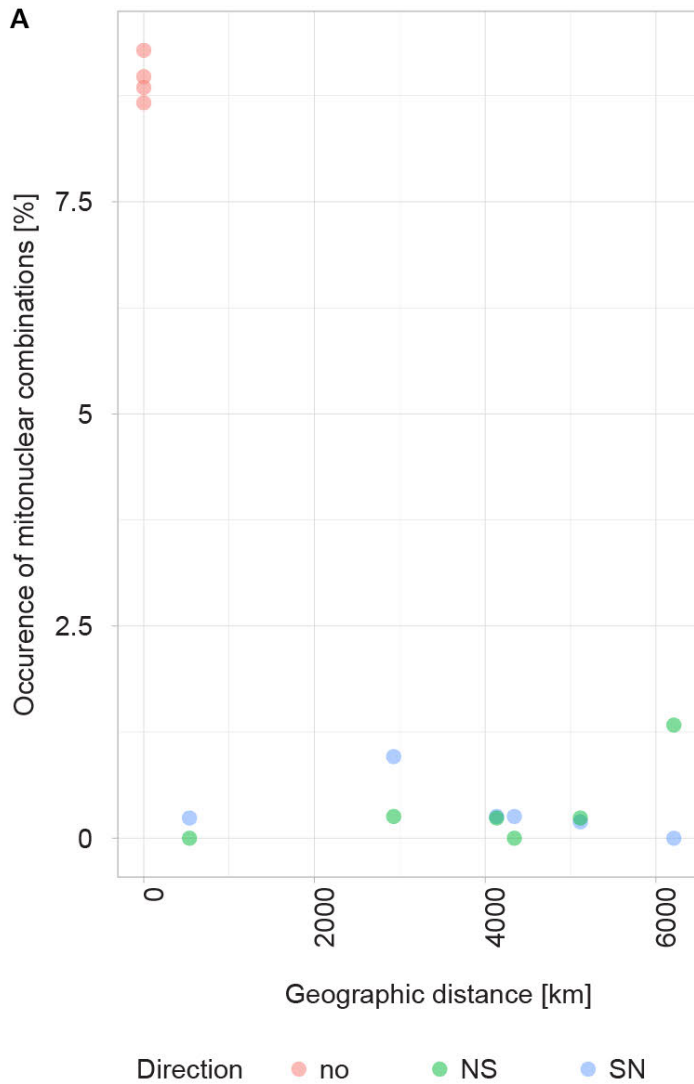
Supplementary Figure S 9 | Assessment of relation of occurrence of mitonuclear combinations with geographic distance and the direction of combination. Data used for visualisation is available in **Supplementary Table S** . A: Occurrence of mitonuclear combinations with regard to geographic distance. Colours correspond to the direction of the combination: no: mitochondrial and nuclear information are congruent, NS: north to south indicating that the mitochondrial clade corresponded to a site southern of where the nucDNA cluster mostly occurs, SN: south to north indicating that the mitochondrial clade corresponded to a site northern of where the nucDNA cluster mostly occurs. One sample with mixed nucDNA cluster and all samples from Broken Spur were excluded from the analysis. B: Statistical pairwise comparison of the occurrence of mitonuclear combinations with different direction. There was no significant difference (ns). NS: north to south, SN: south to north.

B
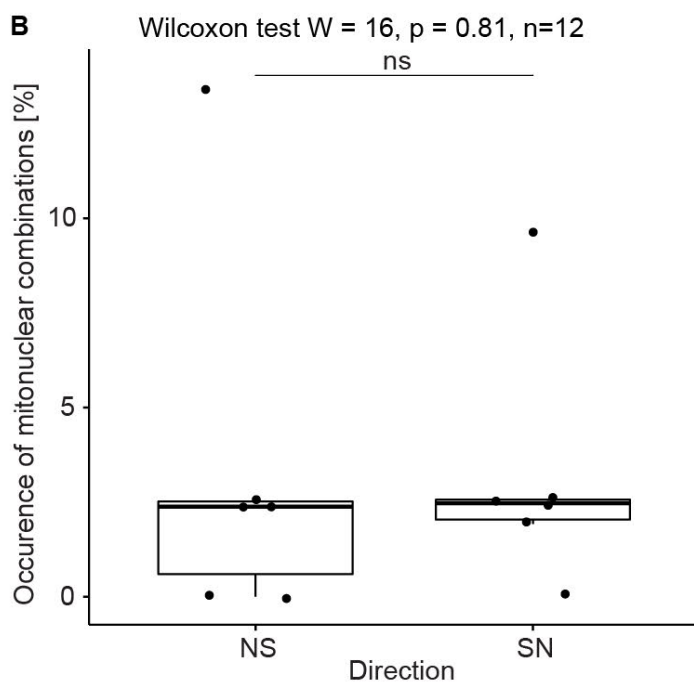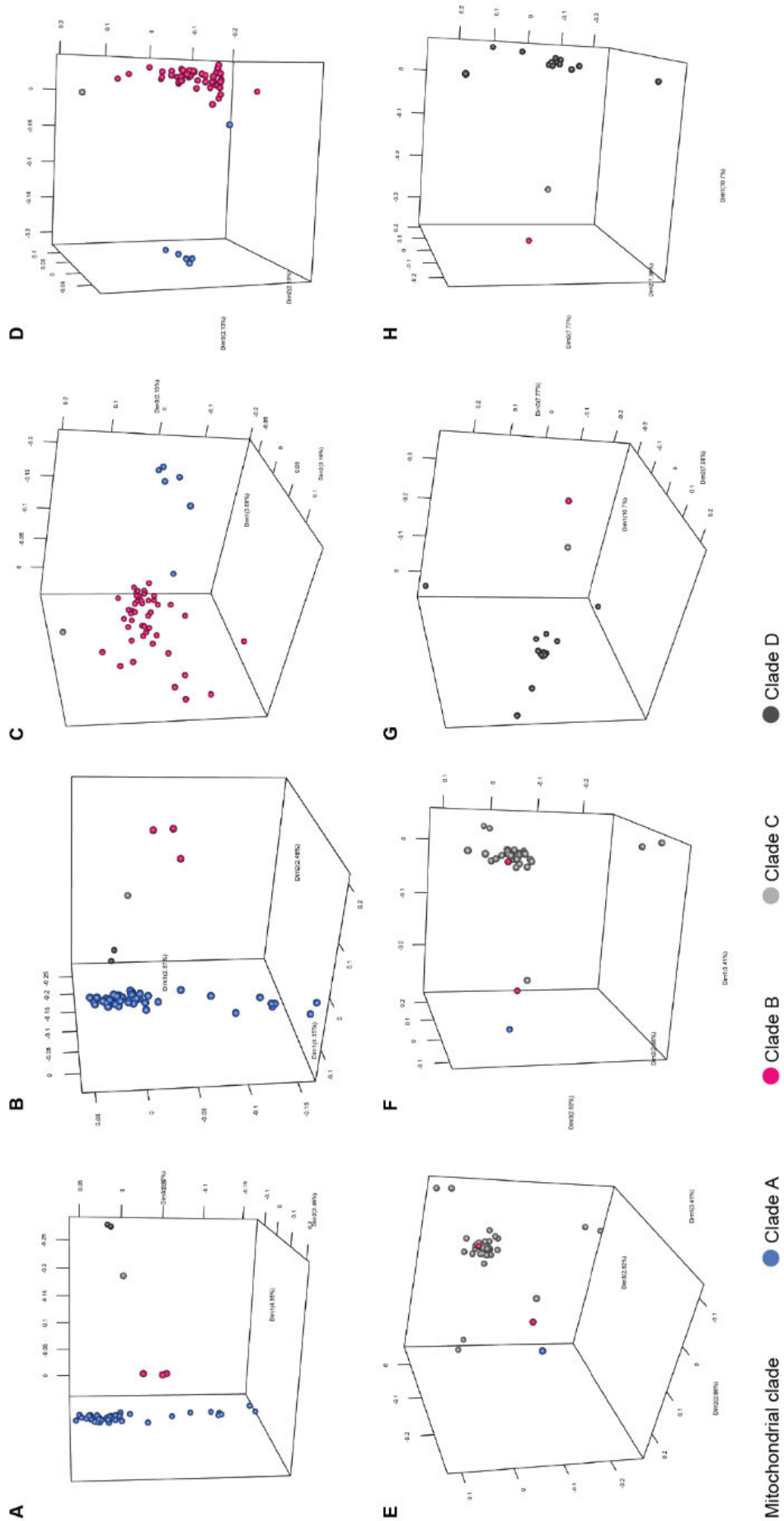
Supplementary Figure S 10 | 3D visualisation of genetic differentiation within nucDNA clusters. A & B: nucDNA cluster 1. C & D: nucDNA cluster 2. E & F: nucDNA cluster 3. G & H: nucDNA cluster 4. Colours represent mitochondrial clades. Samples within each nucDNA cluster grouped according to their mitochondrial clade.

Mitochondrial clade   ● Clade A   ● Clade B   ● Clade C   ● Clade D
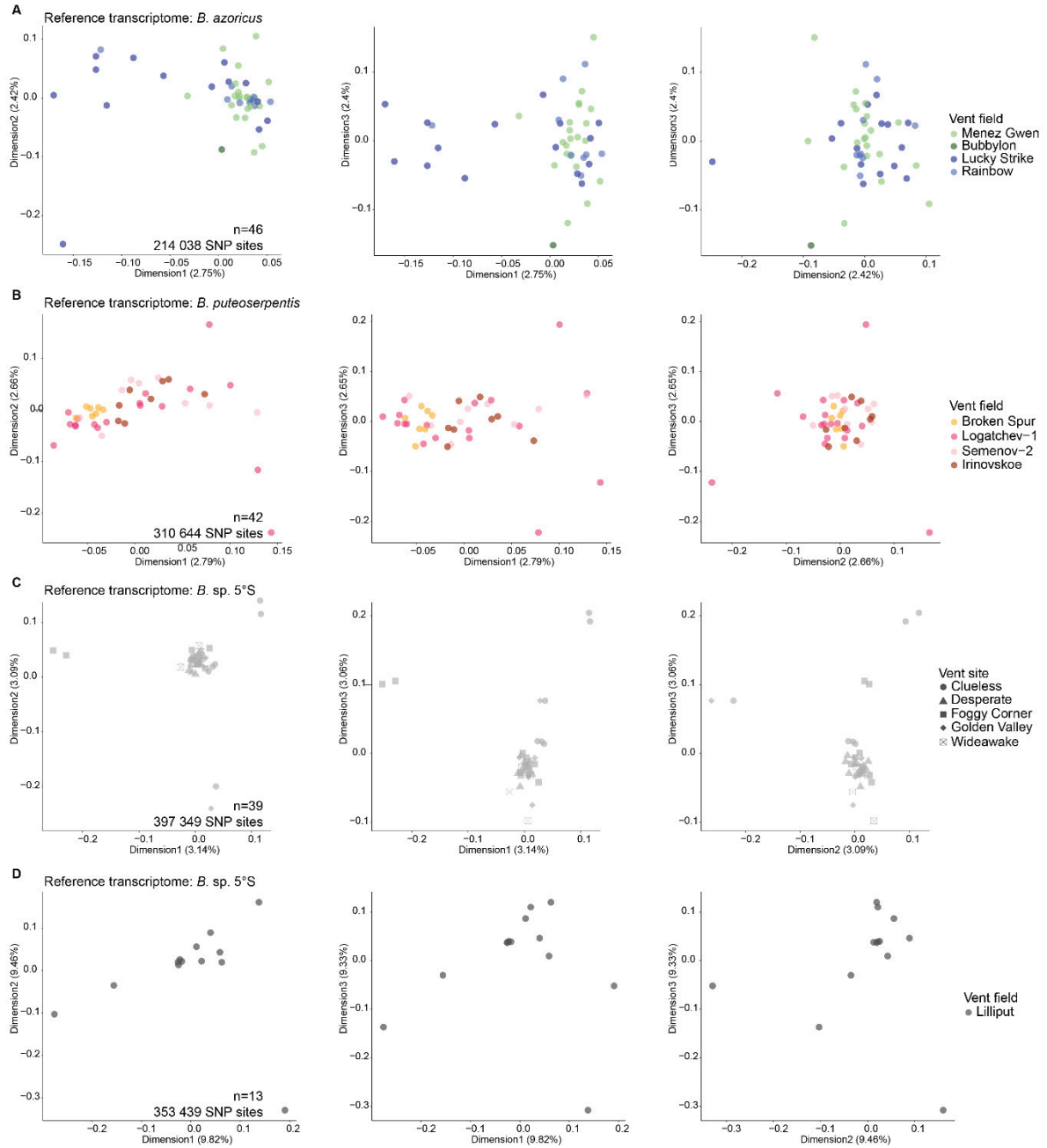
Supplementary Figure S 11 | 3D visualisation of genetic differentiation within nucDNA clusters of mussels from the same geographic region. A: nucDNA cluster 1. B: nucDNA cluster 2. C: nucDNA cluster 3. D: nucDNA cluster 4. Colours represent vent fields or vent sites within a field (when all samples come from the same field, here Comfortless Cove for cluster 3).

Supplementary Tables

Supplementary Table S 1 | Overview of all samples. Lat: Latitute, Lon: Longitude, Length: Read length in bp, Pub: Publication.

ID	Cruise	Date	Lat	Lon	Depth	Library	Length	MusselID	Site	Vent field	Gear	Sample	Fixation	Pub
1586F	Biobaz	2013-08-17	37.28	-32.28	-1689	1586F	100	D10ETb-2	Eiffel Tower	Lucky Strike	ROV Claw	gill	-20°C	[38]
1586H	Biobaz	2013-08-17	37.28	-32.28	-1689	1586H	100	D10ETb-9	Eiffel Tower	Lucky Strike				[39]
1586N	Biobaz	2013-08-09	37.29	-32.28	-1700	1586N	150	D5MS-4	Montsegur	Lucky Strike	ROV Claw	gill	-20°C	[38]
1600F	Biobaz	2013-08-13	36.23	-33.90	-2273	1600F	100	D7RB23	Rainbow	Rainbow	ROV Claw	gill	-20°C	[38]
2065A	M64-2 Logatchev 2005	2005-05-30	14.75	-44.98	-3045	2065A	150	281 ROV-3	Quest	Logatchev				[38]
2065B	M64-2 Logatchev 2005	2005-05-30	14.75	-44.98	-3045	2065B	150		Quest	Logatchev				[38]
2065C	M64-2 Logatchev 2005	2005-05-30	14.75	-44.98	-3045	2065C	150		Quest	Logatchev				[38]
2424D	AT_05/03	2001-07-19	29.17	-43.17	3045	2339_D, 2424_D	150	3676_26	Broken Spur	Broken Spur	Alvin arm/Scoop net	mixed	-80°C	[20]
2424E	AT_05/03	2001-07-19	29.17	-43.17	3045	2339_E, 2424_E	150	3676_28	Broken Spur	Broken Spur	Alvin arm/Scoop net	mixed	-80°C	[20]
2424H	AT_05/03	2001-07-19	29.17	-43.17	3045	2339_H, 2424_H	150	3676_29	Broken Spur	Broken Spur	Alvin arm/Scoop net	mixed	-80°C	[20]
2424N	AT_03/03	1997-07-17	29.17	-43.17	3056	2339_N, 2424_N	150	3125_8	Broken Spur	Broken Spur	Alvin arm/Scoop net	mixed	-80°C	[20]
2487A	M126	2016-04-28	14.75	-44.98	-3047	2487A	150	499ROV/4-2	Quest	Logatchev	Scoop net	gill	RNA later	[38]
2487B	M126	2016-04-28	14.75	-44.98	-3047	2487B	150	499ROV/4-4	Quest	Logatchev	Scoop net	gill	RNA later	[38]
2487C	M126	2016-04-28	14.75	-44.98	-3047	2487C	150	499ROV/4-5	Quest	Logatchev	Scoop net	gill	RNA later	[38]
2487D	M126	2016-05-02	13.51	-44.96	-2447	2487D	150	511ROV/4-1	Ash Lighthouse	Semenov 2	Scoop net	gill	RNA later	[38]
2487E	M126	2016-05-02	13.51	-44.96	-2447	2487E	150	511ROV/4-2	Ash Lighthouse	Semenov 2	Scoop net	gill	RNA later	[38]
3386AJ	AT_03/03	1997-07-17	29.17	-43.17	3056	3386_AJ	150	3125_9	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
3722A	L'Atalante MSM 06-3 2008	2008-01-20	-4.80	-12.37	-2992	3722A	150	ATA52ROV11_3	Clueless	Comfortless Cove	Scoop net	gill		
3722AH	M78-2 (2009)	2009-04-16	-4.80	-12.37	-2987	3722AH	150	267 ROV 4-5	Foggy Corner	Comfortless Cove	Scoop net	gill	RNA later	
3722AI	M78-2 (2009)	2009-04-16	-4.80	-12.37	-2987	3722AI	150	267 ROV 4-6	Foggy Corner	Comfortless Cove	Scoop net	gill	RNA later	
3722AJ	M78-2 (2009)	2009-04-16	-4.80	-12.37	-2987	3722AJ	150	267 ROV 4-7	Foggy Corner	Comfortless Cove	Scoop net	gill	RNA later	

ID	Cruise	Date	Lat	Lon	Depth	Library	Length	MusselID	Site	Vent field	Gear	Sample	Fixation	Pub
3722AP	M78-2 (2009)	2009-04-17	-4.80	-12.37	-2987	3722AP	150	274 ROV 4-2	Golden Valley	Comfortless Cove	Rock sample	gill	RNA later	
3722AY	M126	2016-04-28	14.75	-44.98	-3038	3722AY	150	499ROV/3-5	Irina II	Logatchev	Scoop net	gill	RNA later	
3722BW	M78-2 (2009)	2009-04-29	-9.55	-13.21	-1494	3722BW	150	319 ROV 5-6	Lilliput	Lilliput	Scoop net	gill	RNA later	
3722C	L'Atalante MSM 06-3 2008	2008-01-20	-4.80	-12.37	-2992	3722C	150	ATA52ROV11_5	Clueless	Comfortless Cove	Scoop net	gill		
3722CD	M78-2 (2009)	2009-05-02	-9.55	-13.21	-1494	3722CD	150	335 ROV 5-4	Lilliput	Lilliput	Scoop net	gill	RNA later	
3722CI	Biobaz	2013-08-08	37.29	-32.28	-1700	3722CI	150	D4MS-5	Montsegur	Lucky Strike	ROV Claw	gill	-20°C	
3722CQ	M126	2016-04-28	14.75	-44.98	-3047	3722CQ	150	499ROV/4-3	Quest	Logatchev	Scoop net	gill	RNA later	
3722J	M78-2 (2009)	2009-04-19	-4.80	-12.37	-2988	3722J	150	287 ROV 12-1	Desperate	Comfortless Cove	Scratch shovel	gill	RNA later	
3722K	M78-2 (2009)	2009-04-19	-4.80	-12.37	-2988	3722K	150	287 ROV 12-2	Desperate	Comfortless Cove	Scratch shovel	gill	RNA later	
3722L	M78-2 (2009)	2009-04-19	-4.80	-12.37	-2988	3722L	150	287 ROV 12-3	Desperate	Comfortless Cove	Scratch shovel	gill	RNA later	
3722N	M78-2 (2009)	2009-04-19	-4.80	-12.37	-2988	3722N	150	287 ROV 12-5	Desperate	Comfortless Cove	Scratch shovel	gill	RNA later	
3722O	M78-2 (2009)	2009-04-19	-4.80	-12.37	-2988	3722O	150	287 ROV 12-6	Desperate	Comfortless Cove	Scratch shovel	gill	RNA later	
3722P	M78-2 (2009)	2009-04-19	-4.80	-12.37	-2988	3722P	150	287 ROV 12-7	Desperate	Comfortless Cove	Scratch shovel	gill	RNA later	
3722Q	M78-2 (2009)	2009-04-19	-4.80	-12.37	-2988	3722Q	150	287 ROV 12-8	Desperate	Comfortless Cove	Scratch shovel	gill	RNA later	
3722R	M78-2 (2009)	2009-04-19	-4.80	-12.37	-2988	3722R	150	287 ROV 12-9	Desperate	Comfortless Cove	Scratch shovel	gill	RNA later	
3722S	M78-2 (2009)	2009-04-19	-4.80	-12.37	-2988	3722S	150	287 ROV 12-10	Desperate	Comfortless Cove	Scratch shovel	gill	RNA later	
3722T	M78-2 (2009)	2009-04-19	-4.80	-12.37	-2988	3722T	150	287 ROV 12-11	Desperate	Comfortless Cove	Scratch shovel	gill	RNA later	
3722W	Biobaz	2013-08-17	37.28	-32.28	-1689	3722W	150	D10ETb-3	Eiffel Tower	Lucky Strike	ROV Claw	gill	-20°C	
3722Y	Biobaz	2013-08-17	37.28	-32.28	-1689	3722Y	150	D10ETb-33	Eiffel Tower	Lucky Strike	ROV Claw	gill	-20°C	
3723AB	M82-3	2010-09-20	37.84	-31.52	-839	3723AB	150	719ROV 10-9	Station 719-10	Menez Gwen		gill	-80°C	
3723AC	M82-3	2010-09-23	37.84	-31.52	-840	3723AC	150	729ROV 4-1	Station 729-4	Menez Gwen		gill	RNA later	
3723AD	M82-3	2010-09-23	37.84	-31.52	-840	3723AD	150	729ROV 4-2	Station 729-4	Menez Gwen		gill	RNA later	
3723AE	M82-3	2010-09-26	37.84	-31.52	-836	3723AE	150	736ROV 3-3	Station 736-3	Menez Gwen		gill	RNA later	
3723AS	L'Atalante MSM 06-3 2008	2008-01-21	-4.81	-12.37	-2992	3723AS	150	ATA57ROV7_36	Wideawake	Comfortless Cove	Scoop net	gill	-80°C	

ID	Cruise	Date	Lat	Lon	Depth	Library	Length	MusselID	Site	Vent field	Gear	Sample	Fixation	Pub
3723AV	L'Atalante MSM 06-3 2008	2008-01-21	-4.81	-12.37	-2992	3723AV	150	ATA57ROV7_39	Wideawake	Comfortless Cove	Scoop net	gill	-80°C	
3723B	Biobaz	2013-08-13	36.23	-33.90	-2273	3723B	150	D7RB3	Rainbow	Rainbow	ROV Claw	gill	-20°C	
3723BA	Biobaz	2013-08-06	37.84	-31.52	-828	3723BA	150	D2MG2-7	Woody	Menez Gwen	ROV Claw	gill	-80°C	
3723BE	Biobaz	2013-08-06	37.84	-31.52	-828	3723BE	150	D2MG2-11	Woody	Menez Gwen	ROV Claw	gill	-80°C	
3723BF	Biobaz	2013-08-06	37.84	-31.52	-828	3723BF	150	D2MG2-12	Woody	Menez Gwen	ROV Claw	gill	-80°C	
3723BI	Biobaz	2013-08-06	37.84	-31.52	-828	3723BI	150	D2MG2-15	Woody	Menez Gwen	ROV Claw	gill	-80°C	
3723BJ	Biobaz	2013-08-06	37.84	-31.52	-828	3723BJ	150	D2MG2-16	Woody	Menez Gwen	ROV Claw	gill	-80°C	
3723BL	Biobaz	2013-08-06	37.84	-31.52	-828	3723BL	150	D2MG2-18	Woody	Menez Gwen	ROV Claw	gill	-80°C	
3723BM	Biobaz	2013-08-06	37.84	-31.52	-828	3723BM	150	D2MG2-19	Woody	Menez Gwen	ROV Claw	gill	-80°C	
3723BN	Biobaz	2013-08-06	37.84	-31.52	-828	3723BN	150	D2MG2-21	Woody	Menez Gwen	ROV Claw	gill	-80°C	
3723C	Biobaz	2013-08-13	36.23	-33.90	-2273	3723C	150	D7RB4	Rainbow	Rainbow	ROV Claw	gill	-20°C	
3723E	Biobaz	2013-08-13	36.23	-33.90	-2273	3723E	150	D7RB6	Rainbow	Rainbow	ROV Claw	gill	-20°C	
3723G	Biobaz	2013-08-13	36.23	-33.90	-2273	3723G	150	D7RB8	Rainbow	Rainbow	ROV Claw	gill	-20°C	
3723H	Biobaz	2013-08-13	36.23	-33.90	-2273	3723H	150	D7RB10	Rainbow	Rainbow	ROV Claw	gill	-20°C	
3723R	M126	2016-05-02	13.51	-44.96	-2447	3723R	150	511ROV/4-3	Ash Lighthouse	Semenov 2	Scoop net	gill	RNA later	
3723S	M126	2016-05-02	13.51	-44.96	-2447	3723S	150	511ROV/4-4	Ash Lighthouse	Semenov 2	Scoop net	gill	RNA later	
3723T	M126	2016-05-02	13.51	-44.96	-2447	3723T	150	511ROV/3-8	Ash Lighthouse	Semenov 2	Scoop net	gill	RNA later	
3723V	M82-3	2010-09-20	37.84	-31.52	-839	3723V	150	719ROV 10-3	Station 719-10	Menez Gwen		gill	-80°C	
3723W	M82-3	2010-09-20	37.84	-31.52	-839	3723W	150	719ROV 10-4	Station 719-10	Menez Gwen		gill	-80°C	
3723X	M82-3	2010-09-20	37.84	-31.52	-839	3723X	150	719ROV 10-5	Station 719-10	Menez Gwen		gill	-80°C	
3723Y	M82-3	2010-09-20	37.84	-31.52	-839	3723Y	150	719ROV 10-6	Station 719-10	Menez Gwen		gill	-80°C	
3723Z	M82-3	2010-09-20	37.84	-31.52	-839	3723Z	150	719ROV 10-7	Station 719-10	Menez Gwen		gill	-80°C	
4025AA	Biobaz	2013-08-17	37.28	-32.28	-1689	1586G,4025AA	100	D10ETb-7	Eiffel Tower	Lucky Strike		gill	-80°C	[38]
4025AB	AT_05/03	2001-07-19	29.17	-43.17	3045	3386_I, 4025_AB	150	3676_10	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4025AC	Biobaz	2013-08-17	37.28	-32.28	-1689	1586J,4025AC	100	D10ETb-13	Eiffel Tower	Lucky Strike				[38]
4025AE	M78-2 (2009)	2009-04-17	-4.80	-12.37	-2987	3722AQ,4025AE	150	274 ROV 4-3	Golden Valley	Comfortless Cove	Rock sample	gill	RNA later	
4025AF	Biobaz	2013-08-08	37.29	-32.28	-1700	3722CG,4025AF	150	D4MS-3	Montsegur	Lucky Strike	ROV Claw	gill	-20°C	
4025AG	MSM10-03 Hydromar VII	2009-02-08	14.75	-44.98	-3048	3722CN,4025AG	150	313ROV2-5	Quest	Logatchev	Scoop net	gill	RNA later	
4025AH	MSM10-03 Hydromar VII	2009-02-08	14.75	-44.98	-3048	3722CJ,4025AH	150	313ROV2-1	Quest	Logatchev	Scoop net	gill	RNA later	
4025AI	Biobaz	2013-08-13	36.23	-33.90	-2273	3723D,4025AI	150	D7RB5	Rainbow	Rainbow	ROV Claw	gill	-20°C	
4025AJ	AT_05/03	2001-07-19	29.17	-43.17	3045	3386_AA, 4025_AJ	150	3676_37	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]

ID	Cruise	Date	Lat	Lon	Depth	Library	Length	MusselID	Site	Vent field	Gear	Sample	Fixation	Pub
4025AK	M78-2 (2009)	2009-04-17	-4.80	-12.37	-2986	3722AS,4025AK	150	274 ROV 3-2	Golden Valley	Comfortless Cove	Scoop net	gill	RNA later	
4025AL	M126	2016-04-28	14.75	-44.98	-3038	3722AU,4025AL	150	499ROV/3-1	Irina II	Logatchev	Scoop net	gill	RNA later	
4025AO	AT_03/03	1997-07-17	29.17	-43.17	3056	3386_AD, 4025_AO	150	3125_2	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4025AQ	M126	2016-04-24	13.51	-44.96	-2324	3723L,4025AQ	150	488ROV/1-20	Semenov 2	Semenov 2	Scoop net	gill	RNA later	
4025AT	M126	2016-04-24	13.51	-44.96	-2324	3723K,4025AT	150	488ROV/1-19	Semenov 2	Semenov 2	Scoop net	gill	RNA later	
4025AU	M78-2 (2009)	2009-04-17	-4.80	-12.37	-2987	3722AO,4025A U	150	274 ROV 4-1	Golden Valley	Comfortless Cove	Rock sample	gill	RNA later	
4025AX	M126	2016-04-28	14.75	-44.98	-3038	3722AV,4025AX	150	499ROV/3-2	Irina II	Logatchev	Scoop net	gill	RNA later	
4025AY	M126	2016-04-29	13.33	-44.91	-2791	3722BF,4025AY	150	501ROV/2-6	Irinovskoe	Irinovskoe	Scoop net	gill	RNA later	
4025B	Biobaz	2013-08-09	37.29	-32.28	-1700	1586M,4025B	150	D5MS-3	Montsegur	Lucky Strike	ROV Claw	gill	-20°C	[39]
4025BA	AT_05/03	2001-07-19	29.17	-43.17	3045	3386_P, 4025_BA	150	3676_17	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4025BB	AT_05/03	2001-07-19	29.17	-43.17	3045	3386_V, 4025_BB	150	3676_27	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4025BC	AT_05/03	2001-07-19	29.17	-43.17	3045	3386_AB, 4025_BC	150	3676_38	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4025BE	Biobaz	2013-08-08	37.29	-32.28	-1700	3722CF,4025BE	150	D4MS-2	Montsegur	Lucky Strike	ROV Claw	gill	-20°C	
4025BF	AT_05/03	2001-07-19	29.17	-43.17	3045	3386_C, 4025_BF	150	3676_3	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4025BG	AT_05/03	2001-07-19	29.17	-43.17	3045	3386_Q, 4025_BG	150	3676_18	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4025BH	AT_03/03	1997-07-17	29.17	-43.17	3056	3386_AL, 4025_BH	150	3125_11	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4025BK	M82-3	2010-10-02	37.80	-31.54	-1002	3723BP,4025BK	150		Bubbylon	Bubbylon		gill	PFA/PBS-EtOH	
4025BL	AT_05/03	2001-07-19	29.17	-43.17	3045	3386_H, 4025_BL	150	3676_9	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4025BM	MSM10-03 Hydromar VII	2009-02-08	14.75	-44.98	-3048	3722CO,4025B M	150	313ROV2-6	Quest	Logatchev	Scoop net	gill	RNA later	
4025BN	AT_05/03	2001-07-19	29.17	-43.17	3045	3386_L, 4025_BN	150	3676_12	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4025BO	AT_05/03	2001-07-19	29.17	-43.17	3045	3386_E, 4025_BO	150	3676_5	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4025BP	AT_05/03	2001-07-19	29.17	-43.17	3045	3386_F, 4025_BP	150	3676_6	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4025BR	AT_05/03	2001-07-19	29.17	-43.17	3045	3386_B, 4025_BR	150	3676_2	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4025BS	AT_05/03	2001-07-19	29.17	-43.17	3045	3386_D, 4025_BS	150	3676_4	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]

ID	Cruise	Date	Lat	Lon	Depth	Library	Length	MusselID	Site	Vent field	Gear	Sample	Fixation	Pub
4025BT	AT_05/03	2001-07-19	29.17	-43.17	3045	3386_T, 4025_BT	150	3676_21	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4025E	Biobaz	2013-08-09	37.29	-32.28	-1700	1586K,4025E	100	D5MS-1	Eiffel Tower	Lucky Strike	ROV Claw	gill	-20°C	[38]
4025H	M78-2 (2009)	2009-04-30	-9.55	-13.21	-1494	3722BY,4025H	150	324 ROV 10-2	Lilliput	Lilliput	Scoop net	gill	RNA later	
4025I	M126	2016-04-24	13.51	-44.96	-2324	3723M,4025I	150	488ROV/1-21	Semenov 2	Semenov 2	Scoop net	gill	RNA later	
4025J	M126	2016-04-29	13.33	-44.91	-2791	3722BG,4025J	150	501ROV/2-4	Irinovskoe	Irinovskoe	Scoop net	gill	RNA later	
4025L	M78-2 (2009)	2009-04-17	-4.80	-12.37	-2986	3722AT,4025L	150	274 ROV 3-3	Golden Valley	Comfortless Cove	Scoop net	gill	RNA later	
4025M	MSM10-03 Hydromar VII	2009-02-08	14.75	-44.98	-3048	3722CM,4025M	150	313ROV2-4	Quest	Logatchev	Scoop net	gill	RNA later	
4025N	Odemar	2014-11-22	13.51	-44.96	-2432	1115B,4025N	250	M3-10	Ash Lighthouse	Semenov 2	Scoop net	gill	RNA later	[38]
4025Q	AT_03/03	1997-07-17	29.17	-43.17	3056	3386_AK, 4025_Q	150	3125_10	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4025R	MSM10-03 Hydromar VII	2009-02-08	14.75	-44.98	-3048	3722CL,4025R	150	313ROV2-3	Quest	Logatchev	Scoop net	gill	RNA later	
4025U	AT_05/03	2001-07-19	29.17	-43.17	3045	3386_U, 4025_U	150	3676_22	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4025V	Biobaz	2013-08-10	37.29	-32.28	-1690	1586B,4025V	100	D5ET-2	Eiffel Tower	Lucky Strike	ROV Claw	gill	-20°C	[38]
4132A	AT_03/03	1997-07-17	29.17	-43.17	3056	3386_AC, 4132_A	150	3125_1	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4132AA	L'Atalante MSM 06-3 2008	2008-01-21	-4.81	-12.37	-2992	3723AR,4132AA	150	ATA57ROV7_35	Wideawake	Comfortless Cove	Scoop net	gill	-80°C	
4132AB	L'Atalante MSM 06-3 2008	2008-01-21	-4.81	-12.37	-2992	3723AT,4132AB	150	ATA57ROV7_37	Wideawake	Comfortless Cove	Scoop net	gill	-80°C	
4132AC	L'Atalante MSM 06-3 2008	2008-01-21	-4.81	-12.37	-2992	3723AU,4132AC	150	ATA57ROV7_38	Wideawake	Comfortless Cove	Scoop net	gill	-80°C	
4132AE	Biobaz	2013-08-06	37.84	-31.52	-828	3723BB,4132AE	150	D2MG2-8	Woody	Menez Gwen	ROV Claw	gill	-80°C	
4132AF	Biobaz	2013-08-06	37.84	-31.52	-828	3723BD,4132AF	150	D2MG2-10	Woody	Menez Gwen	ROV Claw	gill	-80°C	
4132AH	Biobaz	2013-08-13	36.23	-33.90	-2273	3723F,4132AH	150	D7RB7	Rainbow	Rainbow	ROV Claw	gill	-20°C	
4132AI	Biobaz	2013-08-13	36.23	-33.90	-2273	3723J,4132AI	150	D7RB11	Rainbow	Rainbow	ROV Claw	gill	-20°C	
4132AJ	Biobaz	2013-08-13	36.23	-33.90	-2273	3723J,4132AJ	150	D7RB13	Rainbow	Rainbow	ROV Claw	gill	-20°C	
4132AK	M126	2016-05-02	13.51	-44.96	-2447	3723Q,4132AK	150	511ROV/3-7	Ash Lighthouse	Semenov 2	Scoop net	gill	RNA later	
4132AL	Biobaz	2013-08-06	37.84	-31.52	-828	3723BH,4132AL	150	D2MG2-14	Woody	Menez Gwen	ROV Claw	gill	-80°C	
4132AO	M126	2016-04-28	14.75	-44.98	-3038	3722AW,4132AO	150	499ROV/3-3	Irina II	Logatchev	Scoop net	gill	RNA later	
4132AQ	M78-2 (2009)	2009-04-29	-9.55	-13.21	-1494	3722BV,4132AQ	150	319 ROV 5-5	Lilliput	Lilliput	Scoop net	gill	RNA later	

ID	Cruise	Date	Lat	Lon	Depth	Library	Length	MusselID	Site	Vent field	Gear	Sample	Fixation	Pub
4132AR	Biobaz	2013-08-17	37.28	-32.28	-1689	3722AB,4132AR	150	D10ETb-10	Eiffel Tower	Lucky Strike	ROV Claw	gill	-20°C	
4132AS	M126	2016-04-29	13.33	-44.91	-2791	3722BI,4132AS	150	501ROV/2-2	Irinovskoe	Irinovskoe	Scoop net	gill	RNA later	
4132AT	M126	2016-04-28	14.75	-44.98	-3038	3722AZ,4132AT	150	499ROV/3-6	Irina II	Logatchev	Scoop net	gill	RNA later	
4132AV	M78-2 (2009)	2009-05-02	-9.55	-13.21	-1494	3722CB,4132AV	150	335 ROV 5-2	Lilliput	Lilliput	Scoop net	gill	RNA later	
4132AX	M126	2016-04-29	13.33	-44.91	-2791	3722BJ,4132AX	150	501ROV/2-1	Irinovskoe	Irinovskoe	Scoop net	gill	RNA later	
4132AY	M78-2 (2009)	2009-04-17	-4.80	-12.37	-2986	3722AR,4132AY	150	274 ROV 3-1	Golden Valley	Comfortless Cove	Scoop net	gill	RNA later	
4132AZ	M126	2016-04-29	13.33	-44.91	-2791	3722BL,4132AZ	150	501ROV/2-8	Irinovskoe	Irinovskoe	Scoop net	gill	RNA later	
4132B	M78-2 (2009)	2009-04-16	-4.80	-12.37	-2987	3722AD,4132B	150	267 ROV 4-1	Foggy Corner	Comfortless Cove	Scoop net	gill	RNA later	
4132BA	AT_03/03	1997-07-17	29.17	-43.17	3056	3386_AE, 4132_BA	150	3125_3	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4132BB	AT_05/03	2001-07-19	29.17	-43.17	3045	3386_A, 4132_BB	150	3676_1	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4132BC	M82-3	2010-09-20	37.84	-31.52	-839	3723AA,4132BC	150	719ROV 10-8	Station 719-10	Menez Gwen		gill	-80°C	
4132BD	M78-2 (2009)	2009-05-02	-9.55	-13.21	-1494	3722CC,4132BD	150	335 ROV 5-3	Lilliput	Lilliput	Scoop net	gill	RNA later	
4132BE	M126	2016-04-29	13.33	-44.91	-2791	3722BP,4132BE	150	501ROV/1-2	Irinovskoe	Irinovskoe	Scoop net	gill	RNA later	
4132BF	Biobaz	2013-08-17	37.28	-32.28	-1689	3722X,4132BF	150	D10ETb-32	Eiffel Tower	Lucky Strike	ROV Claw	gill	-20°C	
4132BG	AT_03/03	1997-07-17	29.17	-43.17	3056	3386_AF, 4132_BG	150	3125_4	Broken Spur	Broken Spur	Alvin arm/Scoop net	gill	-80°C	[20]
4132BH	Biobaz	2013-08-17	37.28	-32.28	-1689	3722ZZ,4132BH	150	D10ETb-12	Eiffel Tower	Lucky Strike	ROV Claw	gill	-20°C	
4132BI	M78-2 (2009)	2009-04-16	-4.80	-12.37	-2987	3722AK,4132BI	150	267 ROV 4-8	Foggy Corner	Comfortless Cove	Scoop net	gill	RNA later	
4132C	M78-2 (2009)	2009-04-16	-4.80	-12.37	-2987	3722AE,4132C	150	267 ROV 4-2	Foggy Corner	Comfortless Cove	Scoop net	gill	RNA later	
4132D	M78-2 (2009)	2009-04-16	-4.80	-12.37	-2987	3722AF,4132D	150	267 ROV 4-3	Foggy Corner	Comfortless Cove	Scoop net	gill	RNA later	
4132E	M78-2 (2009)	2009-04-16	-4.80	-12.37	-2987	3722AG,4132E	150	267 ROV 4-4	Foggy Corner	Comfortless Cove	Scoop net	gill	RNA later	
4132F	M78-2 (2009)	2009-04-16	-4.80	-12.37	-2987	3722AL,4132F	150	267 ROV 4-9	Foggy Corner	Comfortless Cove	Scoop net	gill	RNA later	
4132G	M78-2 (2009)	2009-04-16	-4.80	-12.37	-2987	3722AM,4132G	150	267 ROV 4-10	Foggy Corner	Comfortless Cove	Scoop net	gill	RNA later	
4132H	M78-2 (2009)	2009-04-16	-4.80	-12.37	-2987	3722AN,4132H	150	267 ROV 4-11	Foggy Corner	Comfortless Cove	Scoop net	gill		
4132I	M126	2016-04-28	14.75	-44.98	-3037	3722BB,4132I	150	499ROV/1-5	Irina II	Logatchev				
4132J	M126	2016-04-28	14.75	-44.98	-3037	3722BC,4132J	150	499ROV/1-4	Irina II	Logatchev				
4132K	M126	2016-04-28	14.75	-44.98	-3037	3722BE,4132K	150	499ROV/1-1	Irina II	Logatchev				
4132L	M126	2016-04-29	13.33	-44.91	-2791	3722BQ,4132L	150	501ROV/1-3	Irinovskoe	Irinovskoe	Scoop net	gill	RNA later	
4132M	M126	2016-04-29	13.33	-44.91	-2791	3722BS,4132M	150	501ROV/1-5	Irinovskoe	Irinovskoe	Scoop net	gill	RNA later	

ID	Cruise	Date	Lat	Lon	Depth	Library	Length	MusselID	Site	Vent field	Gear	Sample	Fixation	Pub
4132N	M78-2 (2009)	2009-04-29	-9.55	-13.21	-1494	3722BT,4132N	150	319 ROV 5-2	Lilliput	Lilliput	Scoop net	gill	RNA later	
4132O	M78-2 (2009)	2009-04-29	-9.55	-13.21	-1494	3722BU,4132O	150	319 ROV 5-3	Lilliput	Lilliput	Scoop net	gill	RNA later	
4132P	M78-2 (2009)	2009-04-30	-9.55	-13.21	-1494	3722BX,4132P	150	324 ROV 10-1	Lilliput	Lilliput	Scoop net	gill	RNA later	
4132Q	M78-2 (2009)	2009-04-30	-9.55	-13.21	-1494	3722CA,4132Q	150	324 ROV 10-4	Lilliput	Lilliput	Scoop net	gill	RNA later	
4132R	Biobaz	2013-08-08	37.29	-32.28	-1700	3722CE,4132R	150	D4MS-1	Montsegur	Lucky Strike	ROV Claw	gill	-20°C	
4132S	MSM10-03 Hydromar VII	2009-02-08	14.75	-44.98	-3048	3722CK,4132S	150	313ROV2-2	Quest	Logatchev	Scoop net	gill	RNA later	
4132U	L'Atalante MSM 06-3 2008	2008-01-20	-4.80	-12.37	-2992	3722D,4132U	150	ATA52ROV11_6	Clueless	Comfortless Cove	Scoop net	gill		
4132V	M78-2 (2009)	2009-04-19	-4.80	-12.37	-2988	3722M,4132V	150	287 ROV 12-4	Desperate	Comfortless Cove	Scratch shovel	gill	RNA later	
4132W	Biobaz	2013-08-17	37.28	-32.28	-1689	3722U,4132W	150	D10ETb-34	Eiffel Tower	Lucky Strike			-20°C	
4132X	Biobaz	2013-08-17	37.28	-32.28	-1689	3722V,4132X	150	D10ETb-35	Eiffel Tower	Lucky Strike			-20°C	
4132Z	L'Atalante MSM 06-3 2008	2008-01-21	-4.81	-12.37	-2992	3723AQ,4132Z	150	ATA57ROV7_34	Wideawake	Comfortless Cove	Scoop net	gill	-80°C	
864C	M82-3	2013-08-18	37.84	-31.52	-832	864C	150	D11MG4-19	White Flames	Menez Gwen	ROV Claw	gill	-80°C	
C11-2	M78-2 (2009)	2009-04-22	-4.80	-12.37	-2995	C11-2	150	302/ROV15-2	Clueless	Comfortless Cove	Scratch shovel	gill	RNA later	[38]
C11-3	M78-2 (2009)	2009-04-22	-4.80	-12.37	-2995	C11-3	150	302/ROV15-3	Clueless	Comfortless Cove	Scratch shovel	gill	RNA later	[38]
C11-4	M78-2 (2009)	2009-04-22	-4.80	-12.37	-2995	C11-4	150	302/ROV15-4	Clueless	Comfortless Cove	Scratch shovel	gill	RNA later	[38]
C11-5	M78-2 (2009)	2009-04-22	-4.80	-12.37	-2995	C11-5	150	302/ROV15-5	Clueless	Comfortless Cove	Scratch shovel	gill	RNA later	[38]
C11-6	M78-2 (2009)	2009-04-22	-4.80	-12.37	-2995	C11-6	150	302/ROV15-6	Clueless	Comfortless Cove	Scratch shovel	gill	RNA later	[38]
L10-2	M78-2 (2009)	2009-04-29	-9.55	-13.21	-1494	L10-2	150	319/ROV5-1	Lilliput	Lilliput	Scoop net	gill	RNA later	[38]
L10-4	M78-2 (2009)	2009-04-29	-9.55	-13.21	-1494	L10-4	150	319/ROV5-4	Lilliput	Lilliput	Scoop net	gill	RNA later	[38]
L10-5	M78-2 (2009)	2009-04-29	-9.55	-13.21	-1494	L10-5	150	319/ROV10-2	Lilliput	Lilliput	Scoop net	gill	RNA later	[38]
L5-1	M78-2 (2009)	2009-04-29	-9.55	-13.21	-1494	L5-1	150	319/ROV10-4	Lilliput	Lilliput	Scoop net	gill	RNA later	[38]
L5-4	M78-2 (2009)	2009-04-29	-9.55	-13.21	-1494	L5-4	150	319/ROV10-5	Lilliput	Lilliput	Scoop net	gill	RNA later	[38]

Supplementary Table S 2 | Summary of BUSCO results for transcriptome reference before and after curation.

	B. puteoserpentis		*B. azoricus*		*B.* sp. 5°S	
	Before	After	Before	After	Before	After
Completeness [%]	99.2	91	98.5	92.9	98.1	94.7
Duplication [%]	26	1.5	21.8	0.4	29	0.4

Supplementary Table S 3 | Summary of mitonuclear combinations, their occurrence and direction. nucDNA: nucDNA cluster based on NGSadmix results, Mito: Mitochondrial clade based on COI gene sequences, Total: Total number of samples associated to this mitochondrial clade, Occurrence: Number of times the nucDNA-mitochondrial clade combination occurred in our dataset, Direction: Geographic direction of nucDNA cluster to mitochondrial clade. SN: south to north, NS: north to south. Sample marked in grey was not used for further visual and statistical analysis.

nucDNA	Mito	Combination	Total	Occurrence	Distance	Occurrence [%]	Direction
Cluster1	A	A-Cluster1	52	1	5115	1.92	SN
Cluster2	A	A-Cluster2	52	5	2928	9.62	SN
Cluster3	A	A-Cluster3	52	0	6211	0.00	SN
Cluster4	A	A-Cluster4	52	46	0	88.46	no
Cluster1	B	B-Cluster1	39	1	4134	2.56	SN
Cluster2	B	B-Cluster2	39	35	0	89.74	no
Cluster3	B	B-Cluster3	39	1	4342	2.56	SN
Cluster4	B	B-Cluster4	39	1	2928	2.56	NS
Mix	B	B-Mix	39	1	0	2.56	no
Cluster1	C	C-Cluster1	42	39	0	92.86	no
Cluster2	C	C-Cluster2	42	1	4134	2.38	NS
Cluster3	C	C-Cluster3	42	1	535.5	2.38	SN
Cluster4	C	C-Cluster4	42	1	5115	2.38	NS
Cluster1	D	D-Cluster1	15	0	535.5	0.00	NS
Cluster2	D	D-Cluster2	15	0	4342	0.00	NS
Cluster3	D	D-Cluster3	15	13	0	86.67	no
Cluster4	D	D-Cluster4	15	2	6211	13.33	NS

Supplementary Table S 4 | Results of overall PERMANOVA. Analysis was performed with vegan package in RStudio for pairwise genetic distances of groups from different mitochondrial clades or different vent fields / sites.

Significance: ***: p-value = 0, **: < 0.001, *: < 0.01, .: < 0.05.

		Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)	
Mitochondrial clade	nucDNA cluster 1							
	annot$COI	3	0.5554	0.18512	1.5272	0.08713	0.001	***
	Residuals	48	5.8182	0.12121	0.91287			
	Total	51	6.3735	1				
	nucDNA cluster 2							
	annot$COI	2	0.3498	0.17489	1.4587	0.05513	0.001	***
	Residuals	50	5.9948	0.1199	0.94487			
	Total	52	6.3446	1				
	nucDNA cluster 3							
	annot$COI	2	0.2816	0.14078	1.1245	0.05452	0.003	**
	Residuals	39	4.8827	0.1252	0.94548			
	Total	41	5.1643	1				
	nucDNA cluster 4							
	annot$COI	2	0.34848	0.17424	1.3274	0.18116	0.01	**
	Residuals	12	1.57514	0.13126	0.81884			
	Total	14	1.92363	1				
Vent field	nucDNA cluster 1							
	annot$Field	3	0.402	0.13402	1.0323	0.06867	0.007	**
	Residuals	42	5.4526	0.12982	0.93133			
	Total	45	5.8546	1				
	nucDNA cluster 2							
	annot$Field	3	0.3859	0.12863	1.0152	0.0742	0.011	*
	Residuals	38	4.8149	0.12671	0.9258			
	Total	41	5.2008	1				
Vent site	nucDNA cluster 1 - Menez Gwen							
	annot_MG$Location	4	0.50812	0.12703	0.9771	0.19632	0.946	
	Residuals	16	2.0801	0.13001	0.80368			
	Total	20	2.58822	1				
	nucDNA cluster 1 - Lucky Strike							
	annot_LS$Location	1	0.1325	0.1325	1.0098	0.07208	0.249	
	Residuals	13	1.7058	0.13122	0.92792			
	Total	14	1.8383	1				
	nucDNA cluster 2 - Logatchev-1							
	annot_L$Location	1	0.12883	0.12884	1.0135	0.05957	0.126	
	Residuals	16	2.03397	0.12712	0.94043			
	Total	17	2.16281	1				
	nucDNA cluster 2 - Semenov-2							
	annot_S$Location	1	0.13431	0.13431	1.041	0.12946	0.013	*
	Residuals	7	0.90319	0.12903	0.87054			
	Total	8	1.0375	1				

		Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)	
	nucDNA cluster 3 - Comfortless Cove							
	annot$Location	4	0.5245	0.13111	1.0408	0.10908	0.001	***
	Residuals	34	4.2833	0.12598	0.89092			
	Total	38	4.8077	1				

Supplementary Table S 5 | Results of pairwise post-hoc PERMANOVA. Analysis was performed with pairwiseAdonis in RStudio for pairwise genetic distances of groups from different mitochondrial clades or different vent fields / sites. Bathymodiolusazoricus: Mitochondrial clade A, athymodiolusputeoserpentis: Clade B, BathymodiolusspGoldenvalley: Clade C, BathymodiolusspLilliput: Clade D.

Significance: ***: p-value = 0, **: < 0.001, *: < 0.01, .: < 0.05.

	pairs	Df	SumsOfSqs	F.Model	R2	p.value	p.adjusted	sig
Mitochondrial clade	nucDNA cluster 1							
	Bathymodiolusazoricus vs BathymodiolusspLilliput	1	0.006593558	20.123603	0.30433313	0.002	0.012	.
	Bathymodiolusazoricus vs Bathymodiolusputeoserpentis	1	0.001542449	4.63453	0.08975641	0.002	0.012	.
	Bathymodiolusazoricus vs BathymodiolusspGoldenvalley	1	0.002459228	7.43311	0.14176367	0.025	0.15	
	BathymodiolusspLilliput vs Bathymodiolusputeoserpentis	1	0.002910911	9.308768	0.75627131	0.1	0.6	
	BathymodiolusspLilliput vs BathymodiolusspGoldenvalley	1	0.000459308	2.497765	0.71410318	0.3333333	1	
	Bathymodiolusputeoserpentis vs BathymodiolusspGoldenvalley	1	0.001204433	3.193803	0.61492571	0.25	1	
	nucDNA cluster 2							
	Bathymodiolusputeoserpentis vs Bathymodiolusazoricus	1	0.005110067	13.810778	0.21643331	0.001	0.003	*
	Bathymodiolusputeoserpentis vs BathymodiolusspGoldenvalley	1	0.000537938	1.494194	0.03213722	0.105	0.315	
	Bathymodiolusazoricus vs BathymodiolusspGoldenvalley	1	0.000632731	1.375828	0.2157881	0.286	0.858	
	nucDNA cluster 3							
	BathymodiolusspGoldenvalley vs Bathymodiolusputeoserpentis	1	0.000655678	1.431494	0.03540541	0.051	0.153	
	BathymodiolusspGoldenvalley vs Bathymodiolusazoricus	1	0.001997721	4.398649	0.10374503	0.03	0.09	
	Bathymodiolusazoricus vs BathymodiolusspGoldenvalley	1	0.001047905	1.731743	0.63393341	0.3333333	1	
	nucDNA cluster 4							
	BathymodiolusspLilliput vs BathymodiolusspGoldenvalley	1	0.006667709	1.723023	0.1255571	0.076	0.152	
	BathymodiolusspLilliput vs Bathymodiolusputeoserpentis	1	0.010452642	2.701099	0.1837345	0.073	0.146	
	BathymodiolusspGoldenvalley vs Bathymodiolusputeoserpentis	1	0.00390249	NaN	1	NA	NA	
Vent field	nucDNA cluster 1							
	Lucky Strike vs Rainbow	1	0.000478285	1.2131937	0.05226311	0.162	0.972	
	Lucky Strike vs Menez Gwen	1	0.000469194	1.2101942	0.03437056	0.151	0.906	
	Lucky Strike vs Bubbylon	1	0.000383076	0.9324754	0.06244614	0.457	1	
	Rainbow vs Menez Gwen	1	0.000327976	0.887091	0.03070891	0.584	1	

	pairs	Df	SumsOfSqs	F.Model	R2	p.value	p.adjusted	sig
	Rainbow vs Bubbylon	1	0.000489642	1.3406726	0.14353063	0.215	1	
	Menez Gwen vs Bubbylon	1	0.000444977	1.1977168	0.05650216	0.147	0.882	
	nucDNA cluster 2							
	Logatchev-1 vs Broken Spur	1	0.000961811	2.032904	0.08120926	0.03	0.18	
	Logatchev-1 vs Semenov-2	1	0.000644887	1.382968	0.05241895	0.106	0.636	
	Logatchev-1 vs Irinovskoe	1	0.000468207	1.003903	0.04014984	0.337	1	
	Broken Spur vs Semenov-2	1	0.001506409	3.683109	0.20828402	0.001	0.006	*
	Broken Spur vs Irinovskoe	1	0.001021161	2.522971	0.16253146	0.01	0.06	
	Semenov-2 vs Irinovskoe	1	0.000440298	1.093898	0.06796971	0.219	1	
	nucDNA cluster 2 - Semenov-2							
	Ash Lighthouse vs Semenov-2	1	0.00868953	1.066817	0.1322476	0.036	0.036	.
Vent site	nucDNA cluster 3 - Comfortless Cove							
	Clueless vs Foggy Corner	1	0.001390356	2.9267624	0.16326219	0.002	0.02	.
	Clueless vs Golden Valley	1	0.001325925	2.6115659	0.19186374	0.009	0.09	
	Clueless vs Desperate	1	0.00118833	2.6649132	0.14277662	0.002	0.02	.
	Clueless vs Wideawake	1	0.001600441	3.1834842	0.24147518	0.003	0.03	.
	Foggy Corner vs Golden Valley	1	0.000437374	0.9169904	0.06147288	0.581	1	
	Foggy Corner vs Desperate	1	0.000451574	1.0428502	0.05203103	0.275	1	
	Foggy Corner vs Wideawake	1	0.000557003	1.1831721	0.08342084	0.148	1	
	Golden Valley vs Desperate	1	0.000531956	1.1933563	0.07369419	0.075	0.75	
	Golden Valley vs Wideawake	1	0.000436708	0.8583234	0.08706586	0.626	1	
	Desperate vs Wideawake	1	0.000736425	1.6821719	0.10726651	0.012	0.12	

Supplementary Table S 6 | Summary of mitochondrial clades, nucDNA clusters and symbiont type. mt: Mitochondrial clades are based on COI affiliation. K: number of assumed nucDNA clusters during analysis with ADMIX. Individuals with admixture proportions > 90% were assigned to a cluster, individuals with admixture proportions < 90% for any cluster where classified as mixed (mix). LS: Lucky Strike, ET: Eiffel Tower, M: Montsegur, LOG-Q: Logatchev-1 Quest, BS: Broken Spur, SEM: Semenov-2, MG: Menez Gwen, WF: White Flames, nd: no data.

Sample	Vent Site	Vent Field	mt	K = 3	K = 4	K = 5	Symbiont
1586_F	Eiffel Tower	Lucky Strike	A	1	1	1	SOX_LS-ET
1586_H	Eiffel Tower	Lucky Strike	A	1	1	1	SOX_LS-ET
1586_N	Montsegur	Lucky Strike	A	1	1	1	SOX_LS-ET
1600_F	Rainbow	Rainbow	A	1	1	1	SOX_LS-M
2065_A	Quest	Logatchev-1	B	2	2	2	SOX_LOG-Q
2065_B	Quest	Logatchev-1	B	2	2	2	SOX_LOG-Q
2065_C	Quest	Logatchev-1	B	2	2	2	SOX_LOG-Q
2424_D	Broken Spur	Broken Spur	A	mix	mix	mix	SOX_BS
2424_E	Broken Spur	Broken Spur	A	mix	mix	mix	SOX_BS
2424_H	Broken Spur	Broken Spur	B	2	2	2	SOX_BS
2424_N	Broken Spur	Broken Spur	A	mix	mix	mix	SOX_BS
2487_A	Quest	Logatchev-1	B	2	2	2	SOX_LOG-Q
2487_B	Quest	Logatchev-1	B	2	2	2	SOX_LOG-Q
2487_C	Quest	Logatchev-1	B	2	2	2	SOX_LOG-Q
2487_D	Ash Lighthouse	Semenov-2	B	2	2	2	SOX_SEM
2487_E	Ash Lighthouse	Semenov-2	B	2	2	2	SOX_SEM
3386_AJ	Broken Spur	Broken Spur	A	mix	mix	mix	SOX_BS
3722_A	Clueless	Comfortless Cove	C	mix	3	3	SOX_Clueless
3722_AH	Foggy Corner	Comfortless Cove	C	mix	3	3	SOX_Clueless
3722_AI	Foggy Corner	Comfortless Cove	C	mix	3	3	SOX_Clueless
3722_AJ	Foggy Corner	Comfortless Cove	C	mix	3	3	SOX_Clueless
3722_AP	Golden Valley	Comfortless Cove	C	mix	3	3	SOX_Clueless
3722_AY	Irina II	Logatchev-1	B	2	2	2	SOX_LOG-Q
3722_BW	Lilliput	Lilliput	D	3	4	4	SOX_Lilliput
3722_C	Clueless	Comfortless Cove	C	3	4	4	SOX_Clueless
3722_CD	Lilliput	Lilliput	D	1	1	5	SOX_Lilliput
3722_CI	Montsegur	Lucky Strike	A	2	2	2	SOX_LS-M
3722_CQ	Quest	Logatchev-1	B	mix	3	3	SOX_LOG-Q
3722_J	Desperate	Comfortless Cove	C	mix	3	3	SOX_Clueless
3722_K	Desperate	Comfortless Cove	C	mix	3	3	SOX_Clueless
3722_L	Desperate	Comfortless Cove	C	mix	3	3	SOX_Clueless
3722_N	Desperate	Comfortless Cove	C	mix	3	3	SOX_Clueless
3722_O	Desperate	Comfortless Cove	C	mix	3	3	SOX_Clueless
3722_P	Desperate	Comfortless Cove	C	mix	3	3	SOX_Clueless
3722_Q	Desperate	Comfortless Cove	C	mix	3	3	SOX_Clueless
3722_R	Desperate	Comfortless Cove	C	mix	3	3	SOX_Clueless
3722_S	Desperate	Comfortless Cove	C	mix	3	3	SOX_Clueless

Sample	Vent Site	Vent Field	mt	K = 3	K = 4	K = 5	Symbiont
3722_T	Desperate	Comfortless Cove	C	mix	3	3	SOX_Clueless
3722_W	Eiffel Tower	Lucky Strike	A	1	1	5	SOX_LS-M
3722_Y	Eiffel Tower	Lucky Strike	A	1	1	1	SOX_LS-ET
3723_AB	Station 719-10	Menez Gwen	A	1	1	5	SOX_LS-M
3723_AC	Station 729-4	Menez Gwen	A	1	1	5	SOX_MG-WF
3723_AD	Station 729-4	Menez Gwen	A	1	1	5	SOX_LS-M
3723_AE	Station 736-3	Menez Gwen	A	1	1	1	SOX_LS-M
3723_AS	Wideawake	Comfortless Cove	C	mix	3	3	SOX_Clueless
3723_AV	Wideawake	Comfortless Cove	C	mix	3	3	SOX_Clueless
3723_BA	Woody	Menez Gwen	A	1	1	5	SOX_MG-WF
3723_BE	Woody	Menez Gwen	A	1	1	1	SOX_LS-M
3723_BF	Woody	Menez Gwen	A	1	1	5	SOX_LS-ET
3723_BI	Woody	Menez Gwen	A	1	1	5	SOX_LS-ET
3723_BJ	Woody	Menez Gwen	A	1	1	1	SOX_MG-WF
3723_BL	Woody	Menez Gwen	A	1	1	1	SOX_LS-M
3723_BM	Woody	Menez Gwen	A	1	1	5	SOX_MG-WF
3723_BN	Woody	Menez Gwen	A	1	1	5	SOX_LS-ET
3723_B	Rainbow	Rainbow	A	1	1	1	SOX_MG-WF
3723_C	Rainbow	Rainbow	A	1	1	5	SOX_LS-M
3723_E	Rainbow	Rainbow	A	1	1	5	SOX_LS-M
3723_G	Rainbow	Rainbow	A	1	1	1	SOX_MG-WF
3723_H	Rainbow	Rainbow	A	1	1	5	SOX_LS-M
3723_R	Ash Lighthouse	Semenov-2	B	2	2	2	SOX_LOG-Q
3723_S	Ash Lighthouse	Semenov-2	B	2	2	2	SOX_SEM
3723_T	Ash Lighthouse	Semenov-2	B	2	2	2	SOX_SEM
3723_V	Station 719-10	Menez Gwen	A	1	1	5	SOX_LS-M
3723_W	Station 719-10	Menez Gwen	A	1	1	1	SOX_LS-ET
3723_X	Station 719-10	Menez Gwen	A	1	1	5	SOX_MG-WF
3723_Y	Station 719-10	Menez Gwen	A	1	1	1	SOX_LS-M
3723_Z	Station 719-10	Menez Gwen	A	1	1	1	SOX_MG-WF
4025_AA	Eiffel Tower	Lucky Strike	A	1	1	1	SOX_LS-M
4025_AB	Broken Spur	Broken Spur	B	2	2	2	SOX_BS
4025_AC	Eiffel Tower	Lucky Strike	A	1	1	1	SOX_LS-ET
4025_AE	Golden Valley	Comfortless Cove	C	mix	3	3	SOX_Clueless
4025_AF	Montsegur	Lucky Strike	A	1	1	1	SOX_LS-ET
4025_AG	Quest	Logatchev-1	B	2	2	2	SOX_LOG-Q
4025_AH	Quest	Logatchev-1	B	2	2	2	SOX_LOG-Q
4025_AI	Rainbow	Rainbow	A	1	1	5	SOX_MG-WF
4025_AJ	Broken Spur	Broken Spur	B	mix	mix	mix	SOX_BS
4025_AK	Golden Valley	Comfortless Cove	C	mix	3	3	SOX_Clueless
4025_AL	Irina II	Logatchev-1	B	2	2	2	SOX_LOG-Q
4025_AO	Broken Spur	Broken Spur	A	mix	mix	mix	SOX_BS
4025_AQ	Semenov-2	Semenov-2	B	2	2	2	SOX_SEM
4025_AT	Semenov-2	Semenov-2	B	2	2	2	SOX_SEM
4025_AU	Golden Valley	Comfortless Cove	C	mix	3	3	SOX_Clueless

Sample	Vent Site	Vent Field	mt	K = 3	K = 4	K = 5	Symbiont
4025_AX	Irina II	Logatchev-1	B	2	2	2	SOX_LOG-Q
4025_AY	Irinovskoe	Irinovskoe	B	2	2	2	SOX_LOG-Q
4025_BA	Broken Spur	Broken Spur	B	2	2	2	SOX_BS
4025_BB	Broken Spur	Broken Spur	B	2	2	2	SOX_BS
4025_BC	Broken Spur	Broken Spur	B	2	2	2	SOX_BS
4025_BE	Montsegur	Lucky Strike	A	1	1	1	SOX_LS-ET
4025_BF	Broken Spur	Broken Spur	A	mix	mix	mix	SOX_BS
4025_BG	Broken Spur	Broken Spur	B	2	2	2	SOX_BS
4025_BH	Broken Spur	Broken Spur	A	mix	mix	mix	SOX_BS
4025_BK	Bubbylon	Bubbylon	A	1	1	5	SOX_LS-ET
4025_BL	Broken Spur	Broken Spur	A	mix	mix	mix	SOX_BS
4025_B	Montsegur	Lucky Strike	A	2	2	2	SOX_LS-ET
4025_BM	Quest	Logatchev-1	B	mix	mix	mix	SOX_LOG-Q
4025_BN	Broken Spur	Broken Spur	B	2	2	2	SOX_BS
4025_BO	Broken Spur	Broken Spur	B	2	2	2	SOX_BS
4025_BP	Broken Spur	Broken Spur	B	mix	mix	mix	SOX_BS
4025_BR	Broken Spur	Broken Spur	A	2	2	2	SOX_BS
4025_BS	Broken Spur	Broken Spur	B	1	1	1	SOX_BS
4025_BT	Broken Spur	Broken Spur	A	mix	mix	mix	SOX_BS
4025_E	Eiffel Tower	Lucky Strike	A	1	1	1	SOX_MG-WF
4025_H	Lilliput	Lilliput	D	3	4	4	SOX_Lilliput
4025_I	Semenov-2	Semenov-2	B	2	2	2	SOX_SEM
4025_J	Irinovskoe	Irinovskoe	B	2	2	2	SOX_LOG-Q
4025_L	Golden Valley	Comfortless Cove	C	mix	3	3	SOX_Clueless
4025_M	Quest	Logatchev-1	B	2	2	2	SOX_LOG-Q
4025_N	Ash Lighthouse	Semenov-2	B	2	2	2	SOX_SEM
4025_Q	Broken Spur	Broken Spur	B	2	2	2	SOX_BS
4025_R	Quest	Logatchev-1	B	2	2	2	SOX_LOG-Q
4025_U	Broken Spur	Broken Spur	A	mix	mix	mix	SOX_BS
4025_V	Eiffel Tower	Lucky Strike	A	1	1	1	SOX_MG-WF
4132_AA	Wideawake	Comfortless Cove	C	mix	3	3	SOX_Clueless
4132_A	Broken Spur	Broken Spur	B	mix	3	3	nd
4132_AB	Wideawake	Comfortless Cove	C	mix	3	3	SOX_Clueless
4132_AC	Wideawake	Comfortless Cove	C	1	1	1	SOX_Clueless
4132_AE	Woody	Menez Gwen	A	1	1	1	SOX_MG-WF
4132_AF	Woody	Menez Gwen	A	1	1	1	SOX_LS-M
4132_AH	Rainbow	Rainbow	A	1	1	1	SOX_LS-M
4132_AI	Rainbow	Rainbow	A	1	1	5	SOX_MG-WF
4132_AJ	Rainbow	Rainbow	A	2	2	2	SOX_MG-WF
4132_AK	Ash Lighthouse	Semenov-2	B	1	1	1	SOX_SEM
4132_AL	Woody	Menez Gwen	A	2	2	2	SOX_MG-WF
4132_AO	Irina II	Logatchev-1	B	3	4	4	SOX_LOG-Q
4132_AQ	Lilliput	Lilliput	D	1	1	5	SOX_Lilliput
4132_AR	Eiffel Tower	Lucky Strike	A	2	2	2	SOX_MG-WF
4132_AS	Irinovskoe	Irinovskoe	B	2	2	2	SOX_LOG-Q

Sample	Vent Site	Vent Field	mt	K = 3	K = 4	K = 5	Symbiont
4132_AT	Irina II	Logatchev-1	B	2	2	2	SOX_LOG-Q
4132_AV	Lilliput	Lilliput	D	3	4	4	SOX_Lilliput
4132_AX	Irinovskoe	Irinovskoe	B	2	2	2	SOX_SEM
4132_AY	Golden Valley	Comfortless Cove	C	mix	3	3	SOX_Clueless
4132_AZ	Irinovskoe	Irinovskoe	B	2	2	2	SOX_LOG-Q
4132_BA	Broken Spur	Broken Spur	B	2	2	2	SOX_BS
4132_BB	Broken Spur	Broken Spur	B	2	2	2	SOX_BS
4132_BC	Station 719-10	Menez Gwen	A	1	1	1	SOX_LS-M
4132_BD	Lilliput	Lilliput	D	3	4	4	SOX_Lilliput
4132_BE	Irinovskoe	Irinovskoe	B	2	2	2	SOX_LOG-Q
4132_BF	Eiffel Tower	Lucky Strike	A	1	1	1	SOX_LS-M
4132_B	Foggy Corner	Comfortless Cove	C	2	2	2	SOX_Clueless
4132_BG	Broken Spur	Broken Spur	B	1	1	1	SOX_BS
4132_BH	Eiffel Tower	Lucky Strike	A	mix	3	3	SOX_MG-WF
4132_BI	Foggy Corner	Comfortless Cove	C	mix	3	3	SOX_Clueless
4132_C	Foggy Corner	Comfortless Cove	C	mix	3	3	SOX_Clueless
4132_D	Foggy Corner	Comfortless Cove	C	mix	3	3	SOX_Clueless
4132_E	Foggy Corner	Comfortless Cove	C	mix	3	3	SOX_Clueless
4132_F	Foggy Corner	Comfortless Cove	C	mix	3	3	SOX_Clueless
4132_G	Foggy Corner	Comfortless Cove	C	mix	3	3	SOX_Clueless
4132_H	Foggy Corner	Comfortless Cove	C	mix	3	3	SOX_Clueless
4132_I	Irina II	Logatchev-1	B	2	2	2	SOX_LOG-Q
4132_J	Irina II	Logatchev-1	B	2	2	2	SOX_LOG-Q
4132_K	Irina II	Logatchev-1	B	2	2	2	SOX_LOG-Q
4132_L	Irinovskoe	Irinovskoe	B	2	2	2	SOX_SEM
4132_M	Irinovskoe	Irinovskoe	B	2	2	2	SOX_SEM
4132_N	Lilliput	Lilliput	D	3	4	4	SOX_Lilliput
4132_O	Lilliput	Lilliput	D	3	4	4	SOX_Lilliput
4132_P	Lilliput	Lilliput	D	3	4	4	SOX_Lilliput
4132_Q	Lilliput	Lilliput	D	3	4	4	SOX_Lilliput
4132_R	Montsegur	Lucky Strike	A	1	1	1	SOX_LS-M
4132_S	Quest	Logatchev-1	B	2	2	2	SOX_LOG-Q
4132_U	Clueless	Comfortless Cove	C	mix	3	3	SOX_Clueless
4132_V	Desperate	Comfortless Cove	C	mix	3	3	SOX_Clueless
4132_W	Eiffel Tower	Lucky Strike	A	1	1	1	SOX_MG-WF
4132_X	Eiffel Tower	Lucky Strike	A	1	1	5	SOX_LS-M
4132_Z	Wideawake	Comfortless Cove	C	mix	3	3	SOX_Clueless
864_C	White Flames	Menez Gwen	A	1	1	1	SOX_LS-M
C_11-2R	Clueless	Comfortless Cove	C	mix	3	3	SOX_Clueless
C_11-3R	Clueless	Comfortless Cove	C	mix	3	3	SOX_Clueless
C_11-4R	Clueless	Comfortless Cove	C	mix	3	3	SOX_Clueless
C_11-5R	Clueless	Comfortless Cove	C	mix	3	3	SOX_Clueless
C_11-6R	Clueless	Comfortless Cove	C	mix	3	3	SOX_Clueless
L_10-2R	Lilliput	Lilliput	D	3	4	4	SOX_Lilliput
L_10-4R	Lilliput	Lilliput	D	3	4	4	SOX_Lilliput

Sample	Vent Site	Vent Field	mt	K = 3	K = 4	K = 5	Symbiont
L_10-5R	Lilliput	Lilliput	D	3	4	4	SOX_Lilliput
L_5-1R	Lilliput	Lilliput	D	3	4	4	SOX_Lilliput
L_5-4R	Lilliput	Lilliput	D	3	4	4	SOX_Lilliput

synthesis

[ˈsɪnθɪsɪs] *noun*

the combination of components or elements to form a connected whole

Chapter V | Discussion, future directions, and concluding remarks

Bathymodiolus mussels are foundation species and play a crucial role in hydrothermal vent ecosystems worldwide. Thanks to chemosynthetic symbionts, they can thrive in environments with little to no phototrophic carbon and grow to large mussel beds that provide shelter and habitats to many other vent species.

Since their discovery, *Bathymodiolus* mussels have been well studied. Many scientific advances, be it deep-sea or sequencing technologies, also enhanced our understanding of this symbiotic association and its complexity. In recent years, we learned that the symbiosis can consist of up to five partners on a species level [1] and within some of the species, the diversity on strain level is remarkable [2,3]. Despite the great knowledge on the *Bathymodiolus* symbioses, many fundamental questions remained unresolved. What drives the symbiont community composition? Where do the symbionts come from? How are mussel populations structured and the gene flow maintained?

In this thesis, I investigated symbionts in a mussel hybrid zone, the bacterial community in the environment, and the population genomics of the hosts to better understand the three aspects of the association: the mussels, its symbionts, and their free-living stage. The comparison of nuclear and mitochondrial genome information revealed an intriguing discordance that might be explained by larval migration over large geographical distances (chapter IV). The population genomics results also confirmed the presence of a hybrid zone at the Broken Spur vent field from which I investigated symbionts of hybrid and parental mussels. It seems that location rather than host genetics is driving the symbiont composition in these mussels (chapter II). While the data at hand could not deliver final proof of the existence of a free-living symbiont stage in the water column, it contributed to a better understanding of symbiont transmission and enhanced our knowledge about the geographical distribution of symbiont subspecies in the environment (chapter III).

One great paradox in science is the number of unresolved questions growing as quickly as knowledge itself. This thesis is no exception. In each study, I experienced how investigating one aspect of a symbiotic association directly links to the other aspects, opens up many questions, and poses new challenges, some of which I will discuss in this section.

Comparison of population structure in symbionts and their mussel hosts

Deep-sea hydrothermal vents are scattered along the Mid-Atlantic Ridge (MAR). These environments likely favour animal species with high dispersal and growth rates, and good abilities to colonise, given the large distances between vent fields [4]. Analysis of population structure in these species can reveal much about their population connectivity and evolutionary history. Symbiosis with horizontally transmitted symbionts adds another layer of complexity to the colonisation and distribution of *Bathymodiolus* mussels in the deep sea. *Bathymodiolus* larvae are symbiont-free [5]. The mussels depend on the availability of suited symbiotic bacteria at their site of settlement, besides all the challenges that come with their own dispersal. In my thesis, I analysed the population structure of *Bathymodiolus* mussels (chapter IV) and their symbionts (chapter II). I advanced the knowledge about the geographic ranges of *Bathymodiolus* mussels at the MAR, and what factors are driving the symbiont composition in these mussels. For the host side, I used more than 600 000 exome-wide SNPs for genetic profiling. Even with this resolution, there was no clear subpopulation structure in conspecific mussels from the same geographic region (chapter IV). This implies that gene flow is frequent between mussel populations at different vent sites in the same region. Investigating symbionts from a *Bathymodiolus* hybrid zone, I showed that host genetics have little or no effect on the symbiont composition, even at strain level. Instead, I found that the mussel-associated symbionts were geographically structured (chapter II).

My results contributed new insights on host and symbiont population structure, and revealed the discrepancies in the geographical structuring of both partners. There are several explanations why the patterns between symbiont and host are different. A possible explanation is that hydrothermal vent fields are separated by barriers that hinder the dispersal of marine bacteria but are permissive for mussel larvae. Rift valley walls are often high [6] and could hinder symbiont dispersal. Planktotrophic mussel larvae might be able to overcome such barriers depending on their mode of migration. If larvae at the MAR migrated vertically as described for *'B.' childressi* larvae in the Gulf of Mexico [7], they probably can surmount such topological features. Another result of the topology at the MAR is that buoyant plumes from high-temperature vents are often contained within their rift valley, as the valleys are deeper than the height of the plume [8]. The restriction of hydrothermally influenced water to the vent site is resulting in low concentrations of reduced chemicals in the water column between vent fields. The low concentrations could pose a great challenge to the chemosynthetic symbionts. In contrast, the larvae do not depend on the availability of

chemicals. They do not yet have symbionts but completely rely on filter feeding [5]. These examples illustrate how the geographical setting could pose different dispersal barriers to host and symbionts and thus, influence their population structure differently.

Another explanation for the different population structure in symbionts and hosts is that the symbionts are not structured by the geography per se but rather by the environmental conditions at a given site. Although redundancy analysis indicated that vent type, as a proxy for environmental parameters, had little to no explanatory power of symbiont variation (chapter II), more detailed environmental parameters could contribute to the selection of specific symbionts at each site as shown for other marine bacteria [9,10]. Studies of symbiont strain diversity in *Bathymodiolus* mussels showed that functional variation of the strains correlated to environmental parameters such as the availability of energy sources [2]. The host, in turn, seems to be rather unaffected by short-term changes in the environment as revealed by recent transcriptomic studies [11]. The dependency on resources for chemosynthesis can be a reason why the pressure to adapt is much higher for symbionts than for their host. This could lead to site-specific variation solely in the symbionts, as observed in this thesis.

We need better knowledge about the chemical and physical parameters at and between vent sites to resolve what is driving population structure in vent animals and their symbionts. These parameters can include temperature, flow of hydrothermal fluids, deep-sea water currents, pH, or chemical seawater composition. In addition, we require more studies on the marine bacterial community with higher resolution of symbionts in the free-living stage as proposed in chapter III and parts of this discussion.

Composition and heterogeneity of local mussel populations at the MAR

Analysis of population structure, larvae dispersal, and the geographic ranges of *Bathymodiolus* mussels can contribute to a better understanding of their ecological success. In chapter IV, I analysed *Bathymodiolus* mussels from 10 vent fields in four geographic regions of the MAR. I observed that mussels within the same region are genetically similar while I also found indications for migration of mussels over large geographical distances. This is interesting because if larvae could disperse over long distances regularly, it would be surprising that the majority of the population at a site seems to be genetically similar and "foreign" genotypes the exception. One potential explanation for this pattern would be post-settlement events that are detrimental for foreign mussels, e.g. interaction between

established adults and just-settled larvae [12]. Unfortunately, we do currently have very limited knowledge about the composition and genetic heterogeneity of mussels within one mussel bed that could help resolving this question. In the future, comparative analysis of mussels from different life stages, e.g. larvae, small, medium, and large mussels, could contribute to inferring the colonisation history of a site. Analysing the genetic differentiation across life stages would allow us to investigate whether the population heterogeneity is stable, or changes with age of the mussels. As an example, if foreign genotypes were common in larvae or very small post-settlement mussels compared to larger (older) individuals, this could be indicative of selection mechanisms or some fitness disadvantages of the non-local genotype.

Another motivation to assess the mussel bed composition and genetic heterogeneity is the absence of hybrids and first generation migrants in my population genomic study (chapter IV). I observed mitonuclear discordance, i.e. incongruent information between the mitochondrial and nuclear genome, in 10 % of the investigated mussels. Mitochondrial genomes were always location-specific regardless of the nuclear genome information. The result is best explained by long-distance migration and mitochondrial introgression. Assuming migration, one would also expect to find first-generation migrants. These would be individuals with location-unspecific mitochondrial and nuclear genomes. In addition, larvae might not be able to cross distances between NMAR and SMAR within one generation. A stepwise migration would suggest that hybrids are present at locations other than Broken Spur, which was not observed in my study with one exception in Logatchev. This raises the question whether the first-generation migrants and hybrids are indeed absent and other biological processes underlie the observed pattern. An alternative is that first generation migrants and hybrids were underrepresented in our dataset. Given the knowledge gap on the genetic heterogeneity of individuals at a site, I suggest sampling different patches within and at the edges of a mussel bed as well as scattered individuals close to the mussel bed. Metagenomic reads of these mussels can be investigated using the established pipeline to compare their genetic differentiation and shed some light on the heterogeneity of host individuals at a site.

Lastly, the mussel's location within the mussel bed can have an influence on the symbiont population within individual hosts. At the MAR, symbiont populations from mussels located closely together were more similar than compared to those that came from different patches [2]. Recent studies revealed that the mussel density might be another factor influencing the

symbiont population structure. In *B. brooksi* from the Gulf of Mexico, symbiont populations between mussel individuals were genetically isolated [3]. The authors suggested that symbiont colonisation from the environment happens in early life stages followed by self-infection of newly formed gills with symbionts from older tissues. Continuous uptake of symbionts from the environment is limited in this scenario. In contrast, the SOX symbiont populations of mussels at the MAR were highly similar between mussel individuals from the same site. This indicates ongoing intermixing of the symbionts [2]. One potential explanation is that the mussel density at the cold seep in the Gulf of Mexico was much lower than at hydrothermal vents at the MAR [2,3,13]. Assuming that symbionts can be released for colonisation of other hosts, few mussels at a site would lead to low symbiont abundance in the environment. This would result in lower colonisation by environmental symbionts and a prevalence for self-infection. These studies indicate that the composition of the mussel bed can be of great importance for the dynamics of the associated symbiont populations. This becomes even more striking when we consider the fact that symbiont and host DNA always co-occurred in water metagenomes analysed in chapter III. Based on these results, I hypothesised that symbionts might be regularly transmitted via host particles, a scenario in which proximity of the hosts is key for efficient transmission. However, metadata about the distribution of sampled mussels from a site is often missing or lacks detail. Thorough documentation of the spatial distribution of mussel individuals and quantitative assessment of the mussel density can contribute to linking symbiont population dynamics to the environmental setting.

The potential complexity of Bathymodiolus mitochondria

Since the mitochondrial cytochrome c oxidase I (COI) gene was first proposed for species assignment of animals in 2002 [14], the COI and other mitochondrial markers such as NADH dehydrogenase 4 were widely used to investigate the relations between different animal species. *Bathymodiolus* mussels were studied worldwide using mitochondrial markers to infer the mussels' genetic connectivity, population structure, phylogenies, and evolutionary history [15–21]. I discovered discrepancies in the mitochondrial and nuclear genome information comparing mitochondrial clades to exome-wide SNP data. This indicates that mitochondrial DNA (mtDNA) and nuclear DNA follow different patterns of inheritance. Such discrepancies have been observed in many species including wolf spiders [22], nudibranchs [23], and Australian birds [24], highlighting the need to complement mtDNA by analysis of the nuclear genome for species determination in many species.

Despite the regular use of mitochondrial markers for phylogenetic analyses in *Bathymodiolus* mussels, little is known about the mitochondrial inheritance and genome structure. This is interesting because the family of Mytilidae, to which *Bathymodiolus* mussels belong, are one of the best-described exceptions from the strictly maternal inheritance of mitochondria in animals [25,26]. In shallow-water *Mytilus* mussels, doubly uniparental inheritance (DUI) of mitochondria is common [27] (Figure 1). This means that two types of mtDNA were found to be associated with sex-specific routes of inheritance: The F mtDNA is transmitted from mothers to offspring of both sexes while the M mtDNA is only transmitted from fathers to sons [28–30]. The two types are highly divergent (> 20 % [29]) and evolve at faster rates than expected for animal mtDNA [31]. It has been hypothesised that the sex determination of mussels is linked to the presence of M mtDNA [32]. The influence of the mtDNA on the sex-determination of the developing mussel has implications for the population dynamics, e.g. in species with asymmetric mating preferences. As an example, females of species A select against males of species B, i.e. they do not mate with them. In contrast, females of species B are less strictly selecting against males of species A. In these cases, it becomes relevant for processes such as hybridisation and introgression between species whether an individual is male or female. Further, DUI can influence the genome evolution of mitochondria as the co-existence of F and M mtDNA in the male embryo makes mtDNA recombination more likely [27]. Besides potential recombination, many bivalve mitochondria including those of Mytilidae have remarkable genetic features. These features include extensive repeat regions, complex genome architecture, regions without assigned functional products (RNA or protein), duplicated genes, introns, additional coding genes, and genetic elements [33–35]. Some of the mitochondrial genome characteristics have been proposed to be linked to DUI [36–39].

These bivalve genome characteristics are intriguing when we consider that my attempts to assemble complete mitochondrial genomes from metagenomic reads were successful in only 50 % of the 175 mussel metagenomes (chapter IV). One of the reasons why many of the assemblies were of insufficient quality for tree reconstruction was that a high number of short repeat regions split the contigs in the assembly graph. First screening of the mitochondrial sequences revealed that gene duplications and pseudogenes might also be present in *Bathymodiolus* mitochondrial genomes. An alternative explanation for why the assembly did not work for the whole dataset are mitochondrial-derived sequences that have been incorporated within the nuclear genome for a long time (nuclear mitochondrial DNA or

NUMTs, [40–42]). NUMTs can be long insertions or partial pseudogenes in varying copy numbers that are difficult to separate from nuclear and mitochondrial genes in short-read sequencing data. Some of them have been reported to be flanked by repeat regions in eukaryotic nuclear genomes [43]. NUMTs could explain why 50 % of mitochondrial assemblies in this study were unsuccessful.
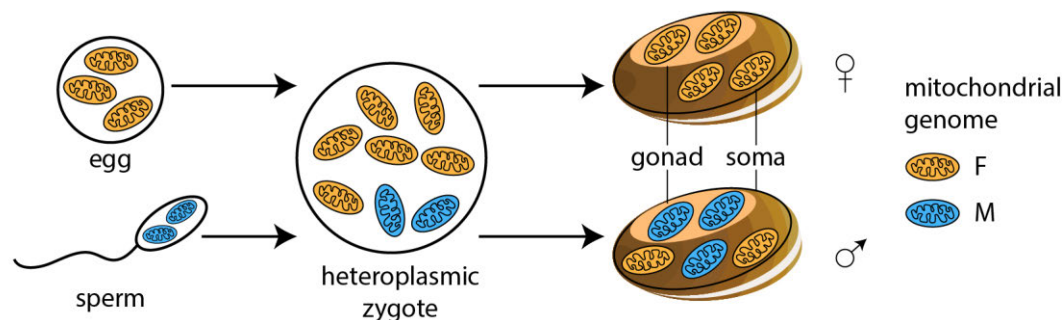
Figure 1 | Doubly uniparental inheritance in *Mytilus* spp. Two divergent mitochondrial genome types are passed on by either the female egg (F type) or the male sperm (M type) to form heteroplasmic zygotes. The M mitochondrial genome is eliminated from embryos that develop into female mussels. Females carry the F type mitochondria in both gonads and soma. Embryos were the M type is retained develop into male mussels. These are heteroplasmic with F and M type mitochondrial genomes in the soma but the gonad only carries the M type.

Altogether, processes such as DUI or recombination can affect phylogenetic and population genetic analyses [25]. So far, no study investigating the presence of different mtDNA types or their inheritance in *Bathymodiolus* mussels has been published. First indications were found that *B. thermophilus* might have different mitochondrial types that are not associated with sex (Maas et al., in prep, cited by [44]). In chapter IV, I suggested different explanations for the observed mitonuclear discordance in *Bathymodiolus* mussels at the MAR. Before investigating processes such as androgenesis, we need a basic understanding of mitochondrial inheritance in *Bathymodiolus* mussels, and whether it is as complex as in their shallow-water relatives. I propose to analyse mitochondrial sequences of mature *Bathymodiolus* mussels to check for divergent mtDNA genomes [45]. In adult mussels, the M type dominates the male germline while the somatic tissue is heteroplasmic to a varying degree and rather dominated by the F type [46]. For detection of the divergent F and M mtDNA, I therefore suggest to analyse the gonads. I recommend to couple any investigation of the mitochondrial inheritance

with more extensive bioinformatics analyses of the mitochondrial genome as some of the genome characteristics outlined above were likely associated with DUI.

The chase goes on

The free-living stage of symbionts is highly interesting to investigate because it has implications for many different aspects of the symbiosis. It could reveal more profound knowledge on the transmission, the specificity of the uptake, and the geographical structuring of these marine bacteria. I applied three diagnostic markers to differentiate symbiont subspecies and to assess their presence in water and mussel metagenomes. Such analyses on the distribution of similar but distinct symbionts across geography and lifestyles, i.e. host-associated or free-living, lay the groundwork to hypothesise about potential drivers of bacterial biogeography and to design future experiments investigating the symbiont uptake.

Expansion of marker gene studies in water metagenomes

I presented evidence that several symbiont subspecies might be present in water and mussel gill metagenomes in different abundances through investigation of the marine bacterial community from different hydrothermal vents at which mussels occur (chapter III). This indicates that more than one symbiont subspecies might be present at each site but that only one type grows to larger abundances, pointing towards an 'everything is everywhere, but the environment selects' scenario [47,48]. While my results provide indications for the co-occurrence of several subspecies at a site, the coverage of each marker gene was very low and the analysis should be expanded. The coverage has to be increased at least 20-fold for reliable results. Ideally, replicate samples should be analysed for each site to account for potential differences during sampling. More diagnostic markers should be applied to the metagenomic data to further enhance the reliability of the results. The screening for marker genes ideally includes all orthologous genes present in the symbionts instead of focusing on marker genes specific to a certain taxonomic lineage, which was used as a starting point in chapter III. Genetic sequences that were previously found to have a high variability in these symbionts, such as restriction modification system genes [49] or CRISPR-Cas [2] genes, could be another angle to obtain a larger number of suitable markers. This might be especially interesting because they were shown to be enriched in symbiotic compared to free-living bacteria from the *Thioglobaceae* [50].

At this stage, marker genes are the most practical approach to investigate the free-living symbiont community, as they are independent of assembly and binning. These procedures

require high sequencing coverage of the target organisms. Such high coverage is hard to obtain in metagenomes with high diversity such as those derived from water samples. Therefore, assembly-based approaches are prone to loosing information about low abundant taxa. When using marker genes, we need to be aware of what they can tell us and where they are limited, especially considering the complexity of the symbiont community on strain level. The symbiont subspecies of interest are highly similar with values of 97 % average nucleotide identity but still distinguishable based on phylogenomics (chapter II, [50]). Within the subspecies, the symbiont population is not clonal but consists of strains with varying gene content and functional capabilities [2,3,51]. Strain-specific genes are detected based on their rather low coverage. If not all strains in a population carry a gene, this gene has lower coverage in sequencing data compared to genes encoded by the whole population [2]. These strain-specific genes are of ecological importance as they have been shown to vary according to environmental parameters. In my study, I did not focus on gene presence and absence but on different sequence types of the same genes. The sequences were derived from symbiont MAGs that represent a consensus assembly of all strains in the population [52]. The markers were chosen to target symbiont subspecies but we cannot distinguish between gene sequences that are specific to a certain symbiont subspecies and those that are specific to a certain symbiont strain within this subspecies. If the marker genes applied in this thesis were strain-specific, these would still show the observed pattern that one marker gene sequence type is dominant at a given site. This is based on the assumption that one of the strains occurs in higher abundance than all other strains at this site. In that case, the low abundant, unexpected sequence type would indicate the presence of non-dominant strains matching this gene sequence. Analysis of strain diversity [2] in mussels compared to the environment could give further insight to understand the relationship between the free-living and symbiotic stage. However, closely related but non-symbiotic bacteria in the water column might bias such analysis.

One advantage of the marker gene approach is that once the markers are established, these could be applied in laboratory-based studies. The spatial distribution of these marker gene sequence types in mussel gills could be investigated using fluorescence *in situ* hybridisation (FISH) with simultaneous visualisation of two probes matching the two marker gene sequence types (geneFISH, [53]) and one probe matching the symbiont 16S rRNA gene [54–57]. This would show whether both types are present in the tissue, their abundances, and their distribution. Similarly, the ribosomal and the marker gene probes could be applied to water

filters to demonstrate the presence of a potential free-living stage of symbionts visually. FISH cannot be applied to the water filters investigated in this thesis as the biomass is far too high due to the large volumes of water that have been filtered. If water was filtered on upcoming cruises, filtering smaller volumes of seawater, and the resulting decrease in biomass, would allow the application of imaging approaches. Once one has a better idea of how specific the markers are and whether only one is present per cell, these could also be used as baits for fluorescence-activated cell sorting (FACS) to obtain fractions of the cells carrying either one or the other marker gene sequence type [58,59]. Targeted sequencing of different subspecies or strains, depending on what the markers are specific to, could be used to assemble their genomes. This would allow the analysis of the highly similar but still different symbiont subspecies or strains in a way that is not feasible with metagenome-assembled genomes (MAGs) representing consensus assemblies from all strains [52].

Aquaria experiments to investigate symbiont uptake

Laboratory experiments can help to better understand the actual symbiont uptake and its specificity. *Bathymodiolus* larvae have been recently shown to be free of symbionts [5]. The colonisation only starts after larvae settlement. To investigate the specificity of symbiont uptake, symbiont-free larvae could be exposed to different bacterial seawater communities. However, obtaining larvae of deep-sea animals is challenging and genomic analyses of the experiment are difficult due to the low DNA yield of the tiny larvae. *Bathymodiolus* mussels were shown to spawn in aquaria, but larvae from aquaria kept mussels showed abnormal development and high lethality [60,61]. In contrast, adult mussels could be maintained in aquaria over longer periods of time [11,60,62,63]. Starvation experiments showed that with reduced availability of chemical compounds needed for chemosynthesis, the amount of symbionts decreased drastically [11,64].

The observed symbiont reduction in starvation experiments and the fact that symbionts are likely acquired throughout the mussel's lifetime [2,65] provide the basis for the following aquaria experiment (Figure 2). The aim of the experiment is to investigate the symbiont uptake with living mussels. In the experimental setup, mussels from two different sites, or ideally two different species, need to be maintained in aquaria. After a period marked by abundance of energy sources, symbiont loss is induced by lowering the concentrations. Mussels from one tank, i.e. one location or species, are placed in the other tank, i.e. the other location or species, and vice versa before increasing substrate concentrations again. Batches of mussels are sampled before decreasing the substrate, after a period of starvation, and after

switching tanks and increasing the substrate again. Two different approaches are thinkable: (1) all mussels from one tank are moved to the other tank, so they are only exposed to the water that the mussels from a different location or species were maintained in, or (2) only half of the mussels are moved so that they are exposed to both the water and neighbouring mussels from the other location or species. This might provide an insight into whether the community in the water or the presence of neighbouring hosts is more important for symbiont transmission. As controls, one needs to maintain tanks with mussels that experience high, stable substrate concentrations throughout the whole experiment and those that experience the starvation step but without switching the tank. Sampled mussels should be analysed with metagenomics, metatranscriptomics, and metabolic and microscopic imaging. In addition, the water should be sampled at the same time points. This allows for comparison of strain diversity in- and outside the mussels, which is important to analyse in case not all symbionts are removed through starvation. Even if some of the original symbionts are retained within in the gills, such comparison of the symbiont populations on strain level can reveal the importance of self-infection compared to environmental acquisition of symbionts. In addition, the analysis of aquaria water samples can shed light on how the *Bathymodiolus* symbiosis influences the marine bacterial community in the environment as previously investigated for *Riftia* tubeworms [66,67].

Various challenges need to be addressed to ensure that such an experiment delivers the desired insights. Challenges of the setup can include the number and size of tanks, the maintenance of stable and non-toxic conditions, and the pressure. In previous aquaria experiments, *Bathymodiolus* mussels were kept in 20, 40 or 50 l tanks at densities of 1 [68], 0.75 [64] or 1.67 mussels/l [62]. This shows that the tanks were rather large, however, many replicates could be maintained in the same tank. The facility, in which the experiment is conducted, should have access to filtered seawater. Changing the water frequently, e.g. daily or later on every 7-10 days, and supplying aeration through air diffusers avoids the accumulation of pseudo-faeces and toxic compounds such as ammonium or hydrogen-sulphide, and oxygen depletion [62,64]. Light and temperature should resemble *in situ* conditions, e.g. 8-10 °C in the dark at around pH 7-9 and 40-60 % oxygen saturation [62,64,68]. Oxygen, substrate concentrations, temperature, and pH should be measured daily to ensure stable conditions throughout the experiment. When the mussels lose their symbionts, they also lose access to their constant nutrient supply. Additional feeding with *Rhodomonas* spp. algae or a food mixture of ocean plankton as previously performed for

'*B.' childressi* and *B. azoricus* [63,69] can help to keep the mussels in a healthy state.

B. azoricus has been previously maintained in aquaria with atmospheric pressure [62,64,68].
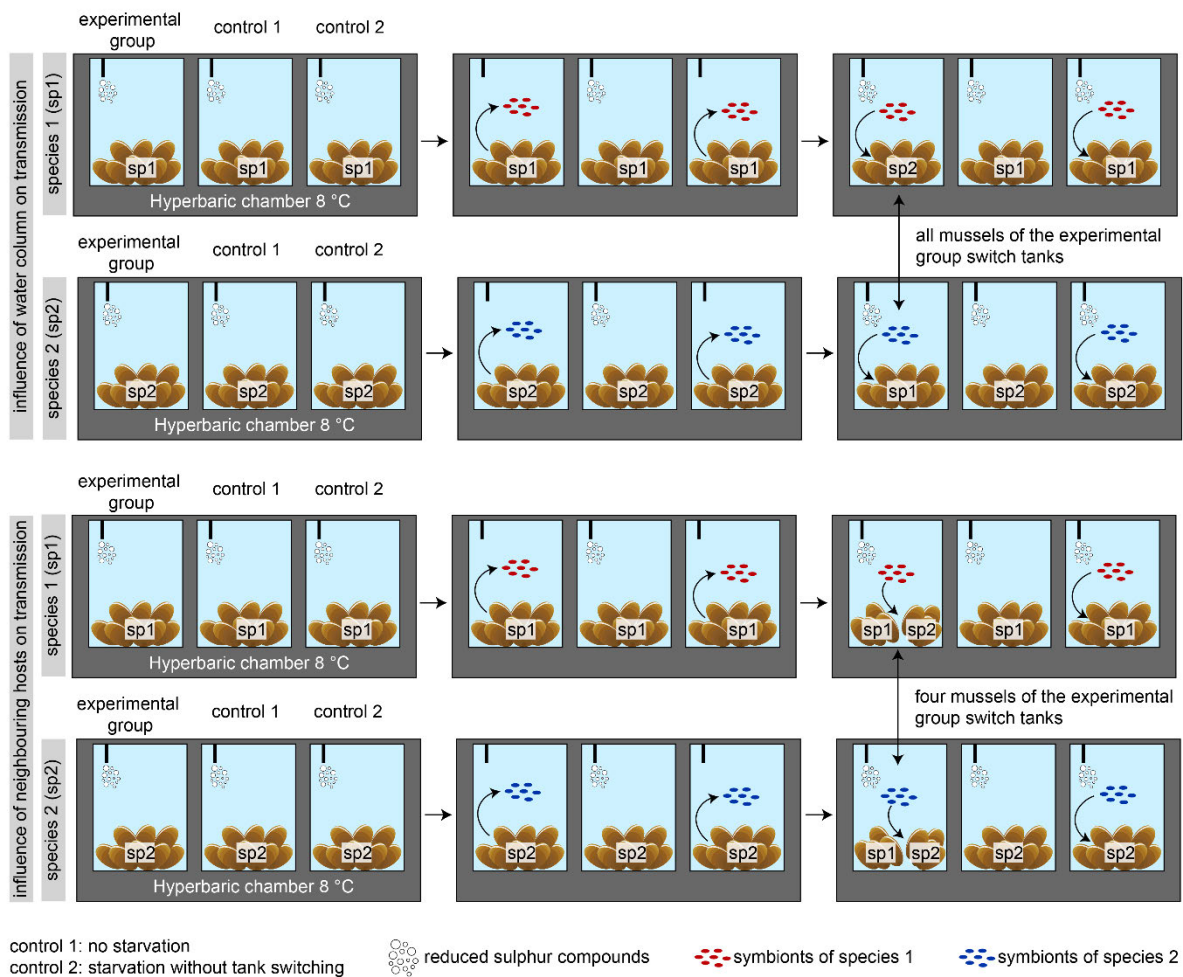
Figure 2 | Setup for aquaria experiments to investigate symbiont uptake in *Bathymodiolus*. Experiments should be conducted at around 8 °C in the dark in hyperbaric chambers approximating *in situ* pressure. Eight mussels can be kept in each tank. In the beginning, reduced sulphur compounds should be added close to *in situ* concentrations and then lowered to induce symbiont loss. After a while of substrate depletion, mussels switch tanks. Either all mussels are placed in the other tank to investigate the influence of the water column community on symbiont transmission, or half of the mussels switch tanks to investigate the influence of neighbouring mussels on the symbiont transmission. In the last step, sulphur concentrations should be increased to induce symbiont re-acquisition. Two controls should be included: (1) no starvation step is applied and (2) mussels are starved but do not switch tanks so that they should re-acquire their original symbionts. Mussels sampled should be prepared for sequencing, proteomics, metabolomics, and imaging. Water should be sampled throughout the experiment to investigate the free-living community.

Experiments with the vent shrimp *Mirocaris fortunata* suggested that vent animals should

better be studied at their *in situ* pressure [70]. Depending on where the mussels were

sampled, atmospheric pressure can put great stress on them, especially for species that occur at greater water depth, such as *B. puteoserpentis*. A way to overcome this challenge is to maintain the mussels at hydrostatic pressure using hyberbaric champers such as IPOCAMP that has been used successfully for *B. azoricus* [63], *Rimicaris* [71] and *Mirocaris* [70] shrimps, and *Hesiolyra* polychaetes from hydrothermal vents [72]. At least five individuals per treatment should be analysed to ensure reliable metagenomics and –transcriptomic results [11]. IPOCAMP has a total volume of 19 l, which would allow to keep 14 to 32 mussels based on the water volume to mussel ratio from previous studies. With four IPOCAMP chambers, each could contain three 5 l tanks with 8 mussels each, one for the experimental group and two for the controls (Figure). The easiest approach to accomplish the outlined setup would be to collaborate with facilities that have gained experience with husbandry of vent animals in the past such as the GEOMAR in Kiel [11] or LabHorta in Portugal [62]. Setting up the experiment probably requires a good amount of time, effort, and troubleshooting. However, it could significantly advance our knowledge about the uptake of symbionts, whether transmission is enabled through the water column or rather neighbouring hosts, and how strain diversity in- and outside the mussels compare. Establishing such an experiment could further serve as the basis for future experiments including the investigation of symbiont release to the environment after the death of the animal host as has been shown for *Riftia* symbionts [73]. These future experiments could shed light on the influence of the *Bathymodiolus* symbiosis on the microbial deep-sea vent community.

Do several SOX symbiont subspecies co-occur within *Bathymodiolus* gills?

Bathymodiolus lineages at the MAR harbour different symbiont subspecies based on phylogenomics and an average nucleotide identity cut-off of 97 % (chapter II, [50]). Analysis of the geographical distribution of these subspecies in- and outside their host can contribute to developing new hypotheses on the symbiont uptake and what factors are driving the symbiont composition. This is crucial because the symbiont composition determines the metabolic capabilities of the symbiosis. I discovered first evidence that different SOX symbiont subspecies might be present in mussel individuals from the MAR based on three marker genes (chapter III). This is intriguing because it has been previously assumed that only one symbiont subspecies is present per mussel lineage at the MAR. Until now, there was only one exception from this assumption for *Bathymodiolus* mussels in general. *B. brooksi* in the Gulf of Mexico has been demonstrated to harbour two divergent but co-occurring types of SOX symbionts [74]. One of the symbionts occurred in much lower abundance. The

discovery raised the question how these similar symbionts can co-occur. Microscopic analysis revealed that the two SOX symbionts were present in different cells. The authors suggested that the spatial distribution is either a sign of competition between the two types or reflects the adaptation to slightly different niches across the gill filament. They highlighted the need to assess the complete symbiont diversity if we want to fully understand the symbiosis including potential interactions between symbionts such as competition.

The results presented in chapter III indicate that several subspecies might be present in *B. puteoserpentis* at the MAR. As outlined above, the results should be confirmed with additional analyses. Mussel metagenomes are ideal to follow up on this finding due to the higher coverage of SOX symbiont. The taxonomic diversity within the mussel gills is much lower than on the water filters, which means that the individual partners have much higher coverage. Many metagenomes derived from mussel gill tissue were deeply sequenced, further increasing the coverage of potential symbionts. This opens up the possibility to use full MAGs instead of 'only' marker genes for a mapping-based investigation of potential additional SOX symbionts.

Such read-map based approaches are increasingly used in metagenomics studies. Problems can arise when the mapping integrity is not checked thoroughly as these approaches come with a great challenge that is to determine when a mapping is trustworthy, i.e. representing a biological finding and being reproducible [75]. Several smaller problems need to be addressed when designing the bioinformatics workflow. First, issues may arise when the references of interest are highly similar so that metagenomic reads could map to different references. The bias in the analysis could even increase if some of the references were less complete than the others, which could lead to higher cross-mapping. In our case, we are interested in different symbiont species and subspecies at the MAR (SOX symbionts of *B. azoricus*, *B. puteoserpentis*, and from Broken Spur, 5°S and 9°S). Symbionts from the northern and southern MAR have ANI values of only around 80 % but the NMAR symbiont subspecies are similar enough to have ANI values of around 97 %. I suggest compiling references of only those orthologous genes that are present in all of the symbiont types to avoid the issues outlined above. Metagenomic reads should be mapped against these references with high stringency, e.g. 98 or 99 % minimum mapping identity and ambiguously mapped reads should be discarded from further analysis.

Second, determining whether a mapped read does indeed correspond to a symbiont type being present rather than showing methodological artefacts needs good quality control. In small datasets, visual inspection of mapped reads can be highly informative. For 'real' mappings, one would expect that the whole gene, and not only sections of conserved sequences, are covered. It can be regarded as red flags when only a few reads map to a gene or only a region of a gene with high coverage while the rest of the sequence has no reds mapped to at all. Such observations indicate that there might not be a biological reason for the mapping. A visual inspection is not feasible when analysing hundreds to thousands of genes, especially when a high number of samples is investigated.

I propose the following bioinformatic procedure (Figure 3): (1) Orthologous genes that are present in all MAGs should be gathered as reference per symbiont type. (2) Genes with a length above 500 bp or even 1 000 bp should be selected to ensure that only genes and not fragments of genes are analysed and to avoid mapping bias based on the sequence length. (3) Metagenomic reads should be mapped with high stringency and ambiguously mapped reads should be discarded. (4) The coverage per base should be extracted from mapped files and filtered for occurrences where at least 95 % of a gene is covered with at least 2x coverage. This would be considered a positive hit. (5) Comparisons of gene numbers that are considered positive or negative hits between different references in different samples will give further insight into the presence of potential symbiont subspecies. As for most analyses, the more replicates are used, the more reliable the results will be and patterns between mussel species will become more apparent compared to patterns that might arise from differences in only a few individuals.

The approach outlined above is rather conservative and might discard genes in which only a variable region is different between the analysed references. However, this should affect all references equally as all ambiguous reads are tossed. The number of genes reported as positive hits might be lowered by this circumstance but the overall ratio between positive and negative hits should be unaffected. The symbiotic association of *B. brooksi* with two SOX symbionts in the presence of additional symbionts is an ideal testing system to assess whether this approach identifies two SOX symbionts although one of them occurs in very low abundance. After successful validation of the method, it could be applied to data from Broken Spur where I detected a second low abundant SOX symbiont during binning (data not shown). Intriguingly, the second SOX symbiont was most closely related to SOX symbionts from 9°S at the SMAR based on phylogenomics and ANI comparisons. This finding

illustrates that the symbiont diversity in mussel hosts might be higher than anticipated and our knowledge about the distribution ranges of symbionts at the MAR needs to be expanded. Finally, the bioinformatics approach described in this section can help to differentiate between symbiont reads coming from contamination during sequencing and those that might be real biological discoveries.
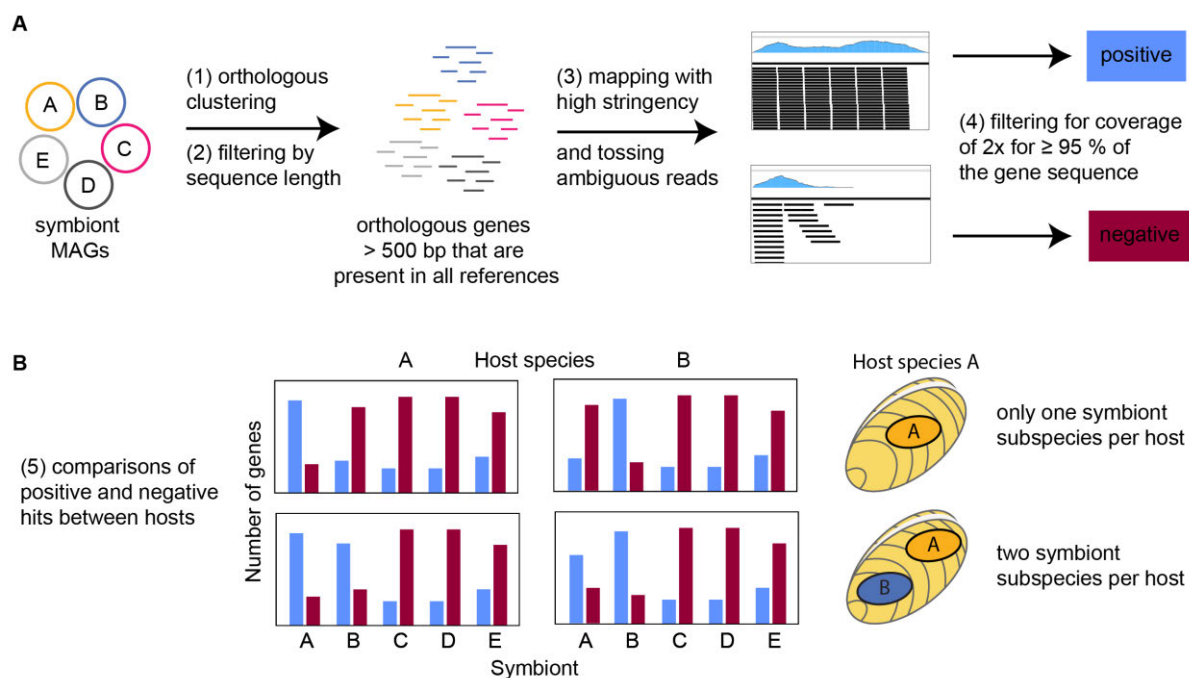
Figure 3 | Bioinformatics workflow to resolve the potential presence of several symbiont subspecies per mussel host. A: Data acquisition from symbiont MAGs to the number of genes that are considered positive (≥ 2x coverage for ≥ 95 % of the gene sequence) or negative hits. B: Data evaluation using bar plots assuming two potential outcomes of the analysis, either only one symbiont subspecies per host (top) or several (here two) symbiont subspecies per host (bottom). Numbering of the steps corresponds to the description in the main text. MAGs: metagenome-assembled genomes.

Metagenomes are just the first step

The thesis at hand is based solely on metagenomics, which was almost unimaginable 20 years ago. Between 2008 and 2016 alone, almost 120 Terabases of metagenome data were deposited in the NCBI database, marking an increase of one order of magnitude relative to the previous four years [76]. The vast amount of sequencing data enables us to not only study new environments and uncultivable taxonomic groups. It also leads to the development of new technologies, new species concepts, and the acquisition of initial knowledge that helps to ask questions that could not have been asked previously. While these advances and

innovations expanded our horizons tremendously, we need to acknowledge that metagenomes and other genomic data are just the first step towards reliable answers to biological questions. Metatranscriptomes and metaproteomes, maybe even followed by laboratory experiments, are needed to assess what part of the genetic potential is actually used. This is important as evolution is driven by selection that only acts on the phenotype-level. The relevance of using metagenomics results and moving further also becomes apparent when one looks at the studies presented in this thesis. When analysing metagenomes from a mussel hybrid zone, I did not observe genetic differences in the SOX symbionts of hybrid and parental mussels (chapter II). However, I do not know whether this observation holds true on the transcriptome and proteome level, or whether there are differences. Bacteria can exhibit phenotypic diversity, meaning that the same genotype can lead to different phenotypes depending on the conditions in their environment [77]. As an example, *Buchnera* symbionts in aphids have been shown to alter their gene expression in response to the genetic background of their hosts [78]. The second question that could benefit from metatranscriptome and -proteome information is the investigation of the free-living symbiont stage in chapter III. I detected symbiont-related genes in the water column. This is an indication for the presence of symbionts, whether they are in a free-living or host-associated stage. However, the genes do not yet reveal if symbionts in the water column are proliferating or rather in a dormant state. In the future, metatranscriptomes and -proteomes should be analysed to move beyond the genotype.

Another limitation of metagenomes, but also -transcriptomes and -proteomes, is the loss of spatial information. Imaging techniques can fill this gap. Powerful tools include light microscopy, especially combined with staining or FISH, electron microscopy, and mass spectrometry to image metabolites. One example where imaging might reveal new insights are the hybrid mussels from Broken Spur. Previous microscopy revealed differences in the tissue appearance of the parental species *B. azoricus* and *B. puteoserpentis*. Gill filaments in *B. azoricus* were rather thick with voluminous bacteriocytes while filaments in *B. puteoserpentis* were much thinner and bacteriocytes appeared flat (personal communication with M. Á. González Porras, Figure 4). Morphological variation can be important to determine an animal's fitness and is therefore often analysed in hybrids [79–81]. Broken Spur mussels had unevenly distributed shell sizes up to 120 mm, a size range that resembles *B. puteoserpentis* rather than the smaller *B. azoricus* [82]. However, nothing is known about their tissue morphology that might be more relevant considering that the gills

are the housing organ for the nutritional symbionts. It cannot be excluded that the morphological differences in the parental species are linked to different sampling procedures. Another potential reason is a difference in health state of the parental mussel species. Thin gills were observed as sign of starving animals [11]. But if the differences were of biological origin and did not result from starving *B. puteoserpentis*, visual assessment of hybrids from Broken Spur and comparison to their parental species could give interesting insights into whether the hybridisation led to ecologically relevant morphological variation.
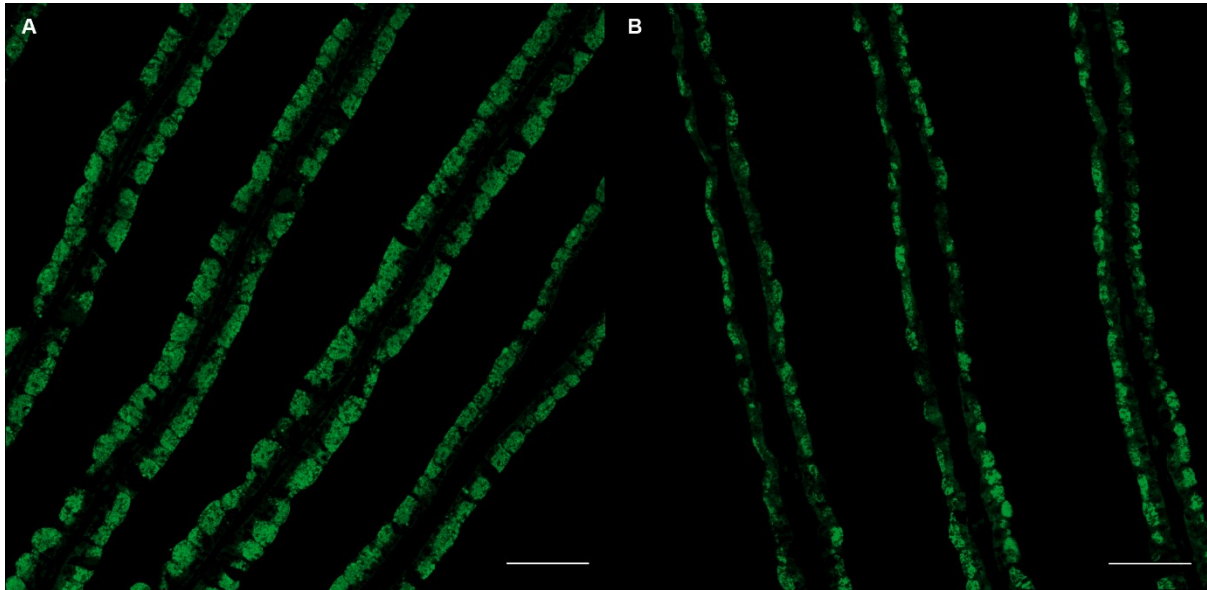
Figure 4 | Comparison of gill tissue morphology of two mussel species: A: *B. azoricus*. B: *B. puteoserpentis*. Green signal represents SOX symbionts. Scale bar = 50 μm. While the gills of *B. azoricus* appear rather thick, the gills of *B. puteoserpentis* seem much thinner. Fluorescence *in situ* hybridisation: M. Á. González Porras.

Symbiosis in ever higher resolution

Symbiont strain diversity is high and ecologically relevant in different *Bathymodiolus* species [2,3,51]. Strain-specific genes have been shown to be important for the functional repertoire of the symbiotic community [2,51]. It is less studied, how different genes are linked to a certain strain and how many and which strains are shared between host individuals [3]. Bacterial strains can be tracked across time and samples by applying isolate-derived marker genes or marker SNP profiles to natural microbial communities. Advances of bioinformatic tools for the analysis of strain-level information, such as ConStrains [83], MIDAS [84], StrainPhlAn [85], inStrain [86], or STRONG [87], will help us to understand the dynamics of microbial communities in the future. One disadvantage is that many of these approaches still rely on at least one representative isolate genome. Emerging techniques such as single-cell,

read cloud, or long read sequencing are promising to resolve which strain-specific genes occur together in a single strain and how strains are distributed across hosts. Single-cell sequencing of *Bathymodiolus* symbionts would allow pinpointing the genetic potential of individual symbiont strains. However, attempts of sequencing single symbiont cells have been unsuccessful so far [13]. Read cloud sequencing links short reads with long-range information. It enables the assembly of high-quality microbial draft genomes from metagenomic data and the investigation of strain diversity [88,89]. Similarly, long read sequencing as offered by Oxford Nanopore Technologies or Pacific Biosciences improves *de novo* assembly of bacterial genomes, even of single strains, and allows the use of methylation motif analysis [90–92]. These sequencing approaches should be considered when investigating the link between genes and strains, and their distribution across hosts.

Leaps forward in resolution are not limited to sequencing technologies, but extend to imaging. Fascinating information is becoming visible in ever higher resolution. Improvements of microscopes, mass spectrometer lasers, or approaches to sample preparation, such as high-pressure freezing, allow us to investigate even the tiniest organisms and organelles. Imaging approaches are useful because they supply spatial information about the subject of interest that can be correlated to phylogenetic affiliation (e.g. [93]). Approaches such as (gene)FISH, transmission electron microscopy, and mass spectrometry imaging [94] will provide information about the distribution of symbiont strains and with it, further insights into the acquisition of symbionts, fine-scale population structure, and competition and niche partitioning of strains with varying genetic repertoire [2,95].

Concluding remarks

The research presented in my thesis provides new insights into the *Bathymodiolus* symbiosis in the Atlantic Ocean from the environment to the host and the symbionts. My results highlight the specificity of the symbiotic association, uncover the importance of geography and environment on symbiont composition, and advance our understanding of how mussel populations are connected in this oceanic region. The discoveries presented in this thesis add to our understanding why *Bathymodiolus* mussels are so successful. In my population genomic study, I hypothesised that larval dispersal can extend over longer distances than previously modelled. The broader distribution of larvae enhances genetic connectivity and contributes to higher genetic diversity of the mussel populations. Higher variability on the genetic level allows better adaptation to new or changing environments, and makes the

population more resistant to stressors. Longer dispersal, potentially enabled through vertical migration and feeding on phototrophic carbon, strengthens the colonisation abilities of these vent animals. Mussel larvae settling at a new site have high chances of acquiring a suited symbiont as revealed by my studies. I showed that SOX symbiont composition was independent of host genetics indicating that symbiont uptake is not hindered by having the 'wrong' genotype. The conclusion is further supported by the discovery of mussels with mitonuclear discordance. In all cases, these mussels harboured location-specific symbionts regardless of their genotype. Analysis of water metagenomes revealed indications for the presence of several SOX symbiont subspecies at the same site. The dominance of one subspecies per site suggests some kind of selection, potentially through environmental conditions. Settling larvae could benefit from the uptake of the well-adapted symbionts at each site. However, the question whether the host or the environment is selecting for specific symbiont subspecies awaits future research. Taken together, these factors likely contribute to the ecological success of *Bathymodiolus* mussels in the deep sea. Throughout the thesis, I discussed the limitations that we currently face, be it with regard to sampling in the remote deep-sea environment or the used metagenomic approaches. For the future, it will be essential to complement our metagenomics insights with other techniques such as metatranscriptomics, -proteomics, and imaging. Integrating these techniques, we will be able to better link spatial information about symbionts on the subspecies or strain level, their genes, and corresponding metabolic products. Where feasible, hypotheses that were put forward should be tested in the laboratory. Experiments and analyses outlined in this discussion can become more powerful when combined with environmental metadata. Combining these will help us understand the distribution ranges of mussels and symbionts, the relationship between symbiotic and free-living stage as well as the impact of the *Bathymodiolus* symbiosis on the hydrothermal vent ecosystem. I hope that this thesis will serve as an inspiration to continue the investigation of symbiotic associations with a holistic approach that includes not only the bacterial symbionts but also considers their eukaryotic host and the environment around them as a crucial part of the system. Following this path, we will make many great discoveries in the future that will improve our understanding of this deep-sea symbiosis and help expanding our knowledge on any symbiotic association in the various branches of life.

References

1. Duperron S, Gaudron SM, Rodrigues CF, Cunha MR, Decker C, Olu K. An overview of chemosynthetic symbioses in bivalves from the north Atlantic and Mediterranean Sea. Biogeosciences. 2013; 10: 3241–67.

2. Ansorge R, Romano S, Sayavedra L, González Porras MÁ, Kupczok A, Tegetmeyer HE, et al. Functional diversity enables multiple symbiont strains to coexist in deep-sea mussels. Nat Microbiol. 2019; 4: 2487–97.

3. Picazo DR, Dagan T, Ansorge R, Petersen JM, Dubilier N, Kupczok A. Horizontally transmitted symbiont populations in deep-sea mussels are genetically isolated. ISME J. 2019; 13: 2954–68.

4. Vrijenhoek RC. Gene flow and genetic diversity in naturally fragmented metapopulations of deep-sea hydrothermal vent animals. J Hered. 1997; 88: 285–93.

5. Franke M, Geier B, Hammel JU, Dubilier N, Leisch N. Becoming symbiotic – the symbiont acquisition and the early development of bathymodiolin mussels. BioRxiv. 2020: 2020.10.09.333211.

6. Macdonald KC. Mid-ocean ridges: fine scale tectonic, volcanic and hydrothermal processes within the plate boundary zone. Annu Rev Earth Planet Sci. 1982; 10: 155–90.

7. Arellano SM, Van Gaest AL, Johnson SB, Vrijenhoek RC, Young CM. Larvae from deep-sea methane seeps disperse in surface waters. Proc Biol Sci. 2014; 281: 20133276.

8. Rudnicki MD, Elderfield H. A chemical model of the buoyant and neutrally buoyant plume above the TAG vent field, 26 degrees N, Mid-Atlantic Ridge. Geochim Cosmochim Acta. 1993; 57: 2939–57.

9. Dick GJ. The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. Nat Rev Microbiol. 2019; 17: 271–83.

10. Fortunato CS, Larson B, Butterfield DA, Huber JA. Spatially distinct, temporally stable microbial populations mediate biogeochemical cycling at and below the seafloor in hydrothermal vent fluids. Environ Microbiol. 2018; 20: 769–84.

11. Tietjen M. Physiology and ecology of deep-sea *Bathymodiolus* symbioses. PhD Thesis. University of Bremen, 2020.

12. Fraschetti S, Giangrande A, Terlizzi A, Boero F. Pre- and post-settlement events in benthic community dynamics. Oceanol Acta. 2002; 25: 285–95.

13. Ansorge R. Strain diversity and evolution in endosymbionts of *Bathymodiolus* mussels. PhD Thesis. University of Bremen, 2019.

14. Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. Proc R Soc B. 2003; 270: 313–21.

15. Miyazaki J-I, Martins L de O, Fujita Y, Matsumoto H, Fujiwara Y. Evolutionary process of deep-sea *Bathymodiolus* mussels. PLOS ONE. 2010; 5: e10363.

16. Lee Y, Kwak H, Shin J, Kim S-C, Kim T, Park J-K. A mitochondrial genome phylogeny of Mytilidae (Bivalvia: Mytilida). Mol Phylogenetics Evol. 2019; 139: 106533.

17. Jones WJ, Won Y-J, Maas PAY, Smith PJ, Lutz RA, Vrijenhoek RC. Evolution of habitat use by deep-sea mussels. Mar Biol. 2006; 148: 841–51.

18. Smith PJ, McVeagh SM, Won Y, Vrijenhoek RC. Genetic heterogeneity among New Zealand species of hydrothermal vent mussels (Mytilidae: *Bathymodiolus*). Mar Biol. 2004; 144: 537–45.

19. Kuo M-Y, Kang D-R, Chang C-H, Chao C-H, Wang C-C, Chen H-H, et al. New records of three deep-sea *Bathymodiolus* mussels (bivalvia: Mytilida: Mytilidae) from

hydrothermal vent and cold seeps in Taiwan. J Mar Sci Technol-Taiwan. 2019; 27: 352–8.

20. Iwasaki H, Kyuno A, Shintaku M, Fujita Y, Fujiwara Y, Fujikura K, et al. Evolutionary relationships of deep-sea mussels inferred by mitochondrial DNA sequences. Mar Biol. 2006; 149: 1111–22.

21. Shen Y, Kou Q, Chen W, He S, Yang M, Li X, et al. Comparative population structure of two dominant species, *Shinkaia crosnieri* (Munidopsidae: *Shinkaia*) and *Bathymodiolus platifrons* (Mytilidae: *Bathymodiolus*), inhabiting both deep-sea vent and cold seep inferred from mitochondrial multi-genes. Ecol Evol. 2016; 6: 3571–82.

22. Ivanov V, Lee KM, Mutanen M. Mitonuclear discordance in wolf spiders: genomic evidence for species integrity and introgression. Mol Ecol. 2018; 27: 1681–95.

23. Layton KKS, Carvajal JI, Wilson NG. Mimicry and mitonuclear discordance in nudibranchs: new insights from exon capture phylogenomics. Ecol Evol. 2020; 10: 11966–82.

24. Pavlova A, Amos JN, Joseph L, Loynes K, Austin JJ, Keogh JS, et al. Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an australian bird. Evolution. 2013; 67: 3412–28.

25. White DJ, Wolff JN, Pierson M, Gemmell NJ. Revealing the hidden complexities of mtDNA inheritance. Mol Ecol. 2008; 17: 4925–42.

26. Birky CW. Uniparental inheritance of mitochondrial and chloroplast genes: Mechanisms and evolution. PNAS. 1995; 92: 11331–8.

27. Zouros E. The exceptional mitochondrial DNA system of the mussel family Mytilidae. Genes Genet Syst. 2000; 75: 313–8.

28. Passamonti M, Scali V. Gender-associated mitochondrial DNA heteroplasmy in the venerid clam *Tapes philippinarum* (Mollusca Bivalvia). Curr Genet. 2001; 39: 117–24.

29. Skibinski DOF, Gallagher C, Beynon CM. Sex-limited mitochondrial DNA transmission in the marine mussel *Mytilus edulis*. Genetics. 1994; 138: 801–9.

30. Zouros E, Ball AO, Saavedra C, Freeman KR. An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. PNAS. 1994; 91: 7463–7.

31. Hoeh WR, Stewart DT, Sutherland BW, Zouros E. Cytochrome c oxidase sequence comparisons suggest an unusually high rate of mitochondrial DNA evolution in *Mytilus* (Mollusca: Bivalvia). Mol Biol Evol. 1996; 13: 418–21.

32. Saavedra C, Reyero MI, Zouros E. Male-dependent doubly uniparental inheritance of mitochondrial DNA and female-dependent sex-ratio in the mussel *Mytilus galloprovincialis*. Genetics. 1997; 145: 1073–82.

33. Dreyer H, Steiner G. The complete sequences and gene organisation of the mitochondrial genomes of the heterodont bivalves *Acanthocardia tuberculata* and *Hiatella arctica* – and the first record for a putative ATPase subunit 8 gene in marine bivalves. Front Zool. 2006; 3: 13.

34. Breton S, Milani L, Ghiselli F, Guerra D, Stewart DT, Passamonti M. A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. Trends Genet. 2014; 30: 555–64.

35. Ghiselli F, Milani L, Guerra D, Chang PL, Breton S, Nuzhdin SV, et al. Structure, transcription, and variability of metazoan mitochondrial genome: Perspectives from an unusual mitochondrial inheritance system. Genome Biol Evol. 2013; 5: 1535–54.

36. Breton S, Beaupré HD, Stewart DT, Hoeh WR, Blier PU. The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? Trends Genet. 2007; 23: 465–74.

37. Passamonti M, Ghiselli F. Doubly uniparental inheritance: two mitochondrial genomes, one precious model for organelle DNA inheritance and evolution. DNA Cell Biol. 2009; 28: 79–89.

38. Zouros E. Biparental inheritance through uniparental transmission: the doubly uniparental inheritance (DUI) of mitochondrial DNA. Evol Biol. 2013; 40: 1–31.

39. Passamonti M, Ricci A, Milani L, Ghiselli F. Mitochondrial genomes and doubly uniparental inheritance: new insights from *Musculista senhousia* sex-linked mitochondrial DNAs (Bivalvia Mytilidae). BMC Genomics. 2011; 12: 442.

40. Bensasson D, Zhang D-X, Hartl DL, Hewitt GM. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol Evol. 2001; 16: 314–21.

41. Richly E, Leister D. NUMTs in sequenced eukaryotic genomes. Mol Biol Evol. 2004; 21: 1081–4.

42. Hazkani-Covo E, Zeller RM, Martin W. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet. 2010; 6: e1000834.

43. Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J Mol Evol. 1994; 39: 174–90.

44. Won Y-J, Young CR, Lutz RA, Vrijenhoek RC. Dispersal barriers and isolation among deep-sea mussel populations (Mytilidae: *Bathymodiolus*) from eastern Pacific hydrothermal vents. Mol Ecol. 2003; 12: 169–84.

45. Plazzi F, Cassano A, Passamonti M. The quest for doubly uniparental inheritance in heterodont bivalves and its detection in *Meretrix lamarckii* (Veneridae: Meretricinae). J Zoolog Syst Evol Res. 2015; 53: 87–94.

46. Garrido-Ramos M, Stewart D, Sutherland B, Zouros E. The distribution of male-transmitted and female-transmitted mitochondrial DNA types in somatic tissues of blue mussels: Implications for the operation of doubly uniparental inheritance of mitochondrial DNA. Genome. 1998; 41: 818–24.

47. Baas Becking LGM. Geobiologie of inleiding tot de milieukunde. Den Haag: W.P. Van Stockum & Zoon; 1934.

48. Wit RD, Bouvier T. 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? Environ Microbiol. 2006; 8: 755–8.

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

50. Ansorge R, Romano S, Sayavedra L, Rubin-Blum M, Gruber-Vodicka HR, Scilipoti S, et al. The hidden pangenome: comparative genomics reveals pervasive diversity in symbiotic and free-living sulfur-oxidizing bacteria. BioRxiv. 2020: 2020.12.11.421487.

51. Ikuta T, Takaki Y, Nagai Y, Shimamura S, Tsuda M, Kawagucci S, et al. Heterogeneous composition of key metabolic gene clusters in a vent mussel symbiont population. ISME J. 2016; 10: 990–1001.

52. Van Rossum T, Ferretti P, Maistrenko OM, Bork P. Diversity within species: interpreting strains in microbiomes. Nat Rev Microbiol. 2020; 18: 491–506.

53. Barrero-Canosa J, Moraru C, Zeugner L, Fuchs BM, Amann R. Direct-geneFISH: a simplified protocol for the simultaneous detection and quantification of genes and rRNA in microorganisms. Environ Microbiol. 2017; 19: 70–82.

54. Duperron S, Bergin C, Zielinski F, Blazejak A, Pernthaler A, McKiness ZP, 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. 2006; 8: 1441–7.

55. Volpi EV, Bridger JM. FISH glossary: an overview of the fluorescence in situ hybridization technique. BioTechniques. 2008; 45: 385–409.

56. Wagner M, Horn M, Daims H. Fluorescence in situ hybridisation for the identification and characterisation of prokaryotes. Curr Opin Microbiol. 2003; 6: 302–9.

57. Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev. 1995; 59: 143–69.

58. Vives-Rego J, Lebaron P, Nebe-von Caron G. Current and future applications of flow cytometry in aquatic microbiology. FEMS Microbiol Rev. 2000; 24: 429–48.

59. Wallner G, Steinmetz I, Bitter-Suermann D, Amann R. Combination of rRNA-targeted hybridization probes and immuno-probes for the identification of bacteria by flow cytometry. Syst Appl Microbiol. 1996; 19: 569–76.

60. Colaço A, Martins I, Laranjo M, Pires L, Leal C, Prieto C, et al. Annual spawning of the hydrothermal vent mussel, *Bathymodiolus azoricus*, under controlled aquarium conditions at atmospheric pressure. J Exp Mar Biol Ecol. 2006; 333: 166–71.

61. Miyake H, Kitada M, Itoh T, Nemoto S, Okuyama Y, Watanabe H, et al. Larvae of deep-sea chemosynthetic ecosystem animals in captivity. Cah Biol Mar. 2010; 51: 441–50.

62. Colaço A, Bettencourt R, Costa V, Lino S, Lopes H, Martins I, et al. LabHorta: a controlled aquarium system for monitoring physiological characteristics of the hydrothermal vent mussel *Bathymodiolus azoricus*. ICES J Mar Sci. 2011; 68: 349–56.

63. Bettencourt R, Costa V, Laranjo M, Rosa D, Pires L, Colaço A, et al. Out of the deep-sea into a land-based aquarium environment: investigating innate immunity in the hydrothermal vent mussel *Bathymodiolus azoricus*. Cah Biol Mar. 2010; 51: 341–50.

64. Kádár E, Bettencourt R, Costa V, Santos RS, Lobo-da-Cunha A, Dando P. Experimentally induced endosymbiont loss and re-acquirement in the hydrothermal vent bivalve *Bathymodiolus azoricus*. J Exp Mar Biol Ecol. 2005; 318: 99–110.

65. Wentrup C, Wendeberg A, Schimak M, Borowski C, Dubilier N. Forever competent: deep-sea bivalves are colonized by their chemosynthetic symbionts throughout their lifetime. Environ Microbiol. 2014; 16: 3699–713.

66. Polzin J, Arevalo P, Nussbaumer T, Polz MF, Bright M. Polyclonal symbiont populations in hydrothermal vent tubeworms and the environment. Proc R Soc B. 2019; 286: 20181281.

67. Harmer TL, Rotjan RD, Nussbaumer AD, Bright M, Ng AW, DeChaine EG, et al. Free-living tube worm endosymbionts found at deep-sea vents. Appl Environ Microbiol. 2008; 74: 3895–8.

68. Bettencourt R, Dando P, Collins P, Costa V, Allam B, Serrão Santos R. Innate immunity in the deep sea hydrothermal vent mussel *Bathymodiolus azoricus*. Comp Biochem Physiol Part A Mol Integr Physiol. 2009; 152: 278–89.

69. Riekenberg P, Carney R, Fry B. Trophic plasticity of the methanotrophic mussel *Bathymodiolus childressi* in the Gulf of Mexico. Mar Ecol Prog Ser. 2016; 547: 91–106.

70. Shillito B, Bris NL, Hourdez S, Ravaux J, Cottin D, Caprais J-C, et al. Temperature resistance studies on the deep-sea vent shrimp *Mirocaris fortunata*. J Exp Biol. 2006; 209: 945–55.

71. Ravaux J, Gaill F, Bris NL, Sarradin P-M, Jollivet D, Shillito B. Heat-shock response and temperature resistance in the deep-sea vent shrimp *Rimicaris exoculata*. J Exp Biol. 2003; 206: 2345–54.

72. Shillito B, Jollivet D, Sarradin P-M, Rodier P, Lallier F, Desbruyères D, et al. Temperature resistance of *Hesiolyra bergi*, a polychaetous annelid living on deep-sea vent smoker walls. Mar Ecol Prog Ser. 2001; 216: 141–9.

73. Klose J, Polz MF, Wagner M, Schimak MP, Gollner S, Bright M. Endosymbionts escape dead hydrothermal vent tubeworms to enrich the free-living population. PNAS. 2015; 112: 11300–5.

74. Duperron S, Sibuet M, MacGregor BJ, Kuypers MMM, Fisher CR, Dubilier N. Diversity, relative abundance and metabolic potential of bacterial endosymbionts in three *Bathymodiolus* mussel species from cold seeps in the Gulf of Mexico. Environ Microbiol. 2007; 9: 1423–38.

75. Menzel P, Frellsen J, Plass M, Rasmussen SH, Krogh A. On the accuracy of short read mapping. In: Shomron N, editor. Deep Sequencing Data Analysis, Totowa, NJ: Humana Press; 2013, p. 39–59.

76. Nayfach S, Pollard KS. Toward accurate and quantitative comparative metagenomics. Cell. 2016; 166: 1103–16.

77. Tadrowski AC, Evans MR, Waclaw B. Phenotypic switching can speed up microbial evolution. Sci Rep. 2018; 8: 8941.

78. Smith TE, Moran NA. Coordination of host and symbiont gene expression reveals a metabolic tug-of-war between aphids and *Buchnera*. PNAS. 2020; 117: 2113–21.

79. Arnold ML, Hodges SA. Are natural hybrids fit or unfit relative to their parents? Trends Ecol Evol. 1995; 10: 67–71.

80. Kohn LAP, Langton LB, Cheverud JM. Subspecific genetic differences in the saddle-back tamarin (*Saguinus fuscicollis*) postcranial skeleton. Am J Primatol. 2001; 54: 41–56.

81. Matthews DG, Albertson RC. Effect of craniofacial genotype on the relationship between morphology and feeding performance in cichlid fishes. Evolution. 2017; 71: 2050–61.

82. Won Y, Hallam SJ, O'Mullan GD, Vrijenhoek RC. Cytonuclear disequilibrium in a hybrid zone involving deep-sea hydrothermal vent mussels of the genus *Bathymodiolus*. Mol Ecol. 2003; 12: 3185–90.

83. Luo C, Knight R, Siljander H, Knip M, Xavier RJ, Gevers D. ConStrains identifies microbial strains in metagenomic datasets. Nat Biotechnol. 2015; 33: 1045–52.

84. Nayfach S, Rodriguez-Mueller B, Garud N, Pollard KS. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. Genome Res. 2016; 26: 1612–25.

85. Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. Microbial strain-level population structure and genetic diversity from metagenomes. Genome Res. 2017; 27: 626–38.

86. Olm MR, Crits-Christoph A, Bouma-Gregson K, Firek BA, Morowitz MJ, Banfield JF. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. Nat Biotechnol. 2021: 1–10.

87. Quince C, Nurk S, Raguideau S, James R, Soyer OS, Summers JK, et al. Metagenomics strain resolution on assembly graphs. BioRxiv. 2020: 2020.09.06.284828.

88. Bishara A, Moss EL, Kolmogorov M, Parada AE, Weng Z, Sidow A, et al. High-quality genome sequences of uncultured microbes by assembly of read clouds. Nat Biotechnol. 2018; 36: 1067–75.

89. Zlitni S, Bishara A, Moss EL, Tkachenko E, Kang JB, Culver RN, et al. Strain-resolved microbiome sequencing reveals mobile elements that drive bacterial competition on a clinical timescale. Genome Med. 2020; 12: 50.

90. Frank JA, Pan Y, Tooming-Klunderud A, Eijsink VGH, McHardy AC, Nederbragt AJ, et al. Improved metagenome assemblies and taxonomic binning using long-read circular consensus sequence data. Sci Rep. 2016; 6: 25373.
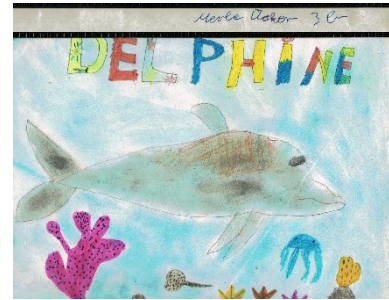
91. Daims H, Lebedeva EV, Pjevac P, Han P, Herbold C, Albertsen M, et al. Complete nitrification by *Nitrospira* bacteria. Nature. 2015; 528: 504–9.

92. Somerville V, Lutz S, Schmid M, Frei D, Moser A, Irmler S, et al. Long-read based *de novo* assembly of low-complexity metagenome samples results in finished genomes and reveals insights into strain diversity and an active phage system. BMC Microbiol. 2019; 19: 143.

93. Geier B, Sogin EM, Michellod D, Janda M, Kompauer M, Spengler B, et al. Spatial metabolomics of *in situ* host–microbe interactions at the micrometre scale. Nat Microbiol. 2020; 5: 498–510.

94. Watrous JD, Dorrestein PC. Imaging mass spectrometry in microbiology. Nat Rev Microbiol. 2011; 9: 683–94.

95. Miguel Ángel González Porras. Molecular biology and evolution of the bacterial intranuclear parasite *Ca.* Endonucleobacter. PhD Thesis. University of Bremen, 2020.

Acknowledgements

This thesis is my childhood dream of being a marine biologist come true. What used to be a naïve dream, turned out to be a good bunch of hard work that required the help of amazing people. In this section, I would like to thank those that provided inspiration, support, and advice along the way.

information, and your help with bureaucratic matters. Thanks to **Harald** for scientific and sequencing advice and the great collaboration during the MarMic bioinformatics practical. Bula, **Miguel**, thank you for being a great travel buddy during the cruise and on the honeymoon island. Thanks to **Max** for our discussions about *Bathy* larvae and for proofreading parts of this thesis, **Benedikt** for inspiring scientific discussions and reading my manuscripts 'for fun', and **Anne** for taking on the task to organise the metagenomic reads.

I must express my gratitude to all my supervisors who accompanied me along the way: **Liz** who supervised me when I was still a bioinformatics newbie and **Juliane** who gave great tips on coding, presenting, and life in general. **Rebecca** and **Yui**, you supervised me for the most part of my PhD, and I am more than grateful for your scientific input and emotional support. Thank you for always having my back, making the time to provide feedback whenever I needed it, keeping up to the pace, and sharing your excitement for science. **Rebecca**, your positivity really brightened up my days and motivated me to keep going. **Yui**, you are a role model for strength, hope and kindness. Thank you both for all the useful advice, I am always hearing your voices asking me: 'Why is it important?'

I am very grateful for the **Max Planck PhDnet** where I could develop new skills and be part of an awesome community! I also thank my **fellow PhD representatives** at the MPIMM, Anni, Anna, Julia, and Miriam for the fun year we had organising events and writing the starter's guide! Thanks to the **MarMic class 2020** aka the Avocado and yoga class for starting out this journey together.

I also appreciate that there is a life besides the PhD and that I have amazing friends reminding me of it. Great thanks to the TeaGang, **Inma**, **Julia**, **Miriam**, **Tina**, and **Ben** for the enjoyable lunch breaks when they were still possible. Thank you as well as **Friederike**, **Jessica**, **Hans**, and **Jörn** for sweetening my life in Bremen. Thanks to my "old" friends **Klara**, **Franka**, **Charlotte**, and **Yasmin** for your incredible friendship. Thank you **Julia** and **Miriam** for proofreading parts of this thesis. Special thanks to you, **Tina**, for being my partner in crime, always writing back on Slack, and reading endless pages about *Bathymodiolus*. This PhD would have not been the same without you.

Before Corona, going to the office was always a joy. Thanks to the best office, **Marius**, **Juliane**, **Målin**, **Janine**, and **Oliver** for making room 2247 a safe and cozy second home. I really miss our daily conversations, and optimising figure colours with you. Thank you **Janine**, **Målin** and **Oli** for your friendship and your valuable feedback on parts of this thesis.

Ich möchte meiner gesamten **Familie** dafür danken, dass sie mir immer das Gefühl gegeben hat, ich könne alles schaffen. Ein Riesendank an meine **Eltern** für ihre Unterstützung ab Tag 0: Meine Begeisterung für die Natur, Neugier und Durchhaltevermögen kommen nicht von ungefähr. **Lara**, danke für die vielen Besuche, Reisen, Telefonate und Überraschungspäckchen, mit denen du mich über die Jahre unterstützt hast – du bist die beste Schwester der Welt! Ich hoffe, du bist für diese Arbeit nur annähernd so stolz auf mich wie ich auf dich bin.

Danke an meine ‚zweite‘ **Familie Rattay Flesch** für eure Unterstützung und den steten Nachschub an süßen und lustigen Kinderfotos, die mich bei Laune halten.

Danke **Thilo**, dass du immer an meiner Seite bist, im richtigen Moment ein Eis kaufst, um die Krise abzuwenden, und dein ehrliches Feedback. Ich freue mich auf alle Herausforderungen, die wir zusammen meistern können (um uns hinterher mit Ramen zu belohnen).

Contribution to manuscripts

Manuscript I (chapter II) | Deep-sea mussels from a hybrid zone on the Mid-Atlantic Ridge host genetically indistinguishable symbionts

Conceptual design: 50 %

Data acquisition and experiments: 70 %

Analysis and interpretation of the results: 70 %

Preparation of figures and tables: 90 %

Writing the manuscript: 70 %

Manuscript II (chapter III) | Chasing free-living stages of mussel symbionts– Metagenomic insights from the environment to the host

Conceptual design: 50 %

Data acquisition and experiments: 60 %

Analysis and interpretation of the results: 90 %

Preparation of figures and tables: 90 %

Writing the manuscript: 90 %

Manuscript III (chapter IV) | Mitonuclear discordance suggests long-distance migration and mitochondrial introgression of deep-sea mussels along the Mid-Atlantic Ridge

Conceptual design: 50 %

Data acquisition and experiments: 40 %

Analysis and interpretation of the results: 80 %

Preparation of figures and tables: 95 %

Writing the manuscript: 95 %

Ort, Datum: Bremen, 02.03.2021

Versicherung an Eides Statt

Ich, **Merle Ücker**

██████████████████

Matrikelnummer: 3028563

.

versichere an Eides Statt durch meine Unterschrift, dass

1. ich die vorstehende Arbeit mit dem Titel „**Metagenomic analyses of a deep-sea mussel symbiosis**" selbständig und ohne fremde Hilfe angefertigt habe.
2. ich alle Stellen, die ich wörtlich dem Sinne nach aus Veröffentlichungen entnommen habe, als solche kenntlich gemacht habe, mich auch keiner anderen als der angegebenen Literatur oder sonstiger Hilfsmittel bedient habe.
3. die elektronische Version der Dissertation identisch mit der abgegebenen gedruckten Version ist.

Ich versichere an Eides Statt, dass ich die vorgenannten Angaben nach bestem Wissen und Gewissen gemacht habe und dass die Angaben der Wahrheit entsprechen und ich nichts verschwiegen habe.

Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

_____

Bremen, 02.03.2021, Merle Ücker