From genomes to communities

Evolution of symbionts associated with globally distributed marine invertebrates

Dissertation

Zur Erlangung des Grades eines

Doktor der Naturwissenschaften

- Dr. rer. nat. -

dem Fachbereich Biologie / Chemie

der Universität Bremen

vorgelegt von

Anna Mankowski

Bremen, September 2021

Diese Doktorarbeit wurde von März 2017 bis September 2021 in der Abteilung Symbiose am Max-Planck-Institut für Marine Mikrobiologie unter Betreuung von Prof. Dr. Nicole Dubilier und unter direkter Betreuung von Dr. Harald R. Gruber-Vodicka durchgeführt.

This PhD thesis was conducted from March 2017 to September 2021 in the Department of Symbiosis at the Max Planck Institute for Marine Microbiology under the supervision of Prof. Dr. Nicole Dubilier and under the direct supervision of Dr. Harald R. Gruber-Vodicka.

Gutachter*innen | Reviewers

Prof. Dr. Nicole Dubilier

Prof. Dr. Alexander J. Probst

Prof. Dr. Hinrich Schulenburg

Tag des Promotionskolloquium | Date of the doctoral defence

10.11.2021

Diese zur Veröffentlichung erstellte Version der Dissertation enthält Korrekturen.

“The Sea, once it casts its spell, holds one in its net of wonder forever.”

Jacques-Yves Cousteau

TABLE OF CONTENTS

“Success is the ability to go from one failure to another with no loss of enthusiasm.”

Winston Churchill

Summary

The association of eukaryotic hosts with diverse bacterial symbiont communities is ubiquitous in nature. The probably most famous and best studied examples of such associations are the gut microbiomes of animals, especially the human gut microbiome. The diversity and composition of human gut microbiota has continuously been shown to tremendously influence host health, highlighting the importance of symbiosis research. Despite the growing recognition of multimember symbioses in nature and their role as a driving force in the ecology and evolution of many organisms, relatively little is known about the evolutionary dynamics in such phylogenetically diverse associations. Here, I studied the evolution of symbionts from taxonomically diverse and globally distributed marine invertebrates. The broad taxonomic and biogeographic range of the hosts allowed me to untangle ecological and host evolutionary effects on symbiont evolution.

In the first part of this thesis (Chapter II), I focused on genome reduction in bacterial symbionts. The research on this topic has been biased towards completely host-restricted symbionts that underwent massive genome reduction and thus, retained strikingly small genomes. In contrast, much less is known about how symbiont genomes evolve when they are not completely isolated from other bacterial populations. To tackle this question, I analyzed genome reduction in 12 closely related lineages of ecto- and endosymbionts. All of these were affected by genome reduction, yet the endosymbiotic lifestyle appeared to accelerate this process.

In the second part of this thesis (Chapter III and IV), I focused on the composition of symbiont communities in two different animal groups. In Chapter III, I analyzed the evolution of symbiont communities in gutless oligochaetes, a group of marine worms that obligately rely on their symbionts for nutrition. Despite their high dependence, these worms are associated with strikingly diverse symbiont communities. This diversity appears to be the result of highly variable levels of symbiont fidelity which appear to balance stability and versatility of the symbiosis. In Chapter IV, I explored the symbiont community composition of placozoans. Due to their simplistic morphology and early divergence during eukaryotic evolution, these animals are often used as model systems for animal evolution and development. Their microbial partners however, have not been intensively studied yet. I showed that placozoans are commonly associated with diverse symbiont consortia, including intra- and intercellular bacteria and postulate that the placozoan symbiosis can be a powerful model to study the ecological and evolutionary dynamics in eukaryotic microbiomes.

Taken together, my thesis sheds light on yet unexplored aspects of symbiont evolution, especially in phylogenetically diverse, multimember associations.

Zusammenfassung

Das Zusammenleben von eukaryotischen Wirten mit vielfältigen bakteriellen Gemeinschaften ist in der Natur weitverbreitet. Die wahrscheinlich bekanntesten und am besten erforschten Beispiele für solche Lebensgemeinschaften sind die Darmbakterien von Wirbeltieren, und besonders die des Menschen. Die Vielfältigkeit und Zusammensetzung der menschlichen Darmbakterien werden fortwährend mit der Gesundheit des Wirtes in Zusammenhang gebracht. Dieser Zusammenhang zwischen der menschlichen Darmflora und der Gesundheit des Wirtsorganismus hebt die Wichtigkeit der Erforschung solcher Lebensgemeinschaften hervor. Obwohl immer mehr solcher diverser Lebensgemeinschaften in der Natur gefunden werden und wir verstehen, dass die Assoziation mit Bakterien großen Einfluss auf die Ökologie und Evolution der Wirte hat, verstehen wir verhältnismäßig wenig von den evolutionären Dynamiken, die diesen phylogenetisch diversen Assoziationen zu Grunde liegen. In dieser Doktorarbeit beschäftige ich mich mit der Evolution von Symbionten die mit taxonomisch diversen und weltweit vorkommenden Gruppen mariner, wirbelloser Tiere assoziiert sind. Die hohe taxonomische und biogeographische Diversität dieser Wirte waren Voraussetzung dafür, die Evolution der Symbionten im Kontext von ökologischen Effekten und Effekten der Wirtsevolution zu beleuchten.

Im ersten Teil dieser Doktorarbeit (Kapitel II), habe ich mich auf die Genomreduzierung in bakteriellen Symbionten fokussiert. Die Forschung auf diesem Gebiet beschäftigt sich hauptsächlich mit Symbionten die ausschließlich in ihrem Wirt und vollkommen von der Umwelt isoliert vorkommen. Diese Symbionten erfahren massive Genomreduktion die zu auffallend kleinen Genomgrößen dieser Organismen führt. Im Gegensatz dazu wissen wir verhältnismäßig wenig darüber wie die Genome von Symbionten evolvieren, wenn diese Symbionten nicht vollständig von anderen bakteriellen Populationen isoliert sind. Um dieser Frage nachzugehen habe ich die Genomreduzierung in 12 nahverwandten Gruppen von Ekto- und Endosymbionten analysiert. Alle diese Gruppen waren von Genomreduzierung betroffen, jedoch scheint die endosymbiontische Assoziierung diesen Prozess zu beschleunigen.

Im zweiten Teil dieser Doktorarbeit (Kapitel III und IV), habe ich mich auf die Zusammensetzung von symbiontischen Gemeinschaften in zwei Tiergruppen fokussiert. In Kapitel III habe ich die Evolution der symbiontischen Gemeinschaften von darmlosen Oligochäten analysiert. Darmlose Oligochäten sind eine Gruppe mariner Ringelwürmer die ihre gesamte Nahrung von ihren Symbionten beziehen und deshalb obligat von diesen abhängig sind. Trotz dieser hohen Abhängigkeit sind diese Würmer mit vielfältigen Gemeinschaften von Symbionten assoziiert. Diese Diversität scheint daraus zu resultieren, dass die Symbionten mit

verschiedener Zuverlässigkeit zwischen den Wirtstieren verbreitet werden was zu einem Ausbalancieren von Stabilität und Flexibilität führt. In Kapitel IV habe ich die Zusammensetzung symbiontischen Partner von Plattentieren untersucht. Plattentiere divergierten früh in der Evolution der Eukaryoten. Deshalb, und auch aufgrund ihrer simplen Morphologie werden sie oft als Modell in der Erforschung von Entwicklung und Evolution von Tieren genutzt. Die bakteriellen Partner der Plattentiere sind jedoch weitestgehend unerforscht. Ich fand heraus, dass Plattentiere häufig mit vielfältigen Gemeinschaften von sowohl inter- als auch intrazellulären Symbionten assoziiert sind. Daher postuliere ich, dass diese Tiere als Modell für die Erforschung der ökologischen und evolutionären Dynamiken zwischen Eukaryoten und ihren bakteriellen Symbionten eignen.

Zusammengenommen beleuchtet meine Doktorarbeit bisher unerforschte Aspekte in der Evolution von bakteriellen Symbionten, mit besonderem Fokus auf phylogenetisch vielfältige Symbiosen zwischen Tieren und mehreren bakteriellen Partnern.

Chapter I | Introduction

1.1 Symbiosis

"No man is an island, entire of itself […]."[1] Although, it probably was not the interpretation that John Donne imagined when he wrote this famous line, changing just one word of it summarizes the importance of symbiotic interactions for life on earth: No organism is an island. Today, we understand that interactions between organisms of different species did not only initiate the evolution of the eukaryotic lineage but also substantially shape the ecology and evolution of likely all organisms[1–8]. Such interactions were first described in 1879 by Anton de Bary who defined the term symbiosis (Greek 'living together') as a long-lasting and intimate association between organisms that belong to different species[9]. Symbioses can occur between organisms of all kingdoms of life and include famous examples such as the symbiosis between clownfish and anemones or the cleaning symbiosis[10,11]. In this thesis, I will focus on the association between eukaryotic, especially animal hosts, and their association with bacterial symbionts. Symbiotic associations are commonly divided into categories, based on the effects of the association on the involved partners, the partners' dependence on each other, the symbiont localization, the transmission mode and the specificity of the association (Box 1).

Throughout this thesis, I will use the term symbiosis according to its original definition, i.e. referring to mutualist, commensalistic and parasitic associations. Also, I will use the term 'symbiont community' to describe any sort of symbiont consortia, not considering the diversity of the bacterial symbionts or the intimacy of the association. However, in Chapter IV, I will refer to the symbiont community of placozoans as their 'microbiome' to emphasize that further research is required to confirm a symbiotic association with all putative symbiont clades.

[1]John Donne, Meditation XVII, Devotions upon Emergent Occasions, 1624

Box 1 | Classification of symbioses

Types of symbioses: Mutualism, commensalism and parasitism

The term symbiosis is often synonymously used to mutualism, but according to its original definition, it refers to three types of associations; i) mutualistic associations, where the partners mutually benefit from each other, ii) commensalistic associations, where one partner benefits without causing any harm to its partner and iii) parasitic associations, where one partner benefits by harming the other partner[12].

Dependence of the partners: Obligate and facultative associations

If one partner completely depends on the association with its counterpart, the association is considered obligate whereas an association where the partners can still survive without each other is considered facultative. Importantly, an association can be obligate for one, and facultative for the other partner.

Symbiont localization: Ecto- and endosymbionts, inter- and intracellular symbionts

Ectosymbionts occur outside of their hosts' bodies while endosymbionts reside within them. In addition, endosymbionts can be intercellular, i.e. occur between host cells or intracellular, i.e. within host cells.

Symbiont transmission: vertical, horizontal or something in between

Vertical transmission describes the direct transfer of symbionts from parental to offspring specimens, often via the maternal germ line. In contrast, horizontal transmission describes other transmission routes, e.g. acquisition of symbionts from the environment or from other host individuals (host switching). If symbionts are neither strictly vertically nor strictly horizontally transmitted, one refers to mixed-mode transmission[13].

Symbiont specificity: from intimate one-on-one associations to complex microbiomes

Symbiont specificity refers to the taxonomic range of symbionts in an association. High specificity thus describes the selective association with a single partner whereas diverse multimember symbioses are per definition less specific[8]. The probably most diverse symbiont communities are found in association with vertebrates, e.g. the human gut symbionts[14,15]. These diverse symbiont communities are often referred to as microbiomes, underpinning their diversity and highlighting that these associations are not necessarily all symbiotic.

1.2 Symbiont evolution

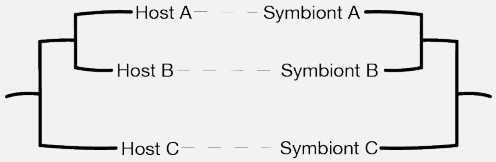
1.2.1 Distinguishing between symbiont transmission and fidelity

The evolution of symbionts is often interpreted in the light of symbiont transmission. For instance, strictly vertical transmission of the aphid symbionts *Buchnera* resulted in the co-divergence between host and symbiont lineages and, over longer evolutionary time scales, in several co-speciation events[16]. However, such effects are not necessarily the result of vertical symbiont transmission but could also result from stringent symbiont selection during horizontal transmission[17]. Thus, I will also interpret the symbiont evolution in the light of symbiont fidelity rather than only symbiont transmission. Symbiont fidelity describes the degree to which a certain host lineage is associated with its particular symbiont lineage over generations and evolutionary scales, either through strict vertical or selective horizontal symbiont transmission[18]. A major advantage of the fidelity-based classification over the transmission-based classification is that the degree of symbiont fidelity can be inferred from phylogenetic data alone (Box 2). High congruence between host and symbiont phylogenies, indicate high symbiont fidelity, whereas largely incongruent phylogenies indicate low symbiont fidelity. Whilst low symbiont fidelity is always the result of at least occasional horizontal transmission events, vertical, horizontal and mixed-mode transmission alike can lead to high symbiont fidelity.

Box 2 | Inferring symbiont fidelity from host and symbiont phylogenies

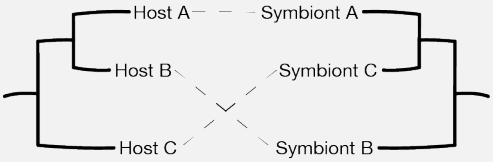
High symbiont fidelity

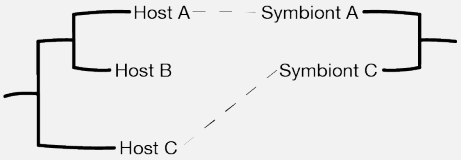
High symbiont fidelity can evolve from strictly vertical symbiont transmission or stringent and host genotype specific acquisition of symbionts from the environment. In both cases, host and symbiont phylogenies would be congruent and the phylogenetic data alone does not indicate the transmission mode.



Low symbiont fidelity

Low symbiont fidelity can evolve from host switching between unrelated host lineages, host genotype independent acquisition of symbionts from the environment or symbiont loss. All cases would lead to at least few incongruencies between the host and the symbiont phylogeny. The degree of incongruence would indicate the degree of fidelity. Similar to the example of high symbiont fidelity, the phylogenetic data alone does not indicate the transmission mode.





1.2.2 Evolution of symbiont genomes

Symbionts often undergo drastic genome reduction that can even lead to the decay of the symbiotic association[19–21]. Massive genome reduction is the result of genetic isolation of symbiont populations, thus it is especially pronounced in strictly vertically transmitted, intracellular symbionts[19,21]. Vertical symbiont transmission leads to extreme bottleneck effects as only a subset of the symbionts that are associated with the parental animals is transferred to the offspring[22]. Genetic bottlenecks as those during vertical symbiont transmission drastically

reduce the effective population size[23]. In small populations, natural selection becomes less effective and slightly deleterious mutations can accumulate. This phenomenon, which is termed Muller's ratchet, leads to extreme genome reduction in several strictly vertically transmitted and obligate insect symbionts (Figure 1B)[24,25]. In contrast, genetic exchange between symbiont populations from different host animals, e.g. through occasional host switches, or between a symbiont population and the free-living bacterial population, due to horizontal symbiont transmission or an extracellular symbiont lifestyle, could prevent massive and detrimental genome reduction (Figure 1A)[26,27].

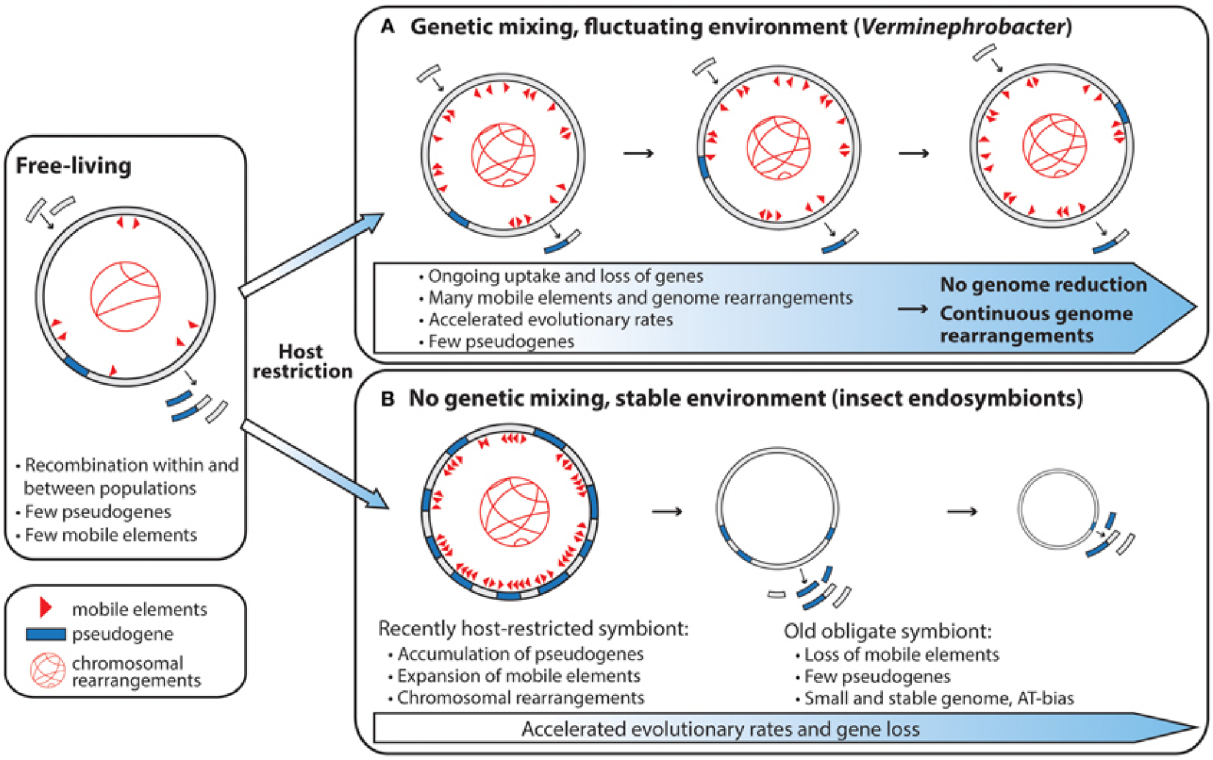


Figure 1 | Genome evolution of symbionts proceeds differently depending on the genetic isolation of the symbionts. The genome evolution of symbionts that are in genetic flux with other bacteria is characterized by ongoing rearrangements, accelerated evolutionary rates, few pseudogenes and repeated loss and uptake of genes **(A)**. Strictly host restricted symbionts undergo massive genome reduction that is characterized by an initial increase of pseudogenes, mobile genetic elements and chromosomal rearrangements and proceeds towards small, stable and AT-biased genomes **(B)**. This figure is taken from Lund *et al.* (2014).

Genome reduction proceeds in several steps. Initially, the accumulation of mutations leads to shifts in open reading frames (ORFs) of coding genes and thus, pseudogenization. In addition, mobile genetic elements (MGEs) proliferate, the genome undergoes rearrangements and chromosomal fragments are deleted. With progressing genome reduction, the pseudogenes and MGEs are degraded, the genomes become smaller and the genome arrangement stabilizes. Extremely reduced genomes are also often, but not always, characterized by a low genomic GC content (Figure 1)[19,21]. The process of genome reduction lead to genomes that are as small as 0.14 Mbp[28,29]. As mentioned above, ongoing genome reduction could lead to the decay of the association, e.g. when the symbiont loses functions that were crucial for its survival. In contrast, given strong co-adaptation between the host and the symbionts and high interdependence and metabolic integration, the symbionts could evolve further into organelle-like organisms[19–21].

1.2.3 Evolution of symbiont community composition

How can genetic diversity evolve in mutualistic associations? is one of the pressing questions in symbiosis research[30]. For the longest time, mutualistic associations were interpreted as a paradox. Evolutionary theory predicts that bacterial lineages that exploit host resources without returning any goods, so called cheaters, would be favored over the mutualistic symbiont by natural selection[30,31]. This, in consequence, would lead to the decay of the association. Several theoretical and experimental studies alike found that mechanisms that ensure the faithful association with the mutualistic symbiont type, e.g. high symbiont specificity and high symbiont fidelity would stabilize the mutualistic associations. Such mechanisms include the stringent selection or strictly vertical transmission of the cooperative symbiont type as well as partner control via e.g. sanctioning of cheaters[32–37]. Based on these findings, mutualistic associations should theoretically be characterized by low genetic diversity of the symbionts[31]. Yet, several examples of genetically diverse mutualism are ubiquitous in nature, including symbiont communities that consist of several strains or symbiont communities that contain a

variety of distinct symbiont lineages from different phyla[15,38,39]. How this genetic variation evolves remains largely elusive[30].

Besides the aim to understand the dynamics that lead to the evolution of genetically diverse consortia of mutualistic symbionts, the research on symbiont communities is also driven by the question *which factors influence the composition of a given symbiont community?* As symbiont community composition can largely impact host functioning and health, as seen for e.g. the human gut microbiome, it is crucial to understand the drivers of symbiont community composition to get a holistic understanding of the host's biology[14]. Symbiont community composition can be influenced by environmental parameters, such as the availability of nutrients. For instance, bathymodioline mussels harbor more bacterial symbiont strains that are capable of hydrogen oxidation at locations where vent fluids contained higher concentrations of hydrogen[39]. In some coral species, the symbionts vary according to water depth, i.e. different light availability[40]. Besides chemical and physical parameters, symbiont community of e.g. sponges was shown to be dependent on the host taxonomy, the lifecycle stage and the composition of the local free-living bacterial populations[41]. In addition, similarity patterns of symbiont community compositions across a variety of hosts have been observed to be correlated to the phylogenetic relations between the host individuals, a pattern that is termed phylosymbiosis[42–49]. Phylosymbiosis can be the result of stochastic effects, ecological or dietary niche variation along the host diversity or disruptive host traits such as antimicrobial peptides and thus, observing a phylosymbiotic signal does not necessarily indicate host evolution as the driver of symbiont community composition; however, high symbiont fidelity in form of strictly vertical symbiont transmission or stringent environmental selection of all symbionts would lead to the observation of phylosymbiosis[45,50–56].

1.3 Nutritional symbioses in the marine environment

1.3.1 Chemosynthetic and chemotrophic symbioses

Chemosynthetic symbioses are wide spread in the marine environment and (subjectively) among the most fascinating associations between animals and bacteria[57,58]. Chemosynthesis describes the metabolic process by which bacteria autotrophically fix CO_2 into organic carbon and biomass. In contrast to photosynthetic organisms that use light energy to fuel carbon fixation, chemosynthetic bacteria generate energy via the oxidation of reduced sulfur compounds, hydrogen or methane. Thus, these organisms can thrive in light depleted deep sea ecosystems that are rich in reduced inorganic compounds, e.g. in proximity of hydrothermal vents, cold seeps and wood and whale falls[59]. Strikingly, it is not only the bacteria themselves that can thrive in these environments but the symbiotic association with chemosynthetic bacteria allows several animal phyla to densely populate these deep sea habitats (Figure 2)[57,58]. Although chemosynthetic symbioses from the deep sea environment are among the most famous, the association between marine invertebrates and chemosynthetic bacteria is not limited to the deep sea. On the contrary, a variety of eukaryotic hosts from several different phyla such as flatworms, nematodes, annelids and bivalves are among the most successful inhabitants of diverse coastal, shallow-water sediments (Figure 2)[57,58].

A special case of chemotrophic, but not chemosynthetic, associations is the symbiosis between *Kentrophoros* ciliates and their symbionts of the genus *Cand.* Kentron. These symbionts were thought to also provide their hosts nutrition through chemosynthesis[60]. However, omics based analyses revealed that these symbionts indeed generate energy via the oxidation of reduced sulfur compounds but they lack genes for autotrophic carbon fixation. Instead, they assimilate carbon via heterotrophic pathways[61].

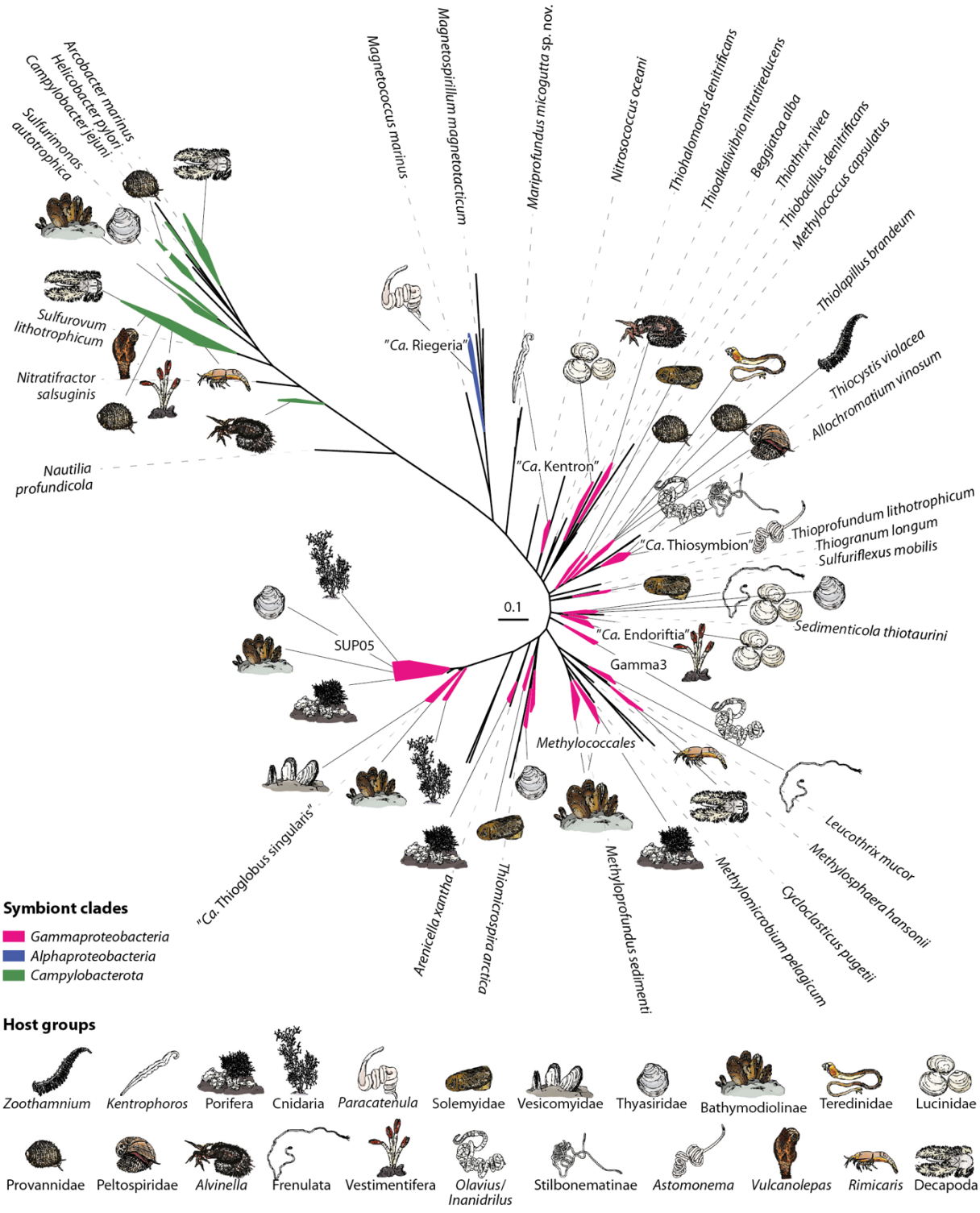


Figure 2 | Phylogenetic diversity of chemotrophic symbionts from ciliate and animal hosts. The majority of chemotrophic symbionts belong to the Gammaproteobacteria but association with symbionts from the Alphaproteobacteria and the Campylobacterota are also common. The hosts of chemotrophic symbiont belong to diverse animal phyla, including ciliates, porifera, cnidarians, flatworms, molluscs, nematodes, annelids and crustaceans. This figure is taken from Sogin *et al.* (2021). Usage of the figure was licensed under the license ID 1146310-1.

1.3.2 Other nutritional symbioses

The research on nutritional symbiosis in the marine environment is dominated by chemo- and photosynthetic symbiosis. Similar to chemosynthetic symbiosis, the widespread association with photosynthetic cyanobacteria or algae provides organic carbon to the host animals, e.g. cnidarians, sponges, flatworms, molluscs and ascidians[62]. However, other forms of nutritional symbiosis between marine animals and bacterial symbionts exist. For instance, marine vertebrates such as fish are associated with diverse gut microbiomes which, as the gut microbiomes of terrestrial vertebrates, could contribute to their hosts nutrition[14,15,63]. In contrast, no examples of marine animals are known where the host dependents on the provision of few limiting nutrients by the symbiont, comparable to the plant sap feeding insects in terrestrial ecosystems that rely on symbionts to supplement their nutrition with only few or even only one vitamins and/or amino acids[28,64–73]. However, intimate association between animals such as bryozoans and placozoans and bacterial symbionts have been reported[74–82]. These symbionts could contribute to their hosts' nutrition[77,81]. Yet, further research is required to fully understand the role of these symbionts.

1.4 Gutless oligochaetes

The monophyletic group of gutless oligochaetes belongs to the subfamily of Phallodrilinae and consists of > 100 described species that belong to two host genera, *Inanidrilus* and *Olavius*. They inhabit globally distributed, coastal, shallow-water marine sediments in tropical to subtropical regions (Figure 3). The association of these worms with chemosynthetic symbionts that provide them with their entire nutrition and waste product recycling lead to the secondary reduction of the worms' digestive system, including mouth, gut and anus, and their excretory organs (hence gutless, Figure 3)[83–86].

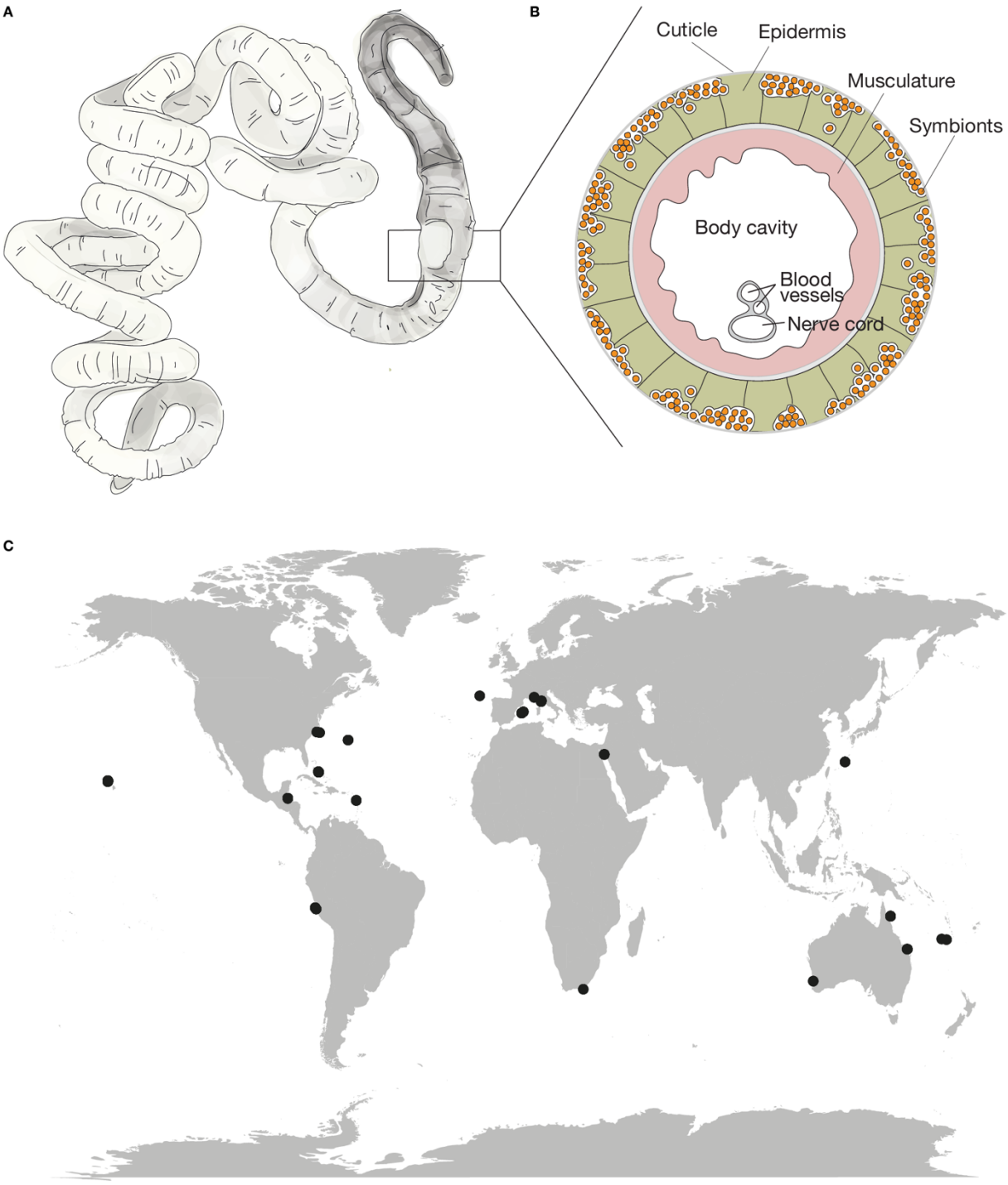


Figure 3 | Gutless oligochaetes lack a digestive system and thrive in shallow-water sediments through nutrition provided by bacterial symbionts that reside below their cuticle. (A) Schematic representation of *O. algarvensis* © Alina Esken. **(B)** Schematic representation of a gutless oligochaete cross-section © Alexander Gruhl. **(C)** Global distribution of gutless oligochaete specimens analyzed in this thesis (Chapter III).

Compared to other chemosynthetic symbioses, gutless oligochaetes are associated with a striking diversity of symbionts. So far, the symbiont communities of seven host species have been studied. In these seven species, 14 symbiont genera from different bacterial phyla were detected. Each host species was associated with a species-specific subset of four to six of these symbionts[83,87–92]. In addition to highly variable symbiont community composition, also the fidelity between host and symbionts appear to be highly flexible[93,94]. For instance, all but one host species were associated with the same primary symbiont, *Cand.* Thiosymbion[87]. *Cand.* Thiosymbion is not only associated with gutless oligochaetes but also with stilbonematine and *Astomonema* nematodes. In the nematode hosts, *Cand.* Thiosymbion displays high symbiont fidelity. In gutless oligochaetes, specificity between a given host species and its associated *Cand.* Thiosymbion phylotype is indeed high but the comparison of the host and symbiont phylogenies revealed several incongruences that indicate frequent host switching[93]. In one host species, *I. exumae*, *Cand.* Thiosymbion was even completely lost and replaced by another, unrelated gammaproteobacterial symbiont[87]. In addition, all the symbionts of the Mediterranean species *O. algarvensis* display varying symbiont fidelity within one host species, i.e. appear to be switched at different rates between host individuals that belong to the same species[94].

Experimental and omics based studies of the symbiont metabolism of the *O. algarvensis* symbionts revealed that the association with different symbionts can optimize the use of resources and widens the nutritional spectrum of the hosts (Figure 4)[83–86]. Syntrophic sulfur cycling between sulfur oxidizing gammaproteobacterial and sulfate reducing deltaproteobacterial symbionts replenishes the main energy source of the primary symbiont[83,84]. Thus, the symbiosis could thrive even if environmental sources of reduced sulfur compounds were limited. In addition, the association with phylogenetically distinct symbionts increases the carbon and energy sources that are used by the symbiotic consortium. For instance, the deltaproteobacterial symbionts can scavenge additional carbon and energy from the oxidation

of hydrogen, carbon monoxide as well as organic carbon sources, e.g. short-chain fatty acids[85]. Fascinatingly, *O. algarvensis* can be simultaneously associated with two different deltaproteobacterial symbionts. These Deltaproteobacteria appeared to be functionally redundant as they both encode and express the same pathways for heterotrophic carbon metabolism[84,85]. However, the carbon signatures of the symbionts' proteins indicated that these symbionts might use the same pathways to metabolize organic carbon from different sources, i.e. host waste products and seagrass exudates and thus, inhabit different microniches within the symbiosis[86].

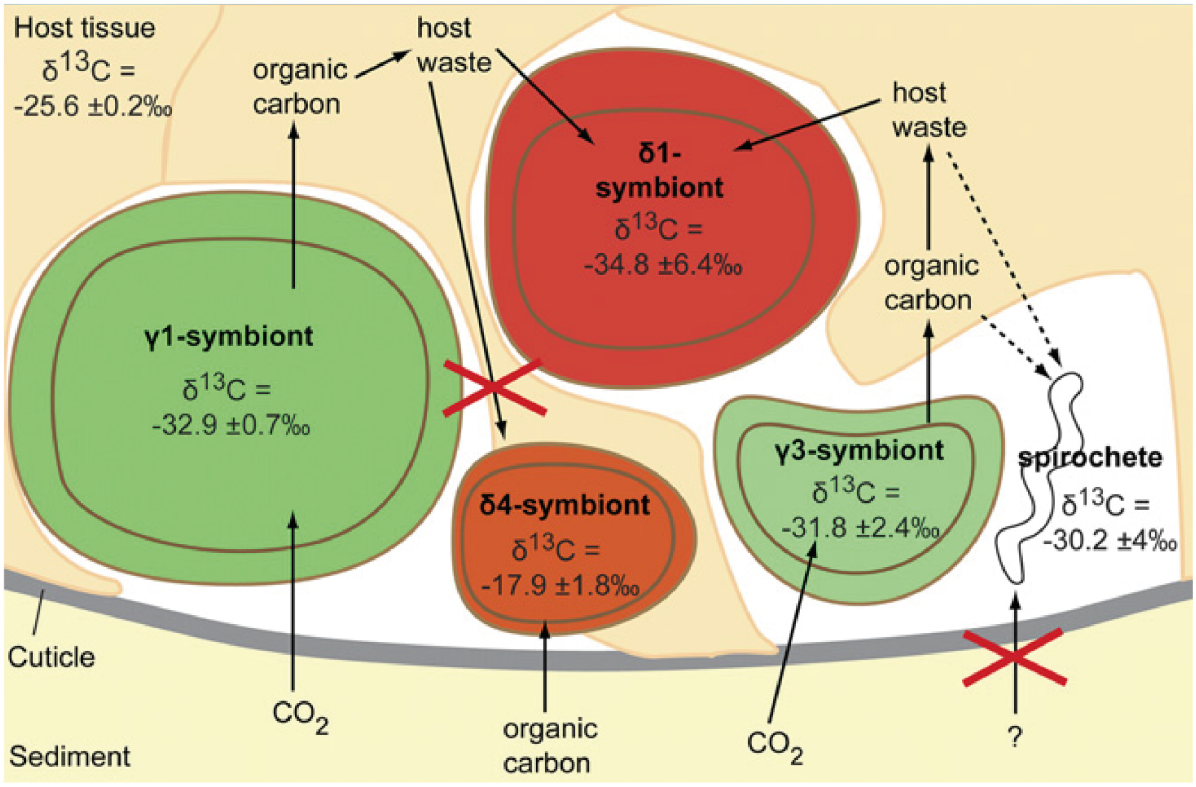


Figure 4 | Stable-isotope fingerprints of the *O. algarvensis* symbionts' proteins indicate that the deltaproteobacterial and Spirochete symbionts use different sources of organic carbon. The lower $\delta^{13}C$ signatures of the Delta1 symbionts indicate that they oxidize host-derived organic compounds whereas the higher $\delta^{13}C$ signatures of the Delta4 symbionts indicate that they oxidize organic carbon that is derived from environmental sources. This figure is taken from Kleiner *et al.* (2018).

1.5 Placozoa

Placozoa is a phylum of the arguably most simple animals that diverged close to the base of the animal phylogeny, most likely as a sister clade to the cnidarians[95,96]. These benthic animals inhabit diverse and globally distributed marine habitats[97]. The phylum Placozoa consists of only three described species from three genera; *Trichoplax adhaerens*, *Hoilungia hongkongensis* and *Polyplacotoma mediterranea* but molecular analyses of the mitochondrial 16S rRNA gene of placozoans indicates that at least 19 different, morphologically indistinguishable genotypes exist (Figure 5)[76,95,98,99]. These mitochondrial genotypes are commonly referred to as host haplotypes.

The bodies of placozoans are millimeter-sized and consist of three distinct cell layers: the dorsal and the ventral epithelium and a connecting meshwork of fiber cells[82,100,101]. The dorsal epithelium consists of ciliated epithelial cells that can be intermixed with fiber cells. The ventral epithelium also consists of ciliated epithelial cells and also glandular and lipophilic cells. As the placozoans do not have an internal digestive system, they excrete digestive enzymes that are used for external digestion of algae and microbial biofilms. The released nutrients are taken up via the animals' ventral epithelia[102,103]. Due to their early divergence during the animal evolution, their simple morphology, their global distribution and the ability to relatively easily sample and cultivate placozoans, they are commonly used as model organism for metazoan evolution, developmental biology and tissue formation[79,95,104].

As most other animals, placozoans live in symbiotic association with bacteria. Despite several mentions of placozoan associated symbionts, little is known about their diversity, their community composition, their phylogenetic relations across host haplotypes and the role for the symbiosis[75–82]. However, previous studies speculated that some of the symbionts could supplement their hosts diet with essential nutrients such as vitamins and amino acids[77,81].

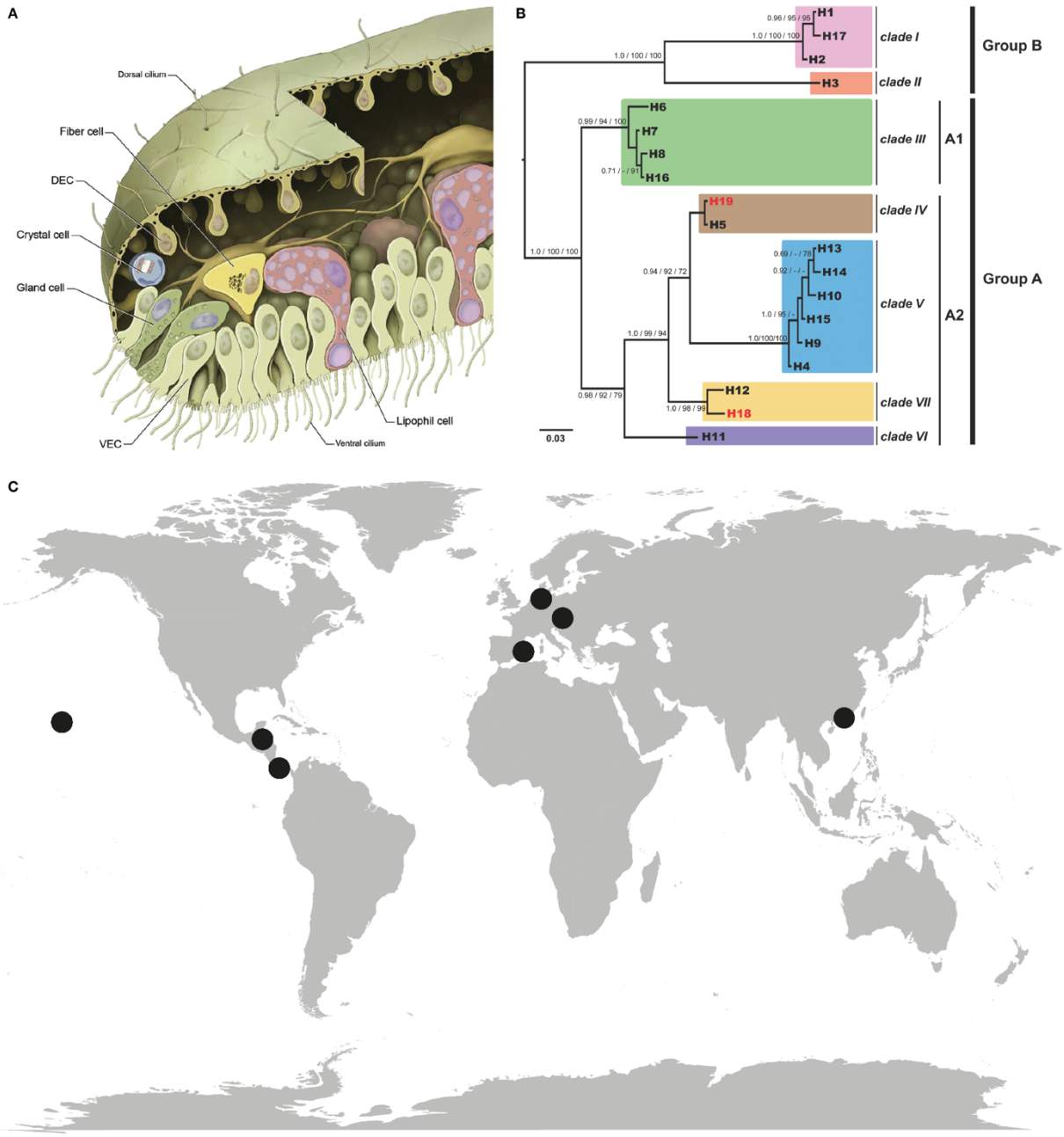


Figure 5 | Morphology, phylogeny and global distribution of Placozoa. (A) Schematic representation of the placozoan morphology. This figure is taken from Smith *et al.* (2014). Usage of this figures was licensed under license ID 5144160912321 **(B)** Placozoan phylogeny based on mitochondrial 16S rRNA genes. This figure is taken from Eitel *et al.* (2018). **(C)** Global distribution of placozoan specimens analyzed in this thesis (Chapter IV).

1.6 Aims of the thesis

This thesis aims to increase our understanding of evolutionary dynamics in symbiotic non-model organisms. I was particularly interested in how symbiont evolution is shaped by environmental parameters and the host evolution. Symbiont evolution and particularly symbiont community composition tremendously influences the host biology, thus understanding the drivers of symbiont community composition is a key part of symbiosis research. In order to untangle the different geographic, environmental and evolutionary effects on symbiont community composition, it is crucial to analyze symbiotic associations that are i) taxonomically diverse, both on the side of the hosts and the symbionts and ii) globally distributed in different habitats. Marine invertebrates that occur in shallow-water habitats and that harbor multimember symbiont consortia are thus predestined study organisms for this field of research as it is relatively easy to sample high numbers of different host species from different sampling locations. A second major aspect of symbiont evolution that I explored is symbiont genome evolution. Studies on genome reduction are biased towards strictly vertically, intracellular symbionts and differences of genome evolution in different types of symbionts remains largely unexplored. One basic example of such unexplored differences is the evolution of genomes of endo- and ectosymbionts. Different symbiont lifestyles could alter the degree of genetic isolation of symbiont populations. Thus, studying such differences will allow us to understand how genetic isolation impacts genome evolution in different associations.

Throughout my thesis, I analyzed symbiont evolution using metagenomic datasets. Metagenomics allow us to assemble host and symbiont marker genes for assessing symbiont (community) evolution in the light of host evolution. In contrast to targeted marker gene sequencing, metagenomics is a primer-free sequencing approach. Thus, no previous knowledge on the host and symbiont gene sequences is required to ensure that all community members are targeted by the sequencing approach. The application of metagenomics also allows us to not

only analyze the symbiont community composition based on marker genes but we can use the exact same datasets to generate symbiont genome drafts. Annotating and analyzing symbiont draft genomes allows us to infer their function and thus, speculate how the observed symbiont community composition influences its phenotype, i.e. the metabolic capabilities of a given community. This knowledge can be further used to understand why certain symbionts are present in their respective hosts or at certain environments. In addition, analyzing symbiont genome drafts allows us to further understand the evolution of symbiont genomes.

How does genome reduction proceed in closely related lineages of ecto- and endosymbionts?

The research on symbiont genome reduction is extremely biased towards symbionts that have lived under severe host restriction for millions of years and thus, already underwent massive genome reduction. Whilst studying these extreme examples is valuable to understand how extremely high interdependence between hosts and their sometimes organelle-like symbionts evolved, it cannot answer how this extreme genome reduction proceeded. In Chapter II, I analyzed genome reduction in a symbiont genus that consists of 12 distinct subclades of which eleven consisted of ectosymbionts whilst the 12th consisted of endosymbionts. Given the two different lifestyles of these symbionts, I could not only compare genome reduction in closely related but distinct symbiont lineages but also compare how the symbiont lifestyle and the resulting differences in the levels of genetic isolation of the symbiont populations influence genome reduction.

How do varying levels of symbiont fidelity influence the evolution of symbiont community composition across a broad host diversity?

In Chapter III, I asked how environmental parameters and the hosts' evolutionary relationships could have influenced the evolution of diverse symbiont communities in gutless oligochaetes. When I started my PhD, only few gutless oligochaetes and their associated symbionts were studied in detail. When I obtained my first results, I was amazed to observe a broad variety of bacterial symbionts that showed different association patterns across the host diversity. I analyzed this associations on a broad evolutionary spectrum; starting at the macroevolutionary level I compared symbiont community composition across host species and zoomed in to analyze the fidelity of individual symbiont genera, from across host species to within host species and even host populations.

What is the symbiont diversity of the arguably most simple animals?

Despite their simplicity, placozoans are an intriguing model system to study eukaryotic evolution and development. I argue, that these organisms could additionally be used to study the ecological and evolutionary mechanisms behind microbiome evolution, however, their symbiont communities have not been intensively studied yet. In Chapter IV, I present a first comprehensive overview of the symbiont diversity across a broad phylogenetic range of placozoan hosts.

References

1. Sapp, J. The Symbiotic Self. *Evol. Biol.* **43**, 596–603 (2016).
2. Margulis, L. & Fester, R. Symbiosis assource of evolutionary innovation: speciation and morphogenesis. *The MIT Press.* 93–101 (1991).
3. 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. *Yale University Press*. (1970).
4. Sagan, L. On the origin of mitosing cells. *J. Theor. Biol.* **14**, 225-IN6 (1967).
5. Gilbert, S. F., Bosch, T. C. G. & Ledón-Rettig, C. Eco-Evo-Devo: developmental symbiosis and developmental plasticity as evolutionary agents. *Nat. Rev. Genet.* **16**, 611–622 (2015).
6. McFall-Ngai, M. *et al.* Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad. Sci.* **110**, 3229 LP – 3236 (2013).
7. Henry, L. P., Bruijning, M., Forsberg, S. K. G. & Ayroles, J. F. The microbiome extends host evolutionary potential. *Nat. Commun.* **12**, 5141 (2021).
8. Douglas, A. The Symbiotic Habit. *Princeton University Press* (2010).
9. de Bary, A. Die Erscheinung der Symbiose. *Karl J. Trübner Verlag* (1879).
10. Collingwood, C. IV. Note on the existence of gigantic sea-anemones in the China Sea, containing within them quasi-parasitic fish. *J. Nat. Hist.* **1**, 31–33 (1868).
11. Gibson, R. N. & Barnes, M. Evolution and ecology of cleaning symbioses in the sea. *Oceanogr. Mar. Biol. An Annu. Rev.* **38**, 311 (2000).
12. Sapp J. Evolution by association : a history of symbiosi*s*. *Oxford University Press, USA*. (1994).
13. Bright, M. & Bulgheresi, S. A complex journey: transmission of microbial symbionts. *Nat. Rev. Microbiol.* **8**, 218–230 (2010).
14. Gomaa, E. Z. Human gut microbiota/microbiome in health and diseases: a review. *Antonie Van Leeuwenhoek* **113**, 2019–2040 (2020).
15. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
16. Munson, M. A. *et al.* Evidence for the establishment of aphid-eubacterium endosymbiosis in an ancestor of four aphid families. *J. Bacteriol.* **173**, 6321–6324 (1991).

17. Nishiguchi, M. K., Ruby, E. G. & McFall-Ngai, M. J. Competitive Dominance among Strains of Luminous Bacteria Provides an Unusual Form of Evidence for Parallel Evolution in Sepiolid Squid-Vibrio Symbioses. *Appl. Environ. Microbiol.* **64**, 3209–3213 (1998).
18. Douglas, A. E. & Werren, J. H. Holes in the Hologenome: Why Host-Microbe Symbioses Are Not Holobionts. *MBio* **7**, e02099-15 (2016).
19. McCutcheon, J. P. & Moran, N. A. Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**, 13–26 (2012).
20. McCutcheon, J. P., Boyd, B. M. & Dale, C. The Life of an Insect Endosymbiont from the Cradle to the Grave. *Curr. Biol.* **29**, R485–R495 (2019).
21. Toft, C. & Andersson, S. G. E. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat. Rev. Genet.* **11**, 465–475 (2010).
22. Mira, A. & Moran, N. A. Estimating Population Size and Transmission Bottlenecks in Maternally Transmitted Endosymbiotic Bacteria . *Microb. Ecol.* **44**, 137–143 (2002).
23. Bobay, L.-M. & Ochman, H. Factors driving effective population size and pan-genome evolution in bacteria. *BMC Evol. Biol.* **18**, 153 (2018).
24. Moran, N. A. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci.* **93**, 2873 LP – 2878 (1996).
25. Rispe, C. & Moran, N. A. Accumulation of Deleterious Mutations in Endosymbionts: Muller's Ratchet with Two Levels of Selection. *Am. Nat.* **156**, 425–441 (2000).
26. Russell, S. L. *et al.* Horizontal transmission and recombination maintain forever young bacterial symbiont genomes. *PLOS Genet.* **16**, e1008935 (2020).
27. Lund, M., Kjeldsen, K. & Schramm, A. The earthworm-*Verminephrobacter* symbiosis: an emerging experimental system to study extracellular symbiosis. *Frontiers in Microbiology* **5** 128 (2014).
28. McCutcheon, J. P., McDonald, B. R. & Moran, N. A. Origin of an alternative genetic code in the extremely small and GC–rich genome of a bacterial symbiont. *PLoS Genet.* **5**, e1000565 (2009).
29. López-Madrigal, S., Latorre, A., Porcar, M., Moya, A. & Gil, R. Complete genome sequence of '*Candidatus* Tremblaya princeps' strain PCVAL, an intriguing translational machine below the living-cell status. *J. Bacteriol.* **193**, 5587–5588 (2011).
30. Heath, K. D. & Stinchcombe, J. R. Explaining Mutualism Variation: A New Evolutionary Paradox? *Evolution (N. Y).* **68**, 309–317 (2014).
31. Leigh Jr, E. G. The evolution of mutualism. *J. Evol. Biol.* **23**, 2507–2528 (2010).

32. Nyholm, S. V & McFall-Ngai, M. The winnowing: establishing the squid–*vibrio* symbiosis. *Nat. Rev. Microbiol.* **2**, 632–642 (2004).
33. Visick, K. L. & McFall-Ngai, M. J. An Exclusive Contract: Specificity in the *Vibrio fischeri-Euprymna scolopes* Partnership. *J. Bacteriol.* **182**, 1779–1787 (2000).
34. Bull, J. J. & Rice, W. R. Distinguishing mechanisms for the evolution of co-operation. *J. Theor. Biol.* **149**, 63–74 (1991).
35. Archetti, M. *et al.* Economic game theory for mutualism and cooperation. *Ecol. Lett.* **14**, 1300–1312 (2011).
36. Westhoek, A. *et al.* Policing the legume-Rhizobium symbiosis: a critical test of partner choice. *Sci. Rep.* **7**, 1419 (2017).
37. Kiers, E. T. & Denison, R. F. Sanctions, cooperation, and the stability of plant-rhizosphere mutualisms. *Annu. Rev. Ecol. Evol. Syst.* **39**, 215–236 (2008).
38. Ellegaard, K. M. & Engel, P. Genomic diversity landscape of the honey bee gut microbiota. *Nat. Commun.* **10**, 446 (2019).
39. Ansorge, R. *et al.* Functional diversity enables multiple symbiont strains to coexist in deep-sea mussels. *Nat. Microbiol.* **4**, 2487–2497 (2019).
40. Cooper, T. F. *et al.* Niche specialization of reef-building corals in the mesophotic zone: metabolic trade-offs between divergent *Symbiodinium* types. *Proc. R. Soc. B Biol. Sci.* **278**, 1840–1850 (2011).
41. Sacristán-Soriano, O. *et al.* Ontogeny of symbiont community structure in two carotenoid-rich, viviparous marine sponges: comparison of microbiomes and analysis of culturable pigmented heterotrophic bacteria. *Environ. Microbiol. Rep.* **11**, 249–261 (2019).
42. Ochman, H. *et al.* Evolutionary Relationships of Wild Hominids Recapitulated by Gut Microbial Communities. *PLOS Biol.* **8**, e1000546 (2010).
43. Phillips, C. D. *et al.* Microbiome analysis among bats describes influences of host phylogeny, life history, physiology and geography. *Mol. Ecol.* **21**, 2617–2627 (2012).
44. Easson, C. G. & Thacker, R. W. Phylogenetic signal in the community structure of host-specific microbiomes of tropical marine sponges. *Frontiers in Microbiology* vol. 5 532 (2014).
45. Moeller, A. H. *et al.* Dispersal limitation promotes the diversification of the mammalian gut microbiota. *Proc. Natl. Acad. Sci.* **114**, 13768 LP – 13773 (2017).

46. Brooks, A. W., Kohl, K. D., Brucker, R. M., van Opstal, E. J. & Bordenstein, S. R. Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. *PLOS Biol.* **14**, e2000225 (2016).
47. Brooks, A. W., Kohl, K. D., Brucker, R. M., van Opstal, E. J. & Bordenstein, S. R. Correction: Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. *PLOS Biol.* **15**, 1–2 (2017).
48. Sanders, J. G. *et al.* Stability and phylogenetic correlation in gut microbiota: lessons from ants and apes. *Mol. Ecol.* **23**, 1268–1283 (2014).
49. Mallott, E. K. & Amato, K. R. Host specificity of the gut microbiome. *Nat. Rev. Microbiol.* (2021) doi:10.1038/s41579-021-00562-3.
50. Franzenburg, S. *et al.* Distinct antimicrobial peptide expression determines host species-specific bacterial associations. *Proc. Natl. Acad. Sci.* **110**, E3730 LP-E3738 (2013).
51. Fraune, S. & Bosch, T. C. G. Long-term maintenance of species-specific bacterial microbiota in the basal metazoan *Hydra*. *Proc. Natl. Acad. Sci.* **104**, 13146 LP – 13151 (2007).
52. Ley, R. E. *et al.* Evolution of Mammals and Their Gut Microbes. *Science (80-.).* **320**, 1647 LP – 1651 (2008).
53. Groussin, M. *et al.* Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nat. Commun.* **8**, 14319 (2017).
54. Mazel, F. *et al.* Is Host Filtering the Main Driver of Phylosymbiosis across the Tree of Life? *mSystems* **3**, e00097-18 (2018).
55. Pollock, F. J. *et al.* Coral-associated bacteria demonstrate phylosymbiosis and cophylogeny. *Nat. Commun.* **9**, 4921 (2018).
56. Grieneisen, L. E. *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 Biol. Sci.* **286**, 20190431 (2019).
57. Sogin, E. M., Kleiner, M., Borowski, C., Gruber-Vodicka, H. R. & Dubilier, N. Life in the Dark: Phylogenetic and Physiological Diversity of Chemosynthetic Symbioses. *Annu. Rev. Microbiol.* (2021) doi:10.1146/annurev-micro-051021-123130.
58. Dubilier, N. & Bergin Claudia & Lott, C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat Rev Micro* **6**, 725–740 (2008).
59. Orcutt, B. N., Sylvan, J. B., Knab, N. J. & Edwards, K. J. Microbial Ecology of the Dark Ocean above, at, and below the Seafloor. *Microbiol. Mol. Biol. Rev.* **75**, 361–422 (2011).

60. Fenchel, T. & Finlay, B. J. Kentrophoros: a mouthless ciliate with a symbiotic kitchen garden. *Ophelia* **30**, 75–93 (1989).
61. Seah, B. K. B. *et al.* Sulfur-Oxidizing Symbionts without Canonical Genes for Autotrophic CO2 Fixation. *MBio* **10**, e01112-19 (2019).
62. Venn, A. A., Loram, J. E. & Douglas, A. E. Photosynthetic symbioses in animals. *J. Exp. Bot.* **59**, 1069–1080 (2008).
63. Ghanbari, M., Kneifel, W. & Domig, K. J. A new view of the fish gut microbiome: Advances from next-generation sequencing. *Aquaculture* **448**, 464–475 (2015).
64. McCutcheon, J. P., McDonald, B. R. & Moran, N. A. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc. Natl. Acad. Sci.* **106**, 15394–15399 (2009).
65. Houk, E. J. & Griffiths, G. W. Intracellular Symbiotes of the Homoptera. *Annu. Rev. Entomol.* **25**, 161–187 (1980).
66. Wu, D. *et al.* Metabolic Complementarity and Genomics of the Dual Bacterial Symbiosis of Sharpshooters. *PLOS Biol.* **4**, e188 (2006).
67. Baumann, P. Biology Of Bacteriocyte-Associated Endosymbionts Of Plant Sap-Sucking Insects. *Annu. Rev. Microbiol.* **59**, 155–189 (2005).
68. McCutcheon, J. P. & Moran, N. A. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc. Natl. Acad. Sci.* **104**, 19392–19397 (2007).
69. Nakabachi, A. *et al.* The 160-kilobase genome of the bacterial endosymbiont *Carsonella. Science (80-.).* **314**, 267 (2006).
70. Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. & Ishikawa, H. Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. *Nature* **407**, 81–86 (2000).
71. Douglas, A. E. Mycetocyte Symbiosis In Insects. *Biol. Rev.* **64**, 409–434 (1989).
72. Sasaki, T. & Ishikawa, H. Production of essential amino acids from glutamate by mycetocyte symbionts of the pea aphid, *Acyrthosiphon pisum*. *J. Insect Physiol.* **41**, 41–46 (1995).
73. Buchner, P. Endosymbiosis of animals with plant microorganisms. *New York: Interscience Publisher.* (1965).
74. Karagodina, N. P., Vishnyakov, A. E., Kotenko, O. N., Maltseva, A. L. & Ostrovsky, A. N. Ultrastructural evidence for nutritional relationships between a marine colonial invertebrate (Bryozoa) and its bacterial symbionts. *Symbiosis* **75**, 155–164 (2018).

75. Driscoll, T., Gillespie, J. J., Nordberg, E. K., Azad, A. F. & Sobral, B. W. Bacterial DNA Sifted from the *Trichoplax adhaerens* (Animalia: Placozoa) Genome Project Reveals a Putative Rickettsial Endosymbiont. *Genome Biol. Evol.* **5**, 621–645 (2013).
76. Eitel, M. *et al.* Comparative genomics and the nature of placozoan species. *PLOS Biol.* **16**, e2005359 (2018).
77. Gruber-Vodicka, H. R. *et al.* Two intracellular and cell type-specific bacterial symbionts in the placozoan *Trichoplax* H2. *Nat. Microbiol.* **4**, 1465–1474 (2019).
78. Kamm, K., Osigus, H.-J., Stadler, P. F., DeSalle, R. & Schierwater, B. *Trichoplax* genomes reveal profound admixture and suggest stable wild populations without bisexual reproduction. *Sci. Rep.* **8**, 11168 (2018).
79. Srivastava, M. *et al.* The *Trichoplax* genome and the nature of placozoans. *Nature* **454**, 955–960 (2008).
80. Guidi, L., Eitel, M., Cesarini, E., Schierwater, B. & Balsamo, M. Ultrastructural analyses support different morphological lineages in the phylum placozoa Grell, 1971. *J. Morphol.* **272**, 371–378 (2011).
81. Klinges, J. G. *et al.* Phylogenetic, genomic, and biogeographic characterization of a novel and ubiquitous marine invertebrate-associated Rickettsiales parasite, Candidatus Aquarickettsia rohweri, gen. nov., sp. nov. *ISME J.* **13**, 2938–2953 (2019).
82. Smith, C. L. *et al.* Novel Cell Types, Neurosecretory Cells, and Body Plan of the Early-Diverging Metazoan *Trichoplax adhaerens*. *Curr. Biol.* **24**, 1565–1572 (2014).
83. Dubilier, N. *et al.* Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. *Nature* **411**, 298–302 (2001).
84. Woyke, T. *et al.* Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**, 950–955 (2006).
85. Kleiner, M. *et al.* Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc. Natl. Acad. Sci.* **109**, E1173 LP-E1182 (2012).
86. Kleiner, M. *et al.* Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. *Proc. Natl. Acad. Sci.* **115**, E5576 LP-E5584 (2018).
87. Bergin, C. *et al.* Acquisition of a Novel Sulfur-Oxidizing Symbiont in the Gutless Marine Worm *Inanidrilus exumae*. *Appl. Environ. Microbiol.* **84**, e02267-17 (2018).

88. Dubilier, N., Giere, O., Distel, D. L. & Cavanaugh, C. M. Characterization of chemoautotrophic bacterial symbionts in a gutless marine worm (Oligochaeta, Annelida) by phylogenetic 16S rRNA sequence analysis and in situ hybridization. *Appl. Environ. Microbiol.* **61**, 2346–2350 (1995).
89. Dubilier, N. *et al.* Phylogenetic diversity of bacterial endosymbionts in the gutless marine oligochete Olavius loisae (Annelida). *Mar. Ecol. Prog. Ser.* **178**, 271–280 (1999).
90. Ruehland, C., Blazejak, A., Lott, C., Loy, A. & Erséus Christer & Dubilier, N. Multiple bacterial symbionts in two species of co-occurring gutless oligochaete worms from Mediterranean sea grass sediments. *Environ. Microbiol.* **10**, 3404–3416 (2008).
91. Blazejak, A., Erseus, C. & Amann Rudolf & Dubilier, N. Coexistence of bacterial sulfide oxidizers, sulfate reducers, and spirochetes in a gutless worm (Oligochaeta) from the Peru margin. *Appl. Environ. Microbiol.* **71**, 1553–1561 (2005).
92. Blazejak, A., Kuever, J., Erséus, C., Amann, R. & Dubilier, N. Phylogeny of 16S rRNA, Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase, and Adenosine 5′-Phosphosulfate Reductase Genes from Gamma- and Alphaproteobacterial Symbionts in Gutless Marine Worms (Oligochaeta) from Bermuda and the Bahamas. *Appl. Environ. Microbiol.* **72**, 5527–5536 (2006).
93. Zimmermann, J. *et al.* Closely coupled evolutionary history of ecto- and endosymbionts from two distantly related animal phyla. *Mol. Ecol.* **25**, 3203–3223 (2016).
94. Sato, Y. *et al.* Fidelity varies in the symbiosis between a gutless marine worm and its microbial consortium. *bioRxiv* 2021.01.30.428904 (2021) doi:10.1101/2021.01.30.428904.
95. Eitel, M., Osigus, H.-J., DeSalle, R. & Schierwater, B. Global Diversity of the Placozoa. *PLoS One* **8**, e57131 (2013).
96. Laumer, C. E. *et al.* Support for a clade of Placozoa and Cnidaria in genes with minimal compositional bias. *Elife* **7**, e36278 (2018).
97. Schierwater, B. *et al.* The enigmatic Placozoa part 1: Exploring evolutionary controversies and poor ecological knowledge. *BioEssays*, 2100080 (2021).
98. Osigus, H.-J., Rolfes, S., Herzog, R., Kamm, K. & Schierwater, B. *Polyplacotoma mediterranea* is a new ramified placozoan species. *Curr. Biol.* **29**, R148–R149 (2019).
99. Schulze, F. E. *Trichoplax adhaerens*, nov. gen., nov. spec. *Zool. Anz.* **6**, 92–97 (1883).
100. Grell, K. G. & Benwitz, G. Ergänzende Untersuchungen zur Ultrastruktur von *Trichoplax adhaerens* F.E. Schulze (Placozoa). *Zoomorphology* **98**, 47–67 (1981).

101. Grell, K. G. & Benwitz, G. Die ultrastruktur von *Trichoplax adhaerens* FE Schulze. *Cytobiologie* **4**, 216–240 (1971).
102. Pearse, V. B. & Voigt, O. Field biology of placozoans (*Trichoplax*): distribution, diversity, biotic interactions. *Integr. Comp. Biol.* **47**, 677–692 (2007).
103. Smith, C. L., Pivovarova, N. & Reese, T. S. Coordinated Feeding Behavior in *Trichoplax*, an Animal without Synapses. *PLoS One* **10**, e0136098 (2015).
104. Sebé-Pedrós, A. *et al.* Early metazoan cell type diversity and the evolution of multicellular gene regulation. *Nat. Ecol. Evol.* **2**, 1176–1188 (2018).

List of publications

Manuscripts included in this thesis:

1. **Genome reduction proceeds differently and independently of phylogenetic relations across closely related symbiont lineages**
   Anna Mankowski, Brandon KB Seah, Rebecca Ansorge, Manuel Kleiner, Cecilia Wentrup, Tanja Woyke, Nicole Dubilier and Harald R. Gruber-Vodicka
   Manuscript in preparation

2. **Forever young symbiont communities in an ancient and obligate symbiosis**
   Anna Mankowski, Manuel Kleiner, Christer Erséus, Nikolaus Leisch, Yui Sato, Jean-Marie Volland, Bruno Hüttel, Cecilia Wentrup, Tanja Woyke, Juliane Wippler, Nicole Dubilier, Harald R. Gruber-Vodicka
   Manuscript in preparation. An early version is available as preprint at bioRxiv under https://doi.org/10.1101/2021.04.28.441735

3. **Simple animals, simple microbiomes? The symbiont diversity of placozoans**
   Anna Mankowski, Nikolaus Leisch, Michael G. Hadfield, Bruno Hüttel, Nicole Dubilier, Harald R. Gruber-Vodicka
   Manuscript in preparation

Collaborations not included in this thesis:

4. **Microorganisms Associate with *Spartina anglica,* an Invasive Dimethylsulfoniopropionate Producing Salt Marsh Plant, are an Unrecognized Sink for Dimethylsulfide**
   Eileen Kröber, **Anna Mankowski,** Hendrik Schäfer
   Manuscript under revision at Microbiome

5. **Genomic and metabolic evolution of a bacterial animal symbiont**
   Marlene Jensen, **Anna Mankowski**, Harald R. Gruber-Vodicka, Manuel Kleiner
   Collaborative project

Chapter II | Genome reduction proceeds differently and independently of phylogenetic relations across closely related symbiont lineages

Anna Mankowski[1], Brandon K. B. Seah[1,2], Rebecca Ansorge[1,3], Manuel Kleiner[4], Cecilia Wentrup[5], Tanja Woyke[6], Nicole Dubilier[*1], Harald R. Gruber-Vodicka[*1]

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

[2]Max Planck Institute for Developmental Biology, Tübingen, Germany

[3]Quadram Institute Bioscience, Norwich, Norfolk, Great Britain

[4]North Carolina State University, Department of Plant and Microbial Biology, Raleigh, North Carolina, USA

[5]University of Vienna, Department of Microbiology and Ecosystem Science, Vienna, Austria

[6]Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, California, USA

[*]Corresponding authors:

Nicole Dubilier: ndubilie@mpi-bremen.de, phone +49 (421) 2028 9320

Harald Gruber-Vodicka: hgruber@mpi-bremen.de, phone +49 (421) 2028 7600

Author contributions: AM, HRGV and ND conceived the study, and AM designed the workflows with input from HRGV. HRGV, MK, CW, and TW acquired gutless oligochaete specimens and generated metagenomic data. BKBS processed and provided the ciliate datasets. RA performed the single nucleotide variant analyses. AM performed all other analyses steps, interpreted the results with input from HRGV and ND and wrote the manuscript with contributions from HRGV.

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

The author order is not fixed.

Abstract

Bacterial symbionts that are genetically isolated from other populations commonly undergo massive genome reduction. Genetic bottlenecks lead to the accumulation of deleterious mutations and subsequent genome reduction. In extreme cases, massive genome reduction can result in the decay of the association. Genome reduction has been intensively studied in insect symbionts that are strictly vertical transmitted and thus, commonly undergo massive genome reduction. In contrast, we have little understanding how genome reduction proceeds in symbionts that are only partly isolated from other bacterial populations. Here, we show that the type of the association, e.g. ecto- vs. endosymbiotic as well as the transmission mode strongly influence the degree of observed genome reduction. In addition, genome reduction appears to be a variable process that proceeds independent of phylogenetic relations across closely related symbiont lineages. These findings provide initial insights into the flexibility of genome reduction in different symbiont lineages and highlight the strong effect of genetic isolation on the degree of genome reduction.

Introduction

Genome reduction is the process of a decrease in genome size as well as the loss and simplification of genetic information. The most extreme examples of genome reduction have been observed in completely host-restricted bacterial symbionts that occur intracellularly and are vertically transmitted between hosts[1,2]. Due to small population sizes, genetic bottlenecks during transmission and genetic isolation from other bacterial populations, such symbiont populations are prone to the accumulation of deleterious mutations and thus, particularly affected by genome reduction[3,4].

The genome reduction of host-restricted bacterial symbionts proceeds in several steps. The initial phase is characterized by the accumulation of mutations that disrupt open reading frames (ORFs) and thus initiate pseudogenization. In addition, the proliferation of mobile genetic elements (MGEs) as well as genome rearrangements and deletion of chromosomal fragments are commonly observed processes during the early stage of genome reduction. As genome reduction proceeds, the abundance of pseudogenes and MGEs decreases, the genomes become smaller and the chromosome architecture stabilizes. With ongoing host restriction, the loss of genes and the decrease of the genome size will continue. In addition to these described processes that occur during genome reduction, reduced symbiont genomes also often exhibit a low GC content In case of strong co-adaptation, interdependence as well as metabolic integration between a host and its symbiont, the symbiont could even evolve further into organelle-like organisms[1,2].

The vast majority of our knowledge on genome reduction in bacterial symbionts is derived from research on insect symbionts that complement their hosts' diet with essential nutrients that are not present in the hosts' nutrition. For example, many insects feed on plant sap that lacks certain amino acids or vitamins which can be synthesized and provided by the insects' symbionts[5–16]. These symbionts are prime examples for extreme genome reduction as they are often intracellular, vertically transmitted symbionts that co-evolved with their host for millions of years and display high metabolic integration. Thus, many of these symbionts underwent massive genome reduction and retained genomes as small as 0.14 Mpb[5,17]. In contrast, bacterial symbionts that occur extracellularly such as the *Verminephrobacter* symbionts of earthworms or symbionts that are at least occasionally transmitted horizontally between hosts such as the chemosynthetic endosymbionts of marine bivalves are in genetic flux with other symbiotic or free-living bacterial populations. The genomes of these symbionts also show indications of ongoing genome reduction, however, the genetic flux between the symbiont populations with

either symbiotic or free-living bacterial populations appear to prevent massive genome reduction[18–20].

Despite growing evidence that both, symbiont localization and transmission can largely impact the degree of genome reduction, we lack exhaustive understanding how these factors affect genome reduction in closely related symbiont lineages. Such analyses are limited by the availability of closely related symbionts that occur in different types of associations. Here, we analyzed genome reduction in such a symbiont clade that occurs in two types of symbiosis: as an ectosymbiont of *Kentrophoros* ciliates and as an endosymbiont of the gutless oligochaete species *Inanidrilus exumae*[21–24]. *I. exumae* lacks the primary gutless oligochaete symbiont *Cand.* Thiosymbion. Instead, all individuals of *I. exumae* were found to be associated with a morphologically similar, yet phylogenetically unrelated lineage of gammaproteobacterial symbionts, called the Gamma4 symbionts[24,25]. Phylogenetic analysis of the closest relatives of the Gamma4 symbionts indicated that the Gamma4 symbionts belong to a clade of symbionts from the *Kentrophoros* ciliates, *Cand.* Kentron[25]. We confirmed that the Gamma4 symbionts phylogenetically belong to the *Cand.* Kentron clade and showed that *I. exumae* acquired it from *Kentrophoros* ciliates. The host switching of a symbiont from *Kentrophoros* ciliates to a gutless oligochaete also resulted in a lifestyle change of the symbionts as they occur as ectosymbionts on the outside of their ciliate hosts and as endosymbionts in the gutless oligochaete hosts, giving us the opportunity to compare the progress of genome reduction in closely related ecto- and endosymbionts.

We analyzed genome sizes and GC content as well as the prevalence of pseudogenization and MGEs across the clade of Gamma4 and *Cand.* Kentron symbionts and compared the progress of genome reduction not only between the ecto- and the endosymbionts but also between several host-species specific lineages of the ciliate ectosymbionts. Our results show that

endosymbionts are indeed more affected by genome reduction than their ectosymbiotic relative. However, we found indications that genome reduction is generally a highly variable process that also proceeds differently and independent of symbiont phylogeny in closely related, yet distinct symbiont lineages.

Results and Discussion

Inanidrilus exumae acquired the Gamma4 symbiont from *Kentrophoros* ciliates

In order to confirm that the Gamma4 symbionts of *I. exumae* form a subclade within the *Cand.* Kentron clade of symbionts from *Kentrophoros* ciliates and in order to compare the genomes of symbionts associated with different host phyla, we obtained metagenome-assembled genomes (MAGs) of three Gamma4 symbionts from *I. exumae* and 36 *Cand*. Kentron symbionts from different species of *Kentrophoros* ciliates (Table S1). Both, phylogenetic analyses of the symbionts' 16S rRNA genes and phylogenomic analyses of clade-specific single-copy orthologues showed that the Gamma4 symbionts of *I. exumae* form a subclade within the *Cand.* Kentron clade (Figure 1, Figure S1). In order to clearly state whether we are referring to the gutless oligochaete or the ciliate symbionts, we will keep calling the symbionts of *I. exumae* the Gamma4 symbionts although they phylogenetically belong to the *Cand.* Kentron clade. The *Cand.* Kentron symbionts formed eleven distinct subclades (Figure 1). For ten of the eleven subclades, all symbionts from a given subclade were always associated with the same host species (Table S1). The symbionts from subclade 11 belonged to two different host species, but as the symbionts from both host species were phylogenetically intermixed, we treated them as one symbiont subclade. To improve the granularity of our analyses, we will not only distinguish between the Gamma4 and the *Cand.* Kentron symbionts in subsequent analyses but also treat the different subclades of *Cand.* Kentron symbionts separately.

The placement of the Gamma4 symbionts within and not as a sister clade to the *Cand.* Kentron clade indicates that *I. exumae* acquired the Gamma4 symbiont through a host switching event of the *Cand.* Kentron symbiont from *Kentrophoros* ciliates. Besides switching host phyla, this switching event also led to a lifestyle change of the symbionts: the *Cand.* Kentron symbionts are ectosymbionts that are attached to the outside of their hosts[21–23]. In contrast, the Gamma4 symbionts, as all other gutless oligochaete symbionts, occur as endosymbionts between the epidermal cells and the cuticle of the worms[24]. Such lifestyle changes from ecto- to endosymbiotic have been rarely observed but offer the unique opportunity to understand how different symbiotic lifestyles could affect symbiont genome evolution.

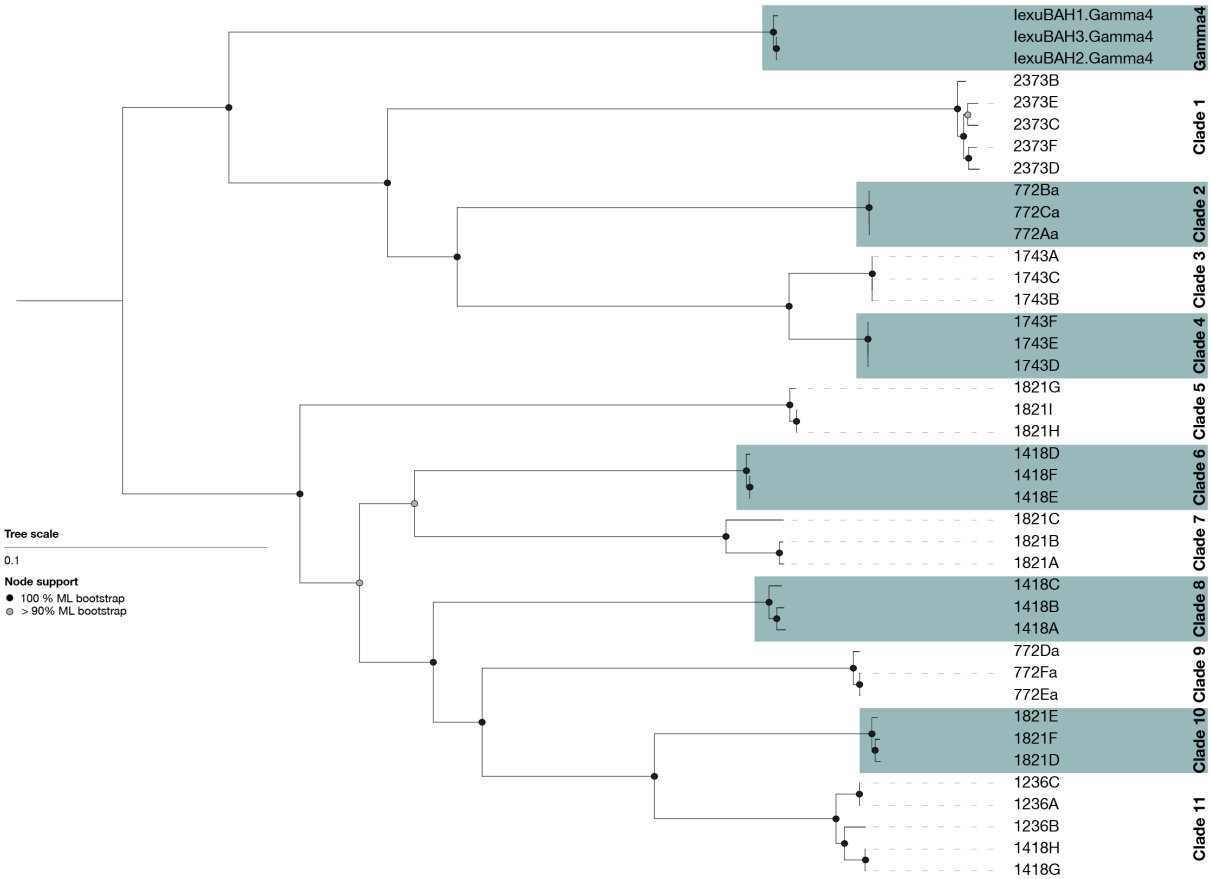


Figure 1 | The Gamma4 symbionts from *I. exumae* are a subclade of the *Cand.* Kentron symbionts. Maximum-likelihood tree that shows the phylogenomic relations between several *Kentrophoros* and *I. exumae* symbionts based on single-copy orthologues. Nodes that are statistically supported by none-parametric bootstrapping are highlighted within the tree. Subclades that are separately discussed throughout the paper are highlighted.

Subclade-specific genome reduction in the Gamma4 and *Cand.* Kentron symbionts

The generated MAGs of Gamma4 and *Cand.* Kentron symbionts had sizes of 3.42-5.02 Mbp and were classified as medium to high quality MAGs according to genome completeness and contamination estimates (81.42-90.95%, Table S1)[26]. In contrast to the reduced completeness, we detected complete rRNA operons in all MAGs and nearly complete sets of tRNAs, carrying 18-20 different amino acids (Table S1). Based on the high completeness of rRNA operons and tRNA sets, we interpret the rather low genome completeness estimates as a result of genome reduction instead of as a sign of reduced MAG quality. In addition to variable genome size and decreased genome completeness, we also observed other signatures of ongoing genome reduction in the Gamma4 and *Cand.* Kentron symbionts, including variations in the GC content, as well as presence of potential pseudogenes and mobile genetic elements[2]. In the following, we analyzed genome size, GC content, pseudogenes and mobile genetic elements across the subclades of Gamma4 and *Cand.* Kentron symbionts. Our data indicated that genome reduction proceeds similarly within host species specific symbiont subclades but differently and independent of symbiont phylogeny between those subclades.

Genome size The genome sizes of the Gamma4 and the *Cand.* Kentron symbionts were relatively similar within symbiont subclades but highly variable between subclades (Figure 2, Table S1). The variability of genome sizes between symbiont subclades was not connected to the symbiont phylogeny. For example, despite being most closely related to the *Cand*. Kentron symbionts with the biggest genomes (subclade 2, 4 and 5), the Gamma4 symbiont genomes were among the smallest of the whole clade (Figure 1). Only the genomes from symbionts of the unrelated *Cand.* Kentron subclade 10 were even smaller. Thus, we assume that the subclade specific variability of genome sizes is the result of lineage specific genome reduction albeit the genomes were still comparatively large among the reported range of genome sizes from thiotrophic symbionts[27–34].

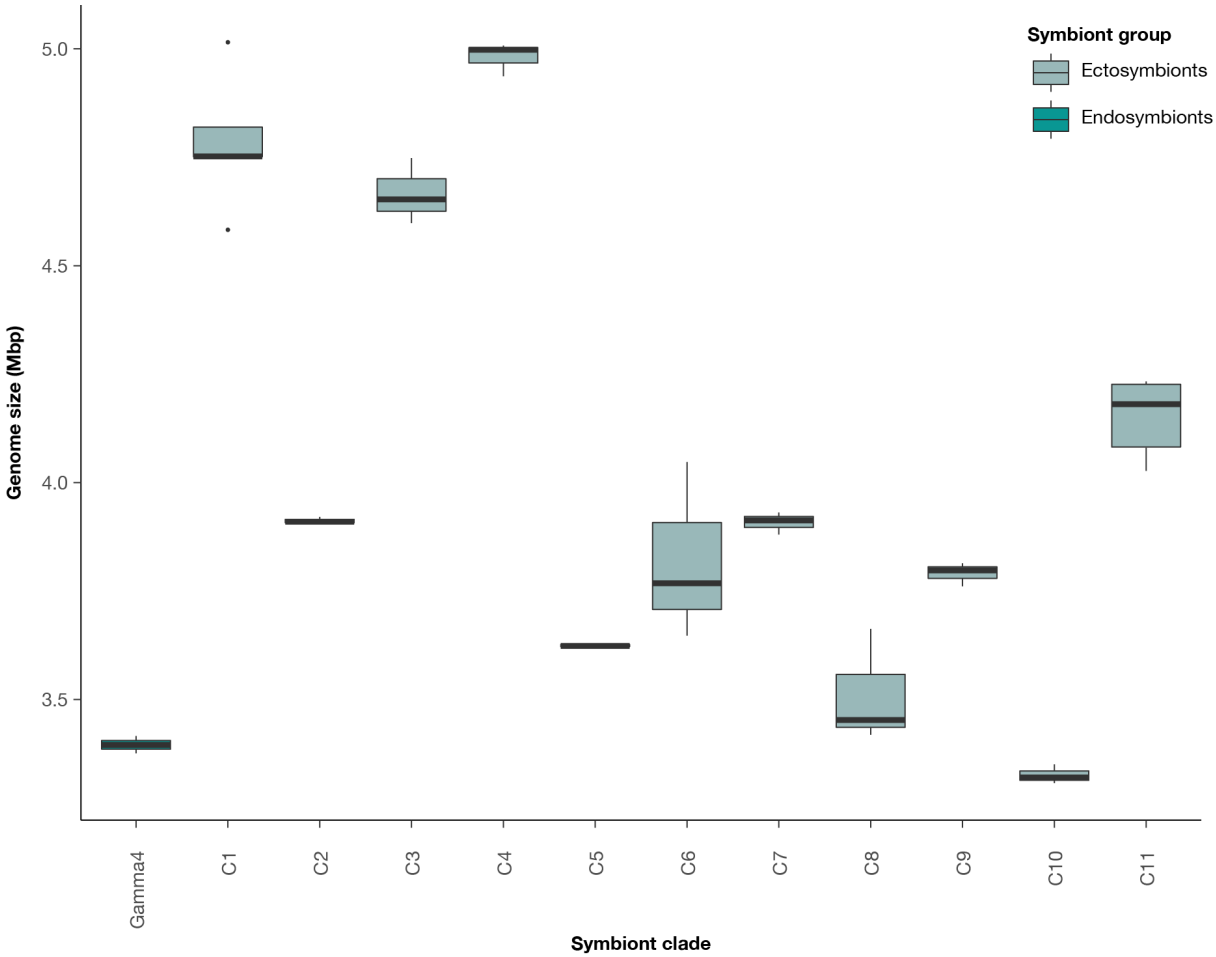


Figure 2 | The genome sizes of ectosymbiotic *Cand.* Kentron symbionts are largely clade specific whilst the endosymbiotic Gamma4 symbionts are exceptionally small. Boxplots show the distribution of genome size within each of the analyzed symbiont subclades. Different colors highlight ecto- and endosymbionts.

GC content As a result of either selection, loss of DNA repair genes or mutational bias in small, asexually reproducing populations, genome reduction in symbiotic bacteria is often linked to a decreased GC content[3,35–38]. To reveal ongoing dynamics of the GC content in the Gamma4 and *Cand.* Kentron symbionts, we analyzed the GC content of the whole genomes, the coding region, the non-coding region as well as the GC content of the third codon positions. As mutations in non-coding regions or at the third codon position are usually synonymous, we would expect that the effects of mutational bias or selection would be more pronounced at these positions of the genomes. Overall, we observed that the GC content of non-coding regions was lower than the GC content of coding regions which is commonly observed in bacterial genomes

and likely the result of relaxed selective pressure (Figure 3)[39]. In contrast, the GC content at the third codon position was higher than the overall GC content of coding sequences, indicating selection for high GC codons and in consequence, selection for high genomic GC content (Figure 3)[40]. Symbiont genomes with lower overall GC contents also exhibited a reduced GC content at the non-coding regions as well as the third codon position, indicating that GC content dynamics that are likely connected to a symbiotic lifestyle equally effect all regions of the symbiont genomes.

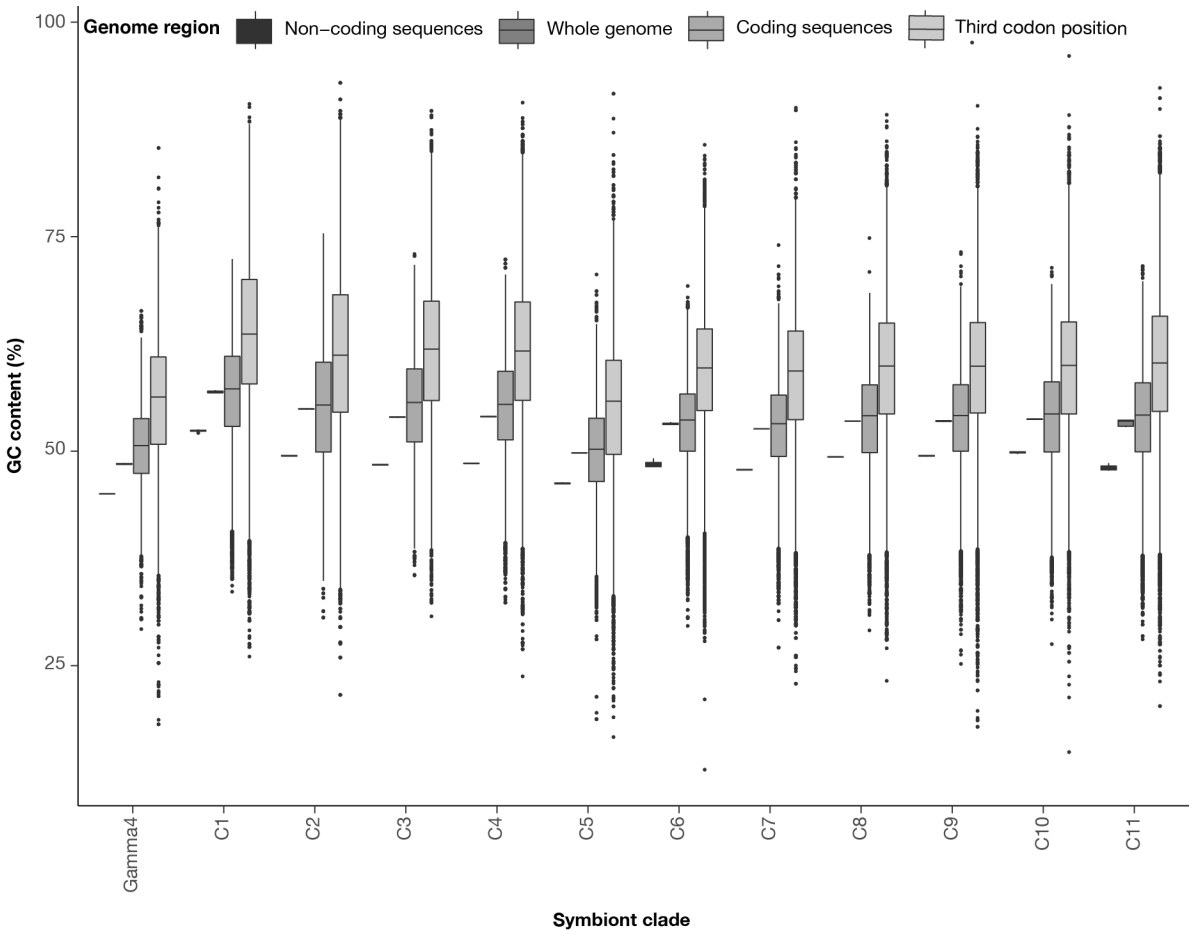


Figure 3 | GC content is variable, reduction affects all genome regions independent of selective processes. Boxplots show the distribution of GC content in different genome regions within each of the analyzed symbiont subclades. Different colors highlight the different genome regions.

Similar to the genome size, also the GC content was relatively stable within genomes from symbionts that belonged to the same phylogenetic subclade but the GC content patterns were

not connected to phylogenetic relations between the symbiont subclades. The genomes of the Gamma4 symbionts were not only among the smallest but also had the lowest GC content despite being most closely related to the symbiont subclades that had the highest GC content (subclade 2-5, Figure 3). Despite a general trend of smaller genomes to display lower GC content, we did not observe a strict correlation between genome size and GC content across the clade of the Gamma4 and *Cand.* Kentron symbionts (Figure 4). However, the endosymbiotic Gamma4 symbionts showed a pronounced reduction in both, genome size and GC content. These results illustrate that genome reduction and decrease in GC content are indeed co-occurring processes in the genome evolution of the endosymbiotic Gamma4 and the ectosymbiotic *Cand.* Kentron symbionts. However overall, the two processes appeared to progress differently in individual symbiont subclades and largely independent from each other.

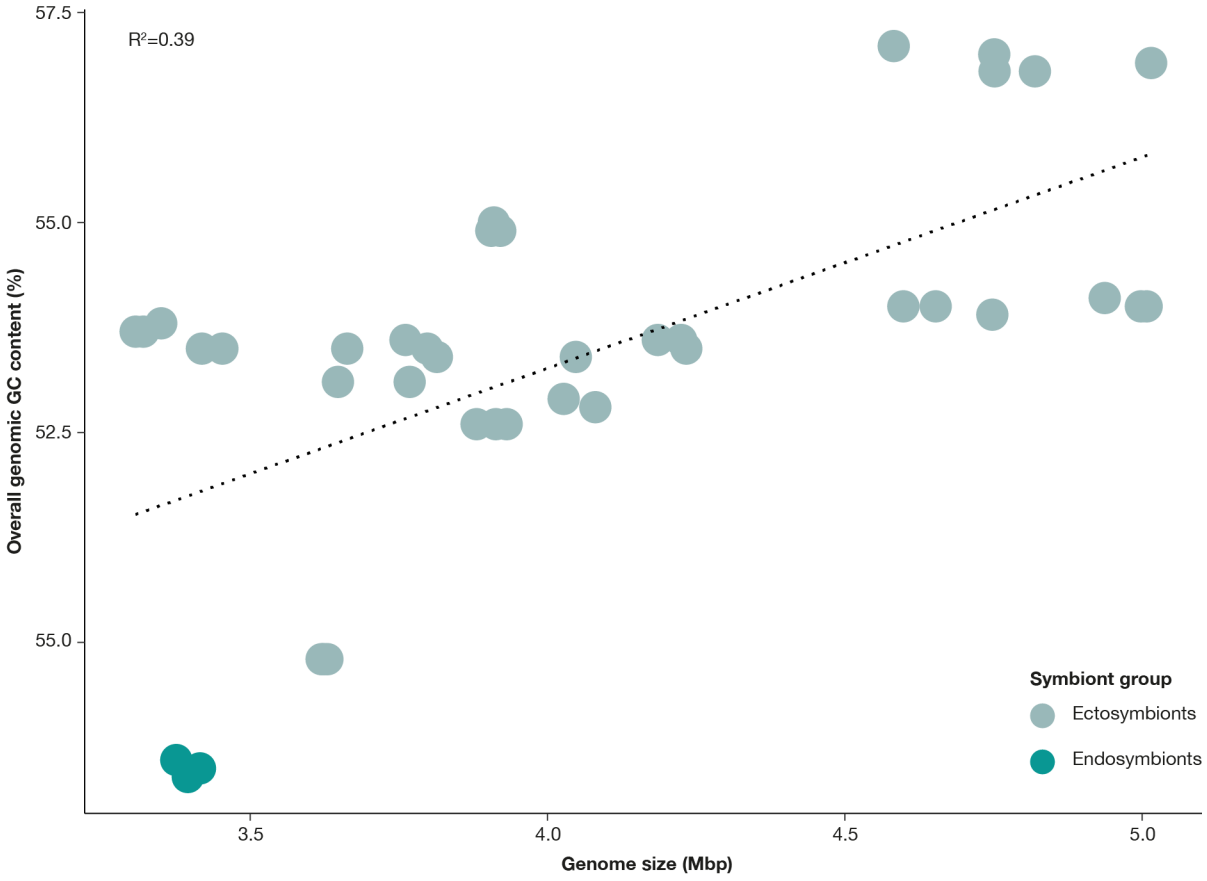


Figure 4 | Smaller genomes tend to display a lower GC content, particularly in the endosymbiotic Gamma4 symbionts. The overall genomic GC content is shown in relation to the genome size. Different colors highlight ecto- and endosymbionts.

Pseudogenization Mutations and genetic drift in small, asexual symbiont populations cannot only alter the GC content of symbiont genomes but the accumulation of non-sense mutations in open reading frames (ORFs) also leads to loss of function, pseudogenization and deletion of affected genes[1,4,41,42]. We used gene length distribution patterns as well as the percentage of predicted genes without functional annotations as proxies to detect pseudogenization. We argue that short genes without functional annotation are likely the result of ORF shifts by non-sense mutations and thus, pseudogenization. Based on these assumptions, we detected several indications for different stages of pseudogenization in the different subclades of the Gamma4 and *Cand.* Kentron symbionts. First, we observed an accumulation of relatively short genes (50-150 bp) in the genomes of all symbionts (Figure 5A). Second, we found that the percentage of predicted genes that were annotated as hypothetical proteins was negatively correlated with the genome size, i.e. smaller genomes contained a higher number of genes with functional annotation (Figure 5C). Thus, we hypothesized that in larger genomes, more pseudogenes were retained and still recognized as predicted genes but could not be annotated. With ongoing removal of pseudogenes, both genome size as well as the number of remnants of pseudogenes that were annotated as hypothetical proteins would decrease.

Strikingly, the gene length distribution as well as number of predicted and functionally annotated genes of the Gamma4 symbionts indicated that only few pseudogenes are retained in this subclade of symbionts. We observed that genes with lengths between 50-150 bp were less accumulated in the Gamma4 symbionts compared to other *Cand.* Kentron subclades. In general, surprisingly little genes were predicted in the genomes of the Gamma4 symbionts. For all other *Cand.* Kentron genomes, the genome size strongly correlated with the number of predicted genes. According to this correlation, we would have expected ~3000 genes to be predicted from the genomes of the Gamma4 symbionts but only ~2000 genes were actually predicted (Figure 5B). In contrast, the percentage of genes that got a functional annotation was much higher than

expected from the correlation between genome size and percentage of annotated genes for the other *Cand.* Kentron subclades (Figure 5C). Thus, we assume that the genomes of the Gamma4 symbionts did not only undergo pseudogenization but that former pseudogenes are also already much more degraded and not detected as genes or gene remnants anymore.

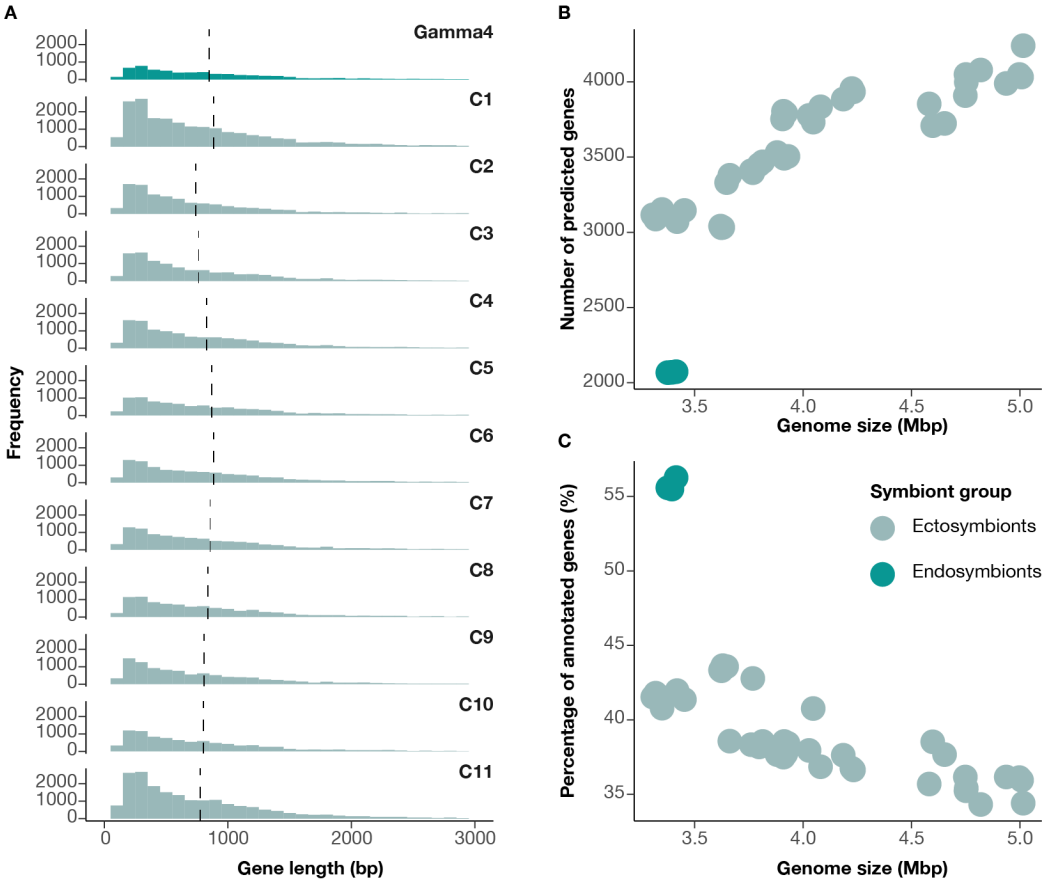


Figure 5 | The accumulation of short hypothetical genes illustrates ongoing pseudogenization in all symbiont subclades, and gene loss has progressed furthest in the endosymbiotic Gamma4 symbionts. (A) Gene length distribution in the different symbiont subclades. **(B)** The number of predicted genes in relation to the genome size. **(C)** The percentage of genes with functional annotation in relation to the genome size. Different colors highlight ecto- and endosymbionts.

Given our previous observations that many processes of genome reduction similarly effect the genomes of symbionts from the same subclade, we analyzed whether also pseudogenization could be a subclade-specific process. Therefore, we compiled a set of all subclade-specific orthologues and analyzed their length and their annotation status (Figure 6). The subclade-specific genes were indeed more often annotated as hypothetical proteins than the other genes

(Figure 6A). Also, the vast majority of subclade-specific genes was shorter than the genes that were not specific to a given symbiont subclade (Figure 6B). Thus, we assume that also pseudogenization is, at least partly, a lineage-specific process of genome reduction.

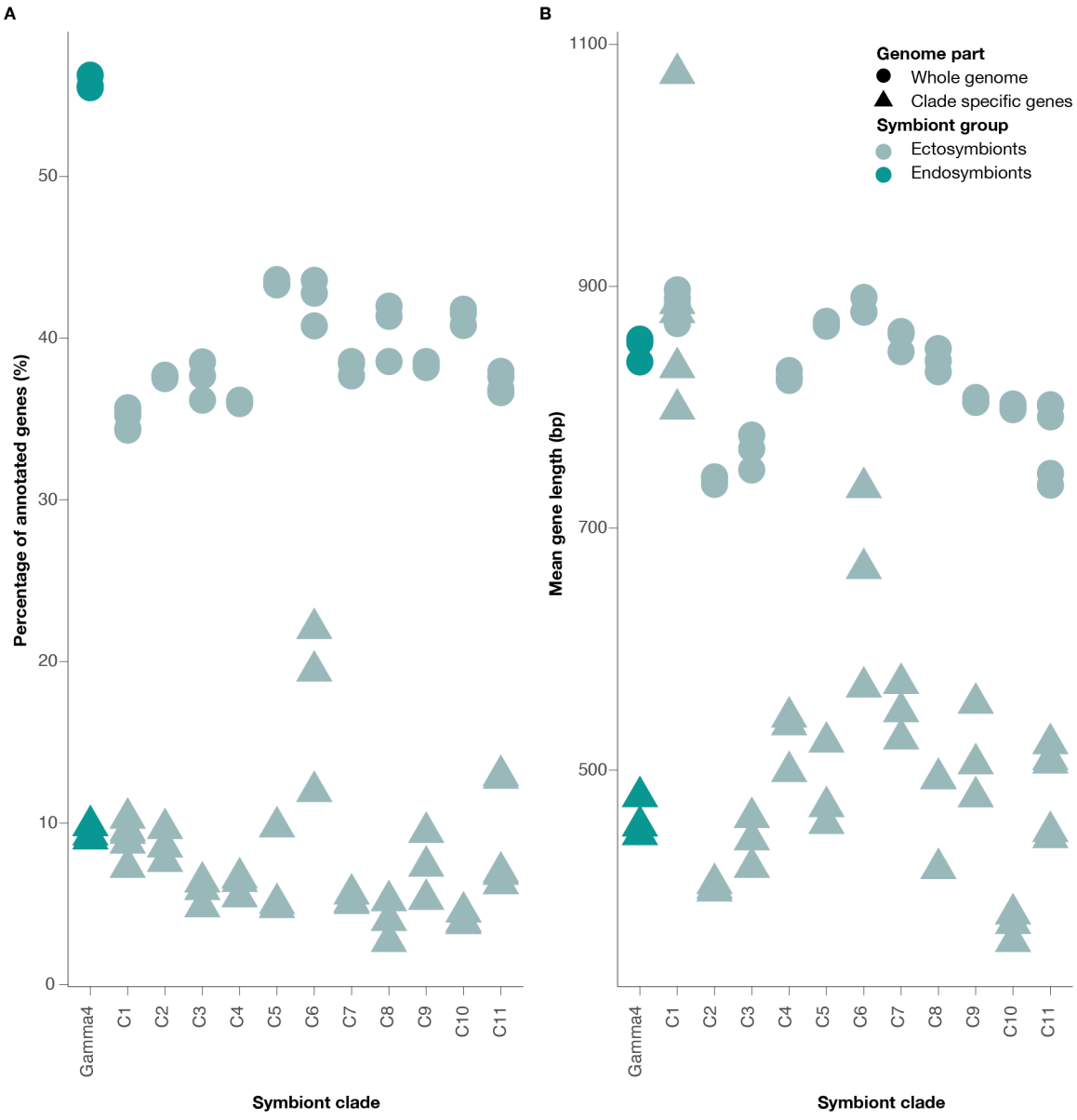


Figure 6 | Subclade specific orthologues are largely hypothetical genes and tend to be shorter than the total gene set in the respective genome. (A) Percentage of genes in the complete genomes and the subclade-specific orthologues of each subclade. **(B)** Length of genes in the complete genomes and the subclade-specific orthologues of each subclade. Different colors highlight ecto- and endosymbionts. Shapes highlight the genome part that was analyzed.

Presence of mobile genetic elements The initial phase of genome reduction in endosymbionts is often linked to the proliferation of mobile genetic elements (MGEs). With proceeding

genome reduction, the MGEs get degraded and are commonly less abundant or even completely absent in highly reduced genomes[1,2]. We analyzed the prevalence of MGEs in relation to the genome size and the phylogenetic relations of the Gamma4 and *Cand.* Kentron symbionts. Similar to the other indications of genome reduction that we discussed above, also the prevalence of MGEs appeared to be highly similar within symbionts from the same subclade but the differences between subclades were not linked to the phylogenetic relationships (Figure 7).

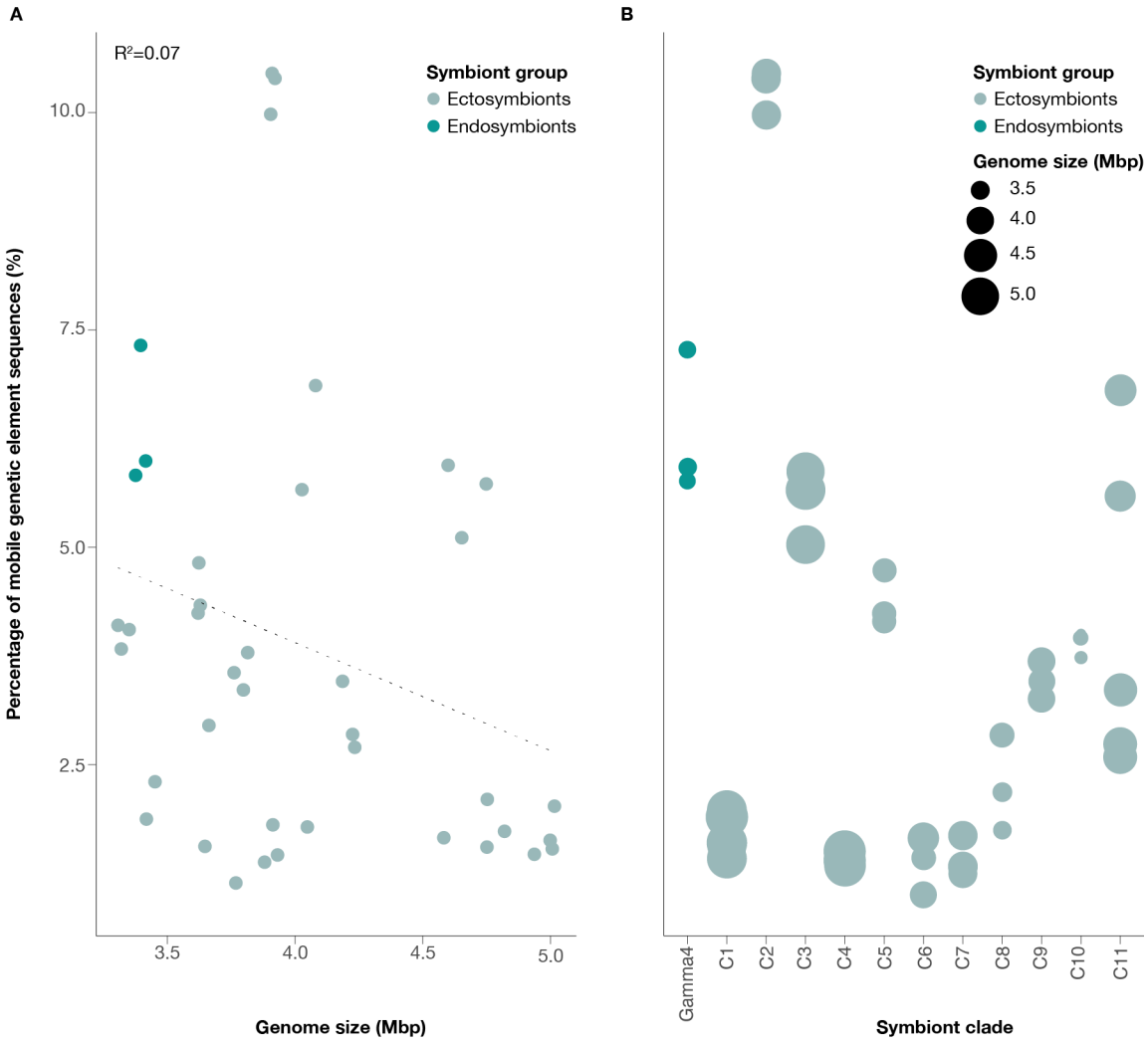


Figure 7 | The prevalence of mobile genetic elements is not correlated with genome size but stable within symbiont subclades. (A) Percentage of mobile genetic elements in each genome in relation to the respective genome size. **(B)** Percentage of mobile gene elements in each genome in relation to the symbiont subclade. Different colors highlight ecto- and endosymbionts. Circle size indicate genome size **(B)**.

In addition, we did not observe a correlation between symbiont genome size and the prevalence of MGEs. We thus assume that genome reduction and degradation of MGEs are not strictly correlated but co-occurring yet partly independent processes (Figure 7A). Extreme cases of genome reduction in e.g. highly specialized insect symbionts, appear to result in both, heavily decreased genome sizes and the absence of MGEs[2]. However, in the process of ongoing genome reduction, the decrease of MGEs appears to be unlinked to the decrease in genome size.

The lifestyle switch from ecto- to endosymbiotic likely accelerated the genome reduction in *Cand.* Kentron symbionts

Comparing twelve symbiont subclades from the same genus that are associated with different host species allowed us to compare the progress of ongoing genome reduction in distinct, yet closely related, phylogenetic lineages. We showed that genome reduction proceeded relatively similar within symbiont subclades but independently of symbiont phylogeny between subclades. Also, we observed that the different processes that are linked to genome reduction in symbionts occurred independent of each other.

The host switch of the *Cand.* Kentron symbionts from a ciliate to a gutless oligochaete host and thus a switch from an ecto- to an endosymbiotic lifestyle represents a rare natural experiment that allowed us to investigate whether the type of the symbiotic association could influence genome reduction. Based on the previously discussed findings on the genome properties of the Gamma4 and the *Cand.* Kentron symbionts, we assume that genome reduction in fact proceeds faster in endosymbionts than in ectosymbionts. The Gamma4 symbionts showed indeed only the second smallest genome sizes but symbionts of this subclade also had the lowest GC content and appeared to be further ahead in the progress of pseudogenization. This combination of comparably small genomes, low GC content and low number of pseudogenes and their

remnants was unique to the Gamma4 symbiont clade. Thus, we assume that genome reduction is accelerated in endosymbionts compared to their ectosymbiotic relatives.

Massive genome reduction is commonly linked to the genetic drift due to bottleneck effects and genetic isolation of symbiont populations, especially in strictly vertically transmitted symbionts that occur within host cells[2,3]. In contrast, symbionts that occur extracellularly and/or are not strictly vertically transmitted, e.g. the nephridial symbionts of earthworms or the chemosynthetic symbionts of marine bivalves are genetically not completely isolated and the genetic flux between these symbionts and other bacterial populations prevents massive genome reduction[18–20]. Similarly, the genomes of the Gamma4 and *Cand.* Kentron symbionts were comparatively large and the observed GC content as well as pseudogene and MGE patterns were not as drastically altered as in many tiny symbiont genomes[1,2]. Thus, we hypothesize that also the populations of the Gamma4 and *Cand.* Kentron symbionts are genetically not completely isolated and in consequence, massive genome reduction is prevented. The *Cand.* Kentron symbionts are attached to the outside of their hosts, therefore, genetic mixing between the symbionts and the free-living bacterial population is likely feasible and could prevent massive genome reduction although the symbionts are mostly vertically transmitted[21–23,43]. In contrast, the Gamma4 symbionts occur below the cuticle of their hosts[24]. This endosymbiotic association indeed also represents and extracellular symbiosis, yet, the genetic flux between the endosymbiotic Gamma4 symbionts and the free-living bacterial populations is likely reduced compared to the ectosymbiotic *Cand.* Kentron symbionts. However, it was recently shown that the symbionts of gutless oligochaetes are not strictly vertically transmitted[25]. Occasional horizontal transmission events could enable the genetic exchange between different symbiont populations and thus, prevent massive genome reduction as it was also shown for the symbionts of chemosynthetic bivalves[19]. In order to confirm our hypothesis that the populations of the Gamma4 and *Cand.* Kentron symbionts are genetically not completely isolated, we analyzed

the prevalence of genetic variability, i.e. the frequency of single nucleotide variants (SNVs) within the symbiont populations of single host individuals. We found that SNVs are indeed present in all symbiont populations, indicating that the symbiont populations of each host individual are not clonal and thus, genetically not completely isolated (Figure 8). Based on these findings, we assume that the genetic flux between different symbiont populations or between symbionts and free-living bacteria prevent massive genome reduction in the Gamma4 and *Cand.* Kentron symbionts. This effect appears to be stronger for ectosymbiotic bacteria that are in genetic flux with diverse, free-living bacterial populations compared to endosymbionts that only exchange genetic material with other symbiont populations.

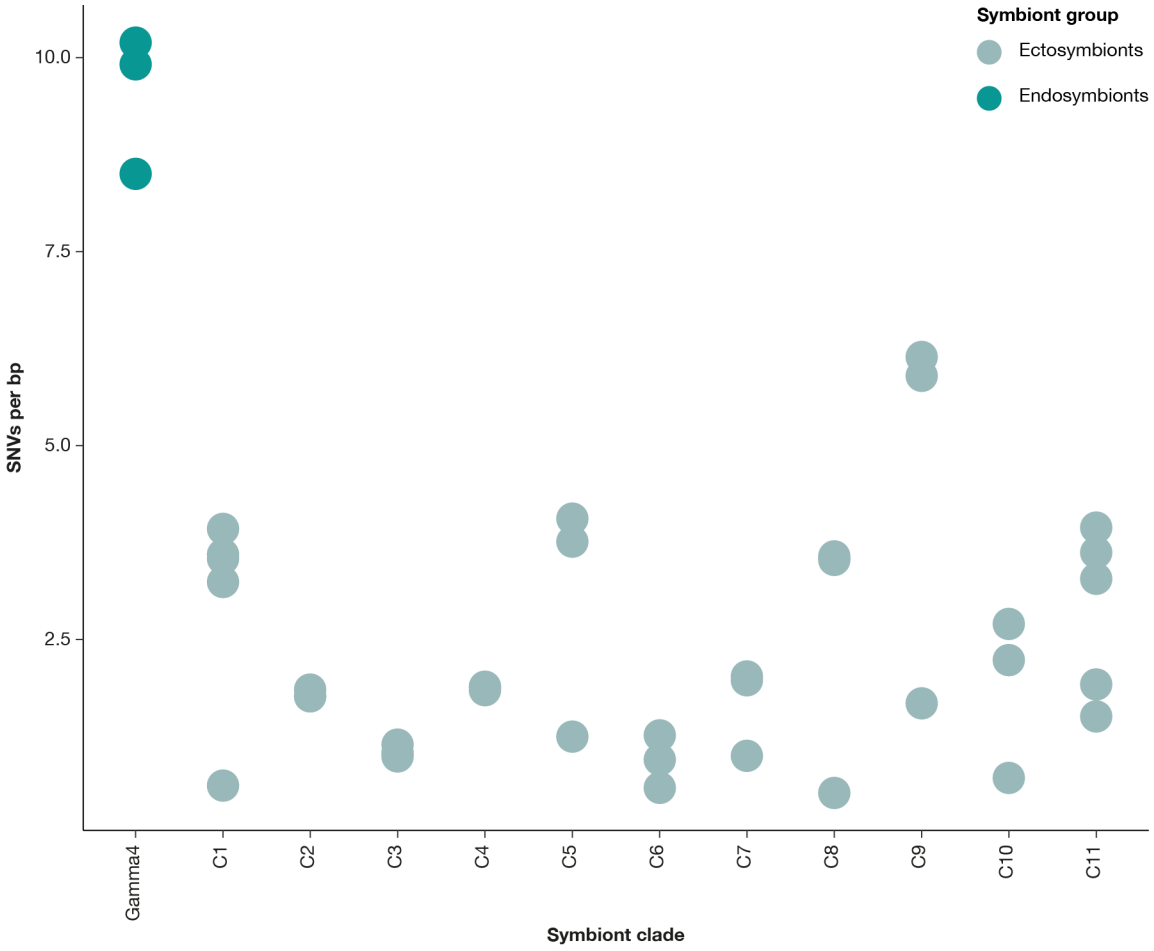


Figure 8 | The prevalence of single nucleotide variants indicates non-clonal symbiont populations with varying degrees on intraspecific diversity. Number of SNVs per kbp in each of the symbiont subclade. Different colors highlight ecto- and endosymbionts.

Genome reduction did not affect the central metabolism of the symbionts

In a next step, we asked how the genome reduction in the Gamma4 and *Cand.* Kentron symbionts effects the metabolism of the symbionts. Therefore, we analyzed the presence of metabolic pathways that are involved in the central carbon, nitrogen and sulfur metabolism as well as the amino acid and energy metabolism of the symbionts (Figure 9, Extended Data S1). In concordance with previous description of the *Cand.* Kentron metabolism, we observed that all symbionts were capable of importing organic carbon sources, heterotrophically assimilating carbon and producing energy by sulfur oxidation as well as aerobic respiration (Figure 9, Extended Data S1)[44]. We also found indications that the Gamma4 and *Cand.* Kentron symbionts could respire nitrate and oxidize DMSO but further evidence is required to confirm that these pathways are in fact functional and expressed in the Gamma4 and *Cand.* Kentron symbionts. In addition, we predicted diverse pathways involved in the biosynthesis and degradation of all amino acids (Figure 9, Extended Data S1).

The majority of the predicted pathways were annotated in all symbiont genomes (Extended Data S1). However, some pathways were at least partially lost in certain subclades of the Gamma4/*Cand.* Kentron symbionts. We hypothesize that the (partial) loss of certain pathways in distinct symbiont subclades is linked to the subclade-specific genome reduction. Interestingly, there were no pathways that were retained in all *Cand.* Kentron symbionts but reduced in all three endosymbiotic Gamma4 symbionts.

In addition, it appears that the central carbon and energy metabolism is hardly affected by the subclade-specific genome reduction and that all symbionts retained the genetic repertoire for a variety of metabolic pathways. Such a broad genetic repertoire is commonly observed in symbionts that represent the sole source of nutrition for their hosts independent of their location in the host and their transmission mode. This includes not only the symbionts of gutless

oligochaetes and *Kentrophoros* ciliates but also the symbionts of other gutless marine invertebrates, e.g. gutless oligochaetes, *Riftia* tubeworms and *Paracatenula* flatworms[34,45–49]. We argue that these symbionts need to retain such a broad genetic repertoire in order to efficiently provide their hosts' nutrition and thus maintain the symbiotic association. In contrast, many insect symbionts with highly reduced genomes fulfill a very specific function, e.g. supplementing their hosts' diet with essential amino acids or vitamins[5–16]. In addition, these symbionts often co-evolved with their hosts over millions of years, which often lead to metabolic integration between the partners[6,50–53]. Thus, these symbionts need to retain a much smaller set of genes that are essential for their own survival and the maintenance of the symbioses. Based on the observed differences in the genome evolution of symbionts that provide few essential nutrients compared to symbionts that provide the entire nutrition for their hosts, we hypothesize that the degree of genome reduction is not only influenced by the location of the symbionts within their hosts or the transmission mode but also by the role of the bacterial symbiont in the association.

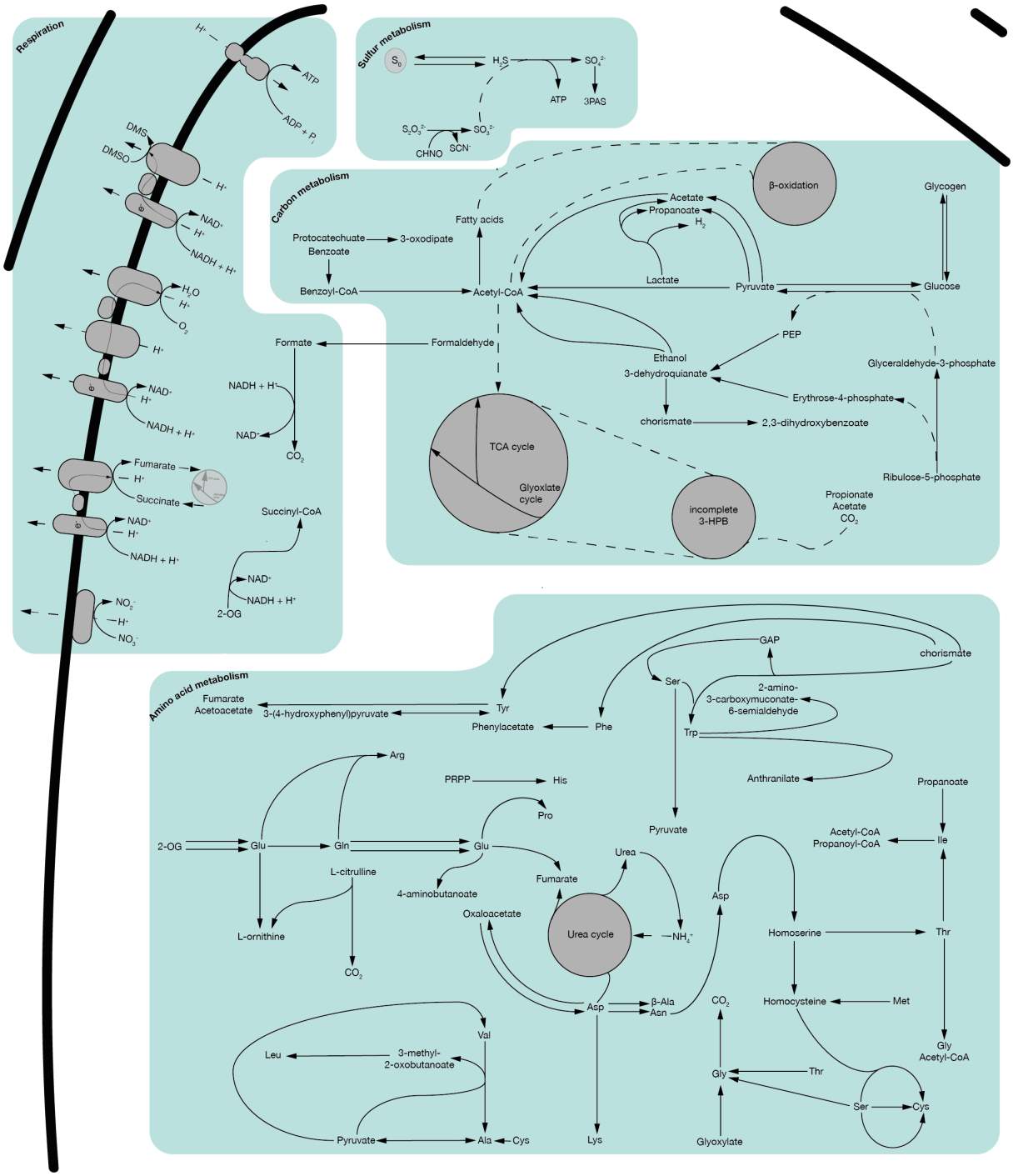


Figure 9 | The Gamma4 and *Cand.* Kentron symbionts retained a broad variety of pathways that are involved in the carbon, amino acid and energy metabolism of the symbionts. Schematic representation of pathways that were annotated and predicted in the symbiont MAGs. Solid arrows represent detected reactions, dashed lines represent likely exchange of metabolites between pathways.

A gutless oligochaete symbiosis without a primary chemoautotrophic symbiont?

Our analyses of the genetic repertoire of the Gamma4 symbionts contradict previous findings that characterized the Gamma4 symbionts as chemoautotrophs that could fix inorganic CO_2 via

the CBB cycle[24]. Therefore, we screened the MAGs of all *I. exumae* symbionts, including those of secondary symbionts, for the presence of the RubisCO gene to validate our pathway predictions (Table S2). Indeed, we did not detect a RubisCO gene in any of the Gamma4 symbiont MAGs. However, we detected genes for the carbon-fixing form of RubisCO in two different secondary symbionts. In one host individual, we detected a RubisCO gene in an alphaproteobacterial symbiont and in another host individual, we detected a RubisCO gene in gammaproteobacterial symbiont that was not related to the Gamma4 symbionts. We did not detect any genes for carbon-fixing RubisCO in any of the symbionts of the third host individual. We also searched all MAGs and the unbinned assemblies for the previously published RubisCO sequence that was assigned to the Gamma4 symbionts[24]. Our analyses indicate that the previously published RubsiCO sequence likely originated from the same alphaproteobacterial symbiont in which we detected a RubisCO gene. The phylogenetic placement of the previously published RubisCO sequence as a gammaproteobacterial gene thus indicates horizontal acquisition of the RubisCO gene from a Gammaproteobacterium by the alphaproteobacterial symbionts of *I. exumae.*

In addition, we screened the MAGs of the *I. exumae* symbionts for marker genes of other pathways that would allow them to autotrophically fix carbon. We found indications that one clade of deltaproteobacterial symbionts could potentially fix carbon via the 3-hydroxypropionate/4-hydroxybutyrate or the dicarboxylate/4-hydroxybutyrate cycle, as we detected a putative 4-hydroxybutanoyl-CoA dehydratase, the key enzyme of these pathways. However, we did not find any indications for autotrophic carbon fixation by the most abundant Gamma4 symbiont. In addition, we even found that in two of the three individuals of *I. exumae* that we analyzed, none of the symbionts was capable of autotrophic carbon fixation. Based on studies of the symbiotic association in another gutless oligochaete species, it was assumed that the autotrophic carbon fixation by thiotrophic symbionts is the main source of organic carbon

that fuels the gutless oligochaete symbiosis[48,49]. However, our results indicate also heterotrophic assimilation of carbon could solely nourish the gutless oligochaete symbiosis.

Conclusion

Despite our extensive knowledge of the processes that are involved in symbiont genome reduction, we still lack comprehensive understanding how these processes occur in different symbiont lineages; simultaneously, in a fixed order or independent from each other. Our comparative analyses of genome reduction in several subclades of one symbiont lineage indicate that genome reduction is a highly variably process that progresses differently and independent from symbiont phylogeny. Whilst pseudogenization and in consequence gene loss are strongly correlated with genome size, GC content dynamics and the prevalence of MGEs are not linked to the decrease in genome size. In addition, our comparison of genome reduction but also the metabolic capabilities in ecto- and endosymbiotic bacteria highlights the importance of the role of the symbiont as well as symbiont localization and transmission and thus, the availability to exchange genetic material with other symbiotic or free-living populations for the degree of genome reduction.

Material and Methods

Acquisition and processing of symbiont genomes

The genomes of genomes from Kentrophoros ciliates were previously published and downloaded from JGI GOLD (study Gs0114545). The genomes of symbionts from *I. exumae* were assembled and binned from single-worm metagenomes. Three specimen of *I. exumae* were sampled on the Bahamas in April 2013 and flash-frozen in liquid nitrogen. DNA was extracted from single worm individuals with the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Construction of sequencing libraries

was performed with the Nextera Low Fragment Input Kit (Illumina Inc., San Diego, California, USA). Library construction, quality control and sequencing were performed at the DOE JGI (Walnut Creek, California, USA).

The metagenomic reads were trimmed and filtered using bbduk v38.90 (https://sourceforge.net/projects/bbmap/); Illumina adapters were removed from both ends of the reads and only reads that were longer than 51 bp and had a quality of 2 or higher were kept (minimum kmer size 11, hamming distance 1). The resulting reads assembled using metaSPAdes v3.12.0 (kmer sizes 33, 55, 77, 99 and 121) and subsequently binned using Autometa v1.0, MaxBin v2.2.6, MetaBAT v2.11.1 and phyloFlash v3.4 discarding contigs that were shorter than 1500 bp[54–58]. Generated MAGs were refined using DAS_Tool v1.1.2 [59]. The identity of the symbiont for each MAG was determined by phylogenomic analyses of all generated bins as well as previously reference MAGs of gutless oligochaetes using the *de novo* workflow from GTDB-Tk v1.3.0[60]. We analysed MAG quality using the lineage-specific workflow from CheckM v1.0.7 and detection of tRNAs with ARAGORN v1.2.38[61,62]. The best bin for each symbiont from each host individual was selected by i) the highest genome completeness value, ii) highest number of amino acids for which a tRNA was encoded and iii) lowest percentage of contamination. The final MAGs were annotated using the RASTtk pipeline implemented in PATRIC[63,64]. We reconstructed metabolic pathways from the PATRIC annotations in Pathway Tools v24.5 using the Metage2Metabo v1.4.1 command line interface[65–67]. In addition, we annotated the genomes with prokka v1.14.5, providing a custom reference file of protein sequences from enzymes that were involved in pathways that are commonly present in thiotrophic symbioses but not always faithfully annotated using automated annotation pipelines (Extended Data S2)[68]. This include sulfate reduction via the Dsr complex, sulfur oxidation via the rDsr pathway, CO_2 fixation via the Calvin-Benson-Bassham cycle, the 3-hydroypropionate bicycle, PP_i-dependent ATP synthesis, trehalose and polyhydroxyalkanoate synthesis, nitrogen fixation. We only considered these pathways present

when the genes for key enzymes were annotated with one of our annotation pipelines. In addition, we used the TCDB database as annotation reference for more detailed annotation of transporter proteins and annotated MGEs using ISEscan v1.7.2.3[69,70]. We additionally screened all symbiont MAGs for the presence of the previously published RubisCO sequence from *I. exumae* by performing a blastn search against a database that contained the predicted RubisCO sequences that resulted from the previous prokka annotation as well as the gene sequence published by Bergin *et al.* (2018)[24]. In addition, we screened all symbiont MAGs for the presence of key genes for other autotrophic carbon fixation pathways by perfoming a blastn search against the database compiled by Seah *et al.* (2019)[44].

For downstream analysis of *Cand.* Kentron specific genes and phylogenomic analysis, we performed and orthologues gene clustering using OrthoFinder v2.5.1 including the MAGs of the Gamma4 and *Cand.* Kentron symbionts as well as a putative *Thiomagarita* sp. symbiont MAG from *I. exumae* individual 2 (IexuBAH2)[71].

Phylogenetic and phylogenomic analyses

For the reconstruction of a phylogenetic tree, we obtained 16S rRNA genes from the Gamma4 and *Cand.* Kentron MAGs using barrnap v0.9 (https://github.com/tseemann/barrnap). We aligned the resulting sequences using mafft-linsi v7.407 and calculated a maximum-likelihood phylogeny iqtree v1.6.10, including automatic selection of the best suited model and generation of 100 none-parametric bootstrap replicates[72–74]. For the reconstruction of a phylogenomic tree, we used the sequences of single copy orthologues that were identified using OrthoFinder. We aligned these sequences using mafft-linsi and calculated a maximum-likelihood phylogeny iqtree, including automatic selection of the best suited model and generation of 100 none-parametric bootstrap replicates. For both, the phylogenetic and phylogenomic analyses, we used the *Thiomargarita* sp. symbiont from *I. exumae* individual 2 as an outgroup.

Analysis of GC content

For the analysis of GC content patterns, we considered the overall GC content as estimated by CheckM (Table S1). In addition, we calculated the GC content coding and non-coding regions as well as the third codon position based on genes predicted by prokka. For the calculations of GC content at specific genomic regions, we used R v3.6.3 and the seqinr package v4.2-8[75,76].

Analysis of pseudogenes

For the detection of pseudogenes, we analyzed gene length distributions as well as the percentage of predicted genes that obtained a functional annotation and were not annotated as 'hypothetical protein' based on prokka gene predictions and annotations. In addition, we used the results of the orthologues clustering to define orthologues gene sets that were unique to one of the symbiont subclades and determined their length and their annotation status.

Analysis of single nucleotide variants

Single nucleotide variants were analyzed with inStrain v1.4.0, using the most complete bin of every symbiont subclade as mapping reference[77].

Code availability

Scripts that were used for the GC content calculations and the characterization of pseudogenes as well as the scripts that were used for data visualization are available under: https://github.com/amankowski/symbiont_genome_reduction. In addition, scripts for our assembly and binning pipeline are available under: https://github.com/amankowski/MG-processing_from-reads-to-bins.

Data availability

Raw metagenomic sequences and symbiont MAGs generated in this study will be deposited in the European Nucleotide Archive (ENA) upon peer-review submission and are currently available upon request.

Acknowledgments

We are thankful for sample collections and field assistance by Christer Erséus, Erich Mueller, Jörg Ott and Pamela Reid, In addition, we would like to thank the crew of the Little Darby Island Research Station This work was supported by the Max Planck Society, a Moore Foundation Marine Microbial Initiative Investigator Award to ND (Grant GBMF3811), a U.S. National Science Foundation award to MK (grant IOS 2003107), the USDA National Institute of Food and Agriculture Hatch project 1014212 (MK), and a Marie-Curie Intra-European Fellowship PIEF-GA-2011-301027 CARISYM (HRGV). The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231.

References

1. Toft, C. & Andersson, S. G. E. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat. Rev. Genet.* **11**, 465–475 (2010).
2. McCutcheon, J. P. & Moran, N. A. Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**, 13–26 (2012).
3. Moran, N. A. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci.* **93**, 2873 LP – 2878 (1996).
4. Andersson, S. G. E. & Kurland, C. G. Reductive evolution of resident genomes. *Trends Microbiol.* **6**, 263–268 (1998).
5. McCutcheon, J. P., McDonald, B. R. & Moran, N. A. Origin of an alternative genetic code in the extremely small and GC–rich genome of a bacterial symbiont. *PLoS Genet.* **5**, e1000565 (2009).
6. McCutcheon, J. P., McDonald, B. R. & Moran, N. A. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc. Natl. Acad. Sci.* **106**, 15394–15399 (2009).
7. Wu, D. *et al.* Metabolic Complementarity and Genomics of the Dual Bacterial Symbiosis of Sharpshooters. *PLOS Biol.* **4**, e188 (2006).
8. Baumann, P. Biology Of Bacteriocyte-Associated Endosymbionts Of Plant Sap-Sucking Insects. *Annu. Rev. Microbiol.* **59**, 155–189 (2005).
9. McCutcheon, J. P. & Moran, N. A. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc. Natl. Acad. Sci.* **104**, 19392–19397 (2007).
10. Nakabachi, A. *et al.* The 160-kilobase genome of the bacterial endosymbiont Carsonella. *Science (80-.).* **314**, 267 (2006).
11. Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. & Ishikawa, H. Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. *Nature* **407**, 81–86 (2000).
12. Douglas, A. E. Mycetocyte Symbiosis In Insects. *Biol. Rev.* **64**, 409–434 (1989).
13. Douglas, A. E. Nutritional Interactions in Insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria *Buchnera*. *Annu. Rev. Entomol.* **43**, 17–37 (1998).
14. Sasaki, T. & Ishikawa, H. Production of essential amino acids from glutamate by mycetocyte symbionts of the pea aphid, *Acyrthosiphon pisum*. *J. Insect Physiol.* **41**, 41–46 (1995).

15. Buchner, P. Endosymbiosis of animals with plant microorganisms. *New York: Interscience Publisher* (1965).
16. Houk, E. J. & Griffiths, G. W. Intracellular Symbiotes of the Homoptera. *Annu. Rev. Entomol.* **25**, 161–187 (1980).
17. López-Madrigal, S., Latorre, A., Porcar, M., Moya, A. & Gil, R. Complete genome sequence of '*Candidatus* Tremblaya princeps' strain PCVAL, an intriguing translational machine below the living-cell status. *J. Bacteriol.* **193**, 5587–5588 (2011).
18. Lund, M., Kjeldsen, K. & Schramm, A. The earthworm-*Verminephrobacter* symbiosis: an emerging experimental system to study extracellular symbiosis . *Frontiers in Microbiology* vol. 5 128 (2014).
19. Russell, S. L. *et al.* Horizontal transmission and recombination maintain forever young bacterial symbiont genomes. *PLOS Genet.* **16**, e1008935 (2020).
20. Russell, S. L., Corbett-Detig, R. B. & Cavanaugh, C. M. Mixed transmission modes and dynamic genome evolution in an obligate animal–bacterial symbiosis. *ISME J.* **11**, 1359–1371 (2017).
21. Foissner, W. Kentrophoros (Ciliophora, Karyorelictea) has Oral Vestiges: a Reinvestigation of *K. fistulosus* (Fauré-Fremiet, 1950) Using Protargol Impregnation. *Arch. für Protistenkd.* **146**, 165–179 (1995).
22. Raikov, I. B. Primitive Never-Dividing Macronuclei of Some Lower Ciliates. *Int. Rev. Cytol.* **95** 267–325 (1985).
23. Sauerbrey, E. Beobachtungen über einige neue oder wenig bekannte marine Ciliaten. *Arch. Für Protistenkd.* ***62, 355-407*** (1928).
24. Bergin, C. *et al.* Acquisition of a Novel Sulfur-Oxidizing Symbiont in the Gutless Marine Worm *Inanidrilus exumae*. *Appl. Environ. Microbiol.* **84**, e02267-17 (2018).
25. Mankowski, A. *et al.* Highly variable fidelity drives symbiont community composition in an obligate symbiosis. *bioRxiv* 2021.04.28.441735 (2021) doi:10.1101/2021.04.28.441735.
26. Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
27. Petersen, J. M. *et al.* Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation. *Nat. Microbiol.* **2**, 16195 (2016).

28. Gardebrecht, A. *et al.* Physiological homogeneity among the endosymbionts of *Riftia pachyptila* and *Tevnia jerichonana* revealed by proteogenomics. *ISME J.* **6**, 766–776 (2012).
29. Moya, A., Pereto, J. & Gil Rosario & Latorre, A. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat. Rev. Genet.* **9**, 218–229 (2008).
30. Newton, I. L. G. *et al.* The *Calyptogena magnifica* Chemoautotrophic Symbiont Genome. *Science (80-.).* **315**, 998 LP – 1000 (2007).
31. Dmytrenko, O. *et al.* The genome of the intracellular bacterium of the coastal bivalve, *Solemya velum*: a blueprint for thriving in and out of symbiosis. *BMC Genomics* **15**, 924 (2014).
32. Kuwahara, H. *et al.* Reduced Genome of the Thioautotrophic Intracellular Symbiont in a Deep-Sea Clam, *Calyptogena okutanii*. *Curr. Biol.* **17**, 881–886 (2007).
33. Sato, Y. *et al.* Fidelity varies in the symbiosis between a gutless marine worm and its microbial consortium. *bioRxiv* 2021.01.30.428904 (2021) doi:10.1101/2021.01.30.428904.
34. Jäckle, O. *et al.* Chemosynthetic symbiont with a drastically reduced genome serves as primary energy storage in the marine flatworm*Paracatenula*. *Proc. Natl. Acad. Sci.* **116**, 8505 LP – 8514 (2019).
35. Rocha, E. P. C. & Danchin, A. Base composition bias might result from competition for metabolic resources. *Trends Genet.* **18**, 291–294 (2002).
36. Moran, N. A. & McCutcheon John P. & Nakabachi, A. Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* **42**, 165–190 (2008).
37. Bentley, S. D. & Parkhill, J. Comparative Genomic Structure of Prokaryotes. *Annu. Rev. Genet.* **38**, 771–791 (2004).
38. Lind, P. A. & Andersson, D. I. Whole-genome mutational biases in bacteria. *Proc. Natl. Acad. Sci.* **105**, 17878 LP – 17883 (2008).
39. Bohlin, J., Skjerve, E. & Ussery, D. W. Investigations of Oligonucleotide Usage Variance Within and Between Prokaryotes. *PLOS Comput. Biol.* **4**, e1000057 (2008).
40. Hildebrand, F., Meyer, A. & Eyre-Walker, A. Evidence of Selection upon Genomic GC-Content in Bacteria. *PLOS Genet.* **6**, e1001107 (2010).
41. Sloan, D. B. & Moran, N. A. Genome Reduction and Co-evolution between the Primary and Secondary Bacterial Symbionts of Psyllids. *Mol. Biol. Evol.* **29**, 3781–3792 (2012).

42. Moran, N. A., McLaughlin, H. J. & Sorek, R. The Dynamics and Time Scale of Ongoing Genomic Erosion in Symbiotic Bacteria. *Science (80-.).* **323**, 379 LP – 382 (2009).
43. Seah, B. K. B. *et al.* Specificity in diversity: single origin of a widespread ciliate-bacteria symbiosis. *Proc. R. Soc. B Biol. Sci.* **284**, 20170764 (2017).
44. Seah, B. K. B. *et al.* Sulfur-Oxidizing Symbionts without Canonical Genes for Autotrophic CO2 Fixation. *MBio* **10**, e01112-19 (2019).
45. Robidart, J. C., Roque, A., Song, P. & Girguis, P. R. Linking Hydrothermal Geochemistry to Organismal Physiology: Physiological Versatility in *Riftia pachyptila* from Sedimented and Basalt-hosted Vents. *PLoS One* **6**, e21692 (2011).
46. Markert, S. *et al.* Status quo in physiological proteomics of the uncultured *Riftia pachyptila* endosymbiont. *Proteomics* **11**, 3106–3117 (2011).
47. Petersen, J. M. *et al.* Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* **476**, 176–180 (2011).
48. Woyke, T. *et al.* Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**, 950–955 (2006).
49. Kleiner, M. *et al.* Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E1173--82 (2012).
50. Wilson, A. C. C. & Duncan, R. P. Signatures of host/symbiont genome coevolution in insect nutritional endosymbioses. *Proc. Natl. Acad. Sci.* **112**, 10255 LP – 10261 (2015).
51. Shigenobu, S. & Wilson, A. C. C. Genomic revelations of a mutualism: the pea aphid and its obligate bacterial symbiont. *Cell. Mol. Life Sci.* **68**, 1297–1309 (2011).
52. Hansen, A. K. & Moran, N. A. Aphid genome expression reveals host–symbiont cooperation in the production of amino acids. *Proc. Natl. Acad. Sci.* **108**, 2849 LP – 2854 (2011).
53. Bing, X. *et al.* Unravelling the relationship between the tsetse fly and its obligate symbiont Wigglesworthia: transcriptomic and metabolomic landscapes reveal highly integrated physiological networks. *Proc. R. Soc. B Biol. Sci.* **284**, 20170360 (2017).
54. Gruber-Vodicka, H. R., Seah, B. K. B. & Pruesse, E. phyloFlash: Rapid Small-Subunit rRNA Profiling and Targeted Assembly from Metagenomes. *mSystems* **5**, e00920-20 (2020).

55. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
56. Kang, D. D., Froula, J. & Egan Rob & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
57. Wu, Y.-W., Tang, Y.-H., Tringe, S. G., Simmons, B. A. & Singer, S. W. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* **2**, 26 (2014).
58. Miller, I. J. *et al.* Autometa: automated extraction of microbial genomes from individual shotgun metagenomes. *Nucleic Acids Res.* **47**, e57–e57 (2019).
59. Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* **3**, 836–843 (2018).
60. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2020).
61. Parks, D. H., Imelfort, M., Skennerton, C. T. & Hugenholtz Philip & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
62. Laslett, D. & Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **32**, 11–16 (2004).
63. Brettin, T. *et al.* RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* **5**, 8365 (2015).
64. Davis, J. J. *et al.* The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Res.* **48**, D606–D612 (2020).
65. Karp, P. D. *et al.* Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology. *Brief. Bioinform.* **17**, 877–890 (2016).
66. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* **40**, D742--53 (2012).
67. Belcour, A. *et al.* Metage2Metabo, microbiota-scale metabolic complementarity for the identification of key species. *Elife* **9**, e61968 (2020).
68. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

69. Xie, Z. & Tang, H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* **33**, 3340–3347 (2017).
70. Saier Jr, M. H. *et al.* The Transporter Classification Database (TCDB): 2021 update. *Nucleic Acids Res.* **49**, D461–D467 (2021).
71. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
72. 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.* **30**, 3059–3066 (2002).
73. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
74. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
75. Charif, D. & Lobry, J. R. SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis BT - Structural Approaches to Sequence Evolution: Molecules, Networks, Populations. in (eds. Bastolla, U., Porto, M., Roman, H. E. & Vendruscolo, M.) 207–232 (Springer Berlin Heidelberg, 2007). doi:10.1007/978-3-540-35306-5_10.
76. R Core Team. R: A Language and Environment for Statistical Computing. (2020).
77. Olm, M. R. *et al.* inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nat. Biotechnol.* **39**, 727–736 (2021).

Chapter II | Supplementary material

Figures

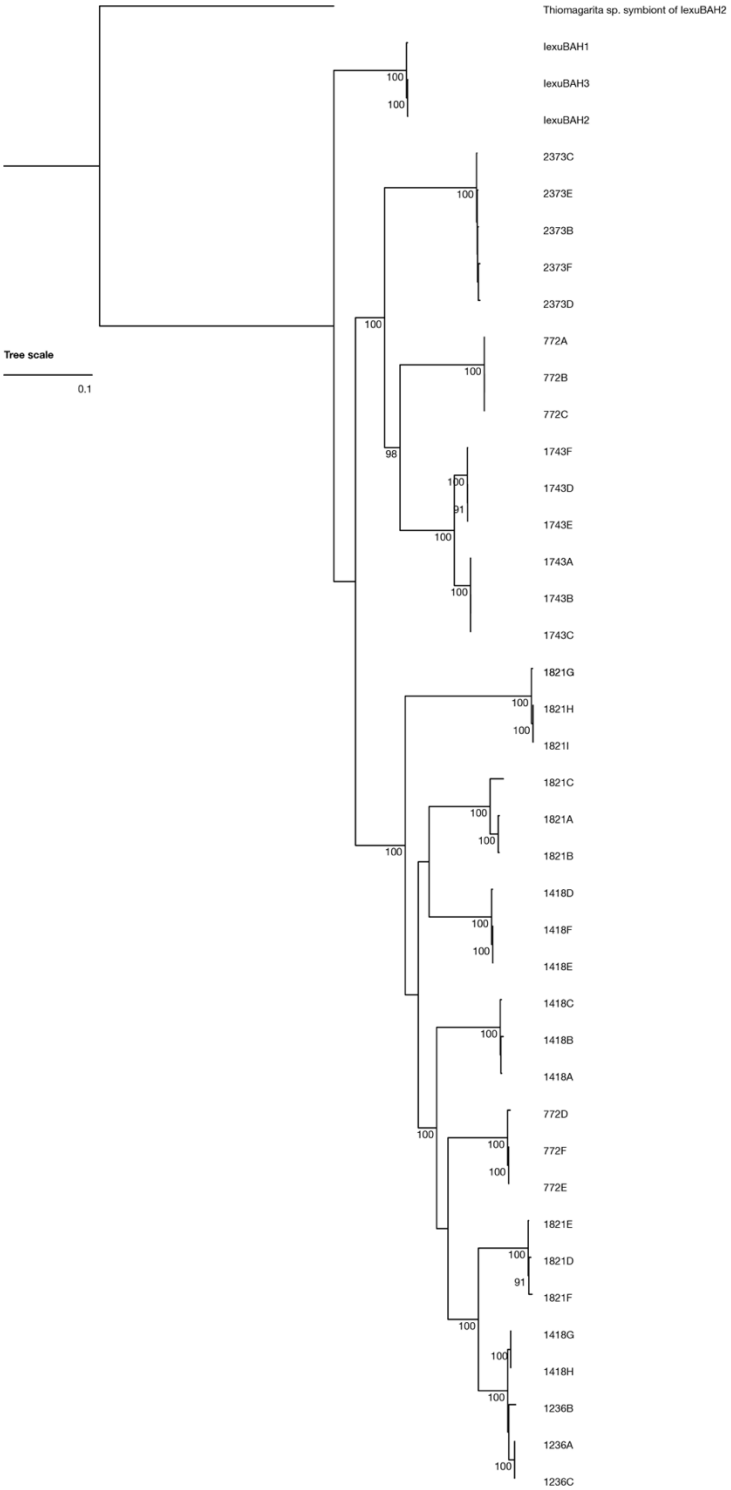


Figure S1 | The Gamma4 symbionts from *I. exumae* are a subclade of the *Cand.* Kentron symbionts. Maximum-likelihood tree that shows the phylogenetic relations between several *Kentrophoros* and *I. exumae* symbionts based on the 16S rRNA gene The scale bar indicates 10% estimated sequence divergence. Nodes that are statistically supported by none-parametric bootstrapping are highlighted within the tree.

Tables

Table S1: Overview of *Cand.* Kentron and Gamma4 symbiont MAGs.

Sample ID	Symbiont clade	Host species	Genome size (Mbp)	Completeness (%)	Contamination (%)	# of amino acids with a present tRNA	GC content (%)	# scaffolds	N50
IexuBAH1	Gamma4	*I. exumae*	3.40	81.43	1.47	18	48.4	4951	2359
IexuBAH2	Gamma4	*I. exumae*	3.38	83.85	2.11	20	48.6	4811	2443
IexuBAH3	Gamma4	*I. exumae*	3.42	85.64	0.88	20	48.5	4983	2394
2373B	Clade1	*K.* cf. *fasciolatus*	5.02	88.98	1.72	20	56.9	542	30714
2373C	Clade1	*K.* cf. *fasciolatus*	4.82	87.25	1.19	19	56.8	507	35107
2373D	Clade1	*K.* cf. *fasciolatus*	4.58	87.28	1.96	20	57.1	449	35933
2373E	Clade1	*K.* cf. *fasciolatus*	4.75	86.97	0.84	20	56.8	389	40851
2373F	Clade1	*K.* cf. *fasciolatus*	4.75	88.09	1.55	20	57.0	469	31687
772Aa	Clade2	*K.* sp. H	3.92	85.67	2.52	21	54.9	1295	10068
772Ba	Clade2	*K.* sp. H	3.91	85.85	2.44	21	55.0	1253	10069
772Ca	Clade2	*K.* sp. H	3.91	85.85	2.44	21	54.9	1293	9926
1743A	Clade3	*K.* sp. FM	5.01	87.71	2.72	20	54.0	1923	5788
1743B	Clade3	*K.* sp. FM	4.94	87.71	2.19	20	54.1	1794	5713
1743C	Clade3	*K.* sp. FM	5.00	87.71	2.72	20	54.0	1828	5786
1743D	Clade4	*K.* sp. G	4.75	87.48	1.58	20	53.9	2741	3539
1743E	Clade4	*K.* sp. G	4.65	87.61	1.55	20	54.0	2723	3440
1743F	Clade4	*K.* sp. G	4.60	86.61	1.55	20	54.0	2677	3342
1821G	Clade5	*K.* sp. MB	3.88	90.35	2.11	20	52.6	520	37471
1821H	Clade5	*K.* sp. MB	3.91	90.72	2.11	20	52.6	473	36412
1821I	Clade5	*K.* sp. MB	3.93	90.70	2.46	20	52.6	467	35532
1418D	Clade6	*K.* sp. TUN	3.62	90.95	1.23	20	49.8	909	13993
1418E	Clade6	*K.* sp. TUN	3.63	90.60	0.59	20	49.8	790	16723
1418F	Clade6	*K.* sp. TUN	3.62	90.60	0.59	20	49.8	772	16667
1821A	Clade7	*K.* sp. FW	3.65	89.48	2.00	20	53.1	386	25283
1821B	Clade7	*K.* sp. FW	3.77	90.32	2.17	20	53.1	522	23320
1821C	Clade7	*K.* sp. FW	4.05	90.82	1.64	20	53.4	454	31676

Sample ID	Symbiont clade	Host species	Genome size (Mbp)	Completeness (%)	Contamination (%)	# of amino acids with a present tRNA	GC content (%)	# scaffolds	N50
1418A	Clade8	*K.* sp. LFY	3.45	89.93	0.97	20	53.5	390	25486
1418B	Clade8	*K.* sp. LFY	3.42	90.31	0.94	20	53.5	420	25968
1418C	Clade8	*K.* sp. LFY	3.66	90.32	1.32	21	53.5	431	26336
772Da	Clade9	*K.* sp. SD	3.81	89.06	2.11	20	53.4	579	23889
772Ea	Clade9	*K.* sp. SD	3.76	89.06	1.94	20	53.6	666	20903
772Fa	Clade9	*K.* sp. SD	3.80	89.06	2.41	20	53.5	642	19955
1821D	Clade10	*K.* sp. TC	3.32	88.55	1.64	20	53.7	584	24925
1821E	Clade10	*K.* sp. TC	3.35	89.08	2.00	20	53.8	504	35962
1821F	Clade10	*K.* sp. TC	3.31	88.72	1.64	20	53.7	558	33527
1236A	Clade11	*K.* sp. LPFa	4.08	88.92	2.50	19	52.8	1187	9983
1236B	Clade11	*K.* sp. LPFa	4.19	89.52	1.29	20	53.6	757	15237
1236C	Clade11	*K.* sp. LPFa	4.03	88.93	2.35	20	52.9	1104	10160
1418G	Clade11	*K.* sp. UNK	4.23	89.54	2.21	20	53.5	610	21136
1418H	Clade11	*K.* sp. UNK	4.22	89.54	2.29	20	53.6	609	21983

Table S2: Overview of MAGs from secondary symbionts of *I. exumae*.

Host individual	Symbiont clade	Genome size (Mbp)	Completeness (%)	Contamination (%)	# of amino acids with a present tRNA	GC content (%)	# scaffolds	N50
IexuBAH1	Alpha12A	4.70	99.13	2.97	19	65.3	231	29576
IexuBAH1	Alpha12B	4.77	99.13	2.61	19	63.0	14	629919
IexuBAH1	Alpha3	4.87	98.63	2.24	20	61.7	2040	12328
IexuBAH1	Chlamydia	1.43	90.03	1.13	20	43.9	167	12505
IexuBAH2	Alpha13	3.36	98.38	1.99	21	66.2	109	54397
IexuBAH2	Delta3	5.40	99.03	0.00	20	52.8	157	55330
IexuBAH2	Delta12	5.69	97.10	0.86	21	52.4	83	108087
IexuBAH2	*Thiomagarita* sp.	5.29	88.46	4.36	17	41.8	580	14057
IexuBAH3	Alpha13	3.46	98.38	1.99	21	66.0	150	37910
IexuBAH3	Delta3	5.71	98.55	0.00	21	52.4	511	46239

Extended data

Extended data sets are available under https://doi.org/10.5281/zenodo.5494635

Extended Data S1

ExtendedDataS1_list-of-pathways-annotated-in-CandKentron-genomes.tsv

Extended Data S2

ExtendendDataS2_sequences-for-custom-annotation.faa

Chapter III | Forever young symbiont communities in an ancient and obligate symbiosis

Anna Mankowski[*1], Manuel Kleiner[2], Christer Erséus[3], Nikolaus Leisch[1], Yui Sato[1], Jean-Marie Volland[4], Bruno Hüttel[5], Cecilia Wentrup[6], Tanja Woyke[4], Juliane Wippler[1], Nicole Dubilier[*1], Harald R. Gruber-Vodicka[*1]

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

[2]North Carolina State University, Department of Plant and Microbial Biology, Raleigh, North Carolina, USA

[3]University of Gothenburg, Department of Biological and Environmental Sciences, Gothenburg, Sweden

[4]Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California, USA

[5]Max Planck Genome Centre Cologne, Max Planck Institute for Plant Breeding Research, Cologne, Germany

[6]University of Vienna, Department of Microbiology and Ecosystem Science, Vienna, Austria

[*]Corresponding authors:

Nicole Dubilier: ndubilie@mpi-bremen.de, phone +49 (421) 2028 9320

Harald R. Gruber-Vodicka: hgruber@mpi-bremen.de, phone +49 (421) 2028 7600

Author contributions AM, MK, HRGV, JW and ND conceived the study and AM designed the workflows with help from HRGV and JW. AM, MK, CÉ, NL, YS, JMV, BH, CW, TW, JW and HRGV acquired specimens and generated metagenomic data. AM analyzed the data. AM, HRGV, MK and ND interpreted the results. AM drafted the manuscript and AM and HRGV edited the manuscript. All authors provided revisions.

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

An earlier version of this manuscript has been deposited as preprint in BioRxiv under https://doi.org/10.1101/2021.04.28.441735.

Abstract

The observation that symbiont communities systematically vary between sampling locations and host taxa is enigmatic when the host animals obligately rely on symbiont services throughout their lifecycle. To explore the evolutionary mechanisms that drive such variable symbiont community compositions, we used gutless oligochaetes that fully depend on chemosynthetic symbiont communities. Our analyses of 231 host individuals which cover 64 host species and 17 globally distributed sampling sites showed that the hosts are associated with 33 symbiont genera of which 21 are specific to gutless oligochaetes. Phylogenetic and multivariate statistical analyses revealed that community composition cannot be predicted by host phylogeny. At the level of symbiont genera, symbiont communities are host species specific but symbiont phylotypes are regularly switched within and between host species. We explain the high evolutionary flexibility of gutless oligochaetes with microevolutionary patterns of varying fidelity for single host – symbiont pairs that drive community variation.

Introduction

Mutualistic associations thrive when stabilizing mechanisms such as high partner specificity and/or high partner fidelity ensure partnership quality[1–3]. High specificity is defined as a low taxonomic range of associated partners and high partner fidelity is the stable association of a specific couple of host and symbiont lineages over several host generations. In consequence, obligate mutualistic associations where a host depends on symbiont services should be characterized by low symbiont diversity and low symbiont variability. Still, obligate and at the same time variable multimember mutualistic associations are ubiquitous in nature, be it that a single animal host species depends on several strains of the same symbiont or that a single host species even harbors many diverse symbionts lineages from multiple phyla[4–6]. How such symbiont diversity and variability can persist in an animal host species that is dependent on its symbionts throughout its life-cycle however remains elusive.

A taxonomically and symbiotically diverse mutualistic association are the gutless oligochaetes and their obligate chemosynthetic symbiont communities. The gutless oligochaetes are a globally distributed group of annelids that do not have nephridia, a mouth or a gut (hence gutless), but rather depend on symbiont communities for their nutrition as well as waste product recycling[7–9]. These symbiont communities display different levels of specificity and fidelity on macro- to microevolutionary scales. At the host species level, the seven species that were analyzed so far were associated with a total of 14 genus-level symbiont clades. The community of each of these host species consisted of a host species specific combination of four to six of these symbiont genera that show only little variation at the individual level[7,10–15]. The main symbiont genus, *Cand.* Thiosymbion, is host species specific, but the incongruent host and symbiont phylogenies indicate frequent symbiont switching, to the point that in one host species, *Cand.* Thiosymbion was even lost and replaced by an unrelated gammaproteobacterium[1516,17]. In *Olavius algarvensis*, the host species researched in the most detail, not all symbionts were present in all samples and symbiont fidelity appeared variable for different members of the symbiont community when compared to mitochondrial evolutionary patterns[17].

With more than 100 described species, and with variable symbiont communities observed for several species, the gutless oligochaetes lend themselves as a model to study evolution of symbiont diversity in mutualistic associations where the host obligately relies on its symbionts. We therefore generated linked datasets of a host nuclear gene (28S rRNA gene), a host mitochondrial gene (mtCOI gene) and symbiont community composition based on 16S rRNA genes by sequencing hundreds of single-individual metagenomes that cover the known diversity of gutless oligochaetes and that were sampled from globally distributed sites. We compared host nuclear and host mitochondrial relationships with community structure and evolution of individual symbiont genera using multivariate statistics, comparisons between the

hosts' and each of the symbionts' phylogenies as well as divergence times and ancestral state reconstructions. Our analyses revealed highly variable levels of symbiont fidelity and flexible community composition that already generate variation over short evolutionary time scales and in single populations.

Results

Incongruent evolution of gutless oligochaete nuclear and mitochondrial genomes

We employ the host evolutionary history as the key framework to interpret symbiont community composition and its evolution. Our dataset provides two genes to reconstruct host evolution, the nuclear encoded 28S rRNA gene, and a mitochondrial marker gene, the mtCOI. Given the fact that several examples of evolutionary incongruencies between the nuclear and mitochondrial genomes have been observed across animal groups[18–21], we first reconstructed and compared their phylogenies. The two phylogenies were largely incongruent, particularly at the more basal nodes, which indicates that the nuclear and mitochondrial genomes in gutless oligochaetes had different evolutionary trajectories (64 out of 216 nodes shared between trees; Figure S1, Extended Data). Thus, in the following analyses, we compared symbiont community composition and symbiont fidelity to both the host 28S rRNA as well as its mtCOI gene phylogeny. Different congruencies between the comparisons to the two genes can then be used to infer potential modes of transmission. Mitochondria are maternally transmitted from parent to offspring[22]. Congruencies between the symbiont phylogeny and the mtCOI based host phylogeny would therefore indicate maternal transmission of the respective symbiont[23]. As an alternative to maternal transmission, symbionts could also be acquired through host switching or from the environment. Selective uptake during host switching or from the environment is largely mediated by traits encoded in the host nuclear genome, e.g. by receptors and effectors that are involved in innate immunity[23]. Congruencies between the symbiont phylogeny and the nuclear 28S based host phylogeny would therefore indicate selective acquisition that is

independent of the source. If the majority of symbionts from a community are transmitted the same way, be it maternally or from other hosts or the environment, this would be reflected in congruence patterns between the dendrogram of symbiont community compositions and one of the host gene phylogenies.

Single-individual metagenomics of gutless oligochaetes reveal constrained and host species specific symbiont communities

The single individual is the unit of selection, with both genetic but also microbiome variation distributed over the individuals of a given population. We therefore analyzed 231 shot-gun metagenomes of single gutless oligochaete specimens to capture the host diversity and the linked symbiont communities for 64 species and from 17 globally distributed sites in the Atlantic, Mediterranean, Indian, Pacific and Red Sea (Figure S2). We reconstructed full-length 16S rRNA genes of the bacteria associated with each specimen and compiled a database of all bacterial phylotypes. Based on previous symbiont descriptions for e.g. *Cand.* Thiosymbion, we then defined clades at the genus level and classified all phylotypes by considering all genera associated with at least 50% of the individuals from at least one species as symbionts (Figure 1, Note S1). Based on this criterion, we observed 33 symbiont genera that belonged to six classes - Actinobacteria, Alpha-, Delta- and Gammaproteobacteria, Marinimicrobia and Spirochaetia (Table S1). 14 of these 33 genera were not known to be associated with gutless oligochaetes before[7,10–15]. We then characterized the composition of the symbiont community in each individual using these genus-level classifications. For each host species, symbiont communities consisted of a subset of two to ten of the 33 symbiont genera. For the symbiont genera, the number of hosts species they were associated with was highly variable, ranging from being present in almost all host species to being present in a single species (for exact host ranges see Figure 1 and Table S2).

We next asked if the gutless oligochaete symbiont genera are also associated with other host groups or occur in other environments. Therefore, we analyzed the phylogenetic relations between all symbionts and their closest relatives from publicly available 16S rRNA gene sequences. In addition, we generated metagenomes of the closest gut-bearing relatives of the gutless oligochaetes that co-occur in the same habitats and screened them for the presence of gutless oligochaete symbionts (Fig. S3, S4-S36, Note S2). We identified 21 symbiont genera that were only associated with gutless oligochaetes. Furthermore, four symbiont genera were also associated with other hosts, such as gutbearing oligochaetes, stilbonematine and *Astomonema* nematodes and *Kentrophoros* ciliates that all co-occur in the same environments as gutless oligochaetes. Eight of the symbiont genera were phylogenetically intermixed with bacteria sampled from other environmental sources.

We were surprised that our analyses of 64 host species revealed only 33 symbiont genera. Although we analyzed the symbiont communities of nine times more host species than previously described, we only discovered roughly three times more symbiont genera[7,10–15]. This discrepancy and our rarefaction analysis showed a saturation of the detection of new symbiont genera, indicating that symbionts are only acquired from a limited pool of bacterial groups (Fig. S37). Given the high diversity of microorganisms in marine sediment communities, our results suggest that distinct symbiont genera are selected for. Using a co-occurrence network analysis, we screened for indications that symbiont-symbiont interactions would select for or against certain symbiont genera combinations. We did not observe any significant symbiont-symbiont exclusion patterns and only found two significant co-occurrence patterns between two pairs of genera (Figure 1, Note S3). In a second step, we compared the influence of host species, of geography, and of chemical and physical parameters of the environment on the symbiont community composition. Host species by far had the highest explanatory power for symbiont community composition based on UniFrac distances of symbiont communities between

specimens (PERMANOVA: host species: 89.62%, p=0.001, ocean basin: 20.26%, p=,0.001 organic input: 6.99, p=0.01, sediment type: 12.90%, p=0.001, Mantel test for geographic distance: R=0.1317, p=0.0001, Figure S38). We only detected minor variations between individuals of the same host species, with highly similar symbiont communities for specimens for each sampled population. In a given population symbiont genera were almost always present in all the overwhelming majority of individuals. At the species level, the community compositions of a given host species from different locations showed more pronounced and location specific variation, but still retained high levels of similarity (Table S2). Thus, we assume that for symbionts, an ability to associate with animal hosts and to interact with host traits such as the immune system are the key selective factors that constrain the overall diversity of gutless oligochaete symbionts[24–28]. Given the species specific patterns, the composition of these host associated bacterial communities then appeared to be defined by host species specific factors.

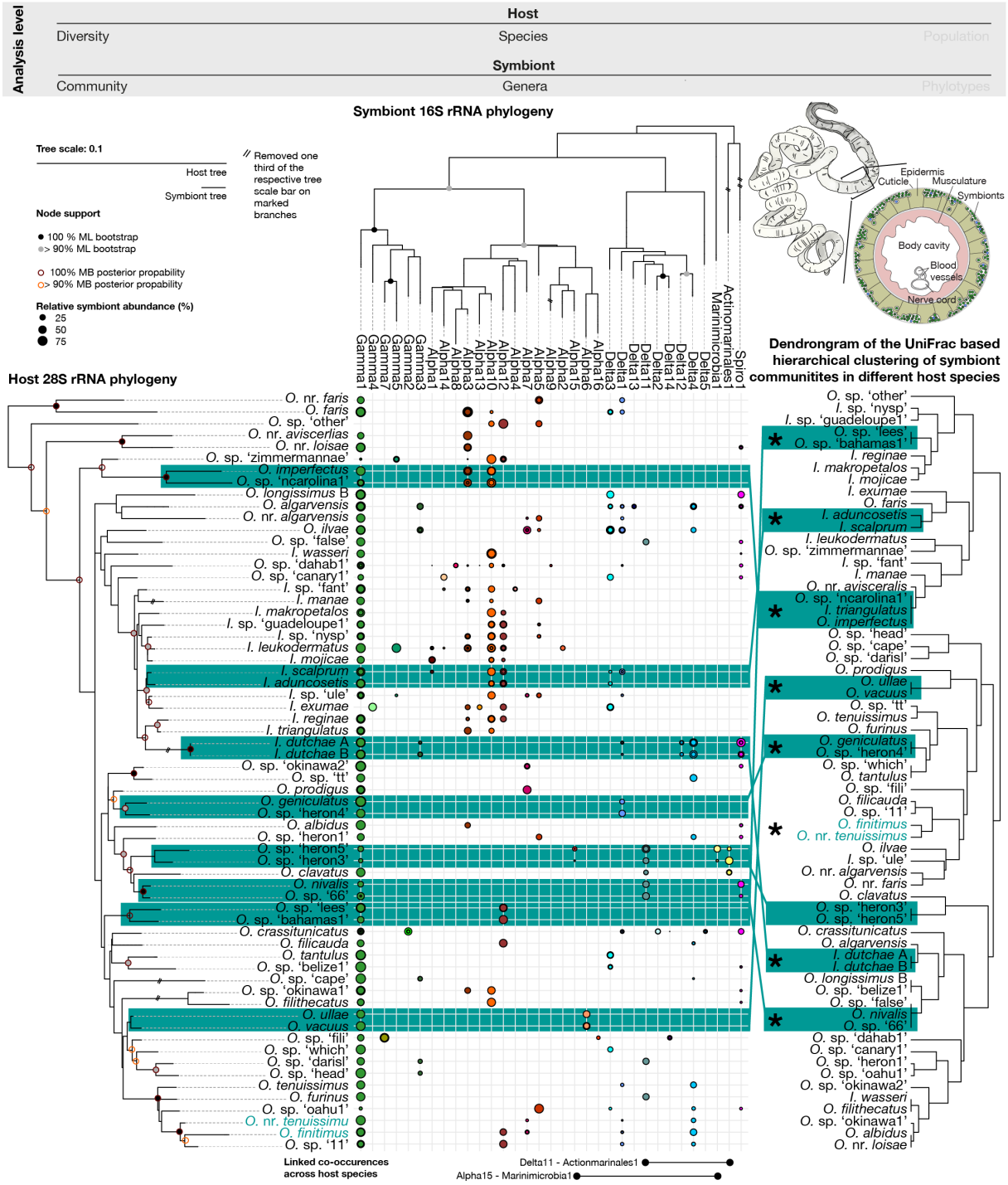


Figure 1 | The composition of symbiont communities across 64 gutless oligochaete species is not linked to host phylogeny. Top right: gutless oligochaete habitus and cross section with the symbionts housed between the cuticle and the epidermis; provided by Alina Esken (habitus) and Alexander Gruhl (cross-section). Top and left tree: Maximum-likelihood phylogeny of the host 28S rRNA gene (left) and the symbiont 16S rRNA gene (top), one representative of each host species/symbiont genus pair is shown. Scale bar indicates 10% estimated sequence divergence. Bootstrap support > 90% is highlighted in both trees; posterior probabilities > 90% estimated with MrBayes are shown for host tree nodes. Middle panel: Relative abundance of symbiont genera per host species estimated by EMIRGE. Right tree: Dendrogram of symbiont community composition representing UniFrac distances between host species. Green boxes highlight congruence between the dendrogram and the 28S rRNA gene phylogeny. *-marks congruence to host mtCOI gene phylogeny.

On macroevolutionary scales symbiont community composition is not linked to host phylogeny

In the light of the observed importance of the host species for community composition, we took a structured approach to reveal links between host and symbiont evolution. As the first step and at a macroevolutionary scale we compared the symbiont community composition to the host phylogeny. Although host species appeared to have a major influence on symbiont community composition, we observed that many host sister species had very different symbiont communities. In addition, some unrelated hosts sampled from very distant locations shared strikingly similar symbiont communities, for example *O. algarvensis* from the Mediterranean and *I. dutchae* from Hawai'i. Given these extreme examples, we statistically tested for congruence between symbiont community composition and both the nuclear and mitochondrial host phylogenies. The statistical analysis revealed that overall, the symbiont community composition is not linked to host phylogeny (Figure 1, 28S rRNA topology comparison: nRF=0.9937, p-value=1.0, nMC= 0.7847, p-value=0.8754; mtCOI topology comparison: nRF=0.8852, p-value=0.0, nMC=0.6126, p-value=0.00252). Although there were no consistent congruencies between host phylogenies and symbiont community composition, we detected nine non-random cases in which host sister species had very similar symbiont communities (t-test p-values for 28S rRNA and mtCOI: $< 2.2^{-16}$, Figure 1). These host sister species were more closely related than other host sister species with divergent symbiont communities (Mann-Whitney-U test p-values: $< 2.2^{-16}$ for 28S rRNA phylogeny and 1.482^{-13} for mtCOI phylogeny). In concordance with this topology-based analysis, the analysis of the relation between phylogenetic distances and the symbiont community composition distances of all host individuals showed only a weak linear correlation (correlation coefficients of linear models: 28S rRNA R^2=0.019, mtCOI: R^2=0.014, Figure 2A and N). The observations that symbiont communities within a host species and between closely related sister species are highly similar, indicate that the host, either via vertical transmission or inheritable traits that convey specificity,

can influence symbiont community composition. The effects of these strong mechanisms that act at the host species level are then apparently lost over macroevolutionary scales.

De novo acquisition, host switching and loss of symbiont genera drive the variable association patterns of symbionts

Considering the complex macroevolutionary patterns that characterize the overall community composition, we focused on each symbiont genus and compared symbiont and host evolution. The key source of variability in symbiont communities is the primary acquisition of new symbionts. To reconstruct when each symbiont genus was first acquired by gutless oligochaetes we used divergence time estimates of the hosts and the symbiont genera and combined these estimates with ancestral state reconstruction of symbiont occurrences (Figure 2, Figure S4-36, Note S3). We estimated that the gutless oligochaetes radiated 256.131 Mya ago. Based on the age and diversification in relation to the divergence time of the gutless oligochaete hosts, symbiont genera diversified i) before the host radiation, ii) alongside the hosts and iii) after host diversification. Our reconstructions show that only the Gamma1/*Cand.* Thiosymbion symbiont was acquired prior to the radiation of gutless oligochaetes. D*e novo* acquisition of all other symbiont genera occurred during or after the diversification of gutless oligochaetes and the rate of acquisition continuously increased over time. This points towards a high importance of the recent time window for primary symbiont acquisition as well as towards the acceleration of evolutionary flexibility and potential specializations.

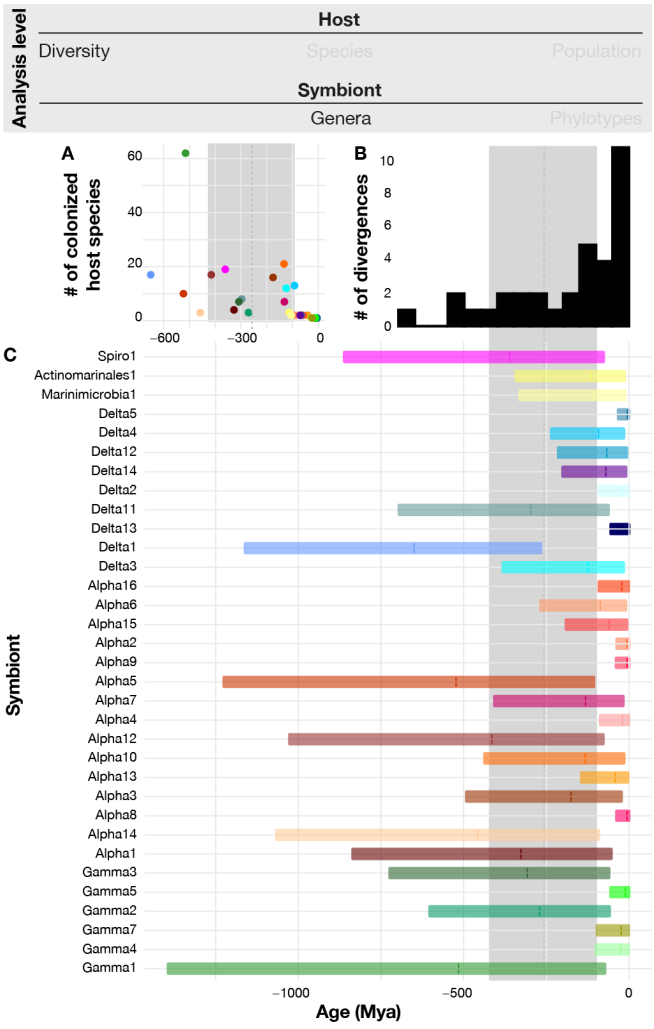


Figure 2 | The estimated divergence times of the host and symbiont clades indicate that the association between the host and the different symbiont clades started at different points of the host evolution. (A) Number of host species colonized by a given symbiont clade in relation to the median estimated divergence time of the clade. **(B)** Histogram showing the numbers of median estimated divergence times of symbiont clades in given 10 Mya time blocks. **(C)** Estimated divergence times of the host clade (gray box) and symbiont clades (colored boxes). Boxes represent the 95% interval of estimated divergence times, the solid bar represent the median estimated divergence time.

Once the association with a symbiont genus is established, repeated uptake in divergent host lineages as well as host switches and loss of this symbiont genus will drive the association pattern across the host diversity. Therefore, we explored the prevalence of uptakes, switches and losses of each symbiont genus. Our ancestral state reconstructions show that the symbiont genera that were acquired after the host diversification were likely acquired only once by a recent last common ancestor of their extant host species. These ‘young’ symbiont genera were only rarely switched between distantly related hosts and thus were mainly found in small,

monophyletic host groups. In contrast, many of the older symbiont genera were not confined to monophyletic groups of hosts but were patchily distributed across the host diversity (Figure 1 and Figure 3). We then tested whether these distributions were based on a single acquisition followed by several losses or multiple acquisitions and switches using ancestral state reconstructions. These analyses showed that the majority of the symbiont genera independently established their symbioses with distinct host lineages several times (Table S3). Our phylogenetic analyses of the symbiont genera and their closest relatives based on publicly available sequence data show that most were not or only partly phylogenetically intermixed with free-living relatives. Thus, we assume that the majority of these repeated acquisitions happened via host switching rather than environmental acquisition (Figure S4-36, Note S2). In addition, the phylogenetic analyses revealed additional hosts for the symbiont genera associated with gutless oligochaetes. Besides the previously published broad host range of the Gamma1/*Cand.* Thiosymbion symbiont, relatives of five other symbiont genera were also associated with other marine invertebrates and ciliates that share the same environment (Figure S4-36, Note S2)[16]. For several symbiont genera, host switching therefore did not only happen between gutless oligochaete individuals but gutless oligochaetes have also repeatedly acquired symbionts from other unrelated hosts.

Figure 3 | Due to low symbiont fidelity between host species symbiont community compositions evolves independent from host evolution. A and N: Each shows the pairwise dissimilarities of the overall symbiont community composition in relation to the pairwise nucleotide dissimilarity of the host 28S rRNA gene **(A)** or the host mtCOI gene **(N)**. **B-M** and **O-Z:** Each shows the16S rRNA nucleotide dissimilarity of one of the eleven most symbiont genera in relation to the pairwise nucleotide dissimilarity of the host 28S rRNA genee **(B-M)** or the host mtCOI gene **(O-Z)** Dotted lines represent the regression curve estimated by applying a linear model. R^2 values of each linear regression are and r values resulting from testing correlation between the pairwise 16S rRNA nucleotide distance versus the pairwise host marker gene nucleotide distance using the Mantel test are shown in each plot panel. Bold r values were statistically significant ($p \leq 0.05$).

On microevolutionary scales, fidelity varies across symbiont genus - host species pairings

In summary, our previous analyses showed that each symbiont genus is characterized by unique patterns of acquisition, switching and loss. Stable associations of a given symbiont genus with a given gutless oligochaete species are contrasted with frequent acquisitions and losses across the diversity of all hosts. To understand microevolutionary dynamics within each symbiont genus, we analyzed the symbiont phylotypes of single specimens and their switches between and within host species they are associated with. We assessed fidelity of every symbiont genus associated with at least five host species by testing for a possible correlation between the host individuals' genetic distances and each of the symbiont phylotypes' genetic distances (Figure 3). We assumed that high fidelity between symbionts from a certain genus and their hosts would lead to a linear correlation between the genetic distance of host pairs and the genetic distance of the respective symbiont phylotypes. We detected significant correlations for a majority of the tested symbiont genera (Mantel test p-value < 0.05: 9/12 symbiont genera vs. 28S rRNA genetic distance and 11/12 vs. mtCOI, Figure3). The low correlation coefficients point to a low predictive power of host relations for symbiont selection and rather illustrate rampant and ongoing symbiont phylotype switching between host individuals from different species (Figure 3).

This pointed us to analyze the symbiont fidelity at the smallest evolutionary scale our data provided – the phylotype association patterns within single host species and populations (Figure 4, Figure 5). Within single host species, only few symbiont genus - host species pairings showed a strong positive linear correlation between host and symbiont genetic distances, indicating high symbiont fidelity (Mantel test R=1; 28S rRNA: 1 of 44, mtCOI: 3 of 49). Other symbiont genus - host species pairings exhibited variable degrees of fidelity, independent of the symbiont genus, host species and host marker gene (Figure 4, Figure 5A and B). Three of the most widely distributed symbiont genera showed very variable degrees of fidelity with

correlations ranging from strong positive to strong negative in different host species (Gamma1, Gamma3, Delta1). Also, some of the symbiont genera had indications for high fidelity to a given host species for only one of the host marker genes (Gamma4 – *I.exumae*,).

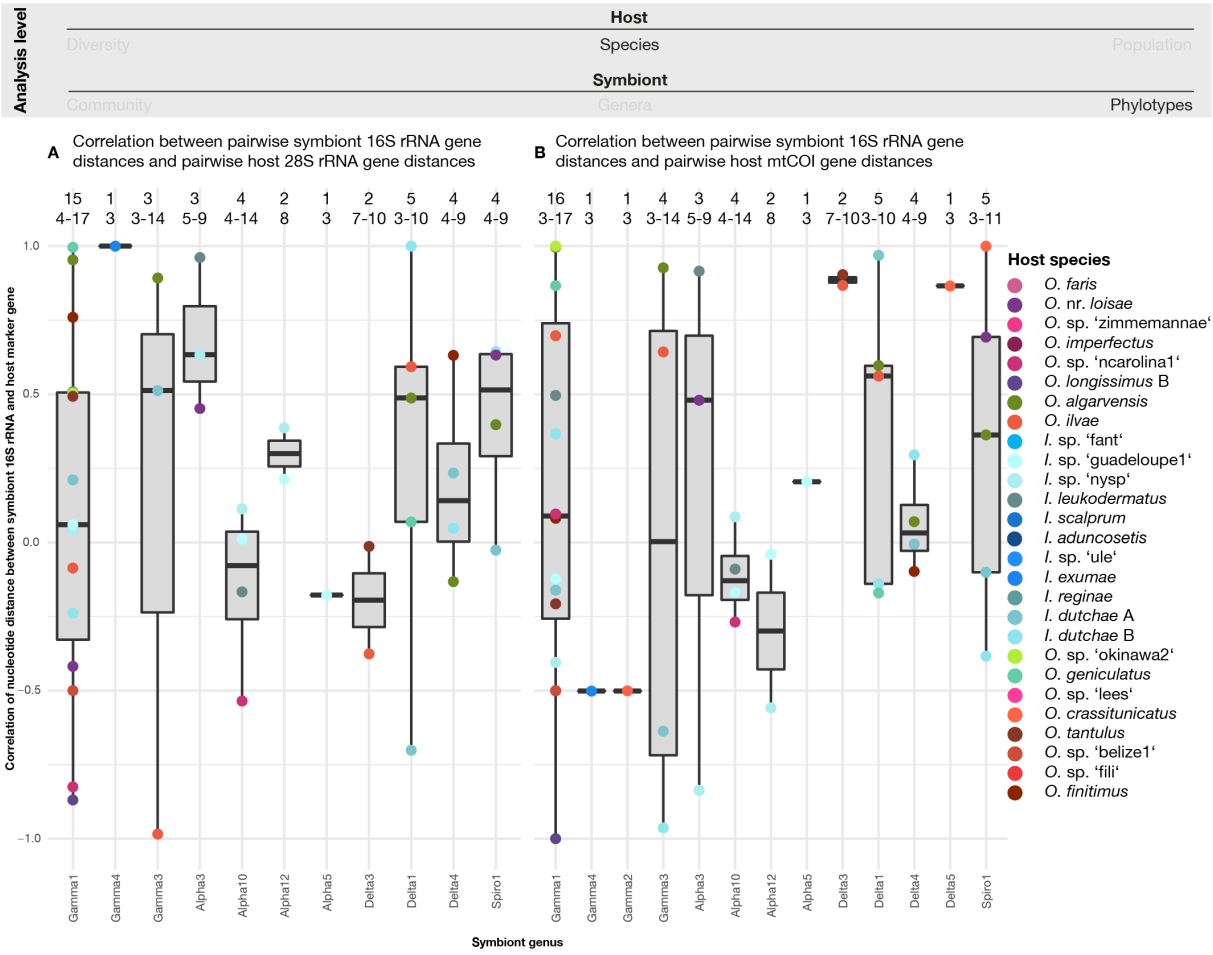


Figure 4 | Symbiont fidelity in gutless oligochaetes varies i) within symbiont genera across their host diversity ii) between symbiont genera in the same host species and iii) between host nuclear or mitochondrial markers. Both panels show symbiont fidelity as correlation coefficients of pairwise 16S rRNA gene distance of different symbiont genera from different host species and the pairwise genetic distance of host marker genes of host individuals the symbionts were associated with. **(A)** Host marker gene: 28S rRNA. **(B)** Host marker gene: mtCOI.

Given the fluctuating degrees of symbiont fidelity between host species, we asked whether the symbiont fidelity patterns were stronger in single host populations compared to the symbiont fidelity patterns across several populations of one host species. Our dataset enabled us to conduct this analysis for three species *Inanidrilus leukodermatus*, *O. algarvensis* and *O. ilvae*

that were each sampled at two different sites (Figure 5). The comparison of symbiont fidelity across populations versus within populations indicated that even within a single host population, symbiont fidelity was highly variable and unique to each symbiont genus – host species pairing. At the single population level, we observed higher congruences between symbiont genetic distances and host mitochondrial genetic distances compared to host nuclear genetic distances, indicating maternal transmission as a major transmission route. However, given the still fluctuating congruence values between symbiont genera and host species, we assume that phylotype switching is common at species-specific rates within a given host populations. This expands on previous findings by Sato *et al.* (2021) who showed varying levels of symbiont for 80 specimens of *O. algarvensis* sampled from a single Mediterranean island[17].

Our results show that the fidelity within symbiont genera is variable across host evolutionary scales; from the microevolutionary level of host genotypes in a given population to different and only distantly related host species. Given the observation that high symbiont fidelity is often correlated with a high degree of dependence of the host on its symbiont, we speculate that the hosts depend more strongly on symbiont genera that showed high symbiont fidelity[29]. Thus, the presence of symbionts that are essential in a given population is selected for and leads to high symbiont fidelity while the variable fidelity of other, non-essential genera could provide evolutionary and metabolic flexibility and adaptability of the whole community.

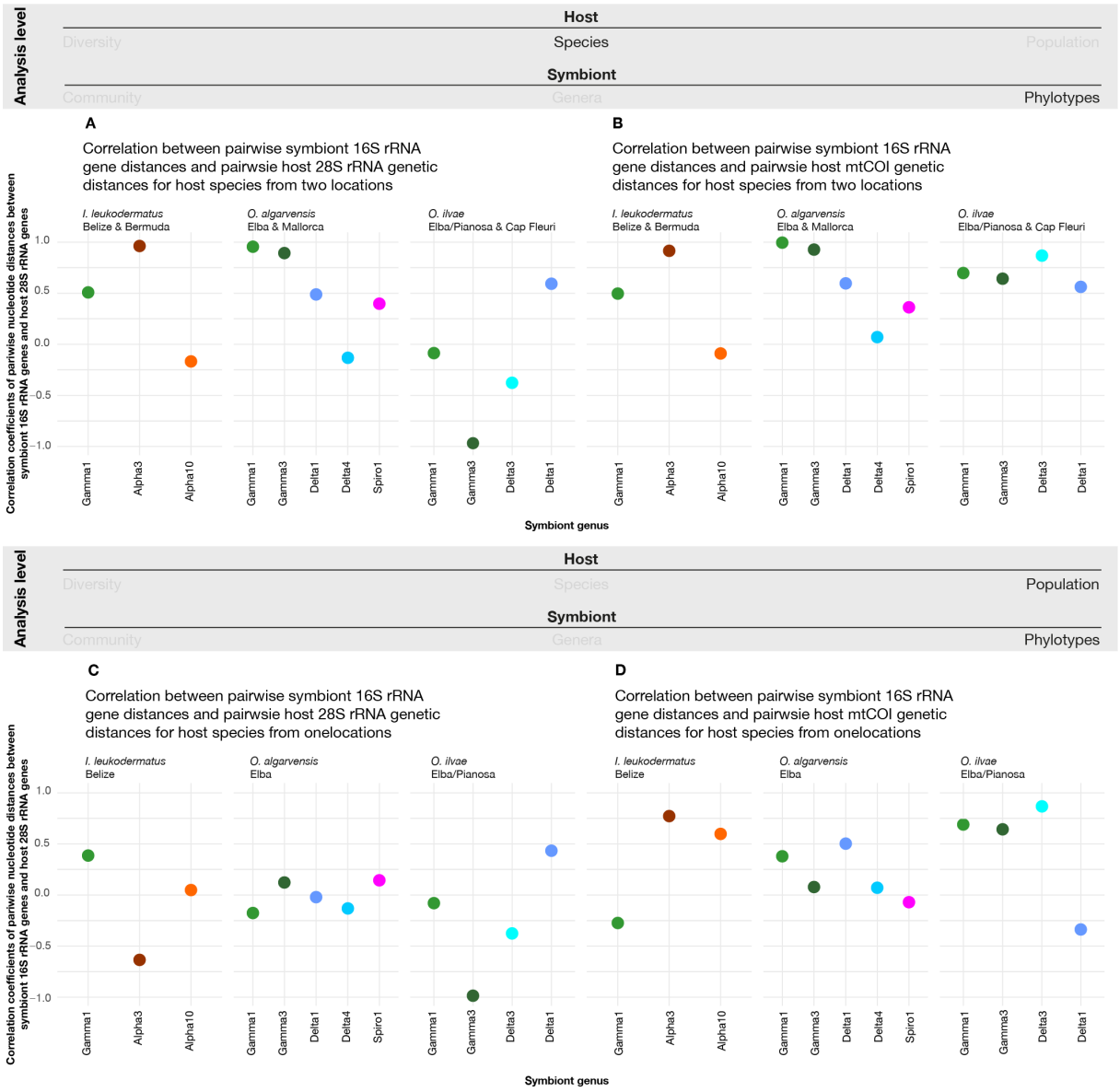


Figure 5 | The fidelity of symbiont genus - host species pairings varies across and within host populations. All panels show the degree of symbiont fidelity between a certain host species and its associated symbiont genera as correlation coefficient estimated by the Mantel test. **(A & B)** Correlation coefficients of 16S rRNA gene distance of different symbiont genera from different host species and the genetic distance of host marker genes of host individuals the symbionts were associated with. **(C & D)** Correlation coefficients of 16S rRNA gene distance of different symbiont genera and gene distance of host marker genes from the three host species where most individuals were sequenced of from one sampling site. **(A & C)** Host marker gene: 28S rRNA. **(B & D)** Host marker gene: mtCOI.

The observed differences in the symbiont fidelity of the same symbiont genus - host species pairing for the nuclear and mitochondrial host genes indicate that high levels of fidelity can be achieved via different means. The observed correlation of symbiont genetic distances to the mitochondrial genetic distances for some symbiont genus - host species pairings suggest that in

some populations high symbiont fidelity appears to be ensured via vertical transmission from mother to offspring. This apparent vertical transmission does not imply long term stability as we did not observe fidelity between symbiont phylotypes of a given symbiont genus and mitochondrial host genotypes across host species. In other populations, high symbiont fidelity appears to be selected for by nuclear host traits as shown by examples of strong correlation between symbiont genetic distance and host nuclear genetic distances. Similar to the linkage patterns between symbiont phylotype and host mitochondrial genotype, the linkage between symbiont phylotypes and host nuclear genotype are also not stable across host species.

Discussion

Microevolutionary patterns explain macroevolutionary stability of symbiont variability

Taken together, the symbiont - host specificity patterns in gutless oligochaetes across evolutionary scales indicate that the symbiont communities are characterized by versatility rather than stringent specificity. Our findings provide a possible answer to one of the critical questions in mutualism evolution: how can genetic variation be maintained in the evolution of mutualistic communities? In gutless oligochaetes, genetic variation is maintained via varying levels of symbiont fidelity. Community analyses that do not operate at a phylotypes level, but are instead limited to e.g. the genus-level would underestimate the versatility of symbioses such as the gutless oligochaete model, as switches of symbiont phylotypes of a given genus are frequent but would not alter genus-level community composition when they happen across and within host species. The observed phylotype switches at the microevolutionary level, that is within host populations, suggest that macroevolutionary patterns are based on population-level flexibility. This population level flexibility also varies between each symbiont genus - host species pairing. In fact, our results from gutless oligochaetes illustrate how important these host – symbiont pairings can be, as our observed symbiont community compositions are the result of the evolutionary dynamics of its single community members. Understanding the dynamics

between host species and symbiont genera is only possible when a broad host diversity is analyzed. Given such a broad taxon sampling and a sufficient phylogenetic resolution, we can start to link microevolutionary patterns of symbiont fidelity of individual symbiont genera in a given host species and macroevolutionary patterns of symbiont community composition across the host diversity.

Symbioses that are only characterized by high symbiont fidelity and specificity generally lead to genome streamlining of the symbionts that often results in deleterious genome reductions and the decay of the symbiotic association[30–38]. At the genomic level, a departure from this one-way from the 'cradle to the grave' scenario was suggested by Russel *et al.* (2020) who showed that symbionts with low fidelity in the form of frequent phylotype exchanges between host individuals have higher homologous recombination rates[39]. These recombinations prevent massive genome erosion and keep symbiont genomes 'forever young' when compared to symbiont genomes of high fidelity symbioses of a similar age[39]. Our results extend this concept to the community level. In gutless oligochaetes, symbiont fidelity varies on an even broader scale. We observed phylotype switches within host populations but also symbiont acquisition, loss and host switching that lead to genus-level variation in symbiont communities across a broad host diversity. Besides homologous recombination within a single symbiont clade described by Russell *et al.* (2020), the observed community versatility could also allow for constant variation in the pool of metabolic functions encoded in a given symbiont community and at the same time likely prevents the loss of key functions[39]. We therefore argue that this level of varying symbiont fidelity might not only keep the genomes of the symbionts 'forever young' but also could enable 'forever young' symbiont communities. Such 'forever young' symbiont communities could promote long-term stability of the association by providing ecological benefits such as a broader spectrum of environmental resources that can be metabolized by a single host individuum, niche partitioning between different host individuals

with different sets of symbionts and adaptability to quickly changing environmental conditions[5,9,40,41].

Material and methods

Sample collection, processing and metagenomic sequencing

231 individuals of gutless and 10 individuals of gut-bearing oligochaetes were sampled between 1991 and 2018 at sites around the world (for overview see Table S4). Individual specimens were either flash-frozen in liquid nitrogen and stored at -80°C or fixed in RNAlater (Thermo Fisher Scientific, Waltham, MA, USA) and stored at 4°C or -20°C. DNA was extracted from single worm individuals with the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Library construction, quality control and sequencing were performed at the DOE JGI (Walnut Creek, California, USA) and the Max Planck Genome Centre (Cologne, Germany). Information on library preparation and sequencing details are listed in Table S5. Some samples were sequenced twice to generate a higher number of reads. In these cases, resulting reads from both sequencing runs were combined for further analyses. Also, two samples were extracted twice using different library preparation methods and individually sequenced. The resulting sequences were also pooled for further analyses.

Assembly of host marker genes

28S rRNA and mtCOI genes of all specimens were assembled by mapping the metagenomic reads to respective databases using bbmap v38.34 (https://sourceforge.net/projects/bbmap/). For the 28S rRNA, we used the SILVA database v138[42,43]. Mapped reads were assembled using SPAdes v3.11.0 setting k-mer sizes to 99, 111 and 127 bp[44]. Final 28S rRNA gene sequences were predicted from the assembled sequences using barrnap v0.9 (https://github.com/tseemann/barrnap). MtCOI genes were assembled by adapting the phyloFlash pipeline to operate on a custom mtCOI reference database and predict mtCOI genes

from assembled sequences[45]. In case that several mtCOI genes were assembled, we only considered the most abundant one.

Identification of host taxa

Species-level host taxa were defined based on mtCOI gene phylogenies that also included the gene sequences of previously identified specimens. Specimens that could not be assigned to published species based on morphological or molecular data were treated as new taxa and were assigned provisional names with consecutive numbers and the sampling location. The gutbearing oligochaete specimens could be identified based on morphological traits.

Host marker gene phylogenies

28S rRNA and mtCOI gene sequences were aligned using mafft-linsi v7.407[46–48] . The mtCOI alignment was manually trimmed in Geneious v11.1.5 and bases 40-695 were kept (https://www.geneious.com). The best suited model for the Bayesian inference based phylogeny was estimated using the MODELTEST function of iqtree v1.6.10[49]. Bayesian inference based phylogenies were calculated using MrBayes 3.2.7a, using 4 chains, running for 4,000,000 generations and applying the GTR+G+I model[50,51]. The sample frequency was set to 1000 and the print frequency was set to 500. 1,000 trees were discarded as initial burn-in. All estimated parameters were controlled to show an effective sampling size (ESS) > 200 in Tracer v1.7.1[52]. Maximum-likelihood based phylogenies were calculated using iqtree, including automatic selection of the best suited model and generation of 100 none-parametric bootstrap replicates. The sequences of one gutbearing oligochaete specimen (*Phallodrilinae* gen. sp. ‘strang’) were used to root the phylogenies. The original tree was calculated on the full alignment, subtrees that were used in subsequent analyses were obtained by manually pruning the tree in iTOL[53]. To compare the 28S rRNA and the mtCOI gene based host phylogenies, we calculated the number of shared nodes between the maximum-likelihood trees calculated for each of the genes using the R package ‘ape’[54].

Symbiont clade definition and quantification

16S rRNA genes were assembled from the metagenomic libraries of gutless oligochaetes using phyloFlash. using the –all option and in addition specifying the read length. For subsequent analyses, we only considered sequences that were i) assembled with SPAdes, ii) longer than 1000 bp and iii) did not contain more than 20 ambiguous bases. The resulting sequences were clustered at 95% sequence similarity using usearch v10.0.240[55]. We used the SINA search and classify algorithm to add the 16S rRNA gene sequences of close relatives from the SILVA database v132 that shared at least 90% sequence similarity for each of our assembled symbiont sequences[56]. All assembled sequences and the SILVA database hits were aligned using mafft-linsi and a phylogenetic tree was calculated from the resulting alignment using FastTree v2.1.1[57]. We mapped the 95% clusters to this tree and manually merged monophyletic clades that consisted of several of the 95% clusters into single symbiont clades. We analyzed the prevalence of all phylogenetically defined symbiont clades across the gutless oligochaete metagenomic libraries. We excluded clades that had distribution patterns that suggested they were contaminations or spurious associations (Note S1). The abundances of the remaining clades (symbiont genera from here on) were quantified across all metagenomic libraries using EMIRGE v.0.61.1 following the standard workflow for custom EMIRGE databases[58].

Phylogeny of all symbionts and their relatives

All sequences included in the symbiont genera defined above were used to obtain sequences from closely related bacteria from the SILVA and the RefSeq public databases[59]. For SILVA, we used the SINA search and classify algorithm to obtain up to 10 relatives for each sequence that shared at least 99% and 95% sequence similarity for each of our input sequences. In addition, we also screened the RefSeq database using BLAST implemented in Geneious v11.1.5 to obtain the ten most similar 16S rRNA genes[60]. Duplicated sequences were removed from the collection of sequences of the symbionts' relatives. In addition, we included the 16S rRNA gene sequence of *Crenarchaeotal* sp. clone JP41 (NCBI accession: L25301.1) as outgroup. The

resulting sequence collection was aligned using mafft-linsi and a phylogenetic tree was calculated using iqtree including automatic selection of the best suited model and generation of 100 none-parametric bootstrap replicates. Subtrees that were used in further analyses were pruned from the resulting tree using iTOL.

Individual symbiont genera phylogenies

For the calculation of trees of individual symbiont genera, the symbiont 16S rRNA gene sequences of each genus were treated individually. We used the SINA search and classify algorithm to obtain up to 10 relatives that shared at least 90% sequence similarity for each of the input sequences from the SILVA database v138.1. In case of the Gamma7 and the Alpha9 symbiont genera, we did not obtain any relative sequence at this threshold. For these genera, we obtained up to 10 relatives that shared at least 85% sequence similarity for each input sequence instead. In addition, we clustered the symbiont sequences at 98% sequence similarity using the cluster_fast algorithm of usearch. We used the resulting centroids of every symbiont genus to obtain the five most similar 16S rRNA gene sequences from the RefSeq database using BLAST implemented in Geneious v11.1.15. Duplicated sequences were removed from the collection of sequences of the symbionts' relatives. In addition, we included the 16S rRNA gene sequence of *Crenarchaeotal* sp. clone JP41 (NCBI accession: L25301.1) as outgroup. The resulting sets of sequences of each symbiont genus were aligned using mafft-linsi. A maximum-likelihood phylogeny was calculated using iqtree including automatic selection of the best suited model and generation of 100 none-parametric bootstrap replicates.

Estimates of divergence times for host and symbiont clades and reconstruction of ancestral states of symbiont association patterns

For the estimation of host divergence times, we used a Bayesian phylogenetic framework and a relaxed molecular clock model. We constructed a matrix of eight 28S rRNA gene sequences of gutless host and eight publicly available 28S rRNA gene sequences of other *Oligochaeta* and one representative of the *Polychaeta*. The oligochaete representatives were selected to i) cover

a broad diversity of the phylum and ii) to include the following calibration points for our molecular clock model: the last common ancestor of the *Goniadidae* (323 Mya), the last common ancestor of the *Hormogastridae* (82 ± 15 Mya), the divergence between *Hirudienea* and *Lumbriculidae* (201 Mya) and the last common ancestor of the *Phyllodocida* (4.85 ± 1.9)[61,62]. The polychaete sequence was included to root the tree and to include the last common ancestor of all *Annelida* (510 ± 10 Mya) as additional calibration point. All calibration points were considered as uniform priors. All sequences were aligned using mafft-linsi and the time calibrated tree was calculated in BEAST v2.6.3[63] using the GTR+G+I model and the relaxed log normal clock model. Besides the mentioned priors for time calibration, we also set Alpha and Beta values of birtRate.Y.t prior to 0.001 and 1,000, respectively. We also defined priors that considered the oligochaetes and the gutless oligochaetes as monophyletic groups. The chain length was set to 100,000,000 generations and a 10% burn-in was defined. All estimated parameters were controlled to show an ESS > 200 in Tracer.

We used the same approach as for the host to estimate the divergence times for symbiont genera based on a subset of our symbiont sequence matrix that combined 2-3 symbiont 16S rRNA gene sequences per symbiont genus with 50 typestrain sequences from RefSeq databases. Typestrains of the *Chromatiaceae* were used to include their previously published divergence estimate as calibration point for the symbiont analysis[64]. In addition, we included the 16S rRNA gene sequence of *Crenarchaeotal* sp. clone JP41 (NCBI accession: L25301.1) as outgroup. The time calibrated tree was calculated in BEAST v2.6.3, using the GTR+G+I model and the relaxed log normal clock model. The divergence time of the *Chromatiaceae* was considered as an exponential prior with a mean value of 0.1 and an offset of 1.64. We additionally constrained the analyses by setting a uniform prior from 3.5-4.5 billion years for the whole dataset to account for the maximum age of life on earth. Additional priors were used to define monophyletic clades for all bacteria, the Delta1, Delta4 and Delta12 genera as well the combined Delta4-Delta12 clade that were observed in the previous phylogenetic analyses of

the symbiont genera. We also set Alpha and Beta values of birtRate.Y.t prior to 0.001 and 1,000, respectively. We ran 4 parallel chains, setting the chain length to 500,000,000 generation and a 10% burn in was defined. All estimated parameters were controlled to show an ESS > 200 in Tracer.

Ancestral states of symbiont presence/absence patterns were performed using a maximum-parsimony based last common ancestor analysis with pastml, using the DOWNPASS prediction method[65].

Analyses and plotting of symbiont community composition

The analyses of symbiont community composition were performed in R v3.6.3 unless differently stated (R Core Team, 2020)[66]. During the analyses, the following packages were used: phyloseq[67], ape, vegan (https://github.com/vegandevs/vegan), plyr[68], MASS[69]
gdata (https://cran.r-project.org/web/packages/gdata/index.html),
reshape2 (https://github.com/hadley/reshape),
forcats (https://github.com/robjhyndman/forecast),
igraph (https://github.com/igraph/rigraph), Hmisc (https://github.com/harrelfe/Hmisc/),
Optparse (https://github.com/trevorld/r-optparse),
data.table (https://github.com/Rdatatable/data.table), ade4 (https://github.com/sdray/ade4)
tidyverse[70] and spa (https://github.com/markvanderloo/rspa).

Plots were generated using ggplot2 from the tidyverse package, gridExtra (https://cran.r-project.org/web/packages/gridExtra/index.html), ggpubr (https://cran.r-project.org/web/packages/ggpubr/index.html), maps (https://www.rdocumentation.org/packages/maps), mapdata (https://www.rdocumentation.org/packages/mapdata) and patchwork (https://github.com/thomasp85/patchwork).

Community composition analyses

The similarity between symbiont communities of host individuals were calculated based on the abundance patterns of the symbiont genera and the symbiont 16S rRNA gene phylogeny using

the UniFrac metric as implemented in the phyloseq package in R. We tested for parameters that could explain differences in symbiont community composition between individuals using PERMANOVA and the Mantel test[71,72]. These parameters included host species, geographic distance, ocean, organic input and sediment type (Table S4). All factors except for geographical distances were treated as categorical data and analyzed using PERMANOVA. Geographical distance was treated as the correlation between the UniFrac distances and actual geographic distances and analyzed using the Mantel test.

Co-occurrence patterns of symbiont genera were analyzed using Spearman's correlations and were corrected using the Benjamini-Hochberg standard false discovery rate correction.

Phylosymbiosis

UniFrac distances on the average symbiont abundances per host species were transformed into a dendrogram using hierarchical clustering. The congruences between the 28S rRNA gene and the mtCOI gene based host tree or and the symbiont community UniFrac dendrogram was assessed separately using the normalized Robinson-Fould metric and the normalized Matching Cluster metric, implemented in TreeCmp v1.0-b291[73–75]. Statistical significance was estimated by comparing the congruence between the host phylogeny vs. 1000 random trees as described by Brooks *et al.* 2016, https://github.com/awbrooks19/phylosymbiosis)[76]. The relation between host phylogenetic distances and the symbiont community composition distances was analyzed using linear regression and the Mantel test.

Correlation between host and symbiont phylogenetic distances

For all hosts with member sequences of a given symbiont genus we calculated pairwise phylogenetic distances of the hosts' 28S rRNA or the mtCOI genes as well as pairwise distances of the 16S rRNA gene sequences using from the respective phylogenies using the R's cophenetic function. We analyzed the correlation between the host and symbiont genetic distances using linear regression and the Mantel test.

Data availability

Raw metagenomic sequences as well as host and symbiont marker genes generated in this study will be deposited in the European Nucleotide Archive (ENA) upon peer-review submission and are currently available upon request.

Code availability

The scripts and data for analyzing symbiont community composition and phylogenetic correlations are available under: https://github.com/amankowski/GO_symbiont-diversity

Acknowledgments

We are thankful for sample collections and field assistance by Alexander Gruhl, Anna Ansebo, Anna Blazejak, Anne-Christin Kreutzmann, Christian Lott, Claudia Bergin, Dolma M. Michellod, Emilia M. Sogin, Erica Mejlon, Erich Mueller, Falk Warnecke, Fred Wells, Jörg Ott, Judith Zimmermann, Katrine Worsaae, Ken Halanych, K. B. Brandon Seah, Lisa Matamoros, Lena Gustavsson, Mario P. Schimak, Michael Hadfield, Miriam Sadowski, Miriam Weber, Olav Giere, Oliver Jäckle, Nicholas Bekkouche, Olivier Gros, Pamela Reid, Philippe Bouchet, Pierre De Wit, Ramon Rosello-Mora, Silke Wetzel, Silvia Bulgheresi, Stefan Sommer, and Tina Enders. In addition, we would like to thank the crew of the Meteor cruise M92 and the Sonne cruise SO147, as well as the Carrie Bow Cay Laboratory, the Heron Island Research Station, the HYDRA Institute Elba, the Mediterranean Institute for Advanced Studies, the Lee Stocking Island Research Station, the Little Darby Island Research Station, the Lizard Island Research Station, and the Okinawa Institute of Science and Technology and their staff for supporting our sampling campaigns. This work was supported by the Max Planck Society, a Moore Foundation Marine Microbial Initiative Investigator Award to ND (Grant GBMF3811), a U.S. National Science Foundation award to MK (grant IOS 2003107), the USDA National Institute of Food and Agriculture Hatch project 1014212 (MK), and a Marie-Curie Intra-European Fellowship PIEF-GA-2011-301027 CARISYM (HRGV). The work

conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231. This work is contribution XXX from the Carrie Bow Cay Laboratory, Caribbean Coral Reef Ecosystem Program, National Museum of History, Washington DC.

References

1. Leigh Jr, E. G. The evolution of mutualism. *J. Evol. Biol.* **23**, 2507–2528 (2010).
2. Archetti, M. *et al.* Economic game theory for mutualism and cooperation. *Ecol. Lett.* **14**, 1300–1312 (2011).
3. Sachs, J. L., Mueller, U. G., Wilcox, T. P. & Bull, J. J. The Evolution of Cooperation. *Q. Rev. Biol.* **79**, 135–160 (2004).
4. Ellegaard, K. M. & Engel, P. Genomic diversity landscape of the honey bee gut microbiota. *Nat. Commun.* **10**, 446 (2019).
5. Ansorge, R. *et al.* Functional diversity enables multiple symbiont strains to coexist in deep-sea mussels. *Nat. Microbiol.* **4**, 2487–2497 (2019).
6. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
7. Dubilier, N. *et al.* Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. *Nature* **411**, 298–302 (2001).
8. Woyke, T. *et al.* Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**, 950–955 (2006).
9. Kleiner, M. *et al.* Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E1173--82 (2012).
10. Dubilier, N. *et al.* Phylogenetic diversity of bacterial endosymbionts in the gutless marine oligochete *Olavius loisae* (Annelida). *Mar. Ecol. Prog. Ser.* **178**, 271–280 (1999).
11. Dubilier, N., Giere, O., Distel, D. L. & Cavanaugh, C. M. Characterization of chemoautotrophic bacterial symbionts in a gutless marine worm (Oligochaeta, Annelida) by phylogenetic 16S rRNA sequence analysis and in situ hybridization. *Appl. Environ. Microbiol.* **61**, 2346–2350 (1995).
12. Blazejak, A., Erseus, C. & Amann Rudolf & Dubilier, N. Coexistence of bacterial sulfide oxidizers, sulfate reducers, and spirochetes in a gutless worm (Oligochaeta) from the Peru margin. *Appl. Environ. Microbiol.* **71**, 1553–1561 (2005).
13. Ruehland, C., Blazejak, A., Lott, C., Loy, A. & Erséus Christer & Dubilier, N. Multiple bacterial symbionts in two species of co-occurring gutless oligochaete worms from Mediterranean sea grass sediments. *Environ. Microbiol.* **10**, 3404–3416 (2008).

14. Blazejak, A., Kuever, J., Erséus, C., Amann, R. & Dubilier, N. Phylogeny of 16S rRNA, Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase, and Adenosine 5′-Phosphosulfate Reductase Genes from Gamma- and Alphaproteobacterial Symbionts in Gutless Marine Worms (Oligochaeta) from Bermuda and the Bahamas. *Appl. Environ. Microbiol.* **72**, 5527–5536 (2006).
15. Bergin, C. *et al.* Acquisition of a Novel Sulfur-Oxidizing Symbiont in the Gutless Marine Worm *Inanidrilus exumae*. *Appl. Environ. Microbiol.* **84**, e02267-17 (2018).
16. Zimmermann, J. *et al.* Closely coupled evolutionary history of ecto- and endosymbionts from two distantly related animal phyla. *Mol. Ecol.* **25**, 3203–3223 (2016).
17. Sato, Y. *et al.* Fidelity varies in the symbiosis between a gutless marine worm and its microbial consortium. *bioRxiv* 2021.01.30.428904 (2021) doi:10.1101/2021.01.30.428904.
18. Ting, N., Tosi, A. J., Li, Y., Zhang, Y.-P. & Disotell, T. R. Phylogenetic incongruence between nuclear and mitochondrial markers in the Asian colobines and the evolution of the langurs and leaf monkeys. *Mol. Phylogenet. Evol.* **46**, 466–474 (2008).
19. Phillips, M. J., Haouchar, D., Pratt, R. C., Gibb, G. C. & Bunce, M. Inferring Kangaroo Phylogeny from Incongruent Nuclear and Mitochondrial Genes. *PLoS One* **8**, e57745 (2013).
20. Cong, Q. *et al.* Complete genomes of Hairstreak butterflies, their speciation and nucleo-mitochondrial incongruence. *Sci. Rep.* **6**, 24863 (2016).
21. Sota, T. & Vogler, A. P. Incongruence of Mitochondrial and Nuclear Gene Trees in the Carabid Beetles *Ohomopterus*. *Syst. Biol.* **50**, 39–59 (2001).
22. Frank, S. A. & Hurst, L. D. Mitochondria and male disease. *Nature* **383**, 224 (1996).
23. Bright Monika & Bulgheresi, S. A complex journey: transmission of microbial symbionts. *Nat. Rev. Microbiol.* **8**, 218–230 (2010).
24. Nakabachi, A. *et al.* Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, *Buchnera*. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 5477 LP – 5482 (2005).
25. Login, F. H. *et al.* Antimicrobial Peptides Keep Insect Endosymbionts Under Control. *Science (80-.).* **334**, 362 LP – 365 (2011).
26. McFall-Ngai, M. Care for the community. *Nature* **445**, 153 (2007).
27. Anselme, C. *et al.* Identification of the Weevil immune genes and their expression in the bacteriome tissue. *BMC Biol.* **6**, 43 (2008).

28. Franzenburg, S. *et al.* Distinct antimicrobial peptide expression determines host species-specific bacterial associations. *Proc. Natl. Acad. Sci.* **110**, E3730 LP-E3738 (2013).
29. Fisher, R. M., Henry, L. M., Cornwallis, C. K., Kiers, E. T. & West, S. A. The evolution of host-symbiont dependence. *Nat. Commun.* **8**, 15973 (2017).
30. Bennett, G. M. & Moran, N. A. Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. *Proc. Natl. Acad. Sci.* **112**, 10169 LP – 10176 (2015).
31. Moran, N. A. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci.* **93**, 2873 LP – 2878 (1996).
32. Rispe, C. & Moran, N. A. Accumulation of Deleterious Mutations in Endosymbionts: Muller's Ratchet with Two Levels of Selection. *Am. Nat.* **156**, 425–441 (2000).
33. Hosokawa, T., Kikuchi, Y., Nikoh, N., Shimada, M. & Fukatsu, T. Strict Host-Symbiont Cospeciation and Reductive Genome Evolution in Insect Gut Bacteria. *PLOS Biol.* **4**, e337 (2006).
34. McCutcheon, J. P. & Moran, N. A. Functional Convergence in Reduced Genomes of Bacterial Symbionts Spanning 200 My of Evolution. *Genome Biol. Evol.* **2**, 708–718 (2010).
35. McCutcheon, J. P. & Moran, N. A. Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**, 13–26 (2012).
36. Sloan, D. B. & Moran, N. A. Genome Reduction and Co-evolution between the Primary and Secondary Bacterial Symbionts of Psyllids. *Mol. Biol. Evol.* **29**, 3781–3792 (2012).
37. Jäckle, O. *et al.* Chemosynthetic symbiont with a drastically reduced genome serves as primary energy storage in the marine flatworm *Paracatenula*. *Proc. Natl. Acad. Sci.* **116**, 8505 LP – 8514 (2019).
38. McCutcheon, J. P., Boyd, B. M. & Dale, C. The Life of an Insect Endosymbiont from the Cradle to the Grave. *Curr. Biol.* **29**, R485–R495 (2019).
39. Russell, S. L. *et al.* Horizontal transmission and recombination maintain forever young bacterial symbiont genomes. *PLOS Genet.* **16**, e1008935 (2020).
40. Feldhaar, H. Bacterial symbionts as mediators of ecologically important traits of insect hosts. *Ecol. Entomol.* **36**, 533–543 (2011).
41. Kleiner, M. *et al.* Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. *Proc. Natl. Acad. Sci.* **115**, E5576 LP-E5584 (2018).

42. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590--6 (2013).
43. Yilmaz, P. *et al.* The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* **42**, D643–D648 (2014).
44. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
45. Gruber-Vodicka, H. R., Seah, B. K. B. & Pruesse, E. phyloFlash: Rapid Small-Subunit rRNA Profiling and Targeted Assembly from Metagenomes. *mSystems* **5**, e00920-20 (2020).
46. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).
47. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
48. 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.* **30**, 3059–3066 (2002).
49. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
50. Altekar, G., Dwarkadas, S., Huelsenbeck, J. P. & Ronquist, F. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* **20**, 407–415 (2004).
51. Ronquist, F. *et al.* MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst. Biol.* **61**, 539–542 (2012).
52. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
53. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).
54. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
55. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).

56. Pruesse, E. & Peplies Jörg & Glöckner, F. O. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. 1823–1829 (2012).
57. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* **5**, e9490 (2010).
58. Miller, C. S., Baker, B. J., Thomas, B. C. & Singer Steven W. & Banfield, J. F. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol.* **12**, R44 (2011).
59. Pruitt, K. D. & Maglott, D. R. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.* **29**, 137–140 (2001).
60. Altschul, S. F., Gish, W., Miller, W. & Myers E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
61. Anderson, F. E. *et al.* Phylogenomic analyses of Crassiclitellata support major Northern and Southern Hemisphere clades and a Pangaean origin for earthworms. *BMC Evol. Biol.* **17**, 123 (2017).
62. Verdes, A. *et al.* Molecular phylogeny of *Odontosyllis* (Annelida, Syllidae): A recent and rapid radiation of marine bioluminescent worms. *bioRxiv* 241570 (2018) doi:10.1101/241570.
63. Bouckaert, R. *et al.* BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLOS Comput. Biol.* **10**, e1003537 (2014).
64. Hugoson, E., Ammunét, T. & Guy, L. Host-adaptation in *Legionellales* is 2.4 Ga, coincident with eukaryogenesis. *bioRxiv* 852004 (2020) doi:10.1101/852004.
65. Ishikawa, S. A., Zhukova, A., Iwasaki, W. & Gascuel, O. A Fast Likelihood Method to Reconstruct and Visualize Ancestral Scenarios. *Mol. Biol. Evol.* **36**, 2069–2085 (2019).
66. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
67. Bartram, A. K. *et al.* Exploring links between pH and bacterial community composition in soils from the Craibstone Experimental Farm. *FEMS Microbiol. Ecol.* **87**, 403–415 (2014).
68. Wickham, H. The Split-Apply-Combine Strategy for Data Analysis. *J. Stat. Softw.* **40**, (2011).
69. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S*. (Springer, 2002).
70. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
71. Mantel, N. The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Res.* **27**, 209 LP – 220 (1967).

72. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**, 32–46 (2001).
73. Robinson, D. F. & Foulds, L. R. Comparison of phylogenetic trees. *Math. Biosci.* **53**, 131–147 (1981).
74. Bogdanowicz, D., Giaro, K. & Wróbel, B. TreeCmp: Comparison of Trees in Polynomial Time. *Evol. Bioinforma.* **8**, EBO.S9657 (2012).
75. Bogdanowicz, D. & Giaro, K. On a matching distance between rooted phylogenetic trees. *Int. J. Appl. Math. Comput. Sci.* **23**, 669–684.
76. Brooks, A. W., Kohl, K. D., Brucker, R. M., van Opstal, E. J. & Bordenstein, S. R. Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. *PLOS Biol.* **14**, e2000225 (2016).

Chapter III | Supplementary material

Figures

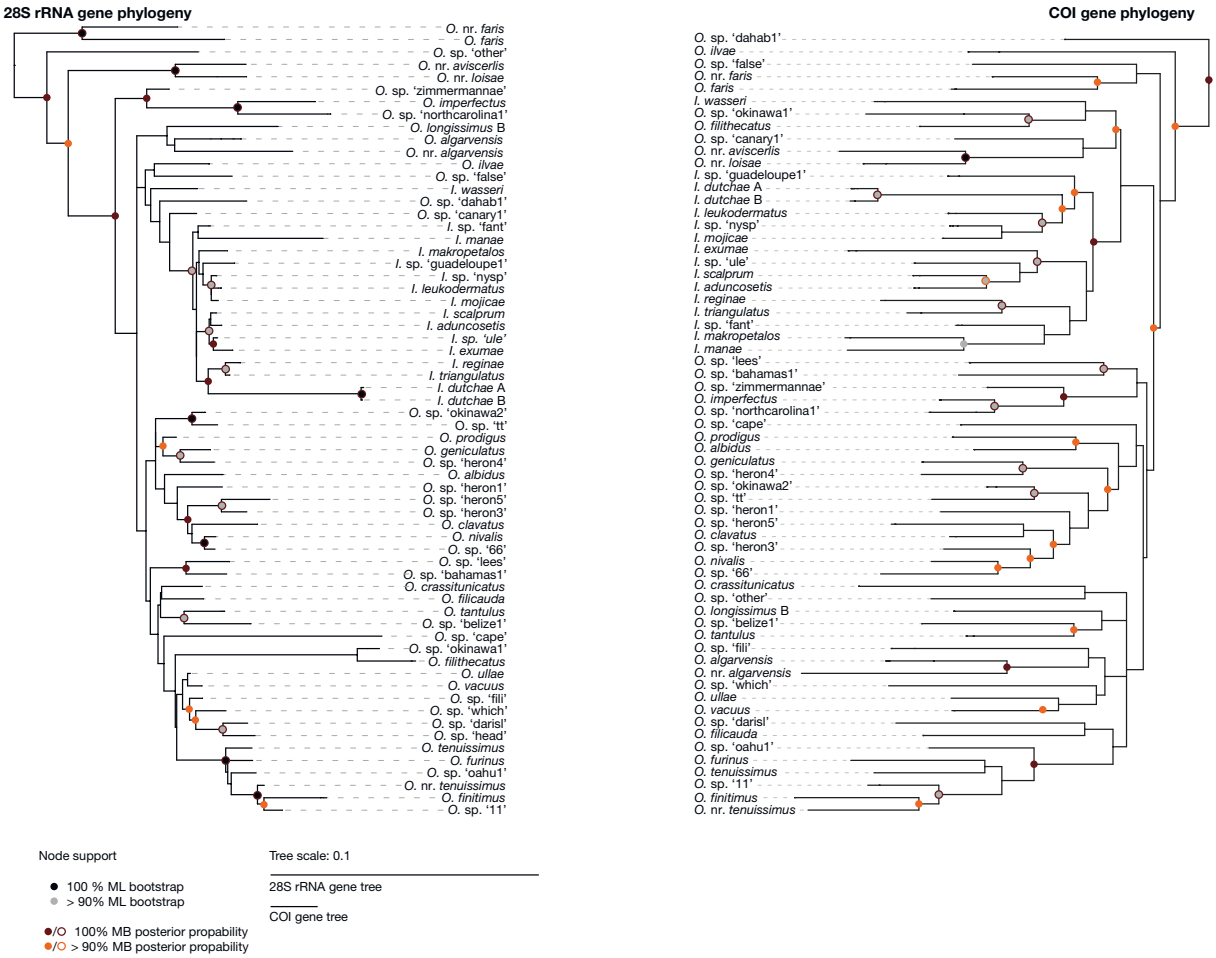


Figure S1 | Nuclear and mitochondrial genomes of gutless oligochaetes evolved differently. Phylogenies of host species shown as maximum-likelihood tree of one representative sequence per host species. Left tree: 28S rRNA, right tree: mtCOI. Nodes with none-parametric bootstrap support > 90% and/or posterior probability > 90% estimated with MrBayes are highlighted. The scale bar indicates 10% estimated sequence divergence.

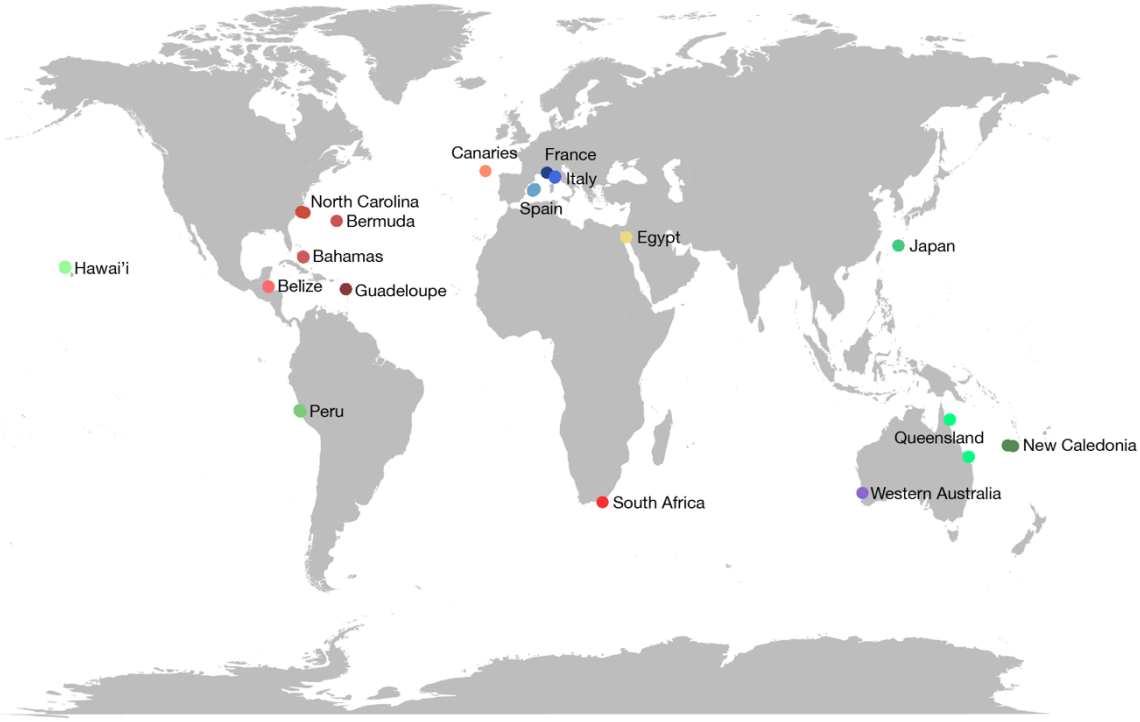


Figure S2 | World map highlighting the field sites where gutless oligochaetes for this study were sampled.

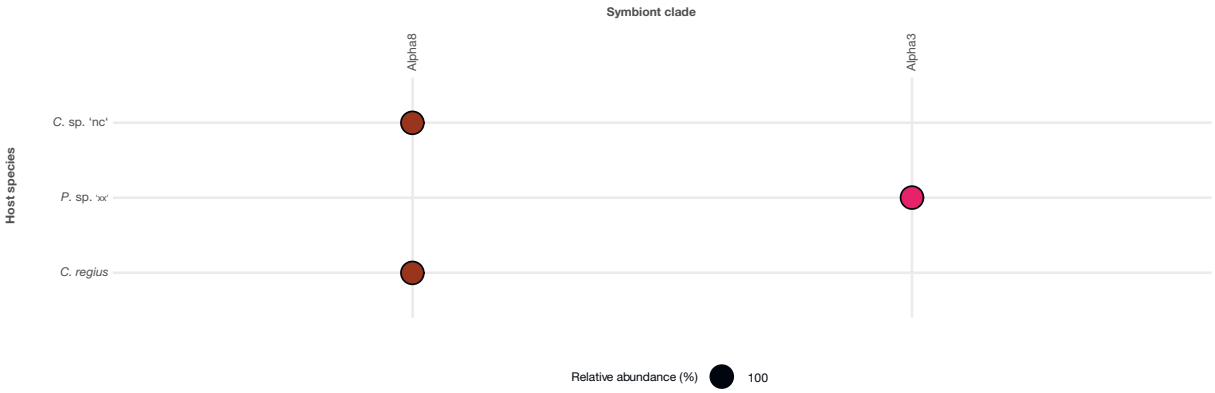


Figure S3 | Abundance plot for gut-bearing relatives: Relative abundance of symbiont clades in individuals of gutbearing oligochaetes estimated with EMIRGE.

Figure S4-36 | Phylogenies of individual symbiont clades. Deposited under https://doi.org/10.5281/zenodo.5494733

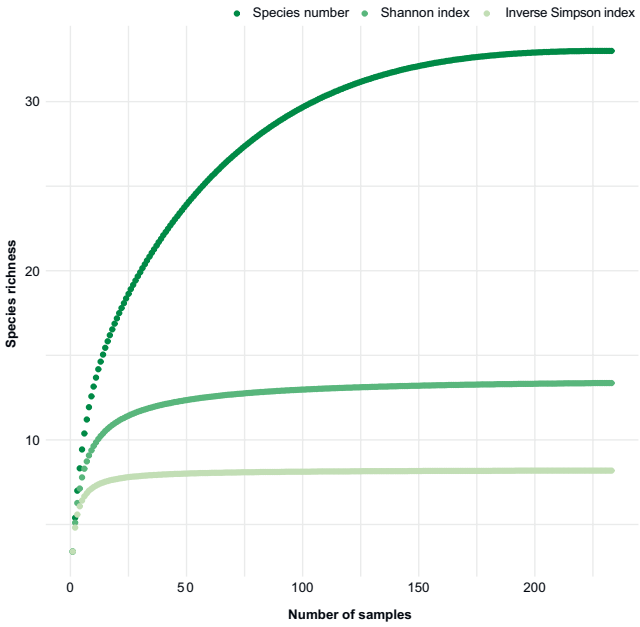


Figure S37 | Gutless oligochaete symbionts consist of a limited pool of bacterial taxa. Rarefaction analysis of the number of symbiont clades over the number of analyzed host individuals. Richness of symbiont clades ('Species richness') was analyzed as absolute number of clades as well as the Shannon Index and the Inverse Simson Index.

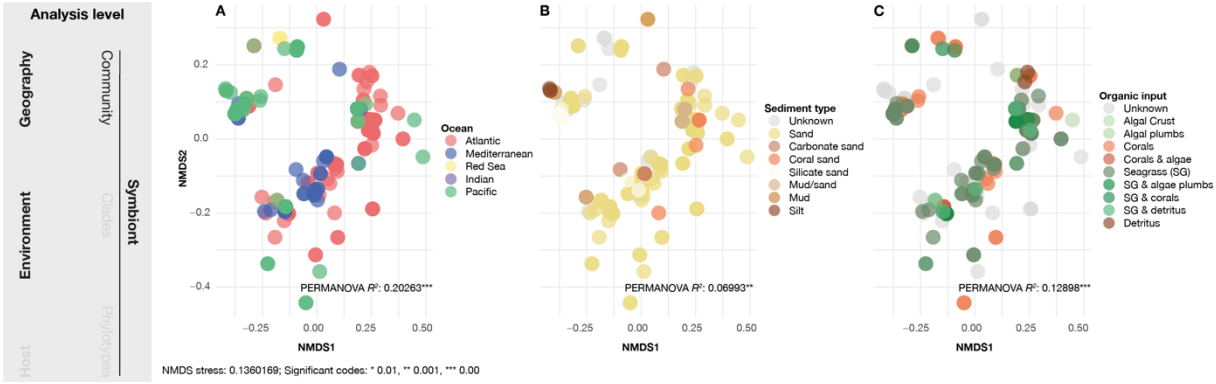


Figure S38 | Symbiont community composition is linked to host species and largely unlinked to environmental parameters. All panels show the same NMDS plot of UniFrac dissimilarity values calculated from estimated relative abundances of symbiont clades in different host individuals. UniFrac stress value: 0.136. Different panels highlight samples differently according to certain metadata categories. For each metadata category, PERMANOVA r^2 values are indicated within the respective plot. r^2 values that are printed in bold were statistically significant.

Tables

Table S1 | Taxonomy of symbiont clades

Symbiont clade	Phylum	Class	Order	Family
Actinomarinales1	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinobacteriota
Alpha1	Proteobacteria	Alphaproteobacteria		
Alpha2	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae
Alpha3	Proteobacteria	Alphaproteobacteria		
Alpha4	Proteobacteria	Alphaproteobacteria	Kiloniellales	Kiloniellaceae
Alpha5	Proteobacteria	Alphaproteobacteria	Defluviicoccales	
Alpha6	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Alpha7	Proteobacteria	Alphaproteobacteria		
Alpha8	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Magnetospiraceae
Alpha9	Proteobacteria	Alphaproteobacteria	Rhizobiales	
Alpha10	Proteobacteria	Alphaproteobacteria		
Alpha12	Proteobacteria	Alphaproteobacteria		
Alpha13	Proteobacteria	Alphaproteobacteria		
Alpha14	Proteobacteria	Alphaproteobacteria		
Alpha15	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae
Alpha16	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Delta1	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfosarcinaceae
Delta2	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfosarcinaceae
Delta3	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfosarcinaceae
Delta4	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfosarcinaceae
Delta5	Proteobacteria	Deltaproteobacteria	Desulfobulbales	Desulfocapsaceae
Delta11	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfosarcinaceae
Delta12	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfosarcinaceae
Delta13	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfosarcinaceae
Delta14	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfosarcinaceae
Gamma1	Proteobacteria	Gammaproteobacteria	Arenicellales	Arenicellaceae
Gamma2	Proteobacteria	Gammaproteobacteria		
Gamma3	Proteobacteria	Gammaproteobacteria		
Gamma4	Proteobacteria	Gammaproteobacteria	Nitrosococcales	Nitrosococcaceae
Gamma5	Proteobacteria	Gammaproteobacteria	Arenicellales	Arenicellaceae
Gamma7	Proteobacteria	Gammaproteobacteria		
Marinimicrobia1	Marinimicrobia (SAR406 clade)			
Spiro1	Spirochaetota	Spirochaetia	Spirochaetales	Spirochaetaceae

Table S2 | Relative symbiont abundances in gutless oligochaete samples

Library	Symbiont clade	Abundance	Library	Symbiont clade	Abundance
2699_C	Gamma1	0.641218	3557_AG	Alpha12	0.08301997
2699_C	Delta13	0.160981	3557_AH	Gamma1	0.568465
2699_C	Spiro1	0.092665	3557_AH	Alpha12	0.255548
2699_C	Gamma3	0.105136	3557_AH	Alpha10	0.126154
2699_D	Gamma1	0.782357	3557_AH	Alpha3	0.049833
2699_D	Delta13	0.104203	3557_AI	Gamma1	0.613349
2699_D	Gamma3	0.066595	3557_AI	Alpha10	0.141278
2699_D	Delta1	0.046845	3557_AI	Alpha12	0.124082
3557_A	Gamma1	0.890965	3557_AI	Alpha1	0.082197
3557_A	Delta3	0.074949	3557_AI	Delta1	0.039094
3557_A	Spiro1	0.034086	3557_AJ	Gamma1	0.557991
3557_AA	Gamma1	0.3555294	3557_AJ	Alpha12	0.351924
3557_AA	Alpha10	0.23962871	3557_AJ	Alpha10	0.044748
3557_AA	Alpha3	0.17918567	3557_AJ	Alpha1	0.045337
3557_AA	Alpha4	0.16457575	3557_AK	Gamma1	0.77290974
3557_AA	Alpha14	0.04320934	3557_AK	Delta3	0.22709026
3557_AA	Alpha14	0.01787113	3557_AL	Gamma1	0.44681255
3557_AB	Gamma1	0.431165	3557_AL	Alpha12	0.38999761
3557_AB	Alpha10	0.431164	3557_AL	Delta4	0.16318984
3557_AB	Alpha12	0.108244	3557_AN	Gamma1	0.837323
3557_AB	Delta1	0.029427	3557_AN	Delta3	0.162677
3557_AC	Alpha10	0.316886	3557_AO	Gamma1	0.67239867
3557_AC	Gamma1	0.270359	3557_AO	Alpha10	0.16794317
3557_AC	Delta1	0.253506	3557_AO	Alpha3	0.09279009
3557_AC	Alpha12	0.159249	3557_AO	Alpha3	0.06686807
3557_AD	Gamma1	0.39914226	3557_AQ	Gamma1	0.473295
3557_AD	Alpha10	0.32785646	3557_AQ	Alpha10	0.294919
3557_AD	Alpha7	0.10322539	3557_AQ	Alpha12	0.160793
3557_AD	Alpha5	0.09018731	3557_AQ	Delta3	0.070993
3557_AD	Alpha5	0.05028061	3557_AR	Gamma1	0.48711558
3557_AD	Delta1	0.02930798	3557_AR	Alpha12	0.40522742
3557_AE	Alpha10	0.450912	3557_AR	Delta3	0.107657
3557_AE	Gamma1	0.424716	3557_AS	Gamma1	0.701413
3557_AE	Alpha12	0.103131	3557_AS	Alpha12	0.186079
3557_AE	Alpha3	0.021241	3557_AS	Delta3	0.08217
3557_AF	Gamma1	0.58239113	3557_AS	Alpha10	0.030338
3557_AF	Alpha10	0.23386871	3557_AT	Gamma1	0.44497289
3557_AF	Alpha3	0.09252207	3557_AT	Alpha10	0.27421055
3557_AF	Alpha1	0.09121809	3557_AT	Alpha3	0.15863232
3557_AG	Gamma1	0.68629826	3557_AT	Alpha12	0.05717311
3557_AG	Alpha3	0.13741579	3557_AT	Alpha1	0.03736007
3557_AG	Alpha10	0.09326598	3557_AT	Alpha14	0.01905604

Library	Symbiont clade	Abundance	Library	Symbiont clade	Abundance
3557_AT	Alpha3	0.00667401	3557_BF	Alpha10	0.24531963
3557_AT	Alpha3	0.001921	3557_BF	Alpha3	0.19293963
3557_AU	Alpha10	0.256244	3557_BF	Alpha12	0.18013253
3557_AU	Gamma1	0.283576	3557_BF	Alpha5	0.02787817
3557_AU	Alpha12	0.178087	3557_BF	Alpha3	0.00120202
3557_AU	Alpha3	0.174815	3557_BG	Gamma1	0.629603
3557_AU	Alpha5	0.09085	3557_BG	Delta3	0.370397
3557_AU	Alpha5	0.015436	3557_BH	Alpha10	0.34987968
3557_AU	Alpha3	0.000992	3557_BH	Alpha12	0.3211833
3557_AV	Alpha10	0.373153	3557_BH	Gamma5	0.29559113
3557_AV	Gamma1	0.337907	3557_BH	Gamma1	0.03091965
3557_AV	Alpha12	0.256026	3557_BH	Alpha3	0.00242624
3557_AV	Alpha5	0.032914	3557_BI	Gamma1	0.547326
3557_AW	Alpha10	0.519752	3557_BI	Alpha10	0.452674
3557_AW	Gamma1	0.257483	3557_BJ	Alpha10	0.54832455
3557_AW	Alpha12	0.106771	3557_BJ	Gamma1	0.24234224
3557_AW	Alpha1	0.054636	3557_BJ	Alpha3	0.17332117
3557_AW	Delta3	0.061358	3557_BJ	Alpha1	0.03601204
3557_AX	Gamma1	0.47748948	3557_BK	Gamma1	0.416704
3557_AX	Alpha12	0.29215329	3557_BK	Alpha10	0.353274
3557_AX	Alpha10	0.23035723	3557_BK	Alpha5	0.173319
3557_AY	Gamma1	0.766844	3557_BK	Alpha7	0.056703
3557_AY	Delta3	0.233156	3557_BL	Alpha10	0.48206621
3557_AZ	Alpha10	0.62670469	3557_BL	Gamma1	0.19700003
3557_AZ	Gamma1	0.37329531	3557_BL	Alpha5	0.15515161
3557_B	Gamma1	0.812963	3557_BL	Alpha7	0.07770162
3557_B	Spiro1	0.187037	3557_BL	Delta1	0.04191416
3557_BA	Alpha10	0.534866	3557_BL	Gamma5	0.04616638
3557_BA	Gamma1	0.465134	3557_BM	Gamma1	0.823274
3557_BB	Gamma1	0.59876	3557_BM	Delta3	0.176726
3557_BB	Alpha10	0.201918	3557_BN	Gamma1	0.47651855
3557_BB	Alpha12	0.108311	3557_BN	Alpha12	0.28650124
3557_BB	Alpha3	0.091011	3557_BN	Alpha10	0.21195142
3557_BC	Alpha3	0.620573	3557_BN	Delta3	0.0250288
3557_BC	Gamma1	0.379427	3557_BO	Gamma1	0.698126
3557_BD	Gamma1	0.746597	3557_BO	Alpha3	0.301874
3557_BD	Delta3	0.253403	3557_BP	Gamma1	0.678521
3557_BE	Gamma1	0.62886863	3557_BP	Delta3	0.321479
3557_BE	Alpha10	0.17063117	3557_BQ	Alpha5	0.48469993
3557_BE	Alpha3	0.11582512	3557_BQ	Gamma1	0.32078562
3557_BE	Alpha12	0.08467508	3557_BQ	Delta1	0.1458589
3557_BF	Gamma1	0.35252802	3557_BQ	Alpha5	0.04865555
3557_BR	Gamma1	0.422146	3557_CF	Delta3	0.18634948

Library	Symbiont clade	Abundance	Library	Symbiont clade	Abundance
3557_BR	Delta1	0.194497	3557_CF	Delta1	0.09458037
3557_BR	Alpha5	0.193485	3557_CF	Alpha7	0.05878795
3557_BR	Alpha5	0.189872	3557_CF	Alpha5	0.06953517
3557_BS	Gamma1	0.796502	3557_CG	Gamma1	0.57703542
3557_BS	Alpha3	0.20347	3557_CG	Delta3	0.28792371
3557_BS	Alpha3	2.80E-05	3557_CG	Delta1	0.13504086
3557_BT	Gamma1	0.59584721	3557_CH	Delta4	0.348352
3557_BT	Delta3	0.23742652	3557_CH	Gamma3	0.258347
3557_BT	Alpha3	0.16672627	3557_CH	Spiro1	0.15794
3557_BU	Gamma1	0.83592126	3557_CH	Gamma1	0.161091
3557_BU	Alpha3	0.16407874	3557_CH	Delta1	0.07427
3557_BV	Gamma1	0.45105755	3557_CI	Gamma1	0.44570455
3557_BV	Alpha3	0.35251665	3557_CI	Gamma3	0.24414476
3557_BV	Delta1	0.14310386	3557_CI	Delta4	0.14905485
3557_BV	Delta3	0.05332195	3557_CI	Delta1	0.1025339
3557_BW	Gamma1	0.57908817	3557_CI	Spiro1	0.05856194
3557_BW	Alpha3	0.42091183	3557_CJ	Gamma1	0.4965365
3557_BX	Alpha3	0.561987	3557_CJ	Delta3	0.17293583
3557_BX	Gamma1	0.438013	3557_CJ	Delta1	0.10818989
3557_BY	Alpha3	0.864841	3557_CJ	Alpha7	0.15875384
3557_BY	Gamma1	0.135159	3557_CJ	Gamma3	0.06358394
3557_BZ	Gamma1	0.47308668	3557_CK	Gamma1	0.606623
3557_BZ	Alpha3	0.35638877	3557_CK	Delta3	0.266111
3557_BZ	Delta3	0.17052454	3557_CK	Alpha7	0.127266
3557_CA	Alpha3	0.63063	3557_CL	Gamma1	0.405439
3557_CA	Gamma1	0.36937	3557_CL	Alpha7	0.293756
3557_CB	Gamma1	0.450205	3557_CL	Delta1	0.219853
3557_CB	Alpha3	0.429197	3557_CL	Gamma3	0.041254
3557_CB	Delta3	0.120598	3557_CL	Delta3	0.039698
3557_CC	Gamma1	0.622421	3557_CM	Delta3	0.56619588
3557_CC	Alpha3	0.313255	3557_CM	Gamma1	0.35267247
3557_CC	Alpha10	0.064324	3557_CM	Gamma3	0.08113165
3557_CD	Gamma3	0.31726032	3557_CN	Delta3	0.38140086
3557_CD	Gamma1	0.46662647	3557_CN	Gamma1	0.1883427
3557_CD	Delta4	0.13319013	3557_CN	Delta1	0.15494356
3557_CD	Spiro1	0.08292308	3557_CN	Alpha7	0.27531287
3557_CE	Gamma1	0.48248	3557_CO	Delta1	0.342511
3557_CE	Gamma3	0.282532	3557_CO	Delta3	0.350517
3557_CE	Delta4	0.129103	3557_CO	Gamma1	0.306972
3557_CE	Spiro1	0.105885	3557_CP	Alpha7	0.490769
3557_CF	Gamma1	0.59074704	3557_CP	Gamma1	0.359745
3557_CP	Delta3	0.149486	3557_O	Alpha12	0.064464
3557_CQ	Gamma1	0.61725338	3557_O	Alpha5	0.047005

Library	Symbiont clade	Abundance	Library	Symbiont clade	Abundance
3557_CQ	Delta1	0.22268778	3557_P	Gamma1	0.421233
3557_CQ	Delta3	0.16005884	3557_P	Alpha10	0.203879
3557_D	Gamma1	0.784211	3557_P	Alpha12	0.15824
3557_D	Delta3	0.189341	3557_P	Alpha3	0.146167
3557_D	Spiro1	0.026448	3557_P	Alpha5	0.070481
3557_E	Gamma1	0.903031	3557_Q	Gamma1	0.53307947
3557_E	Delta3	0.096969	3557_Q	Alpha10	0.2006668
3557_F	Gamma1	0.48018737	3557_Q	Alpha12	0.16700583
3557_F	Alpha3	0.43534195	3557_Q	Delta3	0.0992479
3557_F	Alpha10	0.08447067	3557_R	Gamma1	0.405228
3557_G	Alpha10	0.583216	3557_R	Alpha10	0.256669
3557_G	Gamma1	0.416784	3557_R	Alpha12	0.152808
3557_H	Gamma1	0.4610667	3557_R	Alpha3	0.093425
3557_H	Delta3	0.45736523	3557_R	Alpha5	0.09187
3557_H	Spiro1	0.08156807	3557_S	Gamma1	0.834236
3557_I	Gamma1	0.47683522	3557_S	Delta3	0.165764
3557_I	Alpha3	0.17870717	3557_T	Gamma1	0.641361
3557_I	Alpha10	0.17785087	3557_T	Delta3	0.358639
3557_I	Alpha12	0.10801951	3557_U	Gamma1	0.777802
3557_I	Alpha5	0.05706691	3557_U	Delta3	0.222198
3557_I	Alpha3	0.00152032	3557_V	Gamma1	0.57850402
3557_J	Gamma1	0.72923616	3557_V	Alpha3	0.15451795
3557_J	Alpha3	0.16117666	3557_V	Alpha12	0.14106687
3557_J	Alpha4	0.1085991	3557_V	Alpha10	0.12591115
3557_J	Alpha3	0.00098809	3557_W	Gamma1	0.739933
3557_M	Gamma1	0.391653	3557_W	Delta3	0.260067
3557_M	Alpha10	0.213157	3557_X	Gamma5	0.404608
3557_M	Alpha3	0.181052	3557_X	Gamma1	0.326104
3557_M	Alpha12	0.089613	3557_X	Alpha10	0.174303
3557_M	Alpha5	0.078093	3557_X	Gamma1	0.094985
3557_M	Alpha3	0.046432	3557_Y	Alpha10	0.466134
3557_N	Gamma1	0.65223535	3557_Y	Gamma1	0.411285
3557_N	Alpha12	0.17829982	3557_Y	Alpha12	0.122581
3557_N	Alpha3	0.07692192	3557_Z	Gamma1	0.426557
3557_N	Alpha10	0.07585792	3557_Z	Alpha10	0.351141
3557_N	Alpha3	0.01668498	3557_Z	Alpha4	0.073244
3557_O	Gamma1	0.39197	3557_Z	Alpha3	0.068027
3557_O	Alpha3	0.25375	3557_Z	Alpha3	0.055174
3557_O	Alpha10	0.156992	3557_Z	Alpha14	0.025857
3557_O	Alpha5	0.085819	3585_CK_4019_A	Gamma1	0.697136
3585_CK_4019_A	Alpha3	0.302864	3586_CI_4019_N	Delta4	0.18428924
3585_CL_4019_B	Gamma5	0.690323	3586_CI_4019_N	Gamma3	0.13098009
3585_CL_4019_B	Gamma1	0.243749	3630_C	Gamma1	0.708216

Library	Symbiont clade	Abundance	Library	Symbiont clade	Abundance
3585_CL_4019_B	Alpha10	0.052161	3630_C	Alpha3	0.291784
3585_CL_4019_B	Gamma1	0.013767	3630_D	Gamma1	0.739085
3585_CM_4019_C	Alpha10	0.488587	3630_D	Alpha7	0.260915
3585_CM_4019_C	Gamma1	0.284458	3630_E	Gamma1	0.92830302
3585_CM_4019_C	Alpha12	0.226955	3630_E	Spiro1	0.07169698
3585_CN_4019_D	Alpha10	0.422966	3630_F	Gamma1	0.762232
3585_CN_4019_D	Alpha12	0.277361	3630_F	Alpha7	0.138179
3585_CN_4019_D	Gamma1	0.299673	3630_F	Spiro1	0.099589
3585_CO_4019_E	Alpha10	0.43119	3630_G	Alpha10	0.50985687
3585_CO_4019_E	Gamma1	0.285767	3630_G	Gamma1	0.45715812
3585_CO_4019_E	Alpha3	0.136533	3630_G	Spiro1	0.03298501
3585_CO_4019_E	Alpha12	0.071196	3630_H	Alpha12	0.41665142
3585_CO_4019_E	Alpha5	0.075314	3630_H	Gamma1	0.32541133
3585_CP_4019_F	Alpha12	0.378423	3630_H	Alpha10	0.1956792
3585_CP_4019_F	Alpha10	0.336359	3630_H	Alpha5	0.06225806
3585_CP_4019_F	Gamma1	0.285218	3630_I	Gamma1	0.38439538
3585_CQ_4019_G	Gamma1	0.623459	3630_I	Alpha12	0.32086132
3585_CQ_4019_G	Spiro1	0.376541	3630_I	Alpha10	0.29474329
3586_CC_4019_H	Gamma1	0.62484759	3630_J	Alpha12	0.42597843
3586_CC_4019_H	Delta4	0.20730928	3630_J	Gamma1	0.31852132
3586_CC_4019_H	Spiro1	0.16784313	3630_J	Alpha10	0.22864023
3586_CD_4019_I	Delta4	0.5570268	3630_J	Alpha5	0.02686003
3586_CD_4019_I	Gamma1	0.32773276	3630_K	Gamma1	0.43785744
3586_CD_4019_I	Spiro1	0.11524044	3630_K	Alpha12	0.32387232
3586_CE_4019_J	Gamma1	0.69955173	3630_K	Alpha10	0.21678522
3586_CE_4019_J	Delta4	0.1896103	3630_K	Alpha5	0.02148502
3586_CE_4019_J	Spiro1	0.11083797	3630_L	Alpha10	0.3960236
3586_CF_4019_K	Gamma1	0.633114	3630_L	Alpha12	0.39315161
3586_CF_4019_K	Delta4	0.180584	3630_L	Gamma1	0.17627582
3586_CF_4019_K	Spiro1	0.186302	3630_L	Alpha5	0.03454897
3586_CG_4019_L	Gamma1	0.721915	3630_M	Alpha12	0.350074
3586_CG_4019_L	Delta4	0.15672756	3630_M	Gamma1	0.411482
3586_CG_4019_L	Spiro1	0.12135744	3630_M	Alpha10	0.238444
3586_CH_4019_M	Spiro1	0.64906459	3630_N	Gamma1	0.49446
3586_CH_4019_M	Gamma1	0.2038559	3630_N	Alpha12	0.342992
3586_CH_4019_M	Delta4	0.06821646	3630_N	Alpha10	0.146952
3586_CH_4019_M	Gamma3	0.07886305	3630_N	Alpha5	0.015596
3586_CI_4019_N	Gamma1	0.46884551	3630_O	Gamma1	0.390795
3586_CI_4019_N	Spiro1	0.21588516	3630_O	Alpha12	0.324195
3630_O	Alpha10	0.257481	4148_4289_AV	Gamma1	0.577462
3630_O	Alpha5	0.027529	4148_4289_AV	Delta4	0.422538
3854_4020_A	Alpha10	0.69141789	4148_4289_AZ	Alpha3	0.636674
3854_4020_A	Gamma1	0.30858211	4148_4289_AZ	Gamma1	0.363326

Library	Symbiont clade	Abundance	Library	Symbiont clade	Abundance
3854_4020_C	Alpha3	0.474933	4148_4289_B	Gamma1	0.662489
3854_4020_C	Alpha10	0.214803	4148_4289_B	Alpha7	0.337511
3854_4020_C	Gamma1	0.273047	4148_4289_BA	Alpha3	0.506351
3854_4020_C	Alpha3	0.033883	4148_4289_BA	Gamma1	0.493649
3854_4020_C	Alpha3	0.003334	4148_4289_BB	Delta11	0.53244057
3854_4020_D	Alpha10	0.572035	4148_4289_BB	Marinimicrobia1	0.16820241
3854_4020_D	Gamma1	0.427965	4148_4289_BB	Alpha15	0.1549171
3854_4020_E	Alpha3	0.48968751	4148_4289_BB	Gamma1	0.14443992
3854_4020_E	Gamma1	0.26297974	4148_4289_BC	Alpha12	0.673733
3854_4020_E	Alpha10	0.1995538	4148_4289_BC	Alpha5	0.326267
3854_4020_E	Alpha3	0.04777895	4148_4289_BD	Gamma1	0.582848
4148_4289_AA	Gamma1	0.38719592	4148_4289_BD	Alpha3	0.417152
4148_4289_AA	Delta5	0.13023295	4148_4289_BF	Alpha12	0.750151
4148_4289_AA	Gamma2	0.11987489	4148_4289_BF	Alpha10	0.249849
4148_4289_AA	Delta1	0.11790723	4148_4289_BG	Gamma1	0.509112
4148_4289_AA	Spiro1	0.11635949	4148_4289_BG	Alpha10	0.490888
4148_4289_AA	Gamma1	0.10071275	4148_4289_BH	Gamma1	0.379332
4148_4289_AA	Gamma1	0.02771677	4148_4289_BH	Alpha1	0.385738
4148_4289_AD	Alpha10	0.8375621	4148_4289_BH	Alpha10	0.23493
4148_4289_AD	Gamma1	0.15053572	4148_4289_BJ	Gamma1	0.671438
4148_4289_AD	Spiro1	0.01190219	4148_4289_BJ	Alpha12	0.328562
4148_4289_AI	Alpha10	0.55475165	4148_4289_BK	Gamma1	0.770378
4148_4289_AI	Gamma1	0.4170604	4148_4289_BK	Alpha12	0.229622
4148_4289_AI	Spiro1	0.02818794	4148_4289_BL	Gamma1	0.495912
4148_4289_AJ	Alpha10	0.61787929	4148_4289_BL	Delta11	0.35066
4148_4289_AJ	Gamma1	0.38212071	4148_4289_BL	Gamma3	0.153428
4148_4289_AM	Gamma1	0.79865114	4148_4289_BS	Alpha12	0.52692726
4148_4289_AM	Gamma3	0.18005992	4148_4289_BS	Gamma1	0.47307274
4148_4289_AM	Delta4	0.02128894	4148_4289_BT	Gamma1	0.710871
4148_4289_AO	Gamma1	0.84333404	4148_4289_BT	Alpha12	0.289129
4148_4289_AO	Gamma3	0.15666596	4148_4289_BU	Gamma1	0.50113488
4148_4289_AS	Gamma1	0.60857546	4148_4289_BU	Alpha12	0.49886512
4148_4289_AS	Delta11	0.30763486	4148_4289_BV	Gamma1	0.768206
4148_4289_AS	Spiro1	0.08378968	4148_4289_BV	Delta3	0.231794
4148_4289_AU	Alpha14	0.32770054	4148_4289_BX	Gamma7	0.563131
4148_4289_AU	Delta3	0.33102947	4148_4289_BX	Gamma1	0.240725
4148_4289_AU	Gamma1	0.27142804	4148_4289_N	Gamma2	0.4765825
4148_4289_AU	Spiro1	0.06984194	4148_4289_N	Gamma1	0.18639004
4148_4289_BX	Delta14	0.154847	4148_4289_N	Delta2	0.15471976
4148_4289_BX	Alpha16	0.041297	4148_4289_N	Delta5	0.0850072
4148_4289_BY	Gamma7	0.70117939	4148_4289_N	Delta1	0.06714501
4148_4289_BY	Delta14	0.15849066	4148_4289_N	Spiro1	0.0301555
4148_4289_BY	Gamma1	0.14032996	4148_4289_O	Gamma1	0.57737538

Library	Symbiont clade	Abundance	Library	Symbiont clade	Abundance
4148_4289_BZ	Alpha12	0.618697	4148_4289_O	Alpha12	0.42262462
4148_4289_BZ	Gamma1	0.381303	4148_4289_P	Delta4	0.34101925
4148_4289_CE	Gamma7	0.48526843	4148_4289_P	Gamma1	0.47001276
4148_4289_CE	Gamma1	0.37462779	4148_4289_P	Alpha7	0.13475761
4148_4289_CE	Alpha16	0.09385057	4148_4289_P	Delta1	0.05421039
4148_4289_CE	Delta14	0.04625321	4148_4289_Q	Gamma1	0.639756
4148_4289_CF	Gamma1	0.858265	4148_4289_Q	Alpha5	0.253825
4148_4289_CF	Gamma3	0.141735	4148_4289_Q	Delta1	0.071911
4148_4289_CH	Gamma1	0.67414684	4148_4289_Q	Alpha7	0.034508
4148_4289_CH	Delta4	0.32585316	4148_4289_T	Gamma1	0.76939827
4148_4289_CI	Gamma1	0.44947555	4148_4289_T	Delta4	0.16363545
4148_4289_CI	Alpha1	0.28866071	4148_4289_T	Alpha7	0.05185146
4148_4289_CI	Alpha10	0.17399483	4148_4289_T	Delta1	0.01511482
4148_4289_CI	Alpha12	0.08786891	4148_4289_U	Alpha12	0.402991
4148_4289_CM	Gamma1	0.41008841	4148_4289_U	Gamma1	0.479217
4148_4289_CM	Alpha12	0.31454731	4148_4289_U	Delta4	0.086162
4148_4289_CM	Alpha10	0.27536428	4148_4289_U	Delta1	0.03163
4148_4289_CN	Alpha10	0.57956067	4148_4289_V	Alpha12	0.500773
4148_4289_CN	Alpha12	0.24835802	4148_4289_V	Gamma1	0.298176
4148_4289_CN	Gamma5	0.17208131	4148_4289_V	Delta4	0.11552
4148_4289_CO	Alpha10	0.73390583	4148_4289_V	Delta1	0.085531
4148_4289_CO	Gamma5	0.15907391	4148_4289_W	Gamma1	0.60412481
4148_4289_CO	Alpha12	0.10702026	4148_4289_W	Delta4	0.31419286
4148_4289_D	Gamma1	0.64644056	4148_4289_W	Delta1	0.08168234
4148_4289_D	Delta11	0.35355944	4148_4289_X	Delta4	0.38439215
4148_4289_E	Gamma1	0.583509	4148_4289_X	Gamma1	0.50649328
4148_4289_E	Alpha6	0.416491	4148_4289_X	Alpha7	0.07352149
4148_4289_F	Delta11	0.360682	4148_4289_X	Delta1	0.03559308
4148_4289_F	Spiro1	0.353726	4148_4289_Y	Alpha7	0.579062
4148_4289_F	Gamma1	0.285592	4148_4289_Y	Gamma1	0.420938
4148_4289_H	Gamma1	0.46687375	IexuBAH1	Gamma4	0.45723021
4148_4289_H	Delta11	0.42307196	IexuBAH1	Alpha12	0.33633334
4148_4289_H	Spiro1	0.08545798	IexuBAH1	Alpha3	0.20643645
4148_4289_H	Gamma1	0.02459632	IexuBAH2	Gamma4	0.41425559
IexuBAH2	Alpha13	0.16109984	IexuBAH2	Delta3	0.42464458
IexuBAH3	Gamma4	0.545111	Isp1OAH67	Delta12	0.086489
IexuBAH3	Delta3	0.270808	Isp1OAH67	Delta1	0.072631
IexuBAH3	Alpha13	0.184081	Isp1OAH67	Gamma3	0.027394
Ileu1BER1	Gamma1	0.72497635	Isp1OAH70	Gamma1	0.604408
Ileu1BER1	Alpha2	0.17702953	Isp1OAH70	Delta4	0.257871
Ileu1BER1	Alpha10	0.07150319	Isp1OAH70	Spiro1	0.073535
Ileu1BER1	Alpha3	0.02005271	Isp1OAH70	Gamma3	0.045377
Ileu1BER1	Alpha3	0.00643822	Isp1OAH70	Delta1	0.018809

Library	Symbiont clade	Abundance	Library	Symbiont clade	Abundance
Ileu1BER2	Gamma1	0.76018926	Isp2OAH42	Gamma1	0.557039
Ileu1BER2	Alpha2	0.17352529	Isp2OAH42	Gamma3	0.191861
Ileu1BER2	Alpha10	0.04524803	Isp2OAH42	Spiro1	0.134104
Ileu1BER2	Alpha1	0.02082772	Isp2OAH42	Delta4	0.097831
Ileu1BER2	Alpha3	0.0002097	Isp2OAH42	Delta1	0.019165
ImakBAH1	Gamma1	0.53216708	Isp2OAH49	Gamma1	0.506785
ImakBAH1	Alpha10	0.35136414	Isp2OAH49	Spiro1	0.195329
ImakBAH1	Alpha12	0.11646879	Isp2OAH49	Gamma3	0.143476
ImakBAH2	Alpha10	0.61678325	Isp2OAH49	Delta4	0.086269
ImakBAH2	Alpha12	0.19735675	Isp2OAH49	Delta12	0.054952
ImakBAH2	Gamma1	0.18586	Isp2OAH49	Delta1	0.013189
ImanLIZ1	Gamma1	0.51581844	Isp2OAH54	Gamma1	0.741568
ImanLIZ1	Alpha5	0.2854126	Isp2OAH54	Delta4	0.156452
ImanLIZ1	Alpha10	0.09716898	Isp2OAH54	Spiro1	0.10198
ImanLIZ1	Alpha3	0.06712339	Isp2OAH85	Gamma1	0.400699
ImanLIZ1	Alpha3	0.026827	Isp2OAH85	Delta4	0.186022
ImanLIZ1	Alpha3	0.00764959	Isp2OAH85	Gamma1	0.173789
Isp1OAH13	Gamma1	0.69168902	Isp2OAH85	Spiro1	0.182103
Isp1OAH13	Delta4	0.23770321	Isp2OAH85	Gamma1	0.057387
Isp1OAH13	Spiro1	0.07060778	Isp2OAH91	Delta4	0.5041815
Isp1OAH14	Gamma1	0.60657594	Isp2OAH91	Spiro1	0.32665733
Isp1OAH14	Delta4	0.14640048	Isp2OAH91	Gamma1	0.16916117
Isp1OAH14	Delta12	0.14455513	Isp2OAH98	Gamma1	0.57520158
Isp1OAH14	Spiro1	0.08992196	Isp2OAH98	Delta4	0.16852817
Isp1OAH14	Delta1	0.0125465	Isp2OAH98	Spiro1	0.12772213
Isp1OAH30	Gamma1	0.6030433	Isp2OAH98	Delta12	0.1043991
Isp1OAH30	Delta4	0.27010997	Isp2OAH98	Delta1	0.02414902
Isp1OAH30	Gamma1	0.05842531	ItriBAH1	Gamma1	0.587949
Isp1OAH30	Spiro1	0.05773799	ItriBAH1	Alpha3	0.335908
Isp1OAH30	Gamma3	0.01068342	ItriBAH1	Alpha10	0.076143
Isp1OAH67	Gamma1	0.597728	ItriBAH2	Alpha3	0.41362
Isp1OAH67	Delta4	0.108807	ItriBAH2	Gamma1	0.354584
Isp1OAH67	Spiro1	0.106951	ItriBAH2	Alpha10	0.231796
OalbHER1	Alpha3	0.153894	OalbHER1	Gamma1	0.768165
OalbHER1	Spiro1	0.077941	OalgBSAN2	Gamma1	0.24810249
OalbHER2	Gamma1	0.69155231	OalgBSAN2	Gamma1	0.14619349
OalbHER2	Alpha3	0.25589774	OalgBSAN2	Delta3	0.10529375
OalbHER2	Spiro1	0.05254995	OalgBSAN2	Delta1	0.0793824
OalgACAV1	Gamma1	0.51403513	OalgBSAN2	Gamma1	0.0600385
OalgACAV1	Delta3	0.1858042	OalgBSAN2	Spiro1	0.0600271
OalgACAV1	Gamma3	0.13959061	OalgBSAN2	Gamma1	0.01102035
OalgACAV1	Delta1	0.10775908	OalgBSAN3	Gamma3	0.30438831
OalgACAV1	Spiro1	0.05281098	OalgBSAN3	Gamma1	0.44776203

Library	Symbiont clade	Abundance	Library	Symbiont clade	Abundance
OalgACAV2	Gamma3	0.323656	OalgBSAN3	Delta4	0.13589625
OalgACAV2	Gamma1	0.359537	OalgBSAN3	Delta1	0.06953968
OalgACAV2	Delta4	0.161817	OalgBSAN3	Spiro1	0.04241373
OalgACAV2	Spiro1	0.099586	OclaLIZ1	Gamma1	0.577109
OalgACAV2	Delta1	0.055404	OclaLIZ1	Actinomarinales1	0.257233
OalgACAV3	Gamma1	0.41243941	OclaLIZ1	Delta11	0.165658
OalgACAV3	Gamma3	0.23855524	OclaLIZ2	Gamma1	0.66866267
OalgACAV3	Delta4	0.21454221	OclaLIZ2	Delta11	0.16825617
OalgACAV3	Spiro1	0.08009608	OclaLIZ2	Actinomarinales1	0.16308116
OalgACAV3	Delta1	0.05436705	OcraPER1	Gamma1	0.334756
OalgASAN1	Gamma1	0.4172088	OcraPER1	Spiro1	0.237721
OalgASAN1	Gamma3	0.22557242	OcraPER1	Delta2	0.227661
OalgASAN1	Delta4	0.14062255	OcraPER1	Gamma2	0.08717
OalgASAN1	Spiro1	0.13475805	OcraPER1	Gamma1	0.063516
OalgASAN1	Delta1	0.08183817	OcraPER1	Delta5	0.021528
OalgASAN2	Gamma1	0.47388349	OcraPER1	Gamma1	0.015254
OalgASAN2	Gamma3	0.22521432	OcraPER1	Gamma1	0.010406
OalgASAN2	Delta4	0.1668863	OcraPER1	Delta1	0.001865
OalgASAN2	Spiro1	0.07714035	OcraPER1	Gamma1	6.80E-05
OalgASAN2	Delta1	0.05687554	OcraPER1	Delta14	5.50E-05
OalgASAN3	Gamma3	0.22091641	OcraPER2	Gamma1	0.296739
OalgASAN3	Gamma1	0.18699172	OcraPER2	Spiro1	0.275931
OalgASAN3	Gamma1	0.23120018	OcraPER2	Delta2	0.193648
OalgASAN3	Delta4	0.14750558	OcraPER2	Gamma1	0.118953
OalgASAN3	Spiro1	0.09455739	OcraPER2	Gamma2	0.056889
OalgASAN3	Delta1	0.07208288	OcraPER2	Delta5	0.040098
OalgASAN3	Gamma1	0.04674583	OcraPER2	Gamma1	0.011308
OalgBSAN1	Gamma3	0.30798529	OcraPER2	Gamma1	0.005538
OalgBSAN1	Delta4	0.25056464	OcraPER2	Delta1	0.000896
OalgBSAN1	Gamma1	0.34915786	OfilPIA1	Alpha12	0.48951151
OalgBSAN1	Spiro1	0.0922922	OfilPIA1	Gamma1	0.32266968
OalgBSAN2	Gamma3	0.28994192	OfilPIA1	Alpha12	0.11022989
OfilPIA2	Alpha12	0.538809	OfilPIA1	Delta4	0.07758892
OfilPIA2	Gamma1	0.415453	Osp1DAH1	Spiro1	0.06065145
OfilPIA2	Delta4	0.045738	Osp1DAH1	Alpha12	0.04977381
OgenHER1	Gamma1	0.87875	Osp1DAH1	Delta1	0.01997274
OgenHER1	Delta1	0.12125	Osp1DAH1	Alpha9	0.01875727
OgenHER2	Gamma1	0.95768679	Osp1DAH1	Alpha3	0.01243939
OgenHER2	Delta1	0.04231321	Osp1DAH1	Alpha1	0.0002173
OgenLIZ1	Gamma1	0.901798	Osp1DAH2	Gamma1	0.26474689
OgenLIZ1	Delta1	0.098202	Osp1DAH2	Gamma1	0.26034754
OgenLIZ2	Gamma1	0.803858	Osp1DAH2	Alpha10	0.14935029
OgenLIZ2	Delta1	0.196142	Osp1DAH2	Alpha8	0.1340747

Library	Symbiont clade	Abundance	Library	Symbiont clade	Abundance
OilvPIA1	Gamma1	0.659742	Osp1DAH2	Spiro1	0.07256139
OilvPIA1	Gamma3	0.318684	Osp1DAH2	Gamma1	0.06187872
OilvPIA1	Alpha7	0.021574	Osp1DAH2	Alpha9	0.03378186
OilvPIA2	Gamma1	0.69813172	Osp1DAH2	Delta1	0.01177992
OilvPIA2	Delta4	0.2201613	Osp1DAH2	Alpha12	0.01130893
OilvPIA2	Gamma3	0.06510666	Osp1DAH2	Alpha13	0.00016976
OilvPIA2	Alpha7	0.01660032	Osp1HER1	Gamma1	0.433833
OilvSAN1	Gamma1	0.681199	Osp1HER1	Alpha5	0.274179
OilvSAN1	Delta1	0.159354	Osp1HER1	Delta4	0.212405
OilvSAN1	Gamma3	0.088775	Osp1HER1	Spiro1	0.079583
OilvSAN1	Alpha7	0.070672	Osp1OAH1	Alpha5	0.72233956
OilvSAN2	Gamma1	0.58890641	Osp1OAH1	Delta4	0.09843886
OilvSAN2	Delta3	0.2007198	Osp1OAH1	Gamma1	0.07459827
OilvSAN2	Delta1	0.17103883	Osp1OAH1	Delta3	0.05386421
OilvSAN2	Alpha7	0.03933496	Osp1OAH1	Spiro1	0.0507591
OloiHER1	Gamma1	0.733978	Osp3HER1	Actinomarinales1	0.506967
OloiHER1	Alpha3	0.154888	Osp3HER1	Delta11	0.293015
OloiHER1	Spiro1	0.111134	Osp3HER1	Gamma1	0.143569
OloiHER2	Gamma1	0.52451	Osp3HER1	Marinimicrobia1	0.042528
OloiHER2	Alpha3	0.422862	Osp3HER1	Alpha15	0.013921
OloiHER2	Spiro1	0.052628	Osp4HER1	Gamma1	0.64030504
OloiLIZ1	Gamma1	0.6407967	Osp4HER1	Delta1	0.35969496
OloiLIZ1	Alpha3	0.28985592	Osp5HER1	Marinimicrobia1	0.37588238
OloiLIZ1	Spiro1	0.06934739	Osp5HER1	Gamma1	0.27681628
OloiLIZ2	Gamma1	0.624324	Osp5HER1	Delta11	0.22589423
OloiLIZ2	Alpha3	0.343158	Osp5HER1	Actinomarinales1	0.11837812
OloiLIZ2	Spiro1	0.032518	Osp5HER1	Alpha15	0.003029
Osp1DAH1	Gamma1	0.43365293	OvacBAH1	Gamma1	0.462593
Osp1DAH1	Alpha8	0.17073043	OvacBAH1	Alpha6	0.439255
Osp1DAH1	Alpha10	0.12325509	OvacBAH1	Gamma1	0.073835
Osp1DAH1	Gamma1	0.1105496	OvacBAH1	Gamma1	0.024317
OvacBAH2	Alpha6	0.289457	OvacBAH2	Gamma1	0.710543

Table S3 | Estimated acquisition and losses of symbiont clades in host nuclear or mitochondrial lineages

Symbiont clade	Nuclear host lineages			Mitochondrial host lineages		
	State at the ancestral node	Acquisitions	Losses	State at the ancestral node	Acquisitions	Losses
Actinomarinales1	Absent	1-3	0-2	Absent	1-3	0-2
Alpha1	Absent	7-8	0-1	Absent	7-9	0-2
Alpha2	Absent	1	0	Absent	1	0
Alpha3	Absent	13-15	5	Absent	13-17	8-11
Alpha4	Absent	1	0	Absent	1	0
Alpha5	Absent	10-11	0-1	Absent	9	5
Alpha6	Absent	1	0	Absent	1	0
Alpha7	Absent	7-9	2-3	Absent	6-8	5-6
Alpha8	Absent	1	0	Not resolved	1	0-1
Alpha9	Absent	1	0	Not resolved	1	0-1
Alpha10	Absent	5-7	8-9	Not resolved	5-6	8-10
Alpha12	Absent	11-18	3-9	Not resolved	11-15	5-10
Alpha13	Absent	2-3	0-1	Absent	2	0
Alpha14	Absent	2-3	0-1	Absent	3-4	0-1
Alpha15	Absent	1	0	Absent	2	0
Alpha16	Absent	1	0	Absent	1	1
Delta1	Absent	17-19	9	Not resolved	14-22	12-14
Delta2	Absent	1	0	Absent	1-2	0-1
Delta3	Absent	14-16	2-4	Absent	13-17	3-5
Delta4	Absent	8-9	2-3	Absent	8	3
Delta5	Absent	1	0	Absent	1	0
Delta11	Absent	4	0	Absent	4	0
Delta12	Absent	3-4	0-1	Absent	4	0
Delta13	Absent	1	0	Absent	1	0
Delta14	Absent	1	0	Absent	2	0
Gamma1	Present	0-1	4-5	Present	0-1	4-5
Gamma2	Absent	1	0	Absent	1	0
Gamma3	Absent	11-12	0-1	Absent	12	1
Gamma4	Absent	1	0	Absent	1	0
Gamma5	Absent	3	0	Absent	4	0
Gamma7	Absent	1	0	Absent	1	0
Marinimicrobia1	Absent	1	0	Absent	2	0
Spiro1	Absent	8-23	0-15	Not resolved	16-18	1-5

Table S4 | Sample overview.

Library	Host species	Ocean	Continent	Field site	Latitude	Longitude	Organic input	Sediment type	Water depth (m)	Sampling month	Sampling year
2699_C	*O. algarvensis*	Mediterranean	Europe	Spain	39.503917	2.543103	Seagrass	NA	NA	May	2016
2699_D	*O. algarvensis*	Mediterranean	Europe	Spain	39.503917	2.543103	Seagrass	NA	NA	May	2016
3557_A	*O.* sp. 'belize1'	Atlantic	Central America	Belize	16.82356	-88.10617	Seagrass	Sand	-1.5	April	2017
3557_AA	*I.* sp. 'fant'	Atlantic	Central America	Belize	16.790105	-88.08201	Corals	Sand	-2	April	2017
3557_AB	*I. scalprum*	Atlantic	Central America	Belize	16.790105	-88.08201	Corals	Sand	-2	April	2017
3557_AC	*I. scalprum*	Atlantic	Central America	Belize	16.790105	-88.08201	Corals	Sand	-2	April	2017
3557_AD	*I.* sp. 'ule'	Atlantic	Central America	Belize	16.790105	-88.08201	Corals	Sand	-2	April	2017
3557_AE	*I. reginae*	Atlantic	Central America	Belize	16.82356	-88.10617	Seagrass	Sand	-1.5	April	2017
3557_AF	*I. leukodermatus*	Atlantic	Central America	Belize	NA	NA	NA	Sand	NA	April	2017
3557_AG	*I. leukodermatus*	Atlantic	Central America	Belize	NA	NA	Seagrass	Sand	-1	March	2017
3557_AH	*I. reginae*	Atlantic	Central America	Belize	16.82356	-88.10617	Seagrass	Sand	-1	April	2017
3557_AI	*I. scalprum*	Atlantic	Central America	Belize	NA	NA	Seagrass	Sand	-1	March	2017
3557_AJ	*I. scalprum*	Atlantic	Central America	Belize	NA	NA	Seagrass	Sand	-1	March	2017
3557_AK	*O. tantulus*	Atlantic	Central America	Belize	NA	NA	Seagrass	Sand	-1	March	2017
3557_AL	*O. finitimus*	Atlantic	Central America	Belize	16.8236	-88.105925	Seagrass	Sand	-1.5	March	2017
3557_AN	*O. tantulus*	Atlantic	Central America	Belize	16.8236	-88.1061	Seagrass	Sand	-1.5	April	2017
3557_AO	*I. reginae*	Atlantic	Central America	Belize	16.82356	-88.10617	Seagrass	Sand	-1.5	April	2017
3557_AQ	*I. aduncosetis*	Atlantic	Central America	Belize	16.790105	-88.08201	Corals	Sand	-2	April	2017
3557_AR	*I. aduncosetis*	Atlantic	Central America	Belize	16.790105	-88.08201	Corals	Sand	-2	April	2017
3557_AS	*I. aduncosetis*	Atlantic	Central America	Belize	16.790105	-88.08201	Corals	Sand	-2	April	2017
3557_AT	*I. leukodermatus*	Atlantic	Central America	Belize	16.79	-88.0826	NA	Sand	-1.5	April	2017
3557_AU	*I.* sp. 'nysp'	Atlantic	Central America	Belize	16.79	-88.0826	NA	Sand	-1.5	April	2017
3557_AV	*I.* sp. 'nysp'	Atlantic	Central America	Belize	16.790105	-88.08201	Corals	Sand	-2	April	2017

Library	Host species	Ocean	Continent	Field site	Latitude	Longitude	Organic input	Sediment type	Water depth (m)	Sampling month	Sampling year
3557_AW	*I. scalprum*	Atlantic	Central America	Belize	16.790105	-88.08201	Corals	Sand	-2	April	2017
3557_AX	*I. leukodermatus*	Atlantic	Central America	Belize	16.8265	-88.080695	Seagrass	Sand	-2	April	2017
3557_AY	*O. tantulus*	Atlantic	Central America	Belize	16.823715	-88.081435	Seagrass, algal plumbs	Sand	-8	April	2017
3557_AZ	*O. imperfectus*	Atlantic	Central America	Belize	16.823715	-88.081435	Seagrass, algal plumbs	Sand	-8	April	2017
3557_B	*O. longissimus* B	Atlantic	Central America	Belize	16.804685	-88.0823	Corals	Sand	NA	April	2017
3557_BA	*O. imperfectus*	Atlantic	Central America	Belize	16.823715	-88.081435	Seagrass, algal plumbs	Sand	-8	April	2017
3557_BB	*I. leukodermatus*	Atlantic	Central America	Belize	16.826195	-88.0813	Seagrass	Sand	-1.5	April	2017
3557_BC	*O. imperfectus*	Atlantic	Central America	Belize	16.826195	-88.0813	Seagrass	Sand	-1.5	April	2017
3557_BD	*O. tantulus*	Atlantic	Central America	Belize	16.826195	-88.0813	Seagrass, algal plumbs	Sand	-1.5	April	2017
3557_BE	*I. leukodermatus*	Atlantic	Central America	Belize	16.826195	-88.0813	Seagrass, algal plumbs	Sand	-1.5	April	2017
3557_BF	*I.* sp. 'nysp'	Atlantic	Central America	Belize	16.790135	-88.0818	NA	Coral sand	-2	April	2017
3557_BG	*O. tantulus*	Atlantic	Central America	Belize	16.8236	-88.105925	Seagrass	Sand	-1.5	March	2017
3557_BH	*O.* sp. 'zimmermannae'	Atlantic	Central America	Belize	16.82356	-88.10617	Seagrass	Sand	-1.5	April	2017
3557_BI	*O. imperfectus*	Atlantic	Central America	Belize	16.831375	-88.10835	Seagrass	Sand	-1.5	April	2017
3557_BJ	*I. leukodermatus*	Atlantic	Central America	Belize	16.831375	-88.10835	Seagrass	Sand	-1.5	April	2017
3557_BK	*I.* sp. 'ule'	Atlantic	Central America	Belize	16.790135	-88.0818	NA	Coral sand	-2	April	2017
3557_BL	*I.* sp. 'ule'	Atlantic	Central America	Belize	16.790135	-88.0818	NA	Coral sand	-2	April	2017
3557_BM	*O. longissimus* B	Atlantic	Central America	Belize	16.8045	-88.0823	Corals	Sand	NA	April	2017
3557_BN	*I. aduncosetis*	Atlantic	Central America	Belize	16.8254	-88.08	Seagrass	Sand	-2	April	2017
3557_BO	*O. imperfectus*	Atlantic	Central America	Belize	16.8312	-88.1096	Seagrass	Sand	-1.5	April	2017
3557_BP	*O. tantulus*	Atlantic	Central America	Belize	16.8236	-88.08155	Algal plumbs	Sand	-1	April	2017

Library	Host species	Ocean	Continent	Field site	Latitude	Longitude	Organic input	Sediment type	Water depth (m)	Sampling month	Sampling year
3557_BQ	*O.* nr. *faris*	Mediterranean	Europe	Spain	39.872	3.1545	Seagrass	NA	NA	May	2016
3557_BR	*O.* nr. *faris*	Mediterranean	Europe	Spain	39.872	3.1545	Seagrass	NA	-12	May	2016
3557_BS	*O. faris*	Mediterranean	Europe	Spain	39.872	3.1545	Seagrass	NA	NA	May	2016
3557_BT	*O. faris*	Mediterranean	Europe	Spain	39.5039	2.5431	Seagrass	NA	NA	May	2016
3557_BU	*O. faris*	Mediterranean	Europe	Spain	39.872	3.1545	Seagrass	NA	NA	May	2016
3557_BV	*O. faris*	Mediterranean	Europe	Spain	39.872	3.1545	Seagrass	NA	NA	May	2016
3557_BW	*O. faris*	Mediterranean	Europe	Spain	39.5039	2.5431	Seagrass	NA	NA	May	2016
3557_BX	*O. faris*	Mediterranean	Europe	Spain	39.5041	2.5433	NA	NA	-3	September	2017
3557_BY	*O. faris*	Mediterranean	Europe	Spain	39.872	3.1545	Seagrass	NA	NA	May	2016
3557_BZ	*O. faris*	Mediterranean	Europe	Spain	39.4833	2.5322	Seagrass	Sand	NA	May	2016
3557_CA	*O. faris*	Mediterranean	Europe	Spain	39.5039	2.5431	Seagrass	NA	NA	May	2016
3557_CB	*O. faris*	Mediterranean	Europe	Spain	39.4817	2.5318	NA	NA	NA	August	2017
3557_CC	*O. faris*	Mediterranean	Europe	Spain	39.4817	2.5318	NA	NA	NA	August	2017
3557_CD	*O. algarvensis*	Mediterranean	Europe	Italy	42.7348	10.1865	Seagrass	Sand	-1	November	2016
3557_CE	*O. algarvensis*	Mediterranean	Europe	Italy	42.7348	10.1865	Seagrass	Sand	NA	November	2016
3557_CF	*O. ilvae*	Mediterranean	Europe	Italy	42.7434	10.1194	Seagrass	Sand	-0.3	November	2016
3557_CG	*O. ilvae*	Mediterranean	Europe	Italy	42.7434	10.1194	Seagrass	Sand	-0.3	November	2016
3557_CH	*O. algarvensis*	Mediterranean	Europe	Italy	42.7348	10.1865	Seagrass	Sand	NA	November	2016
3557_CI	*O. algarvensis*	Mediterranean	Europe	Italy	42.7348	10.1865	Seagrass	Sand	NA	November	2016
3557_CJ	*O. ilvae*	Mediterranean	Europe	Italy	42.7434	10.1194	Seagrass	Sand	NA	November	2016
3557_CK	*O. ilvae*	Mediterranean	Europe	Italy	42.7434	10.1194	Seagrass	Sand	-0.3	November	2016
3557_CL	*O. ilvae*	Mediterranean	Europe	Italy	42.7338	10.1927	Seagrass	Sand	NA	October	2016
3557_CM	*O. ilvae*	Mediterranean	Europe	Italy	42.7338	10.1927	Seagrass	Sand	-4	October	2016
3557_CN	*O. ilvae*	Mediterranean	Europe	Italy	42.7434	10.1194	Seagrass	Sand	-0.3	November	2016
3557_CO	*O. ilvae*	Mediterranean	Europe	Italy	42.7434	10.1194	Seagrass	Sand	-0.3	November	2016
3557_CP	*O. ilvae*	Mediterranean	Europe	France	43.7199	7.4009	Seagrass, detritus	Sand	-1.5	June	2017

Library	Host species	Ocean	Continent	Field site	Latitude	Longitude	Organic input	Sediment type	Water depth (m)	Sampling month	Sampling year
3557_CQ	*O. ilvae*	Mediterranean	Europe	France	43.7199	7.4009	Seagrass, detritus	Sand	-1.5	June	2017
3557_D	*O.* sp. 'belize1'	Atlantic	Central America	Belize	16.82356	-88.10617	Seagrass	Sand	-1.5	April	2017
3557_E	*O.* sp. 'belize1'	Atlantic	Central America	Belize	16.82356	-88.10617	Seagrass	Sand	-1.5	April	2017
3557_F	*I. leukodermatus*	Atlantic	Central America	Belize	16.831375	-88.10835	Seagrass	Sand	-1.5	April	2017
3557_G	*I. leukodermatus*	Atlantic	Central America	Belize	16.831375	-88.10835	Seagrass	Sand	-1.5	April	2017
3557_H	*O. longissimus* B	Atlantic	Central America	Belize	16.804685	-88.0823	Corals	Sand	NA	April	2017
3557_I	*I.* sp. 'nysp'	Atlantic	Central America	Belize	16.790135	-88.0818	NA	Coral sand	-2	April	2017
3557_J	*I.* sp. 'fant'	Atlantic	Central America	Belize	16.790135	-88.0818	NA	Coral sand	-2	April	2017
3557_M	*I.* sp. 'nysp'	Atlantic	Central America	Belize	16.7904	-88.0818	Corals	Sand	-1.5	April	2017
3557_N	*I. leukodermatus*	Atlantic	Central America	Belize	16.8254	-88.08	Seagrass	Sand	-2	April	2017
3557_O	*I.* sp. 'nysp'	Atlantic	Central America	Belize	16.8254	-88.08	Seagrass	Sand	-2	April	2017
3557_P	*I.* sp. 'nysp'	Atlantic	Central America	Belize	16.8254	-88.08	Seagrass	Sand	-2	April	2017
3557_Q	*I. aduncosetis*	Atlantic	Central America	Belize	16.8254	-88.08	Seagrass	Sand	-2	April	2017
3557_R	*I.* sp. 'nysp'	Atlantic	Central America	Belize	NA	NA	NA	Sand	NA	April	2017
3557_S	*O. tantulus*	Atlantic	Central America	Belize	16.8025	-88.0822	Seagrass	Sand	NA	April	2017
3557_T	*O. tantulus*	Atlantic	Central America	Belize	16.8025	-88.0822	Seagrass	Sand	NA	April	2017
3557_U	*O. tantulus*	Atlantic	Central America	Belize	16.8025	-88.0822	Seagrass	Sand	NA	April	2017
3557_V	*I. leukodermatus*	Atlantic	Central America	Belize	16.8025	-88.0822	Seagrass	Sand	NA	April	2017
3557_W	*O. tantulus*	Atlantic	Central America	Belize	16.8236	-88.1061	Seagrass	Sand	-1.5	April	2017
3557_X	*I. leukodermatus*	Atlantic	Central America	Belize	16.790105	-88.08201	Corals	Sand	-2	April	2017
3557_Y	*I. leukodermatus*	Atlantic	Central America	Belize	16.790105	-88.08201	Corals	Sand	-2	April	2017
3557_Z	*I.* sp. 'fant'	Atlantic	Central America	Belize	16.790105	-88.08201	Corals	Sand	-2	April	2017
3585_CK_4019_A	*I. leukodermatus*	Atlantic	Central America	Belize	16.8024	-88.0821	Seagrass, detritus	Sand	-0.7	April	2013

Library	Host species	Ocean	Continent	Field site	Latitude	Longitude	Organic input	Sediment type	Water depth (m)	Sampling month	Sampling year
3585_CL_4019_B	*I. leukodermatus*	Atlantic	Central America	Belize	16.8024	-88.0821	Seagrass, detritus	Sand	-0.7	April	2013
3585_CM_4019_C	*I.* sp. 'nysp'	Atlantic	Central America	Belize	16.8033	-88.0818	NA	Sand	-1	April	2013
3585_CN_4019_D	*I.* sp. 'nysp'	Atlantic	Central America	Belize	16.8033	-88.0818	NA	Sand	-1	April	2013
3585_CO_4019_E	*I.* sp. 'nysp'	Atlantic	Central America	Belize	16.8033	-88.0818	NA	Sand	-1	April	2013
3585_CP_4019_F	*I.* sp. 'nysp'	Atlantic	Central America	Belize	16.8033	-88.0818	NA	Sand	-1	April	2013
3585_CQ_4019_G	*O. longissimus* B	Atlantic	Central America	Belize	16.8031	-88.0821	Seagrass	Sand	-0.5	April	2013
3586_CC_4019_H	*I. dutchae* A	Pacific	North America	Hawai'i	21.2608	-157.7867	NA	Sand	-6	September	2016
3586_CD_4019_I	*I. dutchae* A	Pacific	North America	Hawai'i	21.2608	-157.7867	NA	Sand	-6	September	2016
3586_CE_4019_J	*I. dutchae* A	Pacific	North America	Hawai'i	21.2608	-157.7867	NA	Sand	-6	September	2016
3586_CF_4019_K	*I. dutchae* A	Pacific	North America	Hawai'i	21.2608	-157.7867	NA	Sand	-6	September	2016
3586_CG_4019_L	*I. dutchae* A	Pacific	North America	Hawai'i	21.2608	-157.7867	NA	Sand	-4	September	2016
3586_CH_4019_M	*I. dutchae* A	Pacific	North America	Hawai'i	21.2608	-157.7867	NA	Sand	NA	September	2016
3586_CI_4019_N	*I. dutchae* B	Pacific	North America	Hawai'i	21.2608	-157.7867	NA	Sand	NA	September	2016
3630_C	*O.* sp. 'okinawa1'	Pacific	Asia	Japan	26.4676	127.824	Seagrass, corals	Sand	-0.7	November	2017
3630_D	*O.* sp. 'okinawa2'	Pacific	Asia	Japan	26.4687	127.8238	Seagrass, corals	Sand	-0.7	November	2017
3630_E	*O.* sp. 'okinawa2'	Pacific	Asia	Japan	26.4926	127.8421	Seagrass, corals	NA	-2	November	2017
3630_F	*O.* sp. 'okinawa2'	Pacific	Asia	Japan	26.4924	127.8416	Seagrass, corals	NA	-0.5	November	2017
3630_G	*O.* sp. 'okinawa1'	Pacific	Asia	Japan	26.4935	127.8402	Seagrass, corals	NA	-0.7	November	2017
3630_H	*I.* sp. 'guadeloupe1'	Atlantic	Central America	Guadeloupe	16.215238	-61.53657	Seagrass	NA	NA	July	2016
3630_I	*I.* sp. 'guadeloupe1'	Atlantic	Central America	Guadeloupe	16.215238	-61.53657	Seagrass	NA	NA	July	2016
3630_J	*I.* sp. 'guadeloupe1'	Atlantic	Central America	Guadeloupe	16.215238	-61.53657	Seagrass	NA	NA	July	2016
3630_K	*I.* sp. 'guadeloupe1'	Atlantic	Central America	Guadeloupe	16.215238	-61.53657	Seagrass	NA	NA	July	2016
3630_L	*I.* sp. 'guadeloupe1'	Atlantic	Central America	Guadeloupe	16.215238	-61.53657	Seagrass	NA	NA	July	2016

Library	Host species	Ocean	Continent	Field site	Latitude	Longitude	Organic input	Sediment type	Water depth (m)	Sampling month	Sampling year
3630_M	*I.* sp. 'guadeloupe1'	Atlantic	Central America	Guadeloupe	16.215238	-61.53657	Seagrass	NA	NA	July	2016
3630_N	*I.* sp. 'guadeloupe1'	Atlantic	Central America	Guadeloupe	16.215238	-61.53657	Seagrass	NA	NA	July	2016
3630_O	*I.* sp. 'guadeloupe1'	Atlantic	Central America	Guadeloupe	16.215238	-61.53657	Seagrass	NA	NA	July	2016
3854_4020_A	*O.* sp. 'northcarolina1'	Atlantic	North America	North Carolina	34.45694444	-76.81638889	NA	Mud, sand	-35	June	2018
3854_4020_C	*O.* sp. 'northcarolina1'	Atlantic	North America	North Carolina	34.45694444	-76.81638889	NA	Mud, sand	-35	June	2018
3854_4020_D	*O.* sp. 'northcarolina1'	Atlantic	North America	North Carolina	34.45694444	-76.81638889	NA	Mud, sand	-35	June	2018
3854_4020_E	*O.* sp. 'northcarolina1'	Atlantic	North America	North Carolina	34.45694444	-76.81638889	NA	Mud, sand	-35	June	2018
4148_4289_AA	*O. crassitunicatus*	Pacific	South America	Peru	-12.7322	-77.1327	NA	Mud	-3	June	2000
4148_4289_AC	*C. bidentatus*	Pacific	Australia	Queensland	-14.6783	145.47	NA	Sand	-2.5	February	2006
4148_4289_AD	*I. wasseri*	Pacific	Australia	Queensland	-14.69	145.4517	NA	Sand	-3	February	2006
4148_4289_AG	*A.* sp. 'new'	Pacific	Australia	Queensland	-14.575	145.6133	NA	Sand	-6	February	2006
4148_4289_AI	*O. filithecatus*	Pacific	Australia	Queensland	-14.6867	145.4433	NA	Sand	-3.5	February	2006
4148_4289_AJ	*O. filithecatus*	Pacific	Australia	Queensland	-14.6867	145.4433	NA	Sand	-3.5	February	2006
4148_4289_AK	*C.* sp. 'coqu'	Pacific	South America	Chile	-29.9667	-71.3617	NA	Mud, sand	-1.5	February	2009
4148_4289_AM	*O.* sp. 'cape'	Atlantic	North America	North Carolina	34.2939	-75.7682	NA	Mud	NA	May	2011
4148_4289_AO	*O.* sp. 'cape'	Atlantic	North America	North Carolina	34.2939	-75.7682	NA	Mud	NA	May	2011
4148_4289_AS	*O.* sp. 'false	Atlantic	Africa	South Africa	-34.1581	26.4337	NA	Sand	NA	December	2011
4148_4289_AU	*O.* sp. 'canary1'	Atlantic	Europe	Canaries	44.102	-13.7	NA	NA	NA	October	2011
4148_4289_AV	*O.* sp. 'tt'	Pacific	Australia	Queensland	-14.684	145.449	NA	Sand	-1.5	April	2012
4148_4289_AZ	*O.* nr. *avisceralis*	Pacific	Australia	Queensland	-14.6712	145.4424	Corals	Sand	-1.2	April	2012
4148_4289_B	*O. prodigus*	Indian	Australia	Western Australia	-32.0155	115.513	NA	Sand	-1	January	1991
4148_4289_BA	*O.* nr. *loisae*	Pacific	Australia	Queensland	-23.4349	151.945	NA	Sand	-7	September	2012
4148_4289_BB	*O.* sp. 'heron5'	Pacific	Australia	Queensland	-23.4349	151.945	NA	Sand	-7	September	2012
4148_4289_BC	*O.* sp. 'other'	Pacific	Australia	Queensland	-23.4349	151.945	NA	Sand	-7	September	2012
4148_4289_BD	*O.* nr. *loisae*	Pacific	Australia	Queensland	-23.4349	151.945	NA	Sand	-7	September	2012

Library	Host species	Ocean	Continent	Field site	Latitude	Longitude	Organic input	Sediment type	Water depth (m)	Sampling month	Sampling year
4148_4289_BF	*O.* sp. 'other'	Pacific	Australia	Queensland	-23.4349	151.945	Algal crust	Sand	-7	September	2012
4148_4289_BG	*I. wasseri*	Pacific	Australia	Queensland	-23.4349	151.945	Algal crust	Sand	-7	September	2012
4148_4289_BH	*I. mojicae*	Atlantic	Central America	Bahamas	23.856	-76.2248	NA	Sand	-6	April	2013
4148_4289_BJ	*O.* sp. 'lees'	Atlantic	Central America	Bahamas	23.856	-76.2248	NA	Sand	-6	April	2013
4148_4289_BK	*O.* sp. 'lees'	Atlantic	Central America	Bahamas	23.856	-76.2248	NA	Sand	-6	April	2013
4148_4289_BL	*O.* sp. 'darisl'	Atlantic	Central America	Bahamas	23.856	-76.2248	NA	Sand	-6	April	2013
4148_4289_BS	*O.* sp. 'lees'	Atlantic	Central America	Bahamas	23.856	-76.2248	NA	Sand	-6	April	2013
4148_4289_BT	*O.* sp. 'lees'	Atlantic	Central America	Bahamas	23.856	-76.2248	NA	Sand	-6	April	2013
4148_4289_BU	*O.* sp. 'lees'	Atlantic	Central America	Bahamas	23.856	-76.2248	NA	Sand	-6	April	2013
4148_4289_BV	*O.* sp. 'which'	Atlantic	Central America	Bahamas	23.856	-76.2248	NA	Sand	-6	April	2013
4148_4289_BW	*Phallodrilinae* gen. sp. 'strang'	Atlantic	Central America	Bahamas	23.856	-76.2248	NA	Sand	-6	April	2013
4148_4289_BX	*O.* sp. 'fili'	Atlantic	Central America	Bahamas	23.8573	-76.2169	Seagrass	Sand	-3	April	2013
4148_4289_BY	*O.* sp. 'fili'	Atlantic	Central America	Bahamas	23.8573	-76.2169	Seagrass	Sand	-3	April	2013
4148_4289_BZ	*O.* sp. 'bahamas1'	Atlantic	Central America	Bahamas	23.8573	-76.2169	Seagrass	Sand	-3	April	2013
4148_4289_C	*A. wellsi*	Indian	Australia	Western Australia	-31.9981	115.4916	NA	Sand	-1	January	1991
4148_4289_CE	*O.* sp. 'fili'	Atlantic	Central America	Bahamas	23.8582	-76.2177	Seagrass	Sand	-7	April	2013
4148_4289_CF	*O.* sp. 'head'	Atlantic	Central America	Bahamas	23.8565	-76.2158	NA	Sand	-3	April	2013
4148_4289_CH	*O. finitimus*	Atlantic	Central America	Bahamas	23.8561	-76.2254	Seagrass, detritus	Sand	-1	April	2013
4148_4289_CI	*I. mojicae*	Atlantic	Central America	Bahamas	23.7683	-76.1309	Seagrass, detritus	Sand	-0.2	April	2013
4148_4289_CM	*I. reginae*	Atlantic	Central America	Belize	16.7589	-88.1127	NA	Sand	-3.5	April	2013
4148_4289_CN	*O.* sp. 'zimmermannae'	Atlantic	Central America	Belize	16.8032	-88.0821	Seagrass	Sand	-0.5	April	2013
4148_4289_CO	*O.* sp. 'zimmermannae'	Atlantic	Central America	Belize	16.8031	-88.0821	Seagrass	Sand	-0.5	April	2013
4148_4289_D	*O. furinus*	Atlantic	Central America	Bahamas	23.772	-76.1	NA	Sand	-7	April	1999

Library	Host species	Ocean	Continent	Field site	Latitude	Longitude	Organic input	Sediment type	Water depth (m)	Sampling month	Sampling year
4148_4289_E	*O. ullae*	Atlantic	Central America	Bahamas	23.7683	-76.13	NA	Sand	-7	March	1998
4148_4289_F	*O. nivalis*	Pacific	Asia	New Caledonia	-20.7517	165.2667	NA	Sand	-1	September	1993
4148_4289_H	*O.* sp. '66'	Pacific	Asia	New Caledonia	-20.9183	167.1283	NA	Sand	-10	November	2000
4148_4289_K	*C.* sp. 'nc'	Pacific	Asia	New Caledonia	-21.0333	167.425	NA	Sand	-7	November	2000
4148_4289_N	*O. crassitunicatus*	Pacific	South America	Peru	-12.7322	-77.1327	NA	Mud	NA	June	2000
4148_4289_O	*O.* sp. 'lees'	Atlantic	Central America	Bahamas	23.7679	-76.1317	NA	Sand	NA	April	2002
4148_4289_P	*O. finitimus*	Atlantic	Central America	Bahamas	23.7617	-76.1236	NA	Sand	-7	April	2002
4148_4289_Q	*O.* nr. *algarvensis*	Atlantic	Central America	Bahamas	23.762	-76.1227	NA	Sand	-6	April	2002
4148_4289_R	*P.* sp. 'xx'	Atlantic	Central America	Bahamas	23.7681	-76.1311	NA	NA	-6	April	2002
4148_4289_S	*P.* nr. *vicinus*	Atlantic	Central America	Bahamas	23.7733	-76.1046	NA	Sand	-6	April	2002
4148_4289_T	*O.* nr. *tenuissimus*	Atlantic	Central America	Bahamas	23.7897	-76.1388	NA	Sand	-6	April	2002
4148_4289_U	*O.* sp. '11'	Atlantic	Central America	Bahamas	23.7665	-76.0974	Seagrass	Sand	-6	April	2002
4148_4289_V	*O.* sp. '11'	Atlantic	Central America	Bahamas	23.7665	-76.0974	Seagrass	Sand	-3	April	2002
4148_4289_W	*O. tenuissimus*	Atlantic	Central America	Bahamas	23.773	-76.105	NA	Sand	-3	April	2002
4148_4289_X	*O. finitimus*	Atlantic	Central America	Bahamas	23.773	-76.105	NA	Sand	-3	April	2002
4148_4289_Y	*O. prodigus*	Indian	Australia	Western Australia	-32.0155	115.513	NA	Sand	-4	January	1991
4148_4289_Z	*C. regius*	Indian	Australia	Western Australia	-33.9947	122.2208	NA	Sand	-2.5	February	2003
IexuBAH1	*I. exumae*	Atlantic	Central America	Bahamas	23.774533	-76.09835	Corals	Sand	-2.5	April	2013
IexuBAH2	*I. exumae*	Atlantic	Central America	Bahamas	23.774533	-76.09835	Corals	Sand	-2.5	April	2013
IexuBAH3	*I. exumae*	Atlantic	Central America	Bahamas	23.774533	-76.09835	Corals	Sand	-2.5	April	2013
Ileu1BER1	*I. leukodermatus*	Atlantic	Central America	Bermuda	32.324033	-64.73855	Detritus	NA	-1	November	2009
Ileu1BER2	*I. leukodermatus*	Atlantic	Central America	Bermuda	32.324033	-64.73855	Detritus	NA	-1	November	2009
ImakBAH1	*I. makropetalos*	Atlantic	Central America	Bahamas	23.774533	-76.09835	Corals	Sand	-2.5	April	2013

Library	Host species	Ocean	Continent	Field site	Latitude	Longitude	Organic input	Sediment type	Water depth (m)	Sampling month	Sampling year
ImakBAH2	*I. makropetalos*	Atlantic	Central America	Bahamas	23.774533	-76.09835	Seagrass	Sand	-4	April	2013
ImanLIZ1	*I. manae*	Pacific	Australia	Queensland	-14.69	145.451667	NA	Sand	-3	February	2006
Isp1OAH13	*I. dutchae* A	Pacific	North America	Hawai'i	21.260833	-157.786698	Corals	NA	-1.6	October	2015
Isp1OAH14	*I. dutchae* A	Pacific	North America	Hawai'i	21.260833	-157.786698	Corals	NA	-1.6	October	2015
Isp1OAH30	*I. dutchae* A	Pacific	North America	Hawai'i	21.270441	-157.774833	NA	NA	-1.75	October	2015
Isp1OAH67	*I. dutchae* A	Pacific	North America	Hawai'i	21.283312	-157.846999	Corals	NA	-3	October	2015
Isp1OAH70	*I. dutchae* A	Pacific	North America	Hawai'i	21.283312	-157.846999	Corals	NA	-3	October	2015
Isp2OAH42	*I. dutchae* B	Pacific	North America	Hawai'i	21.468661	-157.818443	NA	NA	-8	October	2015
Isp2OAH49	*I. dutchae* B	Pacific	North America	Hawai'i	21.468661	-157.818443	Corals	NA	-1.5	October	2015
Isp2OAH54	*I. dutchae* B	Pacific	North America	Hawai'i	21.468661	-157.818443	Corals	NA	-1.5	October	2015
Isp2OAH85	*I. dutchae* B	Pacific	North America	Hawai'i	21.393579	-157.715667	Corals	NA	-1.8	October	2015
Isp2OAH91	*I. dutchae* B	Pacific	North America	Hawai'i	21.393579	-157.715667	Corals	NA	-1.8	October	2015
Isp2OAH98	*I. dutchae* B	Pacific	North America	Hawai'i	21.393579	-157.715667	Corals	NA	-1.8	October	2015
ItriBAH1	*I. triangulatus*	Atlantic	Central America	Bahamas	23.72595	-76.178733	NA	Sand	-6	April	2013
ItriBAH2	*I. triangulatus*	Atlantic	Central America	Bahamas	23.72595	-76.178733	NA	Sand	-6	April	2013
OalbHER1	*O. albidus*	Pacific	Australia	Queensland	-23.44344	151.91307	NA	Sand	-0.5	August	2012
OalbHER2	*O. albidus*	Pacific	Australia	Queensland	-23.44344	151.91307	NA	Sand	-0.5	August	2012
OalgACAV1	*O. algarvensis*	Mediterranean	Europe	Italy	42.734418	10.185232	Seagrass	Silicate sand	-7	May	2014
OalgACAV2	*O. algarvensis*	Mediterranean	Europe	Italy	42.734418	10.185232	Seagrass	Silicate sand	-7	May	2014
OalgACAV3	*O. algarvensis*	Mediterranean	Europe	Italy	42.734418	10.185232	Seagrass	Silicate sand	-7	May	2014
OalgASAN1	*O. algarvensis*	Mediterranean	Europe	Italy	42.807222	10.141111	Seagrass	Silicate sand	-6	March	2013
OalgASAN2	*O. algarvensis*	Mediterranean	Europe	Italy	42.807222	10.141111	Seagrass	Silicate sand	-6	March	2013
OalgASAN3	*O. algarvensis*	Mediterranean	Europe	Italy	42.807222	10.141111	Seagrass	Silicate sand	-6	March	2013
OalgBSAN1	*O. algarvensis*	Mediterranean	Europe	Italy	42.807222	10.141111	Seagrass	Silicate sand	-6	March	2013

Library	Host species	Ocean	Continent	Field site	Latitude	Longitude	Organic input	Sediment type	Water depth (m)	Sampling month	Sampling year
OalgBSAN2	*O. algarvensis*	Mediterranean	Europe	Italy	42.807222	10.141111	Seagrass	Silicate sand	-6	March	2013
OalgBSAN3	*O. algarvensis*	Mediterranean	Europe	Italy	42.807222	10.141111	Seagrass	Silicate sand	-6	March	2013
OclaLIZ1	*O. clavatus*	Pacific	Australia	Queensland	-14.686667	145.443333	NA	Sand	-3	February	2006
OclaLIZ2	*O. clavatus*	Pacific	Australia	Queensland	-14.686667	145.443333	Seagrass, corals	Sand	-3	February	2006
OcraPER1	*O. crassitunicatus*	Pacific	South America	Peru	-12.4265	-77.42	NA	Silt	-296	January	2013
OcraPER2	*O. crassitunicatus*	Pacific	South America	Peru	-12.4265	-77.42	NA	Silt	-296	January	2013
OfilPIA1	*O. filicauda*	Mediterranean	Europe	Italy	42.574217	10.066183	NA	Carbonate sand	-6	May	2012
OfilPIA2	*O. filicauda*	Mediterranean	Europe	Italy	42.574217	10.066183	NA	Carbonate sand	-6	May	2012
OgenHER1	*O. geniculatus*	Pacific	Australia	Queensland	-23.44575	151.91333	Corals, algae	Sand	-0.3	August	2012
OgenHER2	*O. geniculatus*	Pacific	Australia	Queensland	-23.44575	151.91333	Corals, algae	Sand	-0.3	August	2012
OgenLIZ1	*O. geniculatus*	Pacific	Australia	Queensland	-14.641667	145.491667	NA	Sand	-10	February	2006
OgenLIZ2	*O. geniculatus*	Pacific	Australia	Queensland	-14.641667	145.491667	Corals	Sand	-10	February	2006
OilvPIA1	*O. ilvae*	Mediterranean	Europe	Italy	42.574217	10.066183	NA	Carbonate sand	-6	May	2012
OilvPIA2	*O. ilvae*	Mediterranean	Europe	Italy	42.574217	10.066183	NA	Carbonate sand	-6	May	2012
OilvSAN1	*O. ilvae*	Mediterranean	Europe	Italy	42.80816	10.142104	Seagrass	Silicate sand	-6	June	2014
OilvSAN2	*O. ilvae*	Mediterranean	Europe	Italy	42.80816	10.142104	Seagrass	Silicate sand	-6	June	2014
OloiHER1	*O.* nr. *loisae*	Pacific	Australia	Queensland	-23.4349	151.945	Corals	Sand	-5	September	2012
OloiHER2	*O.* nr. *loisae*	Pacific	Australia	Queensland	-23.4349	151.945	Corals	Sand	-5	September	2012
OloiLIZ1	*O.* nr. *loisae*	Pacific	Australia	Queensland	-14.648333	145.451667	Corals	Sand	-3	February	2006
OloiLIZ2	*O.* nr. *loisae*	Pacific	Australia	Queensland	-14.648333	145.451667	NA	Sand	-3	February	2006
Osp1DAH1	*O.* sp. 'dahab1'	Red Sea	Africa	Egypt	28.505376	34.521874	Corals	NA	-6.5	October	2009
Osp1DAH2	*O.* sp. 'dahab1'	Red Sea	Africa	Egypt	28.505376	34.521874	Corals	NA	-6.5	October	2009
Osp1HER1	*O.* sp. 'heron1'	Pacific	Australia	Queensland	-23.4349	151.945	Corals	Sand	-7	September	2012
Osp1OAH1	*O.* sp. 'oahu1'	Pacific	North America	Hawai'i	21.468661	-157.818443	Corals	NA	-8	October	2015
Osp3HER1	*O.* sp. 'heron3'	Pacific	Australia	Queensland	-23.4349	151.945	Corals	Sand	-7	September	2012

Library	Host species	Ocean	Continent	Field site	Latitude	Longitude	Organic input	Sediment type	Water depth (m)	Sampling month	Sampling year
Osp4HER1	*O.* sp. 'heron4'	Pacific	Australia	Queensland	-23.4349	151.945	Corals	Sand	-7	September	2012
Osp5HER1	*O.* sp. 'heron5'	Pacific	Australia	Queensland	-23.4349	151.945	Corals	Sand	-7	September	2012
OvacBAH1	*O. vacuus*	Atlantic	Central America	Bahamas	23.770117	-76.130067	NA	Sand	-0.2	April	2013
OvacBAH2	*O. vacuus*	Atlantic	Central America	Bahamas	23.770117	-76.130067	NA	Sand	-0.2	April	2013

Table S5 | Library preparation, read length and sequencing depth.

Sample ID	Library type	Kit/protocol	Read length (bp)	Total number of reads
2699_C	DNA FS library	NEBNext Ultra™ II FS DNA Library Prep Kit for Illumina (New England Biolabs)	15-151	29,652,576
2699_D	DNA FS library	NEBNext Ultra™ II FS DNA Library Prep Kit for Illumina (New England Biolabs)	15-151	30,735,940
3557_A	TPase-based DNA library	Nextera[1]	150	40,047,495
3557_AA	TPase-based DNA library	Nextera[1]	150	27,659,643
3557_AB	TPase-based DNA library	Nextera[1]	150	29,622,588
3557_AC	TPase-based DNA library	Nextera[1]	150	34,035,721
3557_AD	TPase-based DNA library	Nextera[1]	150	28,490,778
3557_AE	TPase-based DNA library	Nextera[1]	150	29,861,937
3557_AF	TPase-based DNA library	Nextera[1]	150	28,925,848
3557_AG	TPase-based DNA library	Nextera[1]	150	28,767,847
3557_AH	TPase-based DNA library	Nextera[1]	150	32,910,337
3557_AI	TPase-based DNA library	Nextera[1]	150	28,658,785
3557_AJ	TPase-based DNA library	Nextera[1]	150	40,262,852
3557_AK	TPase-based DNA library	Nextera[1]	150	27,286,732
3557_AL	TPase-based DNA library	Nextera[1]	150	27,571,589
3557_AN	TPase-based DNA library	Nextera[1]	150	29,587,283
3557_AO	TPase-based DNA library	Nextera[1]	150	46,691,285
3557_AQ	TPase-based DNA library	Nextera[1]	150	26,128,896
3557_AR	TPase-based DNA library	Nextera[1]	150	27,876,817
3557_AS	TPase-based DNA library	Nextera[1]	150	34,272,688
3557_AT	TPase-based DNA library	Nextera[1]	150	32,855,300
3557_AU	TPase-based DNA library	Nextera[1]	150	27,059,681
3557_AV	TPase-based DNA library	Nextera[1]	150	28,992,675
3557_AW	TPase-based DNA library	Nextera[1]	150	38,822,941
3557_AX	TPase-based DNA library	Nextera[1]	150	28,510,164
3557_AY	TPase-based DNA library	Nextera[1]	150	27,246,999
3557_AZ	TPase-based DNA library	Nextera[1]	150	30,171,146
3557_B	TPase-based DNA library	Nextera[1]	150	27,415,254
3557_BA	TPase-based DNA library	Nextera[1]	150	34,391,476
3557_BB	TPase-based DNA library	Nextera[1]	150	28,133,352
3557_BC	TPase-based DNA library	Nextera[1]	150	33,545,030
3557_BD	TPase-based DNA library	Nextera[1]	150	35,916,006
3557_BE	TPase-based DNA library	Nextera[1]	150	31,932,933
3557_BF	TPase-based DNA library	Nextera[1]	150	28,160,431
3557_BG	TPase-based DNA library	Nextera[1]	150	27,284,749
3557_BH	TPase-based DNA library	Nextera[1]	150	29,458,703

Sample ID	Library type	Kit/protocol	Read length (bp)	Total number of reads
3557_BI	TPase-based DNA library	Nextera[1]	150	33,932,307
3557_BJ	TPase-based DNA library	Nextera[1]	150	26,901,444
3557_BK	TPase-based DNA library	Nextera[1]	150	28,080,579
3557_BL	TPase-based DNA library	Nextera[1]	150	32,792,752
3557_BM	TPase-based DNA library	Nextera[1]	150	28,890,076
3557_BN	TPase-based DNA library	Nextera[1]	150	27,415,015
3557_BO	TPase-based DNA library	Nextera[1]	150	26,707,895
3557_BP	TPase-based DNA library	Nextera[1]	150	32,904,747
3557_BQ	TPase-based DNA library	Nextera[1]	150	26,379,552
3557_BR	TPase-based DNA library	Nextera[1]	150	31,371,498
3557_BS	TPase-based DNA library	Nextera[1]	150	61,199,905
3557_BT	TPase-based DNA library	Nextera[1]	150	31,437,547
3557_BU	TPase-based DNA library	Nextera[1]	150	29,269,577
3557_BV	TPase-based DNA library	Nextera[1]	150	27,562,136
3557_BW	TPase-based DNA library	Nextera[1]	150	27,959,679
3557_BX	TPase-based DNA library	Nextera[1]	150	27,412,002
3557_BY	TPase-based DNA library	Nextera[1]	150	26,143,306
3557_BZ	TPase-based DNA library	Nextera[1]	150	25,474,515
3557_CA	TPase-based DNA library	Nextera[1]	150	34,116,895
3557_CB	TPase-based DNA library	Nextera[1]	150	33,448,043
3557_CC	TPase-based DNA library	Nextera[1]	150	29,498,251
3557_CD	TPase-based DNA library	Nextera[1]	150	27,526,191
3557_CE	TPase-based DNA library	Nextera[1]	150	27,381,070
3557_CF	TPase-based DNA library	Nextera[1]	150	36,783,504
3557_CG	TPase-based DNA library	Nextera[1]	150	37,742,276
3557_CH	TPase-based DNA library	Nextera[1]	150	32,576,089
3557_CI	TPase-based DNA library	Nextera[1]	150	25,487,402
3557_CJ	TPase-based DNA library	Nextera[1]	150	31,812,449
3557_CK	TPase-based DNA library	Nextera[1]	150	30,470,472
3557_CL	TPase-based DNA library	Nextera[1]	150	27,694,255
3557_CM	TPase-based DNA library	Nextera[1]	150	29,247,350
3557_CN	TPase-based DNA library	Nextera[1]	150	30,039,547
3557_CO	TPase-based DNA library	Nextera[1]	150	27,725,324
3557_CP	TPase-based DNA library	Nextera[1]	150	32,555,982
3557_CQ	TPase-based DNA library	Nextera[1]	150	28,548,310
3557_D	TPase-based DNA library	Nextera[1]	150	28,165,652
3557_E	TPase-based DNA library	Nextera[1]	150	26,975,648

Sample ID	Library type	Kit/protocol	Read length (bp)	Total number of reads
3557_F	TPase-based DNA library	Nextera[1]	150	30,343,070
3557_G	TPase-based DNA library	Nextera[1]	150	29,970,510
3557_H	TPase-based DNA library	Nextera[1]	150	32,240,084
3557_I	TPase-based DNA library	Nextera[1]	150	27,987,189
3557_J	TPase-based DNA library	Nextera[1]	150	29,102,865
3557_M	TPase-based DNA library	Nextera[1]	150	28,502,749
3557_N	TPase-based DNA library	Nextera[1]	150	26,976,480
3557_O	TPase-based DNA library	Nextera[1]	150	29,422,418
3557_P	TPase-based DNA library	Nextera[1]	150	29,171,965
3557_Q	TPase-based DNA library	Nextera[1]	150	28,412,340
3557_R	TPase-based DNA library	Nextera[1]	150	27,174,873
3557_S	TPase-based DNA library	Nextera[1]	150	27,901,558
3557_T	TPase-based DNA library	Nextera[1]	150	28,977,512
3557_U	TPase-based DNA library	Nextera[1]	150	29,163,333
3557_V	TPase-based DNA library	Nextera[1]	150	28,937,322
3557_W	TPase-based DNA library	Nextera[1]	150	26,842,360
3557_X	TPase-based DNA library	Nextera[1]	150	29,049,358
3557_Y	TPase-based DNA library	Nextera[1]	150	34,576,047
3557_Z	TPase-based DNA library	Nextera[1]	150	27,635,935
3585_CK_4019_A	TPase-based DNA library	Nextera[1]	150	88,274,983
3585_CL_4019_B	TPase-based DNA library	Nextera[1]	150	86,692,003
3585_CM_4019_C	TPase-based DNA library	Nextera[1]	150	82,352,251
3585_CN_4019_D	TPase-based DNA library	Nextera[1]	150	82,353,539
3585_CO_4019_E	TPase-based DNA library	Nextera[1]	150	94,705,999
3585_CP_4019_F	TPase-based DNA library	Nextera[1]	150	91,692,483
3585_CQ_4019_G	TPase-based DNA library	Nextera[1]	150	81,251,827
3586_CC_4019_H	TPase-based DNA library	Nextera[1]	150	81,761,548
3586_CD_4019_I	TPase-based DNA library	Nextera[1]	150	80,392,342
3586_CE_4019_J	TPase-based DNA library	Nextera[1]	150	113,753,316
3596_CF_4019_K	TPase-based DNA library	Nextera[1]	150	74,005,196
3586_CG_4019_L	TPase-based DNA library	Nextera[1]	150	128935071
3586_CH_4019_M	TPase-based DNA library	Nextera[1]	150	86,365,252
3586_CI_4019_N	TPase-based DNA library	Nextera[1]	150	82,633,856
3630_C	TPase-based DNA library	Nextera[1]	150	39,704,844
3630_D	TPase-based DNA library	Nextera[1]	150	20,911,336
3630_E	TPase-based DNA library	Nextera[1]	150	52,396,753
3630_F	TPase-based DNA library	Nextera[1]	150	52,341,902

Sample ID	Library type	Kit/protocol	Read length (bp)	Total number of reads
3630_G	TPase-based DNA library	Nextera[1]	150	93,431,784
3630_H	TPase-based DNA library	Nextera[1]	150	22,426,953
3630_I	TPase-based DNA library	Nextera[1]	150	33,042,927
3630_J	TPase-based DNA library	Nextera[1]	150	54,833,175
3630_K	TPase-based DNA library	Nextera[1]	150	50,400,755
3630_L	TPase-based DNA library	Nextera[1]	150	51,359,581
3630_M	TPase-based DNA library	Nextera[1]	150	60,551,693
3630_N	TPase-based DNA library	Nextera[1]	150	68,603,218
3630_O	TPase-based DNA library	Nextera[1]	150	69,511,061
3854_4020_A	Ultra Low Input DNA library	Ovation® Ultralow Library Systems V2 (NuGEN)	150	111,211,294
3854_4020_C	DNA FS library	NEBNext Ultra™ II FS DNA Library Prep Kit for Illumina (New England Biolabs)	150	94,981,136
3854_4020_D	DNA FS library	NEBNext Ultra™ II FS DNA Library Prep Kit for Illumina (New England Biolabs)	150	88,886,899
3854_4020_E	DNA FS library	NEBNext Ultra™ II FS DNA Library Prep Kit for Illumina (New England Biolabs)	150	91,980,226
4148_4289_AA	TPase-based DNA library	Nextera[1]	150	72,185,937
4148_4289_AC	TPase-based DNA library	Nextera[1]	150	72,310,643
4148_4289_AD	TPase-based DNA library	Nextera[1]	150	74,992,626
4148_4289_AG	TPase-based DNA library	Nextera[1]	150	66,708,749
4148_4289_AI	TPase-based DNA library	Nextera[1]	150	77,986,887
4148_4289_AJ	TPase-based DNA library	Nextera[1]	150	75,635,935
4148_4289_AK	TPase-based DNA library	Nextera[1]	150	71,187,507
4148_4289_AM	TPase-based DNA library	Nextera[1]	150	70,539,848
4148_4289_AO	TPase-based DNA library	Nextera[1]	150	75,335,621
4148_4289_AS	TPase-based DNA library	Nextera[1]	150	76,378,700
4148_4289_AU	TPase-based DNA library	Nextera[1]	150	72,031,807
4148_4289_AV	TPase-based DNA library	Nextera[1]	150	77,382,844
4148_4289_AZ	TPase-based DNA library	Nextera[1]	150	81,318,544
4148_4289_B	TPase-based DNA library	Nextera[1]	150	82,260,122
4148_4289_BA	TPase-based DNA library	Nextera[1]	150	70,320,251
4148_4289_BB	TPase-based DNA library	Nextera[1]	150	72,392,712
4148_4289_BC	TPase-based DNA library	Nextera[1]	150	73,231,301
4148_4289_BD	TPase-based DNA library	Nextera[1]	150	74,658,666
4148_4289_BF	TPase-based DNA library	Nextera[1]	150	71,692,295
4148_4289_BG	TPase-based DNA library	Nextera[1]	150	81,803,569
4148_4289_BH	TPase-based DNA library	Nextera[1]	150	23,578,734
4148_4289_BJ	TPase-based DNA library	Nextera[1]	150	62,560,299
4148_4289_BK	TPase-based DNA library	Nextera[1]	150	74,106,795
4148_4289_BL	TPase-based DNA library	Nextera[1]	150	75,361,279

Sample ID	Library type	Kit/protocol	Read length (bp)	Total number of reads
4148_4289_BS	TPase-based DNA library	Nextera[1]	150	73,105,983
4148_4289_BT	TPase-based DNA library	Nextera[1]	150	70,308,897
4148_4289_BU	TPase-based DNA library	Nextera[1]	150	77,706,367
4148_4289_BV	TPase-based DNA library	Nextera[1]	150	80,959,443
4148_4289_BW	TPase-based DNA library	Nextera[1]	150	74,048,036
4148_4289_BX	TPase-based DNA library	Nextera[1]	150	72,549,537
4148_4289_BY	TPase-based DNA library	Nextera[1]	150	66,549,010
4148_4289_BZ	TPase-based DNA library	Nextera[1]	150	73,800,046
4148_4289_C	TPase-based DNA library	Nextera[1]	150	82,219,420
4148_4289_CE	TPase-based DNA library	Nextera[1]	150	81,744,875
4148_4289_CF	TPase-based DNA library	Nextera[1]	150	74,067,870
4148_4289_CH	TPase-based DNA library	Nextera[1]	150	75,279,473
4148_4289_CI	TPase-based DNA library	Nextera[1]	150	75,487,698
4148_4289_CM	TPase-based DNA library	Nextera[1]	150	61,466,733
4148_4289_CN	TPase-based DNA library	Nextera[1]	150	78,414,783
4148_4289_CO	TPase-based DNA library	Nextera[1]	150	68,771,024
4148_4289_D	TPase-based DNA library	Nextera[1]	150	59,638,816
4148_4289_E	TPase-based DNA library	Nextera[1]	150	88,882,071
4148_4289_F	TPase-based DNA library	Nextera[1]	150	89,938,092
4148_4289_H	TPase-based DNA library	Nextera[1]	150	87,992,617
4148_4289_K	TPase-based DNA library	Nextera[1]	150	83,939,191
4148_4289_N	TPase-based DNA library	Nextera[1]	150	84,797,380
4148_4289_O	TPase-based DNA library	Nextera[1]	150	84,319,368
4148_4289_P	TPase-based DNA library	Nextera[1]	150	65,687,609
4148_4289_Q	TPase-based DNA library	Nextera[1]	150	82,274,232
4148_4289_R	TPase-based DNA library	Nextera[1]	150	66,306,156
4148_4289_S	TPase-based DNA library	Nextera[1]	150	66,753,040
4148_4289_T	TPase-based DNA library	Nextera[1]	150	77,751,084
4148_4289_U	TPase-based DNA library	Nextera[1]	150	76,647,505
4148_4289_V	TPase-based DNA library	Nextera[1]	150	74,225,111
4148_4289_W	TPase-based DNA library	Nextera[1]	150	84,279,992
4148_4289_X	TPase-based DNA library	Nextera[1]	150	71,716,514
4148_4289_Y	TPase-based DNA library	Nextera[1]	150	69,564,169
4148_4289_Z	TPase-based DNA library	Nextera[1]	150	75,648,469
IexuBAH1	Low Input (DNA)	Illumina Low Input Fragment - Nextera	150	194,332,856
IexuBAH2	Low Input (DNA)	Illumina Low Input Fragment - Nextera	150	113,332,930
IexuBAH3	Low Input (DNA)	Illumina Low Input Fragment - Nextera	150	77,344,456
Ileu1BER1	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	152,817,786

Sample ID	Library type	Kit/protocol	Read length (bp)	Total number of reads
Ileu1BER2	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	104,567,430
ImakBAH1	Low Input (DNA)	Illumina Low Input Fragment - Nextera	150	137,078,218
ImakBAH2	Low Input (DNA)	Illumina Low Input Fragment - Nextera	150	141,144,546
ImanLIZ1	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	98,404,426
Isp1OAH13	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	102,357,352
Isp1OAH14	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	159,402,400
Isp1OAH30	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	142,940,116
Isp1OAH67	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	75,848,112
Isp1OAH70	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	72,627,816
Isp2OAH42	Regular (DNA)	Illumina Regular Fragment, 270 bp	151	120,735,000
Isp2OAH49	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	110,912,064
Isp2OAH54	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	60,526,134
Isp2OAH85	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	190,486,762
Isp2OAH91	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	111,049,958
Isp2OAH98	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	67,777,428
ItriBAH1	Low Input (DNA)	Illumina Low Input Fragment - Nextera	150	104,748,702
ItriBAH2	Low Input (DNA)	Illumina Low Input Fragment - Nextera	150	63,243,218
OalbHER1	Regular (DNA)	Illumina Regular Fragment, 270 bp	150	89,450,896
OalbHER2	Regular (DNA)	Illumina Regular Fragment, 270 bp	150	71,514,112
OalgACAV1	Regular (DNA)	Illumina Regular Fragment, 270 bp	150	74,627,694
OalgACAV2	Regular (DNA)	Illumina Regular Fragment, 270 bp	150	172,626,642
OalgACAV3	Regular (DNA)	Illumina Regular Fragment, 270 bp	150	78,280,702
OalgASAN1	Regular (DNA)	Illumina Regular Fragment, 270 bp	150	110,986,108
OalgASAN2	Regular (DNA)	Illumina Regular Fragment, 270 bp	151	76,859,984
OalgASAN3	Regular (DNA)	Illumina Regular Fragment, 270 bp	150	169,567,918
OalgBSAN1	Regular (DNA)	Illumina Regular Fragment, 270 bp	151	84,347,270
OalgBSAN2	Regular (DNA)	Illumina Regular Fragment, 270 bp	150	107,228,142
OalgBSAN3	Regular (DNA)	Illumina Regular Fragment, 270 bp	151	89,370,480
OclaLIZ1	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	105,774,608
OclaLIZ2	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	93,691,302
OcraPER1	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	120,435,470
OcraPER2	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	84,591,598
OfilPIA1	Regular (DNA)	Illumina Regular Fragment, 270 bp	150	73,015,632
OfilPIA2	Low Input (DNA)	Illumina Low Input Fragment - Nextera	150	129,740,838
OgenHER1	Low Input (DNA)	Illumina Low Input Fragment - Nextera	150	189,254,650
OgenHER2	Low Input (DNA)	Illumina Low Input Fragment - Nextera	150	166,592,028
OgenLIZ1	Low Input (DNA)	Illumina Low Input Fragment - Nextera	150	139,402,742
OgenLIZ2	Low Input (DNA)	Illumina Low Input Fragment - Nextera	54-150	155,046,972
OilvPIA1	Regular (DNA)	Illumina Regular Fragment, 270 bp	150-151	290,322,406
OilvPIA2	Regular (DNA)	Illumina Regular Fragment, 270 bp	151	124,905,618
OilvSAN1	Regular (DNA)	Illumina Regular Fragment, 270 bp	150	74,577,460
OilvSAN2	Regular (DNA)	Illumina Regular Fragment, 270 bp	150	72,972,930
OloiHER1	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	127,327,778

Sample ID	Library type	Kit/protocol	Read length (bp)	Total number of reads
OloiHER2	Regular (DNA) and Low Input (DNA)	Illumina Regular Fragment, 270 bp and Illumina Low Input Fragment - Nextera	150-151	135,825,690
OloiLIZ1	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	84,230,968
OloiLIZ2	Low Input (DNA)	Illumina Low Input Fragment - Nextera	54-150	57,644,806
Osp1DAH1	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	116,652,266
Osp1DAH2	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	141,336,876
Osp1HER1	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	57,973,242
Osp1OAH1	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	207,606,362
Osp3HER1	Low Input (DNA)	Illumina Low Input Fragment - Nextera	151	88,350,288
Osp4HER1	Low Input (DNA)	Illumina Low Input Fragment - Nextera	54-150	109,652,754
Osp5HER1	Low Input (DNA)	Illumina Low Input Fragment - Nextera	54-150	62,619,502
OvacBAH1	Regular (DNA) and Low Input (DNA)	Illumina Regular Fragment, 270 bp and Illumina Low Input Fragment - Nextera	151	255,667,936
OvacBAH2	Regular (DNA)	Illumina Regular Fragment, 270 bp	150	75,165,886

Notes

Note S1: Curation of symbiont clades

To exclude potential contamination in our dataset, we only considered clades for further analyses that were found in the majority of individuals in at least one host species. That means only clades that were detected in at least 50% of at least two host individuals of one host species were considered for further analyses. Additionally, we also excluded singletons and clades that only reached the 'majority status' in species with just two host individuals sequenced but not in other host species of which more individuals were available. Also, we excluded clades that only reached the 'majority status' in libraries generated in one out of the collaborating sequencing centers but not in the same species sequenced at the other facility. In contrast, we also considered two clades that only reached the majority status in their given host species at one but not all sampling locations.

Note S2: Phylogenetic background of gutless oligochaete symbiont clades

To identify other hosts or environments that members of these symbiont clades are associated with, we screened publicly available 16S rRNA gene sequence data, and analyzed the phylogenetic relations between the symbionts and their closest relatives (Fig. S4-S36). We found that 23 of the 33 gutless oligochaete symbiont clades consisted exclusively of gutless oligochaete symbionts. Two out of these 23 clades represented sister clades to potential symbionts of Stilbonematinae nematodes (Gamma5 and Gamma7). In addition, two symbiont clades were phylogentically intermixed with symbionts from animals of the co-existing meiofauna: as previously reported, the Gamma1 symbionts were also associated with Stilbonematinae and *Astomonema* nematodes and in addition, the Gamma4 symbionts were also associated with *Kentrophoros* ciliates. The symbionts from the remaining eight clades were phylogenetically intermixed with bacteria that were previously detected in environmental samples (Alpha12, Alpha14, Delta1, Delta3, Delta4, Delta11, Gamma3 and Spiro1). The degree

to which these gutless oligochaete symbionts were phylogenetically intermixed with free-living relatives varied between the clades and could point towards different modes of acquisition and inheritance across the host diversity (Note S4).

To understand whether gutless oligochaete symbiont clades were associated specifically with gutless hosts or also occurred in related marine oligochaetes in general we screened nine specimens that were morphologically identified as members of the closely related gut-bearing Phallodrilinae. Two symbiont clades, the Alpha3 and Alpha8, could also be detected in association with these gut-bearing relatives, indicating that most of the symbiont clades are linked to the gutless lifestyle of their hosts (Figure S3). Based on these results, we grouped the symbiont clades into three categories: i) symbionts only associated with gutless oligochaetes, ii) symbionts also associated with other marine invertebrates and iii) symbionts that are phylogenetically intermixed with populations of free-living bacteria.

Note S3: Symbiont-symbiont interactions

Besides host traits, symbiont-symbiont interactions could alter community composition as shown in other highly specific symbiont consortia in e.g. plant hosts[2]. Therefore, we tested for linked co-occurrences as well as linked exclusion patterns of symbiont clades using an unbiased network analysis. We found no examples for symbiont exclusion suggesting that community composition is not based on symbiont-symbiont competition. In addition, we found six linked co-occurrences of which four were detected between symbiont clades that co-existed in only a single host species. The other two were detected between symbiont clades that co-existed in closely related sister species of hosts. The low number of stable positive interactions that were limited to symbiont clades present in a single clade of hosts indicate that overall symbiont-symbiont interactions only play a very minor role in mediating community composition across gutless oligochaetes.

Note S4: *De novo* symbiont acquisition became more frequent over time

We analyzed the evolutionary history of gutless oligochaetes based on their nuclear phylogeny and by dating the last common ancestor of gutless oligochaetes in the larger phylogenetic context of annelids using a Bayesian phylogenetic framework and a relaxed molecular clock model. We analyzed the evolutionary history of the hosts within the annelida based on the nuclear genome due to better resolution of basal nodes. Our data suggest that gutless oligochaetes diverged between 422.7252–97.1542 Mya years ago (median estimated divergence time: 256.131 Mya , Figure 2). We assume that from then on, the clade underwent a rapid radiation that led to very low phylogenetic resolution at basal nodes of the tree (Suppl. Figure 1).

Similarly, we used a Bayesian phylogenetic framework and a relaxed molecular clock model to estimate the ages of last common ancestors of each symbiont clade. The comparison of host and symbiont clade divergence times indicated two larger time windows for the primary acquisition of each symbiont clade (Figure 2). Primary symbiont acquisition could have happened either alongside host radiation or after host radiation. As symbionts are known to enable the radiation of their hosts, clades that were acquired alongside host radiation could have played a major role in the gutless oligochaete radiation and could be in the progress of co-evolution[3]. We grouped symbiont clades based on their age and diversification in relation to the divergence time of the hosts: symbiont clades that diversified before the host radiation, symbiont clades that diversified with the host and symbiont clades that diversified after host radiation.

Clades that had already diversified before the host radiation were Gamma1, Delta1, Alpha5 and Alpha14 and the hosts likely sampled these bacteria from an already given diversity (Figure 2). For these clades, we cannot estimate the timing of the association with gutless oligochaetes

based on age analyses alone. However, ancestral state reconstructions suggest that the Gamma1 clade was the first to establish the symbiosis, potentially initiated the host radiation (Table S3). The other three clades likely entered the symbiosis at a later point of the host radiation (Table S3).

Eleven clades diversified alongside the host radiation, suggesting the possibility that the last common ancestors of these clades were taken up by gutless oligochaete ancestors and co-evolved with the hosts by a combination of inheritance, host switching into other host lineages and potential losses (Figure 2, Table S3). Alternatively, the symbiont clades could have evolved independently in the same time frame and members of the clades could have independently associated with gutless oligochaetes several times. However, this appears unlikely for the majority of the clades as their exclusive associations with gutless oligochaetes indicate host-associated lifestyles of these clades (9 of 11). One notable exception and a good case for independent evolution is the Gamma3 clade that is phylogenetically intermixed with environmental relatives (Figure S31).

The remaining 16 symbiont clades are estimated to have diverged after the gutless oligochaete radiation. These symbiont clades were likely taken up by distinct host lineages at a later point of the evolution of gutless oligochaetes and, if at all, only rarely switched between hosts (Figure 2, Table S3). The only exception was the Delta4 symbiont which was associated with a much broader variety of host species than all the other ‘young’ symbiont clades, suggesting either frequent uptake or host switching events.

Overall, it appears that the mode of acquisition and transfer of symbionts in gutless oligochaetes changed over time. We observed a low number of old symbiont clades of which all have a broad distribution across the host diversity. This is contrasted with an increasing number of *de novo*

acquisitions of young symbiont clades with narrow host distributions. The increase of symbiont acquisitions might seem counter intuitive as they happened after the loss of the digestive and excretory organs including the oral opening. The oral opening is one of the most common entering point for symbionts, however, symbiont acquisition via other routes such as sexual reproduction or the skin are common, even in animals that bear a gut as an adult[4]. One factor that might have favored increased symbiont acquisition could be that the obligate symbiosis altered host traits such as the immune response, making it more feasible for new bacterial clades to enter the association[5,6]. In addition, symbiont acquisition of such young clades might allow co-occurring hosts to exploit different microhabitats, thus reducing interspecies competition[7,8].

The low host range of new symbiont clades indicates little to no host switching of these clades. As hosts diverged over time, divergent host traits could more and more prevent host switching of such a new clade from one species to a distantly related one. In addition, geological changes confound the ability of host switches over the long evolutionary history of the hosts. The gutless oligochaetes likely radiated around 256.131 Mya ago and at that time the major oceans were circumtropically connected facilitating gene and symbiont flow (Figure 2). This connectivity started to deteriorate with the closure of the Neotethys in the Neogene and was finally interrupted with the closure of the Isthmus of Panama at the beginning of the Pleistocene. Therefore, recently acquired symbiont clades can only reach a limited geographic and are likely limited in the phylogenetic spectrum of host that they can infect.

The higher number of young clades underlines the importance of the recent time window for symbiont clade establishment, suggesting accelerating evolutionary flexibility and potential specializations (Figure 2B).

Extended data

The host and symbiont phylogenies that were generated and used during this study are available as newick files under https://itol.embl.de/shared/cdlzpFI3Qqt4

References

1. Hennig, B. P. *et al.* Large-Scale Low-Cost NGS Library Preparation Using a Robust Tn5 Purification and Tagmentation Protocol. *G3 Genes|Genomes|Genetics* **8**, 79–89 (2018).
2. Uroz, S., Courty, P. E. & Oger, P. Plant Symbionts Are Engineers of the Plant-Associated Microbiome. *Trends Plant Sci.* **24**, 905–916 (2019).
3. Mitter, C., Farrell, B. & Wiegmann, B. The Phylogenetic Study of Adaptive Zones: Has Phytophagy Promoted Insect Diversification? *Am. Nat.* **132**, 107–128 (1988).
4. Bright, M. & Bulgheresi, S. A complex journey: transmission of microbial symbionts. *Nat. Rev. Microbiol.* **8**, 218–230 (2010).
5. Gerardo, N. M. *et al.* Immunity and other defenses in pea aphids, Acyrthosiphon pisum. *Genome Biol.* **11**, R21 (2010).
6. Douglas, A. E., Bouvaine, S. & Russell, R. R. How the insect immune system interacts with an obligate symbiotic bacterium. *Proc. R. Soc. B Biol. Sci.* **278**, 333–338 (2011).
7. Feldhaar, H. Bacterial symbionts as mediators of ecologically important traits of insect hosts. *Ecol. Entomol.* **36**, 533–543 (2011).
8. Wall, C. B., Kaluhiokalani, M., Popp, B. N., Donahue, M. J. & Gates, R. D. Divergent symbiont communities determine the physiology and nutrition of a reef coral across a light-availability gradient. *ISME J.* **14**, 945–958 (2020).

Chapter IV | Simple animals, simple microbiomes? The symbiont diversity of placozoans

Anna Mankowski[1], Nikolaus Leisch[1], Michael G. Hadfield[2], Bruno Hüttel[3], Nicole Dubilier[*1], Harald R. Gruber-Vodicka[*1]

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

[2]Kewalo Marine Laboratory, Pacific Biosciences Research Centre, University of Hawai'i at Mānoa, Honolulu, HI, USA

[3] Max Planck Genome Centre Cologne, Max Planck Institute for Plant Breeding Research, Cologne, Germany

[*]Corresponding authors:

Nicole Dubilier: ndubilie@mpi-bremen.de, phone +49 (421) 2028 9320

Harald R. Gruber-Vodicka: hgruber@mpi-bremen.de, phone +49 (421) 2028 7600

Author contributions: AM, HRGV and ND conceived the study, and AM designed the workflows with input from HRGV. HRGV, NL, MGH and BH acquired placozoan specimens and generated metagenomic data. AM analyzed the data with input from HRGV and wrote the manuscript with contributions from HRGV.

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

The author order is not fixed.

Abstract

Placozoa is the phylum of the arguably simplest animals that diverged early during metazoan evolution. Due to their phylogenetic placement and their simplistic bodies, they are considered important models to understand metazoan evolution and development. Several studies noted the association of placozoans with bacterial symbionts, yet the phylogenetic composition of the placozoan microbiomes, their evolution as well as the interactions with the hosts remain mostly unknown. Here, we show that phylogenetically diverse and globally distributed placozoans are associated with variable microbiomes. Several of the detected partners appear to be intimate, intracellular symbionts that could largely influence their hosts. Thus, we argue that understanding placozoan biology requires the analysis of the associated microbiota. In addition, we hypothesize that placozoan microbiomes could be used as a model to understand the evolution of diversity in animal microbiomes.

Introduction

Placozoa is a phylum of globally distributed, benthic marine invertebrates that diverged close to the base of the animal phylogeny[1–3]. A recent study suggests that they, together with the Cnidaria, are the sister clade to the Bilateria[1]. The phylum Placozoa includes three described species, *Trichoplax adhaerens*, *Hoilungia hongkongensis* and *Polyplacotoma mediterranea,* but analysis of mitochondrial haplotypes indicated a diversity of 19 cryptic species that are morphologically indistinguishable[3–6]. The bodies of placozoans are millimeter-sized, disc-shaped and consist of only six cell types that form three distinct layers. The ventral epithelium consists of ciliated epithelial cells and occasionally lipophilic and glandular cells. The dorsal epithelium also consists of ciliated epithelial cells as well as crystal cells. The epithelial layers are connected by a meshwork of fiber cells[7–9]. Placozoans do not have an internal digestive system. Instead, they externally digest algae or microbial biofilms and take up the released nutrients via their ventral epithelium[10,11]. Given their early divergence in the evolution of

animals, their simplistic bodies, their global distribution as well as the fact that they can easily be sampled and cultivated under laboratory conditions, they are a powerful model system to understand metazoan evolution, developmental biology and tissue formation[3,12,13].

As most other animals, placozoans were found to live in association with symbiotic bacteria. Across the placozoan host diversity, the association with intracellular Rickettsiales symbionts appear to be strikingly common[5,9,12,14–18]. In addition, a distinct haplotype from Hawai'i was associated with a second, unrelated intracellular symbiont that belongs to the Margulisbacteria (*Cand.* Ruthmannia) and metagenome screening for bacterial 16S rRNA genes indicated that placozoans could be associated with various, yet undescribed, symbiotic bacteria[14]. So far, little is known about the microbiome composition of placozoans as well as the phylogenetic relations between symbionts from different host types and their interactions with the hosts.

Here, we screened 109 placozoan metagenomes from eight haplotypes and eight sampling sites for bacterial symbionts. In addition, we treated some of the placozoan hosts with antibiotics to test whether the placozoan microbiome could be manipulated and which effects these manipulations have on the microbiome composition. In total, we identified 24 bacterial clades associated with placozoan. The microbiome of each host individual consisted of one to 13 of these bacterial clades that commonly included five symbiont clades that are most likely intracellular symbionts, including the previously recognized Rickettsiales and *Cand.* Ruthmannia symbionts as well as members of the Chlamydiae and Coxiellaceae[5,9,12,14–21]. Our data suggests that the microbiome composition between individuals is a result of several factors, including sampling location, host haplotype, environment as well as targeted manipulation with antibiotics. Each of these factors appeared to differently influence the presence/absence of individual symbiont clades and in consequence, the overall microbiome composition.

Results and discussion

The *Placozoa* are associated with diverse microbiomes

In order to describe the microbiome of placozoans across a broad biogeographic host range, we analyzed metagenomes of 109 host individuals from eight different host mitochondrial 16S rRNA genotypes (m16S rRNA genotypes, host haplotypes from here on, Note S1) and eight globally distributed sampling locations (Figure S1, Table S1). We identified host haplotypes by reconstructing full-length mitochondrial 16S rRNA gene sequences and inferring phylogenetic relationships between our samples and reference sequences (Figure S2)[3]. For the description of the associated microbiomes, we reconstructed full-length symbiont 16S rRNA genes and defined genus-level symbiont clades by clustering the sequences at 95% sequence identity. Our initial profiling resulted in 77 genus-level symbiont clades. We reconstructed 16S rRNA gene sequences of all 77 clades, analyzed their phylogenetic background and relations and estimated their relative abundances in each host individual (Extended Data, Table S2). For the final description of the symbiont community composition, we excluded two groups of clades: clades that are likely human or flow cell contamination and singletons, i.e. clades that were only detected a single time across all samples (Table S2). After contamination removal, we identified 63 bacterial clades that were associated with Placozoan hosts, including the previously described intracellular Rickettsiales and Margulisbacteria symbionts (Figure 1, Table S3)[5,9,12,14–18].

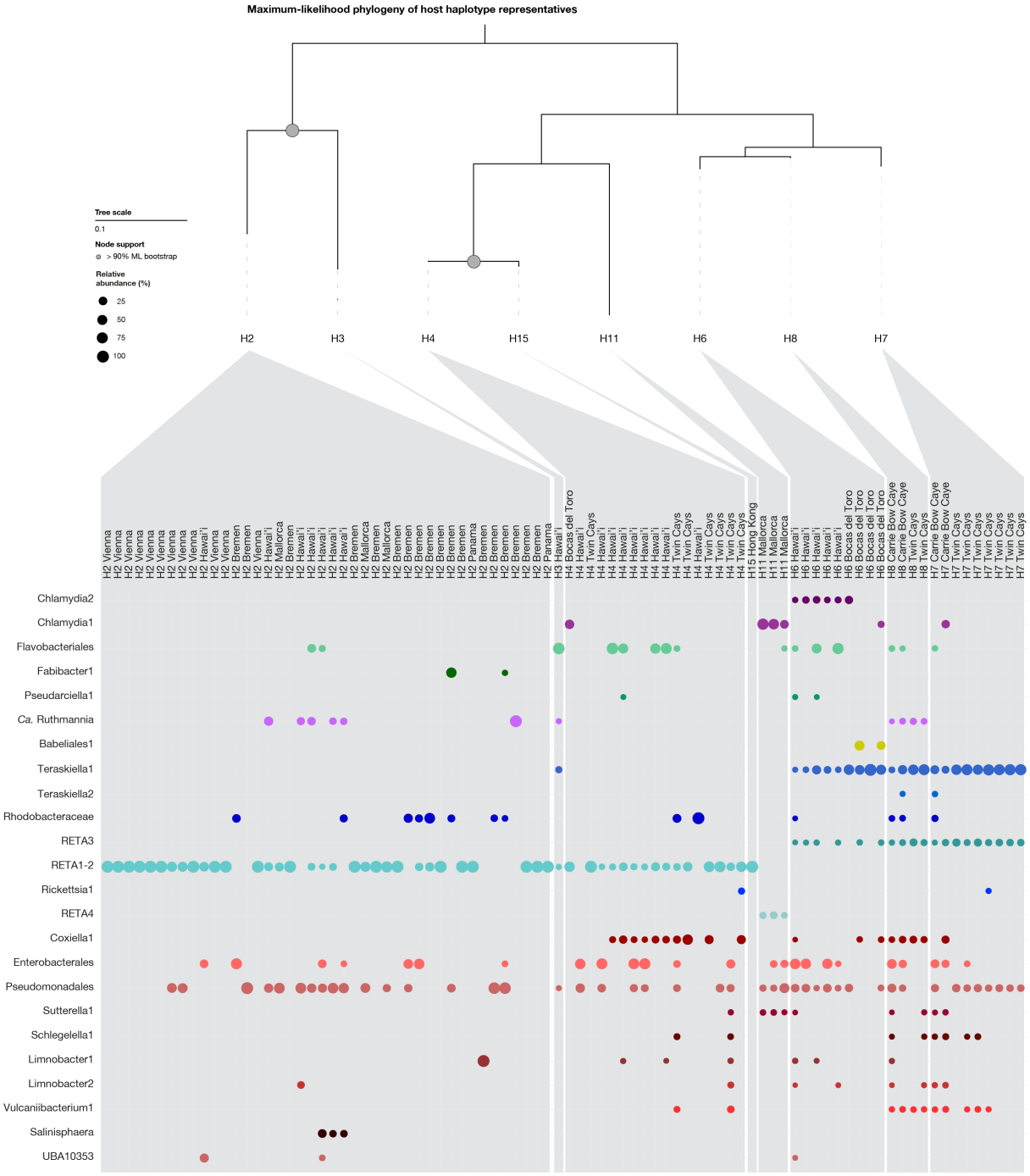


Figure 1 | 24 bacterial groups form variable microbiomes across the diversity of placozoan haplotypes. Top tree: Maximum-likelihood trees of the host m16S rRNA gene phylogeny including one individual per host haplotype. The scale bar indicates 10% estimated sequence divergence. Nodes with none-parametric bootstrap support > 90% are highlighted. Middle panel: Averaged relative abundance of symbiont clades per host individual as estimated with EMIRGE.

When analyzing the phylogenetic relations between the symbiont clades, we observed that several clades belonged to the orders of Flavobacteriales, Pseudomonadales and Enterobacterales as well as the family of Rhodobacteraceae (Table S2). The clades from each

of these groups did not form separate, monophyletic groups but were phylogenetically intermixed (Extended Data). This lack of phylogenetic clustering is likely caused by the presence of several bacterial phylotypes of each of these taxa in every sample that could not be sufficiently resolved with our short-read sequencing approach. In contrast, we usually observe high specificity between a certain genotype of each of the symbiont clades and the host individual. Thus, we assume that members of the Flavobacteriales, Pseudomonadales, Enterobacterales and Rhodobacteraceae are not intimate Placozoan symbionts but rather members of the diverse free-living microbial community that could attach to the host individuals and commensalistically feed on the nutrients that are produced during the hosts' external digestion of algae and biofilms[10,11]. Due to the lack of phylogenetic clustering of the clades from these taxa, we respectively combined these clades for further analyses (Figure 1).

Diverse ecological and evolutionary dynamics influence symbiont community composition

One central question of microbiome research is *which factors drive the phylogenetic composition of microbiomes?* Here, we tested the effects of host phylogeny, sampling location, habitat (aquarium vs. free-living) as well as antibiotic treatment on the composition of the placozoans' microbiomes. Overall, it appeared that the tested factors only weakly influence the variability of the placozoan microbiome composition. Among these four factors, sampling site had the highest explanatory power for variability in community composition (18.682%, p-value=0.001), followed by host haplotype (12.14%, p-value=0.029), antibiotics treatment (9.594%, p-value=0.01) and habitat (2.914%, p-value=0.034, Figure 2).

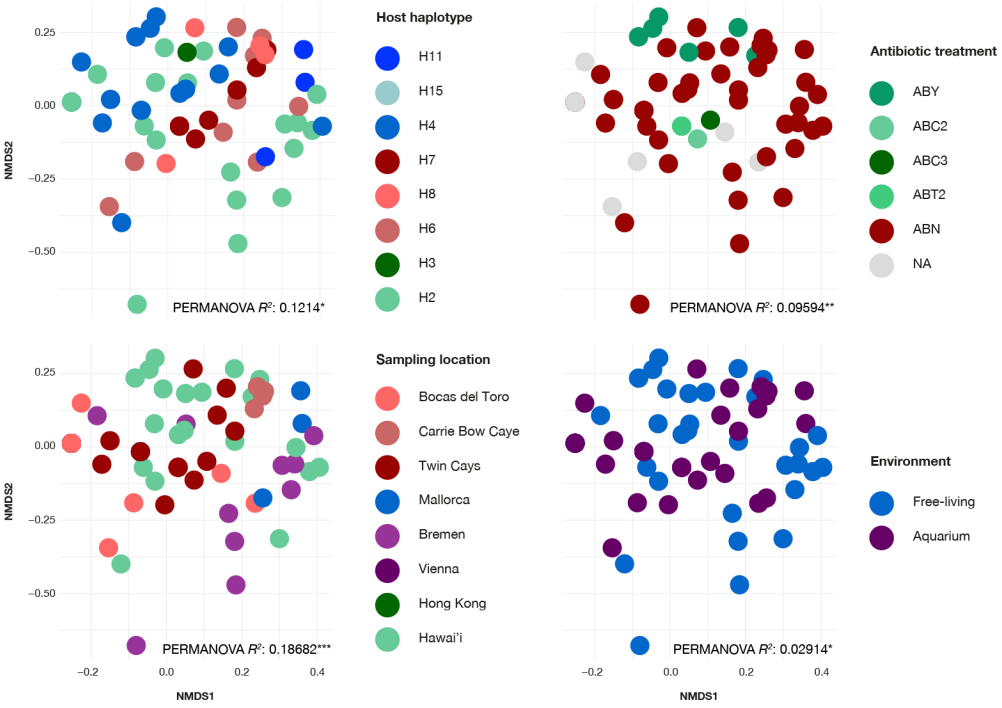


Figure 2: The microbiome composition of placozoans is the result of several ecological and evolutionary factors. All panels show the same NMDS plot of UniFrac dissimilarity values calculated from estimated relative abundances of symbiont clades in different host individuals. UniFrac stress value: 0.1999552. Different panels highlight samples differently according to certain metadata categories. For each metadata category, PERMANOVA r^2 values are indicated within the respective plot. Statistical significance by p-values is indicated by the following significant codes: * 0.01 ** 0.001 *** 0.00.

In addition, we tested whether more closely related host individuals shared more similar symbiont communities than more distantly related host species by calculating the congruence between the host phylogenetic tree and a dendrogram representing the dissimilarities of microbiome composition. In concordance with the rather low explanatory power of host haplotype for microbiome composition, we observed several incongruences between the trees, indicating that the variability of microbiome composition is not linked to the phylogenetic relations between host individuals (nRF=0.8146, p-value=0.0, nMC=0.6125, p-value=0.0).

Although the tested factors appeared to have only minor effects on the overall microbiome composition, we found that sampling site, environment and host taxonomy appeared to differently affect the presence/absence patterns of individual members of the microbiome. We

found that eight clades were unique to a given sampling site, five clades were constrained to a single host haplotype and eight clades appeared either in the aquaria or the natural environment (Table S4). In addition, we tested whether antibiotics treatment could alter the abundance of individual symbiont clades (Figure S3). We found that the majority of symbiont clades appear to be less abundant in antibiotic treated samples albeit the decreased abundance was only statistically significant in the RETA1-2, the Limnobacter1 and the Pseudomonadales clades. This observed ability to artificially alter the placozoan microbiome using antibiotics treatment could be of interest for future *in vitro* studies. In contrast to the decreased abundance of most of the clades, members of the Flavobacteriales were significantly more abundant in antibiotic treated samples. Two potential scenarios could explain the increased abundance of these bacteria in the antibiotics treated samples. First, the Flavobacteriales could be constrained by microbe-microbe interactions, e.g. antimicrobial metabolites that could be produced by other members of the symbiont communities. If these members were affected by the antibiotic treatment and thus, less abundant, the abundance of Flavobacteriales could be less constrained. As we did not observe any exclusion patterns between the Flavobacteriales and the other symbiont clades, it remains elusive if and which symbionts could constrain the presence of Flavobacteriales (Table S5). Secondly, the antibiotics treatment could lead to cell lysis of microbes that were present in the culture medium of the placozoan cultures. This cell lysis would lead to higher concentrations of organic molecules, i.e. increased concentration of food sources for heterotrophic bacteria such as the Flavobacteriales and in consequence, an increased abundance in antibiotic treated samples.

Taken together, our results illustrate that the composition of the placozoans' microbiome likely cannot be explained by a single ecological or evolutionary factor, a phenomenon that is often observed when studying microbiome composition. Instead, different partners usually underlie different ecological and evolutionary dynamics. For example, several factors such as the host

immune system, age, diet as well as delivery pattern, treatment with anti-, pre and probiotics and fecal microbial transplants have different effects on certain members of the human gut microbiome [22]. In addition, the composition of obligate symbiont communities in a group of gutless worms is likely the result of a mixture of specific environmental acquisition, faithful inheritance, partner switches, symbiont replacement as well as loss of individual symbiont clades [23]. We assume that similarly also the microbiome composition of placozoans, and potentially many other host groups, cannot be explained by the influence of individual factors on the overall composition. Rather, the presence/absence patterns and relative abundance of the different partners is individually influenced by different factors and the overall microbiome composition is largely the result of these individual dynamics.

Despite our effort to sample a broad biogeographic diversity of host individuals, the factors we tested for are not completely unlinked, e.g. certain host haplotypes were derived from only one sampling site and we acquired host specimens either from aquaria or the natural habitat at each sampling site. Thus, further sampling of placozoan hosts is required to untangle the ecological and evolutionary factors that influence the presence/absence patterns of different microbiome members and in consequence, the microbiome composition. In addition, improving the granularity of the analyses by further characterizing the sampling sites via biochemical parameters could increase our understanding of which environmental properties could affect certain members of the placozoan microbiome.

The Placozoa are commonly associated with intracellular bacteria

In total, we detected 24 different bacterial groups in the metagenomes of placozoan hosts. So far, we can only speculate about the type of the association based on phylogenetic data and image-based analyses are required for further insights into the different association types. As discussed above, we assume that bacterial clades that showed a high sequence variability within

a single sample are likely free-living commensalists that do not live in intimate association with placozoans but externally attach to the host animals and scavenge nutrients from the hosts' external digestion of bacteria and algae. In contrast, there are four groups of bacteria that are likely intracellular symbionts: Chlamydiae and Coxiellaceae, that are known to only occur as intracellular symbionts as well as Rickettsiales and *Cand.* Ruthmannia symbionts that have been previously described to intracellularly occur in placozoan hosts[5,9,12,14–18,20,21]. In the following paragraphs, we will discuss the association patterns between each of these groups and the placozoan hosts and speculate about their potential role in the placozoan microbiome.

Coxiellaceae Members of the Coxiella1 clade belong to the genus *Coxiella* (Coxiellaceae*,* Legionellales, Figure 3A). The Legionellales are known to only consist of intracellular, host-associated bacteria [21]. Thus, we assume that also the placozoan-associated *Coxiella* occur as intracellular symbionts. We detected *Coxiella* symbionts in five host haplotypes from four sampling sites (Figure 1, Figure 3A). As the symbiont phylogeny did neither reflect the hosts' phylogenetic relation nor their sampling sites, we assume that the association between *Coxiella* symbionts and placozoan hosts was either independently established several times with individual hosts or that *Coxiella* symbionts could be switched between distantly related host individuals.

We can only speculate about the metabolism of the *Coxiella* symbiont of placozoans. Nevertheless, given the strict host-associated lifestyle of the whole order of Legionellales, we assume that also the placozoan symbionts are adapted to a host-associated lifestyle and potentially rely on the hosts to provide certain resources. Whether the *Coxiella* symbionts are detrimental to the placozoan hosts' health will require further studies of this association.

Chlamydiae We identified two different clades of Chlamydiae symbionts that both belong to the Simkaniaceae (Figure 3B, Table S2). Members of the Chlamydia1 clade are most closely related to the cultured representative species *Simkania nevegensis* and members of this clade were detected in a broader host range including host haplotypes H7 from Carrie Bow Caye, H6 from Bocas del Toro and H11 from Mallorca (Figure 1, Figure 3B). Members of the Chlamydia2 clade are most closely related to *Cand.* Syngnamydia salmonis and they were only detected in the host haplotype H6 from two sampling sites, Bocas del Toro as well as Hawai'i (Figure 1, Figure 3B). So far, it remains elusive whether the observed difference in the host ranges of the two Chlamydiae clades are a result of a sampling bias or indicate different levels of specificity. Despite the potential differences in the specificity between host haplotypes and genus-level symbiont clades, both Chlamydiae clades show high specificity between host haplotypes and symbiont phylotypes. Each host haplotype from one sampling location was associated with a distinct lineage of Chlamydiae symbionts, indicating that even despite potential different host ranges of the Chlamydiae clades, distinct symbionts lineages of both clades are well adapted to a distinct haplotype and are not or only rarely switched between host haplotypes.

Considering that all Chlamydiae described today live in intracellular association with eukaryotic hosts, we assume that also the two placozoan associated Chlamydiae occur intracellularly[20]. In addition, we speculate that similarly to the majority of known Chlamydiae, also the placozoan associated representatives live of host derived resources such as the commonly scavenged glucose-6-phosphate[24]. In contrast, it remains elusive how the association with Chlamydiae could affect the placozoan hosts. Understanding whether this association is of a parasitic, commensalistic or mutualistic nature will require further studies of the two Chlamydiae groups.

***Cand.* Ruthmannia** Previously, bacteria belonging to the novel candidatus genus *Cand.* Ruthmannia within the Margulisbacteria were only detected in H2 haplotypes from Hawai'i[14]. Other studies that assessed the placozoan microbiome did not observe this symbiont type in other host haplotypes or at other sites, posing the question how widespread the association between *Cand.* Ruthmannia symbionts and placozoans is. Here, we detected relatives of the H2 *Cand.* Ruthmannia that shared at least 95% identity on the 16S level in the H2 haplotype from Hawai'i also in the H3 haplotype from Hawai'i, the H2 haplotype from Bremen and the H8 haplotypes from Carrie Bow Caye and Twin Cays (both Belize, Figure 1, Figure 3C). The association between *Cand.* Ruthmannia symbionts and distinct host haplotype lineages from certain locations appear to be relatively stable. Yet this association does not appear to be wide spread across the diversity of placozoan hosts as we detected *Cand.* Ruthmannia in only three out of eight haplotypes and 11 out of 109 host individuals.

We posit that the association pattern between *Cand.* Ruthmannia and placozoans indicates a rather low dependence of the hosts on these symbionts. Based on genome-based metabolic modeling, the *Cand.* Ruthmannia symbionts could potentially supplement the hosts' diet with essential amino acids. However, it is unlikely that the placozoan hosts solely depend on amino acids provided by *Cand.* Ruthmannia symbionts as the digestion of algae and bacterial biofilms likely provides all essential nutrients. In contrast, no free-living relatives of the intracellular *Cand.* Ruthmannia symbionts are known today, indicating that they evolved to an exclusively intracellular lifestyle and gain their nutrition from host digestive products[14]. Based on these assumptions, we conclude that the association between placozoans and *Cand.* Ruthmannia symbionts is obligate for the symbionts. In contrast, placozoan host do likely not dependent on the association with *Cand.* Ruthmannia, and further studies are required to understand whether the symbionts could provide any benefit to the hosts.

Rickettsiales Previous analyses of the placozoan microbiome revealed consistent association with Rickettsiales bacteria across a broad host diversity. Rickettsiales bacteria are a clade of obligate intracellular symbionts, often pathogens, of diverse eukaryotes including humans[25]. Also in placozoans, the Rickettsiales symbionts were found to occur intracellularly[5,9,14,18]. Here, we detected four clades of Rickettsiales symbionts that were associated with distinct host lineages (Figure 1, Figure 3D). The H2, H3, H4 and H15 haplotype were associated with relatives of the previously described midichloriacean symbionts of placozoa (Figure 3D). Our clustering of 16S rRNA genes at 95% sequence identity as well as previous studies considered these symbionts a single symbiont genus (RETA1-2, also described as *Cand.* Aquarickettsia) that includes the RETA1 clade described for the H1 haplotype, the *Cand.* Grellia/RETA2 clade described for the H2 haplotype as well as symbionts from other placozoans as well as corals. However, members of this *Cand.* Aquarickettsia clade formed two distinct phylogenetic clades that based on amino acid sequence identity represent two different genera (Figure 3D)[14]. Thus, we will consider them two different symbiont clades, namely RETA1 and *Cand.* Grellia/RETA2 for the following discussion. The *Cand.* Grellia/RETA2 symbionts were only detected in the H2 haplotype from Bremen, Hawai'i and Vienna. In contrast, the RETA1 symbionts were detected in H2 haplotypes from Mallorca and Panama, the H3 haplotype from Hawai'i, the H4 haplotypes from Belize, Hawai'i and Panama and H15 haplotype from Hong Kong. Neither members of the RETA1 nor the *Cand.* Grellia/RETA2 symbiont clades were detected in the host haplotypes H6, H7 and H8 from Belize that form a divergent host clade. Host individuals of this clade were associated with another clade of midichloriacean symbionts that also contains a symbiont of the coral *Montastrea faveolata* (Figure 3D, RETA3). In addition, we did not detect any midichloriacean symbionts in association with the H11 haplotypes. The H11 haplotypes were instead associated with relatives of coral symbionts from another clade within the *Rickettsiales* that most likely belong to the candidatus genus *Cand.* Megaira (Figure 3D, RETA4).

Summing up, the association between placozoan hosts and members of the Rickettsiales, especially Midichloriaceae, is strikingly common as these symbionts were detected in all host haplotypes and nearly all host individuals. Within the Placozoa, it appears that the association with the RETA1 symbionts is the most wide spread as members of this symbiont clade were detected in most host haplotypes across the full phylogenetic host range. In concordance, a last common ancestor analysis of the placozoan hosts and the Rickettsiales symbionts indicated that the RETA1 symbiont is the ancestral Rickettsiales symbiont of the placozoa and that the *Cand.* Grellia/RETA2, RETA3 and RETA symbionts likely replaced the RETA1 symbiont in distinct host lineages (Figure S4). As all of the Rickettsiales symbionts were most closely related to coral-associated bacteria, we hypothesize that the symbionts can efficiently infect both cnidarian as well as placozoan hosts and relatively flexibly switch between these two closely related animal phyla. Successful host switches from corals to co-occurring placozoans could have initiated the replacement of the RETA1 symbiont in a given host lineage (Figure 3D).

Genomic analyses of the RETA1 and *Cand.* Grellia/RETA2 symbionts showed that these bacteria are dependent on host provided nutrients such as amino acids and nucleotides that they cannot synthesize themselves[14,16]. Thus, we assume that similar to the *Cand.* Ruthmannia symbionts, also the RETA1 and *Cand*. Grellia/RETA2 symbionts obligately rely on nutrients provided by their eukaryotic hosts. So far, genomic information on the metabolism of the RETA3 and RETA4 clades is lacking but considering the similar association patterns, we hypothesize that they also live in an obligate association with their hosts. Similar to other Rickettsiales, the *Cand.* Grellia/RETA2 symbionts encode for genes of typical energy parasites but as they did not express them, we assume a non-parasitic relationship between the Rickettsiales symbionts and the placozoan hosts [14,19,26]. Given that all host haplotypes and nearly all host individuals were associated with Rickettsiales symbionts, we rather assume a mutualistic association. It remains elusive what the Rickettsiales bacteria could provide to their hosts that leads to such a widespread association.

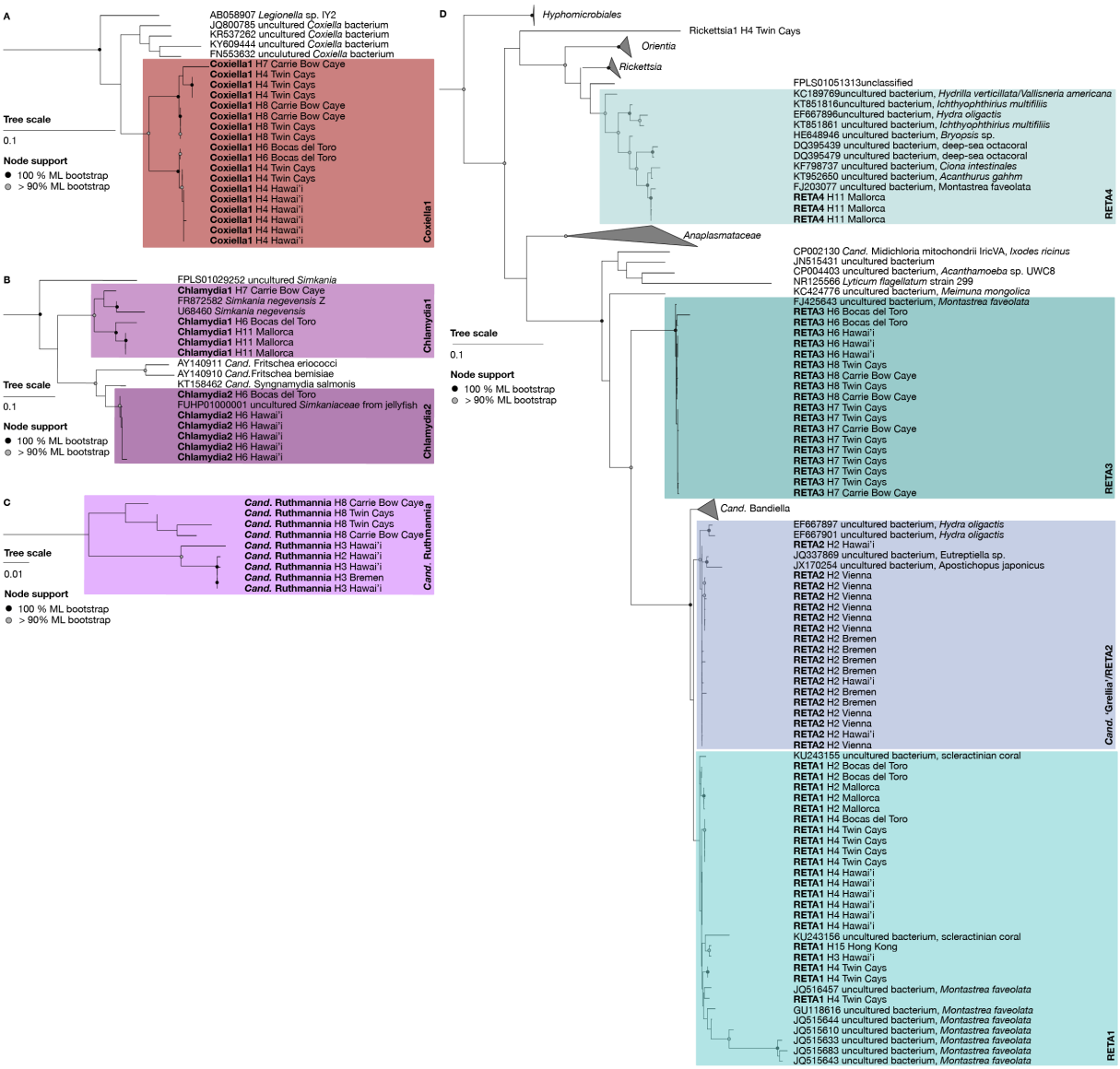


Figure 3 | Putative intracellular symbionts of placozoans show different association patterns across the phylogenetic host diversity. (A) Phylogeny of Coxiella1 symbionts and close relatives, **(B)** Phylogeny of Chlamyida1 and Chlamydia2 symbionts and close relatives, **(C)** Phylogeny of *Cand* Ruthmannia symbionts, **(D)** Phylogeny of RETA1, *Cand.* Grellia/RETA2, RETA3 and RETA4 symbionts and close relatives. Each panel shows the maximum-likelihood phylogeny of a given symbiont clade based on the symbionts' 16S rRNA gene sequences. The scale bar indicates 10% estimated sequence divergence **(A, B and D)** and 1% estimated sequence **(C)** respectively. Nodes with none-parametric bootstrap support > 90% are highlighted in all trees. Colored boxed highlight clades that contain the placozoan-associated bacteria.

Conclusion

The association with complex microbiomes is wide spread across the animal diversity and microbiome composition largely influences the ecology and evolution of their host. Despite our growing understanding of the effects of microbiome composition on their hosts, we still lack an exhaustive understanding of what drives the microbiome composition. Placozoans, arguably the simplest animals known, are associated with a phylogenetically diverse microbiome. In contrast to many other eukaryotic microbiomes the placozoan microbiome consists of a comparably low number of bacterial partners. Such a reduced diversity enables us to investigate the ecological and evolutionary factors that do not only drive overall microbiome composition but also untangle which members are influenced by different factors such as host phylogeny or environmental parameters. Initial analyses indicate that indeed, several ecological and evolutionary parameters have different effects on the association between placozoan hosts and each symbiont lineage, indicating that overall microbiome composition is the result of several independent dynamics.

Our data emphasize the prevalence and variability of the placozoan associated microbiome. Members of the microbiome in general and especially the intimately associated intracellular symbionts could largely influence their hosts' ecology and evolution. Thus, we argue that future research on placozoans should also consider the associated bacteria and take their effects on the host biology into account.

Material & methods

Sample collection, processing and metagenomic sequencing

108 individuals of placozoans were sampled at various field sites (for overview see Figure S1, Figure S2 and Table S1). Placozoans were identified using a dissection microscope and afterwards kept in culture in 400 ml glass beakers. The culture medium was 34.5‰ artificial

seawater and cultures were fed weekly with $2x10^6$ cells ml^{-1} of *Isochrysis galbana*. Cultures were kept at 25°C and a 16:8 hours light:dark cycle. DNA was extracted from single individuals with the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions except the following exceptions: the proteinase K digest was performed overnight, elution volumes were halved, all samples were eluted twice reusing the first eluate and all elutions were performed after ten minutes of waiting prior to centrifugation. Library construction, quality control and sequencing were performed at the Max Planck Genome Centre (Cologne, Germany). If needed, DNA concentration was increased using the MinElute PCR Purification kit (Qiagen). Libraries for Illumina sequencing were prepared using the Ovation Ultralow Library Systems Kit (NuGEN) according to the manufacturer's instructions. The libraries were size selected by agarose gel electrophoresis. The quality and quantity of selected fragments were analysed by fluorometry. Paired end reads of 100, 150 or 250 bp length were sequenced on an Illumina HiSeq 4000 sequencer (Table S1). In addition, we included three previously published placozoan metagenomes into our analysis. They were downloaded from the NCBI short read archive (accession numbers SRR5311040, SRR5934055 and SRR5934125).

Assembly of host marker genes, phylogenetic inference and identification of host haplotypes

Mitochondrial 16S rRNA genes of all specimens were assembled by adapting the phyloFlash pipeline to operate on a custom m16S rRNA gene reference database and predict m16S rRNA genes from assembled sequences[27]. The detected m16S gene sequences were aligned to m16S rRNA gene sequences of previously identified specimens using mafft-linsi v7.407[3,28–30]. Maximum-likelihood based phylogenies were calculated using iqtree, including automatic selection of the best suited model and generation of 100 none-parametric bootstrap replicates [31]. The sequences of *Trichoplax* H0 was used to root the phylogenies. Host haplotypes were defined based on the phylogenetic relations to the previously identified specimens (Note S1).

Symbiont clade definition and quantification

16S rRNA genes were assembled from the metagenomic libraries using phyloFlash. using the –all option and in addition specifying the read length. For subsequent analyses, we only considered sequences that were i) assembled with SPAdes, ii) longer than 1000 bp and iii) did not contain more than 20 ambiguous bases. The resulting sequences were clustered at 95% sequence similarity using usearch v10.0.240[32]. We used the SINA search and classify algorithm to add the 16S rRNA gene sequences of close relatives from the SILVA database v132 that shared at least 90% sequence similarity for each of our assembled symbiont sequences[33]. All assembled sequences and the SILVA database hits were aligned using mafft-linsi and a phylogenetic tree was calculated from the resulting alignment using FastTree v2.1.1[34]. The abundances of all cluster were quantified across all metagenomic libraries using EMIRGE v.0.61.1 following the standard workflow for custom EMIRGE databases[35]. Subsequently, we excluded cluster that appear to be contaminations or appeared only once (Table S2) and we merged cluster that were phylogenetically intermixed. After contamination removal and merging of cluster, we normalized the relative abundances of the remaining clades to 100% (Table S3).

Phylogeny of all symbionts and their relatives

Sequences of all initially detected clades were used to obtain sequences from closely related bacteria from the SILVA and the RefSeq public databases[36]. For SILVA, we used the SINA search and classify algorithm to obtain up to 10 relatives for each sequence that shared at least 99% and 95% sequence similarity for each of our input sequences. In addition, we also screened the RefSeq database using BLAST implemented in Geneious v11.1.5 to obtain the ten most similar 16S rRNA genes[37]. Duplicated sequences were removed from the collection of sequences of the symbionts' relatives. Chloroplast 16S rRNA gene sequences were used as outgroup. The resulting sequence collection was aligned using mafft-linsi and a phylogenetic

tree was calculated using iqtree including automatic selection of the best suited model generation of 100 none-parametric bootstrap replicates.

Analyses and plotting of symbiont community composition

The analyses of symbiont community composition were performed in R v3.6.3 unless differently stated. During the analyses, the following packages were used: phyloseq[38], ape[39], vegan (https://github.com/vegandevs/vegan), plyr[40], MASS[41], gdata (https://cran.r-project.org/web/packages/gdata/index.html), reshape2 (https://github.com/hadley/reshape), forcats (https://github.com/robjhyndman/forecast), igraph (https://github.com/igraph/rigraph), Hmisc (https://github.com/harrelfe/Hmisc/), optparse (https://github.com/trevorld/r-optparse), data.table (https://github.com/Rdatatable/data.table), ade4 (https://github.com/sdray/ade4), tidyverse[42] and spa (https://github.com/markvanderloo/rspa). Plots were generated using ggplot2 from the tidyverse package, gridExtra (https://cran.r-project.org/web/packages/gridExtra/index.html), ggpubr (https://cran.r-project.org/web/packages/ggpubr/index.html), maps (https://www.rdocumentation.org/packages/maps), mapdata (https://www.rdocumentation.org/packages/mapdata), and patchwork (https://github.com/thomasp85/patchwork).

Community composition analyses

The similarity between symbiont communities of host individuals were calculated based on the abundance patterns of the symbiont clades and the symbiont 16S rRNA gene phylogeny using the UniFrac metric as implemented in the phyloseq package in R. We tested for parameters that could explain differences in microbiome composition between individuals using PERMANOVA[43]. These parameters were host haplotype, sampling site, environment and antibiotics treatment (Table S1). All factors were treated as categorical data and analyzed using PERMANOVA. Co-occurrence patterns of symbiont clades were analyzed using a probabilistic model of co-occurrences[44]. The effects of the antibiotic treatment on the abundance of individual symbiont clades were assessed using the Kruskal Wallis rank sum test[45].

Phylosymbiosis

UniFrac distances on the average symbiont abundances per host species were transformed into a dendrogram using hierarchical clustering. The congruences between the m16S rRNA gene based host tree or and the symbiont community UniFrac dendrogram was assessed separately using the normalized Robinson-Fould metric and the normalized Matching Cluster metric, implemented in TreeCmp v1.0-b291[46–48]. Statistical significance was estimated by comparing the congruence between the host phylogeny vs. 1000 random trees as described by Brooks *et al.* 2016, https://github.com/awbrooks19/phylosymbiosis)[49].

Last common ancestor analysis of the Rickettsiales symbionts

In order to determine which of the Rickettsiales symbionts was the ancestral one, we performed a maximum-parsimony based last common ancestor analysis with pastml, using the DOWNPASS prediction method[50].

Data availability

Raw metagenomic sequences as well as host and symbiont marker genes generated in this study will be deposited in the European Nucleotide Archive (ENA) upon peer-review submission and are currently available upon request.

Code availability

The scripts and data for analyzing symbiont community composition and phylogenetic correlations are available under https://github.com/amankowski/Tplax_symbiont-diversity.

Acknowledgments

We are thankful for sample collections and field assistance by Chris Keller, Daniel Abed-Navandi, Ramon Rosello-Mora and Tomas Wilkop. In addition, we would like to thank the Carrie Bow Cay Laboratory and the Bocas del Toro Research Station of the Smithsonian Tropical Research Institute, the Haus des Meeres Vienna, the Las Palma Aquarium and their staff for supporting our sampling campaigns. This work was supported by the Max

Planck Society, a Moore Foundation Marine Microbial Initiative Investigator Award to ND (Grant GBMF3811), and a Marie-Curie Intra-European Fellowship PIEF-GA-2011-301027 CARISYM (HRGV). This work is contribution XXX from the Carrie Bow Cay Laboratory, Caribbean Coral Reef Ecosystem Program, National Museum of History, Washington DC.

References

1. Laumer, C. E. *et al.* Support for a clade of Placozoa and Cnidaria in genes with minimal compositional bias. *Elife* **7**, e36278 (2018).
2. Simion, P. *et al.* A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals. *Curr. Biol.* **27**, 958–967 (2017).
3. Eitel, M., Osigus, H.-J., DeSalle, R. & Schierwater, B. Global Diversity of the Placozoa. *PLoS One* **8**, e57131 (2013).
4. Schulze, F. E. i, nov. gen., nov. spec. *Zool. Anz.* **6**, 92–97 (1883).
5. Eitel, M. *et al.* Comparative genomics and the nature of placozoan species. *PLOS Biol.* **16**, e2005359 (2018).
6. Osigus, H.-J., Rolfes, S., Herzog, R., Kamm, K. & Schierwater, B. *Polyplacotoma mediterranea*n is a new ramified placozoan species. *Curr. Biol.* **29**, R148–R149 (2019).
7. Grell, K. G. & Benwitz, G. Die ultrastruktur von *Trichoplax adhaerens* FE Schulze. *Cytobiologie* **4**, 216–240 (1971).
8. Grell, K. G. & Benwitz, G. Ergänzende Untersuchungen zur Ultrastruktur von *Trichoplax adhaerens* F.E. Schulze (Placozoa). *Zoomorphology* **98**, 47–67 (1981).
9. Smith, C. L. *et al.* Novel Cell Types, Neurosecretory Cells, and Body Plan of the Early-Diverging Metazoan *Trichoplax adhaerens*. *Curr. Biol.* **24**, 1565–1572 (2014).
10. Smith, C. L., Pivovarova, N. & Reese, T. S. Coordinated Feeding Behavior in *Trichoplax*, an Animal without Synapses. *PLoS One* **10**, e0136098 (2015).
11. Pearse, V. B. & Voigt, O. Field biology of placozoans (*Trichoplax*): distribution, diversity, biotic interactions. *Integr. Comp. Biol.* **47**, 677–692 (2007).
12. Srivastava, M. *et al.* The Trichoplax genome and the nature of placozoans. *Nature* **454**, 955–960 (2008).
13. Sebé-Pedrós, A. *et al.* Early metazoan cell type diversity and the evolution of multicellular gene regulation. *Nat. Ecol. Evol.* **2**, 1176–1188 (2018).
14. Gruber-Vodicka, H. R. *et al.* Two intracellular and cell type-specific bacterial symbionts in the placozoan *Trichoplax* H2. *Nat. Microbiol.* **4**, 1465–1474 (2019).
15. Kamm, K., Osigus, H.-J., Stadler, P. F., DeSalle, R. & Schierwater, B. *Trichoplax* genomes reveal profound admixture and suggest stable wild populations without bisexual reproduction. *Sci. Rep.* **8**, 11168 (2018).
16. Klinges, J. G. *et al.* Phylogenetic, genomic, and biogeographic characterization of a novel and ubiquitous marine invertebrate-associated Rickettsiales parasite, *Candidatus* Aquarickettsia rohweri, gen. nov., sp. nov. *ISME J.* **13**, 2938–2953 (2019).

17. Driscoll, T., Gillespie, J. J., Nordberg, E. K., Azad, A. F. & Sobral, B. W. Bacterial DNA Sifted from the *Trichoplax adhaerens* (Animalia: Placozoa) Genome Project Reveals a Putative Rickettsial Endosymbiont. *Genome Biol. Evol.* **5**, 621–645 (2013).
18. Guidi, L., Eitel, M., Cesarini, E., Schierwater, B. & Balsamo, M. Ultrastructural analyses support different morphological lineages in the phylum placozoa Grell, 1971. *J. Morphol.* **272**, 371–378 (2011).
19. Driscoll, T. P. *et al.* Wholly Rickettsia! Reconstructed Metabolic Profile of the Quintessential Bacterial Parasite of Eukaryotic Cells. *MBio* **8**, e00859-17 (2017).
20. Horn, M. Chlamydiae as Symbionts in Eukaryotes. *Annu. Rev. Microbiol.* **62**, 113–131 (2008).
21. Duron, O., Doublet, P., Vavre, F. & Bouchon, D. The Importance of Revisiting Legionellales Diversity. *Trends Parasitol.* **34**, 1027–1037 (2018).
22. Hasan, N. & Yang, H. Factors affecting the composition of the gut microbiota, and its modulation. *PeerJ* **7**, e7502 (2019).
23. Mankowski, A. *et al.* Highly variable fidelity drives symbiont community composition in an obligate symbiosis. *bioRxiv* 2021.04.28.441735 (2021) doi:10.1101/2021.04.28.441735.
24. Köstlbacher, S. *et al.* Pangenomics reveals alternative environmental lifestyles among chlamydiae. *Nat. Commun.* **12**, (2021).
25. Thomas, S. Rickettsiales: Biology, Molecular Biology, Epidemiology, and Vaccine Development. *Springer International Publishing* (2016).
26. Stephan, S.-E. *et al.* ATP/ADP Translocases: a Common Feature of Obligate Intracellular Amoebal Symbionts Related to Chlamydiae and Rickettsiae. *J. Bacteriol.* **186**, 683–691 (2004).
27. Gruber-Vodicka, H. R., Seah, B. K. B. & Pruesse, E. phyloFlash: Rapid Small-Subunit rRNA Profiling and Targeted Assembly from Metagenomes. *mSystems* **5**, e00920-20 (2020).
28. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).
29. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
30. 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.* **30**, 3059–3066 (2002).

31. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
32. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
33. Pruesse, E. & Peplies Jörg & Glöckner, F. O. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. 1823–1829 (2012).
34. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* **5**, e9490 (2010).
35. Miller, C. S., Baker, B. J., Thomas, B. C. & Singer Steven W. & Banfield, J. F. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol.* **12**, R44 (2011).
36. Pruitt, K. D. & Maglott, D. R. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.* **29**, 137–140 (2001).
37. Altschul, S. F., Gish, W., Miller, W. & Myers E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
38. Bartram, A. K. *et al.* Exploring links between pH and bacterial community composition in soils from the Craibstone Experimental Farm. *FEMS Microbiol. Ecol.* **87**, 403–415 (2014).
39. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
40. Wickham, H. The Split-Apply-Combine Strategy for Data Analysis. *J. Stat. Softw.* **40**, (2011).
41. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S*. (Springer, 2002).
42. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
43. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**, 32–46 (2001).
44. Veech, J. A. A probabilistic model for analysing species co-occurrence. *Glob. Ecol. Biogeogr.* **22**, 252–260 (2013).
45. Kruskal, W. H. & Wallis, W. A. Use of Ranks in One-Criterion Variance Analysis. *J. Am. Stat. Assoc.* **47**, 583–621 (1952).
46. Robinson, D. F. & Foulds, L. R. Comparison of phylogenetic trees. *Math. Biosci.* **53**, 131–147 (1981).

47. Bogdanowicz, D., Giaro, K. & Wróbel, B. TreeCmp: Comparison of Trees in Polynomial Time. *Evol. Bioinforma.* **8**, EBO.S9657 (2012).
48. Bogdanowicz, D. & Giaro, K. On a matching distance between rooted phylogenetic trees. *Int. J. Appl. Math. Comput. Sci.* **23**, 669–684.
49. Brooks, A. W., Kohl, K. D., Brucker, R. M., van Opstal, E. J. & Bordenstein, S. R. Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. *PLOS Biol.* **14**, e2000225 (2016).
50. Ishikawa, S. A., Zhukova, A., Iwasaki, W. & Gascuel, O. A Fast Likelihood Method to Reconstruct and Visualize Ancestral Scenarios. *Mol. Biol. Evol.* **36**, 2069–2085 (2019).

Chapter IV | Supplementary material

Figures

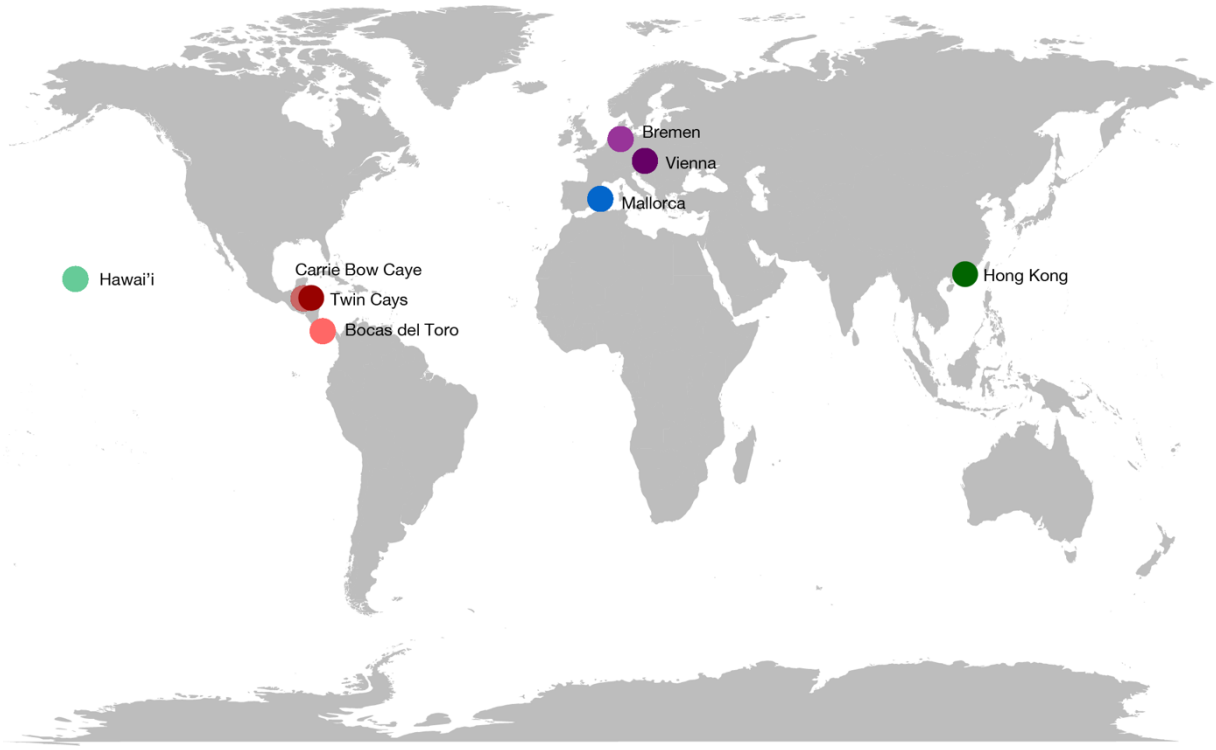


Figure S1 | Sampling sites of *Trichoplax* individuals. Globally distributed sampling sites are highlight according to the color scheme that was used throughout the manuscript.

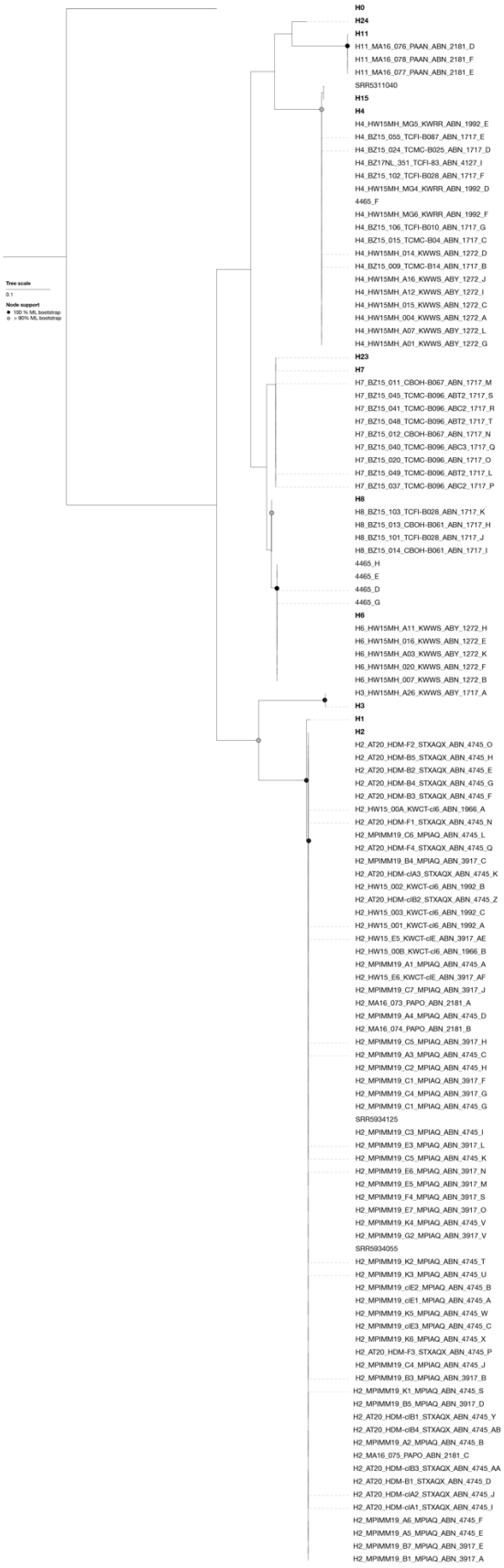


Figure S2 | Host mitochondrial 16S rRNA gene phylogeny. Maximum-likelihood trees of the host m16S rRNA gene phylogeny including all analyzed host individuals. The scale bar indicates 10% estimated sequence divergence. Nodes with none-parametric bootstrap support > 90% are highlighted.

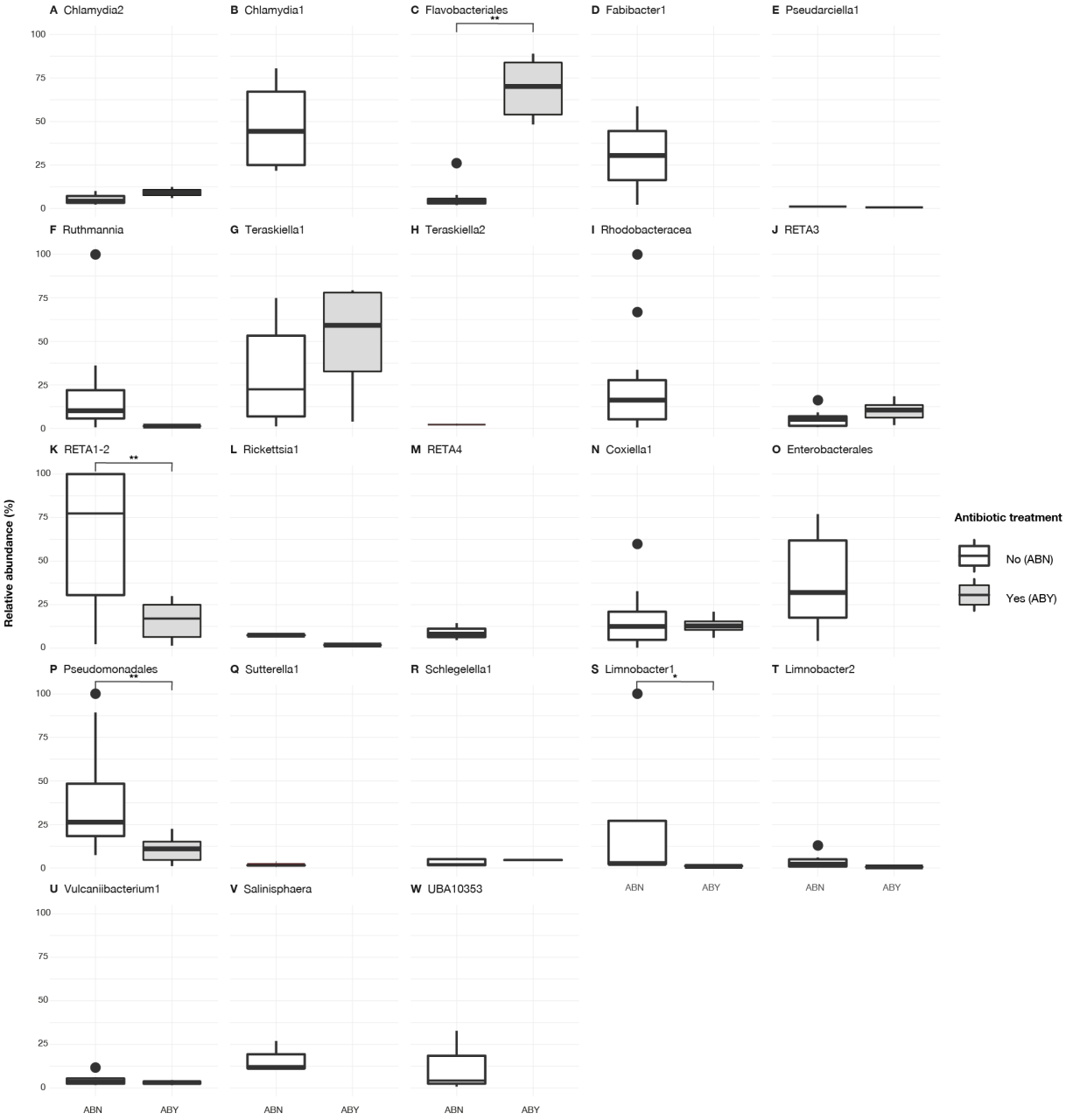


Figure S3 | Antibiotic treatment differently effect the members of the placozoan microbiome. Boxplots of the relative abundance of individual symbionts clades in none-treated (ABN, white boxes) and antibiotics treated (ABY, gray boxes) samples. Significant changes identified by the p-values generated of the Kruskal Wallis rank sum test are highlighted by the following significant codes: * 0.01 ** 0.001 *** 0.00.

Figure S4 | The last common ancestor of the Placozoa was most likely associated with the RETA1 symbiont. Each circle represents a node in the phylogeny of placozoan hosts. The upper circle represents the most basal node, i.e. the last common ancestor of all analyzed individuals. The color and text indicate which Rickettsiales symbiont was most likely present at this node of the tree. If a node had more than one color, either of the respective symbionts could have been present in the respective last common ancestor. The bold numbers indicate the number of leaves in which the respective symbiont type was found without any changes of the symbiont type in between the node and the leaves. Gray errors indicate changes of the Rickettsiales symbiont between two nodes of the tree. Numbers next to the errors show the frequency of the respective change.

Tables

Table S1 | Sample overview.

Library	Host haplotype	Country	Sampling site	Description of sampling site	Environment	Latitude	Longitude	# of PE reads	Read length (bp)
4465_D	H6	Panama	Bocas del Toro		Free-living	9.35001	-82.2617	10652083	150
4465_E	H6	Panama	Bocas del Toro		Free-living	9.35001	-82.2617	7607021	150
4465_F	H4	Panama	Bocas del Toro		Free-living	9.35001	-82.2617	8535616	150
4465_G	H6	Panama	Bocas del Toro		Free-living	9.35001	-82.2617	10600409	150
4465_H	H6	Panama	Bocas del Toro		Free-living	9.35001	-82.2617	10899108	150
H11_MA16_076_PAAN_ABN_2181_D	H11	Spain	Mallorca	Las Palmas Aquarium, Anthias tank	Aquarium	39.53151	2.72949	29093765	250
H11_MA16_077_PAAN_ABN_2181_E	H11	Spain	Mallorca	Las Palmas Aquarium, Anthias tank	Aquarium	39.53151	2.72949	28955897	250
H11_MA16_078_PAAN_ABN_2181_F	H11	Spain	Mallorca	Las Palmas Aquarium, Anthias tank	Aquarium	39.53151	2.72949	28494669	250
H2_AT20_HDM-B1_STXAQX_ABN_4745_D	H2	Austria	Vienna	Haus des Meeres, STXAQX	Aquarium	48.19766	16.35292	10615901	150
H2_AT20_HDM-B2_STXAQX_ABN_4745_E	H2	Austria	Vienna	Haus des Meeres, STXAQX	Aquarium	48.19766	16.35292	10370688	150
H2_AT20_HDM-B3_STXAQX_ABN_4745_F	H2	Austria	Vienna	Haus des Meeres, STXAQX	Aquarium	48.19766	16.35292	10423679	150
H2_AT20_HDM-B4_STXAQX_ABN_4745_G	H2	Austria	Vienna	Haus des Meeres, STXAQX	Aquarium	48.19766	16.35292	10452208	150
H2_AT20_HDM-B5_STXAQX_ABN_4745_H	H2	Austria	Vienna	Haus des Meeres, STXAQX	Aquarium	48.19766	16.35292	10331096	150
H2_AT20_HDM-clA1_STXAQX_ABN_4745_I	H2	Austria	Vienna	Haus des Meeres, STXAQX	Aquarium	48.19766	16.35292	10497857	150
H2_AT20_HDM-clA2_STXAQX_ABN_4745_J	H2	Austria	Vienna	Haus des Meeres, STXAQX	Aquarium	48.19766	16.35292	10393142	150
H2_AT20_HDM-clA3_STXAQX_ABN_4745_K	H2	Austria	Vienna	Haus des Meeres, STXAQX	Aquarium	48.19766	16.35292	10520516	150
H2_AT20_HDM-clB1_STXAQX_ABN_4745_Y	H2	Austria	Vienna	Haus des Meeres, STXAQX	Aquarium	48.19766	16.35292	11346357	150
H2_AT20_HDM-clB2_STXAQX_ABN_4745_Z	H2	Austria	Vienna	Haus des Meeres, STXAQX	Aquarium	48.19766	16.35292	9383297	150
H2_AT20_HDM-clB3_STXAQX_ABN_4745_AA	H2	Austria	Vienna	Haus des Meeres, STXAQX	Aquarium	48.19766	16.35292	12084262	150
H2_AT20_HDM-clB4_STXAQX_ABN_4745_AB	H2	Austria	Vienna	Haus des Meeres, STXAQX	Aquarium	48.19766	16.35292	14031588	150
H2_AT20_HDM-F1_STXAQX_ABN_4745_N	H2	Austria	Vienna	Haus des Meeres, STXAQX	Aquarium	48.19766	16.35292	13665069	150
H2_AT20_HDM-F2_STXAQX_ABN_4745_O	H2	Austria	Vienna	Haus des Meeres, STXAQX	Aquarium	48.19766	16.35292	10545509	150
H2_AT20_HDM-F3_STXAQX_ABN_4745_P	H2	Austria	Vienna	Haus des Meeres, STXAQX	Aquarium	48.19766	16.35292	12285778	150
H2_AT20_HDM-F4_STXAQX_ABN_4745_Q	H2	Austria	Vienna	Haus des Meeres, STXAQX	Aquarium	48.19766	16.35292	9647462	150
H2_HW15_001_KWCT-cl6_ABN_1992_A	H2	USA	Hawai‘i	Kewalo Marine Laboratory, coral tank cl6	Aquarium	21.19133	-157.86106	23596198	100
H2_HW15_002_KWCT-cl6_ABN_1992_B	H2	USA	Hawai‘i	Kewalo Marine Laboratory, coral tank cl6	Aquarium	21.19133	-157.86106	15011100	100
H2_HW15_003_KWCT-cl6_ABN_1992_C	H2	USA	Hawai‘i	Kewalo Marine Laboratory, coral tank cl6	Aquarium	21.19133	-157.86106	14458446	100

Library	Host haplotype	Country	Sampling site	Description of sampling site	Environment	Latitude	Longitude	# of PE reads	Read length (bp)
H2_HW15_00A_KWCT-cl6_ABN_1966_A	H2	USA	Hawai'i	Kewalo Marine Laboratory, coral tank cl6	Aquarium	21.19133	-157.86106	21707829	100
H2_HW15_00B_KWCT-cl6_ABN_1966_B	H2	USA	Hawai'i	Kewalo Marine Laboratory, coral tank cl6	Aquarium	21.19133	-157.86106	21556940	100
H2_HW15_E5_KWCT-clE_ABN_3917_AE	H2	USA	Hawai'i	Kewalo Marine Laboratory, coral tank clE	Aquarium	21.19133	-157.86106	12548915	150
H2_HW15_E6_KWCT-clE_ABN_3917_AF	H2	USA	Hawai'i	Kewalo Marine Laboratory, coral tank clE	Aquarium	21.19133	-157.86106	12297722	150
H2_MA16_073_PAPO_ABN_2181_A	H2	Spain	Mallorca	Las Palmas Aquarium, Posidonia tank	Aquarium	39.53151	2.72949	28365072	250
H2_MA16_074_PAPO_ABN_2181_B	H2	Spain	Mallorca	Las Palmas Aquarium, Posidonia tank	Aquarium	39.53151	2.72949	24611406	250
H2_MA16_075_PAPO_ABN_2181_C	H2	Spain	Mallorca	Las Palmas Aquarium, Posidonia tank	Aquarium	39.53151	2.72949	20699200	250
H2_MPIMM19_A1_MPIAQ_ABN_4745_A	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	11739285	150
H2_MPIMM19_A2_MPIAQ_ABN_4745_B	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	13343311	150
H2_MPIMM19_A3_MPIAQ_ABN_4745_C	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	13750426	150
H2_MPIMM19_A4_MPIAQ_ABN_4745_D	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	8843151	150
H2_MPIMM19_A5_MPIAQ_ABN_4745_E	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	8743591	150
H2_MPIMM19_A6_MPIAQ_ABN_4745_F	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	9329943	150
H2_MPIMM19_B1_MPIAQ_ABN_3917_A	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	9058421	150
H2_MPIMM19_B3_MPIAQ_ABN_3917_B	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	9857235	150
H2_MPIMM19_B4_MPIAQ_ABN_3917_C	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	13033392	150
H2_MPIMM19_B5_MPIAQ_ABN_3917_D	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	8615494	150
H2_MPIMM19_B7_MPIAQ_ABN_3917_E	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	12470712	150
H2_MPIMM19_C1_MPIAQ_ABN_3917_F	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	11986400	150
H2_MPIMM19_C1_MPIAQ_ABN_4745_G	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	10482409	150
H2_MPIMM19_C2_MPIAQ_ABN_4745_H	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	12504594	150
H2_MPIMM19_C3_MPIAQ_ABN_4745_I	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	13282816	150
H2_MPIMM19_C4_MPIAQ_ABN_3917_G	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	12375449	150
H2_MPIMM19_C4_MPIAQ_ABN_4745_J	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	5762414	150
H2_MPIMM19_C5_MPIAQ_ABN_3917_H	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	10604471	150
H2_MPIMM19_C5_MPIAQ_ABN_4745_K	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	12769802	150
H2_MPIMM19_C6_MPIAQ_ABN_4745_L	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	7166450	150
H2_MPIMM19_C7_MPIAQ_ABN_3917_J	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	10855061	150
H2_MPIMM19_clE1_MPIAQ_ABN_4745_A	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	10495941	150
H2_MPIMM19_clE2_MPIAQ_ABN_4745_B	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	10495566	150

Library	Host haplotype	Country	Sampling site	Description of sampling site	Environment	Latitude	Longitude	# of PE reads	Read length (bp)
H2_MPIMM19_clE3_MPIAQ_ABN_4745_C	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	10595476	150
H2_MPIMM19_E3_MPIAQ_ABN_3917_L	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	10752528	150
H2_MPIMM19_E5_MPIAQ_ABN_3917_M	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	12896195	150
H2_MPIMM19_E6_MPIAQ_ABN_3917_N	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	9894349	150
H2_MPIMM19_E7_MPIAQ_ABN_3917_O	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	12759499	150
H2_MPIMM19_F4_MPIAQ_ABN_3917_S	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	12536417	150
H2_MPIMM19_G2_MPIAQ_ABN_3917_V	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	11665245	150
H2_MPIMM19_K1_MPIAQ_ABN_4745_S	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	10485504	150
H2_MPIMM19_K2_MPIAQ_ABN_4745_T	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	14239889	150
H2_MPIMM19_K3_MPIAQ_ABN_4745_U	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	10751345	150
H2_MPIMM19_K4_MPIAQ_ABN_4745_V	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	13185335	150
H2_MPIMM19_K5_MPIAQ_ABN_4745_W	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	13695609	150
H2_MPIMM19_K6_MPIAQ_ABN_4745_X	H2	Germany	Bremen	MPI Bremen, marine aquarium	Aquarium	53.10965	8.84755	9047643	150
H3_HW15MH_A26_KWWS_ABY_1717_A	H3	USA	Hawai‘i	Kewalo Marine Laboratory, water system	Aquarium	21.19133	-157.86106	34218125	150
H4_BZ15_009_TCMC-B14_ABN_1717_B	H4	Belize	Twin Cays	Twin Cays, main channel	Free-living	16.81796	-88.10004	31970805	150
H4_BZ15_015_TCMC-B04_ABN_1717_C	H4	Belize	Twin Cays	Twin Cays, main channel	Free-living	16.81796	-88.08192	32578326	150
H4_BZ15_024_TCMC-B025_ABN_1717_D	H4	Belize	Twin Cays	Twin Cays, main channel	Free-living	16.81796	-88.08192	31807883	150
H4_BZ15_055_TCFI-B087_ABN_1717_E	H4	Belize	Twin Cays	Twin Cays, Fishery’s Bay	Free-living	16.81796	-88.08192	31706697	150
H4_BZ15_102_TCFI-B028_ABN_1717_F	H4	Belize	Twin Cays	Twin Cays, Fishery’s Bay	Free-living	16.81796	-88.08192	31599096	150
H4_BZ15_106_TCFI-B010_ABN_1717_G	H4	Belize	Twin Cays	Twin Cays, Fishery’s Bay	Free-living	16.81796	-88.08192	32492743	150
H4_BZ17NL_351_TCFI-83_ABN_4127_I	H4	Belize	Twin Cays	Twin Cays, Fishery’s Bay	Free-living	16.81796	-88.08192	13649745	150
H4_HW15MH_004_KWWS_ABN_1272_A	H4	USA	Hawai‘i	Kewalo Marine Laboratory, water system	Aquarium	21.19133	-157.86106	10182493	100
H4_HW15MH_014_KWWS_ABN_1272_D	H4	USA	Hawai‘i	Kewalo Marine Laboratory, water system	Aquarium	21.19133	-157.86106	12138886	100
H4_HW15MH_015_KWWS_ABN_1272_C	H4	USA	Hawai‘i	Kewalo Marine Laboratory, water system	Aquarium	21.19133	-157.86106	15451522	100
H4_HW15MH_A01_KWWS_ABY_1272_G	H4	USA	Hawai‘i	Kewalo Marine Laboratory, water system	Aquarium	21.19133	-157.86106	67428942	100
H4_HW15MH_A07_KWWS_ABY_1272_L	H4	USA	Hawai‘i	Kewalo Marine Laboratory, water system	Aquarium	21.19133	-157.86106	69341124	100
H4_HW15MH_A12_KWWS_ABY_1272_I	H4	USA	Hawai‘i	Kewalo Marine Laboratory, water system	Aquarium	21.19133	-157.86106	14631584	100
H4_HW15MH_A16_KWWS_ABY_1272_J	H4	USA	Hawai‘i	Kewalo Marine Laboratory, water system	Aquarium	21.19133	-157.86106	77152469	100
H4_HW15MH_MG4_KWRR_ABN_1992_D	H4	USA	Hawai‘i	Kewalo Marine Laboratory, RR	Aquarium	21.19133	-157.86106	21910567	100
H4_HW15MH_MG5_KWRR_ABN_1992_E	H4	USA	Hawai‘i	Kewalo Marine Laboratory, RR	Aquarium	21.19133	-157.86106	21192161	100

Library	Host haplotype	Country	Sampling site	Description of sampling site	Environment	Latitude	Longitude	# of PE reads	Read length (bp)
H4_HW15MH_MG6_KWRR_ABN_1992_F	H4	USA	Hawai‘i	Kewalo Marine Laboratory, RR	Aquarium	21.19133	-157.86106	21808174	100
H6_HW15MH_007_KWWS_ABN_1272_B	H6	USA	Hawai‘i	Kewalo Marine Laboratory, water system	Aquarium	21.19133	-157.86106	73054094	100
H6_HW15MH_016_KWWS_ABN_1272_E	H6	USA	Hawai‘i	Kewalo Marine Laboratory, water system	Aquarium	21.19133	-157.86106	13820597	100
H6_HW15MH_020_KWWS_ABN_1272_F	H6	USA	Hawai‘i	Kewalo Marine Laboratory, water system	Aquarium	21.19133	-157.86106	70894165	100
H6_HW15MH_A03_KWWS_ABY_1272_K	H6	USA	Hawai‘i	Kewalo Marine Laboratory, water system	Aquarium	21.19133	-157.86106	97867178	100
H6_HW15MH_A11_KWWS_ABY_1272_H	H6	USA	Hawai‘i	Kewalo Marine Laboratory, water system	Aquarium	21.19133	-157.86106	68457622	100
H7_BZ15_011_CBOH-B067_ABN_1717_M	H7	Belize	Carrie Bow Cay	Carrie Bow Caye, outhouse	Free-living	16.80262	-88.08192	32404210	150
H7_BZ15_012_CBOH-B067_ABN_1717_N	H7	Belize	Carrie Bow Cay	Carrie Bow Caye, outhouse	Free-living	16.80262	-88.08192	32846370	150
H7_BZ15_020_TCMC-B096_ABN_1717_O	H7	Belize	Twin Cays	Twin Cays, main channel	Free-living	16.81796	-88.08192	32788503	150
H7_BZ15_037_TCMC-B096_ABC2_1717_P	H7	Belize	Twin Cays	Twin Cays, main channel	Free-living	16.81796	-88.08192	31918890	150
H7_BZ15_040_TCMC-B096_ABC3_1717_Q	H7	Belize	Twin Cays	Twin Cays, main channel	Free-living	16.81796	-88.08192	31870220	150
H7_BZ15_041_TCMC-B096_ABC2_1717_R	H7	Belize	Twin Cays	Twin Cays, main channel	Free-living	16.81796	-88.08192	32809560	150
H7_BZ15_045_TCMC-B096_ABT2_1717_S	H7	Belize	Twin Cays	Twin Cays, main channel	Free-living	16.81796	-88.08192	31658842	150
H7_BZ15_048_TCMC-B096_ABT2_1717_T	H7	Belize	Twin Cays	Twin Cays, main channel	Free-living	16.81796	-88.08192	31739739	150
H7_BZ15_049_TCMC-B096_ABT2_1717_L	H7	Belize	Twin Cays	Twin Cays, main channel	Free-living	16.81796	-88.08192	32421981	150
H8_BZ15_013_CBOH-B061_ABN_1717_H	H8	Belize	Carrie Bow Cay	Carrie Bow Caye, outhouse	Free-living	16.80262	-88.08192	31592157	150
H8_BZ15_014_CBOH-B061_ABN_1717_I	H8	Belize	Carrie Bow Cay	Carrie Bow Caye, outhouse	Free-living	16.80262	-88.08192	32716595	150
H8_BZ15_101_TCFI-B028_ABN_1717_J	H8	Belize	Twin Cays	Twin Cays, Fishery’s Bay	Free-living	16.81796	-88.08192	31511563	150
H8_BZ15_103_TCFI-B028_ABN_1717_K	H8	Belize	Twin Cays	Twin Cays, Fishery’s Bay	Free-living	16.81796	-88.08192	32203819	150
SRR5311040	H15	China	Hong Kong		Free-living	22.35273	114.2517	120429967	
SRR5934055	H2	Panama	Bocas del Toro		Free-living	9.35001	-82.2617	28214222	
SRR5934125	H2	Panama	Bocas del Toro		Free-living	9.35001	-82.2617	42695180	

Table S2: Taxonomy of all detected bacterial clades.

Clade number	Clade name	Kingdom	Phylum	Class	Order	Family
Clade0	RETA1-2	Bacteria	Proteobacteria	Alphaproteobacteria	Rickettsiales	Fokiniaceae
Clade1	Chlamydia1	Bacteria	Verrucomicrobiota	Chlamydiae	Chlamydiales	Simkaniaceae
Clade2	Contamination	Bacteria	Actinobacteriota	Actinobacteria	Corynebacteriales	Corynebacteriaceae
Clade3	Pseudomonadales	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Saccharospirillaceae
Clade4	Enterobacterales	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales	Alteromonadaceae
Clade5	Rhodobacteraceae	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Clade6	Singleton	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae
Clade7	Enterobacterales	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales	Pseudoalteromonadaceae
Clade8	Contamination	Bacteria	Actinobacteriota	Actinobacteria	Micrococcales	Microbacteriaceae
Clade9	RETA3	Bacteria	Proteobacteria	Alphaproteobacteria	Rickettsiales	Fokiniaceae
Clade10	Schlegelella1	Bacteria	Proteobacteria	Gammaproteobacteria	Burkholderiales	Comamonadaceae
Clade11	Pseudomonadales	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Endozoicomonadaceae
Clade12	Flavobacteriales	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae
Clade13	Coxiella1	Bacteria	Proteobacteria	Gammaproteobacteria	Coxiellales	Coxiellaceae
Clade14	Teraskiella1	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Terasakiellaceae
Clade15	Contamination	Bacteria	Actinobacteriota	Actinobacteria	Actinomycetales	Actinomycetaceae
Clade16	Contamination	Bacteria	Actinobacteriota	Actinobacteria	Propionibacteriales	Propionibacteriaceae
Clade17	Fabibacter1	Bacteria	Bacteroidota	Bacteroidia	Cytophagales	Cyclobacteriaceae
Clade18	Pseudomonadales	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Alcanivoracaceae1
Clade19	Pseudomonadales	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Marinobacteraceae
Clade20	Flavobacteriales	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae
Clade21	Chlamydia2	Bacteria	Verrucomicrobiota	Chlamydiae	Chlamydiales	Simkaniaceae
Clade22	Pseudomonadales	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae
Clade23	Limnobacter1	Bacteria	Proteobacteria	Gammaproteobacteria	Burkholderiales	Comamonadaceae
Clade24	Ruthmannia	Bacteria	Margulisbacteria			
Clade25	Contamination	Bacteria	Firmicutes	Bacilli	Staphylococcales	Staphylococcaceae
Clade26	Flavobacteriales	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae
Clade27	Singleton	Bacteria	Proteobacteria	Alphaproteobacteria	Parvibaculales	
Clade28	Pseudomonadales	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Alcanivoracaceae1
Clade29	Pseudomonadales	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Spongiibacteraceae
Clade30	Pseudomonadales	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Nitrincolaceae

Clade number	Clade name	Kingdom	Phylum	Class	Order	Family
Clade31	Limnobacter2	Bacteria	Proteobacteria	Gammaproteobacteria	Burkholderiales	Burkholderiaceae
Clade32	Vulcaniibacterium1	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae
Clade33	Rhodobacteraceae	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Clade34	Singleton	Bacteria	Spirochaetota	Leptospirae	Leptospirales	Leptospiraceae
Clade35	Enterobacterales	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales	Alteromonadaceae
Clade36	Pseudomonadales	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Endozoicomonadaceae
Clade37	Singleton	Bacteria	Bacteroidota	Rhodothermia	Balneolales	Balneolaceae
Clade38	Flavobacteriales	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae
Clade39	Contamination	Bacteria	Bdellovibrionota	Bdellovibrionia	Bacteriovoracales	Bacteriovoracaceae
Clade40	Contamination	Bacteria	Proteobacteria	Alphaproteobacteria	Rickettsiales	Mitochondria
Clade41	Pseudarciella1	Bacteria	Bacteroidota	Bacteroidia	Cytophagales	Spirosomaceae
Clade42	Singleton	Bacteria	Myxococcota	Myxococcia	Myxococcales	Myxococcaceae
Clade43	Contamination	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae
Clade44	Singleton	Bacteria	Proteobacteria	Gammaproteobacteria	Burkholderiales	Neisseriaceae
Clade45	Singleton	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae
Clade46	UBA10353	Bacteria	Proteobacteria	Gammaproteobacteria	UBA10353 marine group	
Clade47	Salinisphaera	Bacteria	Proteobacteria	Gammaproteobacteria	Salinisphaerales	Salinisphaeraceae
Clade48	Flavobacteriales	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae
Clade49	Pseudomonadales	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Spongiibacteraceae
Clade50	RETA4	Bacteria	Proteobacteria	Alphaproteobacteria	Rickettsiales	Rickettsiaceae
Clade51	Pseudomonadales	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Spongiibacteraceae
Clade52	Singleton	Bacteria	Proteobacteria	Alphaproteobacteria	Caulobacterales	
Clade53	Contamination	Bacteria	Cyanobacteria	Cyanobacteriia	Chloroplast	
Clade54	Contamination	Bacteria	Cyanobacteria	Cyanobacteriia	Chloroplast	
Clade55	Pseudomonadales	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudohongiellaceae
Clade56	Enterobacterales	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales	Alteromonadaceae
Clade57	Teraskiella2	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Terasakiellaceae
Clade58	Babeliales1	Bacteria	Dependentiae	Babeliae	Babeliales	
Clade59	Contamination	Bacteria	Proteobacteria	Gammaproteobacteria	Methylococcales	Methylomonadaceae
Clade60	Contamination	Bacteria	Cyanobacteria	Cyanobacteriia	Chloroplast	
Clade61	Pseudomonadales	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Saccharospirillaceae
Clade62	Ricketsia1	Bacteria	Proteobacteria	Alphaproteobacteria	Rickettsiales	Rickettsiaceae

Clade number	Clade name	Kingdom	Phylum	Class	Order	Family
Clade63	Enterobacterales	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales	Alteromonadaceae
Clade64	Rhodobacteraceae	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Clade65	Contamination	Bacteria	Proteobacteria	Alphaproteobacteria	Rickettsiales	Mitochondria
Clade66	Sutterella1	Bacteria	Proteobacteria	Gammaproteobacteria	Burkholderiales	Sutterellaceae
Clade67	Singleton	Bacteria	Bacteroidota	Rhodothermia	Balneolales	Balneolaceae
Clade68	Contamination	Bacteria	Cyanobacteria	Cyanobacteriia	Chloroplast	
Clade69	Pseudomonadales	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Nitrincolaceae
Clade70	Pseudomonadales	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Halomonadaceae
Clade71	Enterobacterales	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales	Vibrionaceae
Clade72	Rhodobacteraceae	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
Clade73	Contamination	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales	Sedimenticolaceae
Clade74	Pseudomonadales	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Halomonadaceae
Clade75	Enterobacterales	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales	Colwelliaceae
Clade76	Contamination	Bacteria	Cyanobacteria	Cyanobacteriia	Chloroplast	
Clade77	Enterobacterales	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales	Idiomarinaceae

Table S3 | Relative abundance of putative symbiont clades in analyzed metagenomic datasets.

Library	Symbiont clade	Relative abundance (%)
4465_D	Coxiella1	3.14
4465_D	Teraskiella1	44.04
4465_D	Babeliales1	49.50
4465_D	RETA3	3.32
4465_E	Teraskiella1	100.00
4465_F	Chlamydia1	39.17
4465_F	RETA1-2	60.83
4465_G	Teraskiella1	59.11
4465_G	Chlamydia2	19.70
4465_G	Pseudomonadales	21.19
4465_H	Chlamydia1	7.82
4465_H	Coxiella1	4.15
4465_H	Teraskiella1	51.68
4465_H	Babeliales1	27.85
4465_H	RETA3	3.12
4465_H	Pseudomonadales	5.39
H11_MA16_076_PAAN_ABN_2181_D	Chlamydia1	80.43
H11_MA16_076_PAAN_ABN_2181_D	RETA4	8.08
H11_MA16_076_PAAN_ABN_2181_D	Sutterella1	3.18
H11_MA16_076_PAAN_ABN_2181_D	Pseudomonadales	8.30
H11_MA16_077_PAAN_ABN_2181_E	Chlamydia1	62.68
H11_MA16_077_PAAN_ABN_2181_E	RETA4	14.34
H11_MA16_077_PAAN_ABN_2181_E	Sutterella1	3.95
H11_MA16_077_PAAN_ABN_2181_E	Enterobacterales	11.55
H11_MA16_077_PAAN_ABN_2181_E	Pseudomonadales	7.49
H11_MA16_078_PAAN_ABN_2181_F	Chlamydia1	26.08
H11_MA16_078_PAAN_ABN_2181_F	RETA4	4.68
H11_MA16_078_PAAN_ABN_2181_F	Sutterella1	1.25
H11_MA16_078_PAAN_ABN_2181_F	Enterobacterales	16.76
H11_MA16_078_PAAN_ABN_2181_F	Flavobacteriales	2.07
H11_MA16_078_PAAN_ABN_2181_F	Pseudomonadales	49.16
H2_AT20_HDM-B1_STXAQX_ABN_4745_D	RETA1-2	100.00
H2_AT20_HDM-B2_STXAQX_ABN_4745_E	RETA1-2	100.00
H2_AT20_HDM-B3_STXAQX_ABN_4745_F	Pseudomonadales	46.44
H2_AT20_HDM-B3_STXAQX_ABN_4745_F	RETA1-2	53.56
H2_AT20_HDM-B4_STXAQX_ABN_4745_G	Pseudomonadales	48.43
H2_AT20_HDM-B4_STXAQX_ABN_4745_G	RETA1-2	51.57
H2_AT20_HDM-B5_STXAQX_ABN_4745_H	RETA1-2	100.00
H2_AT20_HDM-clA1_STXAQX_ABN_4745_I	RETA1-2	100.00
H2_AT20_HDM-clA2_STXAQX_ABN_4745_J	RETA1-2	100.00
H2_AT20_HDM-clA3_STXAQX_ABN_4745_K	RETA1-2	100.00
H2_AT20_HDM-clB2_STXAQX_ABN_4745_Z	RETA1-2	100.00
H2_AT20_HDM-clB3_STXAQX_ABN_4745_AA	RETA1-2	100.00
H2_AT20_HDM-F1_STXAQX_ABN_4745_N	RETA1-2	100.00
H2_AT20_HDM-F3_STXAQX_ABN_4745_P	RETA1-2	100.00
H2_HW15_001_KWCT-cl6_ABN_1992_A	Ruthmannia	18.41
H2_HW15_001_KWCT-cl6_ABN_1992_A	Limnobacter2	12.96
H2_HW15_001_KWCT-cl6_ABN_1992_A	Pseudomonadales	68.63
H2_HW15_002_KWCT-cl6_ABN_1992_B	Ruthmannia	36.13

Library	Symbiont clade	Relative abundance (%)
H2_HW15_002_KWCT-cl6_ABN_1992_B	Pseudomonadales	36.08
H2_HW15_002_KWCT-cl6_ABN_1992_B	RETA1-2	27.79
H2_HW15_003_KWCT-cl6_ABN_1992_C	Ruthmannia	23.15
H2_HW15_003_KWCT-cl6_ABN_1992_C	Flavobacteriales	26.05
H2_HW15_003_KWCT-cl6_ABN_1992_C	Pseudomonadales	37.39
H2_HW15_003_KWCT-cl6_ABN_1992_C	RETA1-2	13.41
H2_HW15_00A_KWCT-cl6_ABN_1966_A	UBA10353	32.75
H2_HW15_00A_KWCT-cl6_ABN_1966_A	Enterobacterales	25.20
H2_HW15_00A_KWCT-cl6_ABN_1966_A	RETA1-2	42.05
H2_HW15_00B_KWCT-cl6_ABN_1966_B	UBA10353	4.14
H2_HW15_00B_KWCT-cl6_ABN_1966_B	Salinisphaera	26.88
H2_HW15_00B_KWCT-cl6_ABN_1966_B	Enterobacterales	25.08
H2_HW15_00B_KWCT-cl6_ABN_1966_B	Flavobacteriales	7.73
H2_HW15_00B_KWCT-cl6_ABN_1966_B	Pseudomonadales	33.83
H2_HW15_00B_KWCT-cl6_ABN_1966_B	RETA1-2	2.33
H2_HW15_E5_KWCT-clE_ABN_3917_AE	Ruthmannia	11.22
H2_HW15_E5_KWCT-clE_ABN_3917_AE	Salinisphaera	10.76
H2_HW15_E5_KWCT-clE_ABN_3917_AE	Pseudomonadales	64.96
H2_HW15_E5_KWCT-clE_ABN_3917_AE	RETA1-2	13.06
H2_HW15_E6_KWCT-clE_ABN_3917_AF	Ruthmannia	9.10
H2_HW15_E6_KWCT-clE_ABN_3917_AF	Salinisphaera	11.69
H2_HW15_E6_KWCT-clE_ABN_3917_AF	Enterobacterales	4.21
H2_HW15_E6_KWCT-clE_ABN_3917_AF	Pseudomonadales	57.42
H2_HW15_E6_KWCT-clE_ABN_3917_AF	Rhodobacteraceae	17.58
H2_MA16_073_PAPO_ABN_2181_A	Pseudomonadales	46.46
H2_MA16_073_PAPO_ABN_2181_A	RETA1-2	53.54
H2_MA16_074_PAPO_ABN_2181_B	Pseudomonadales	18.84
H2_MA16_074_PAPO_ABN_2181_B	RETA1-2	81.16
H2_MA16_075_PAPO_ABN_2181_C	Pseudomonadales	68.56
H2_MA16_075_PAPO_ABN_2181_C	RETA1-2	31.44
H2_MPIMM19_A1_MPIAQ_ABN_4745_A	RETA1-2	100.00
H2_MPIMM19_A2_MPIAQ_ABN_4745_B	RETA1-2	100.00
H2_MPIMM19_A6_MPIAQ_ABN_4745_F	RETA1-2	100.00
H2_MPIMM19_B1_MPIAQ_ABN_3917_A	Enterobacterales	44.17
H2_MPIMM19_B1_MPIAQ_ABN_3917_A	Pseudomonadales	22.20
H2_MPIMM19_B1_MPIAQ_ABN_3917_A	Rhodobacteraceae	33.63
H2_MPIMM19_B4_MPIAQ_ABN_3917_C	Enterobacterales	75.28
H2_MPIMM19_B4_MPIAQ_ABN_3917_C	Rhodobacteraceae	24.72
H2_MPIMM19_B5_MPIAQ_ABN_3917_D	Pseudomonadales	100.00
H2_MPIMM19_B7_MPIAQ_ABN_3917_E	Enterobacterales	62.29
H2_MPIMM19_B7_MPIAQ_ABN_3917_E	Rhodobacteraceae	17.50
H2_MPIMM19_B7_MPIAQ_ABN_3917_E	RETA1-2	20.21
H2_MPIMM19_C1_MPIAQ_ABN_3917_F	Rhodobacteraceae	66.72
H2_MPIMM19_C1_MPIAQ_ABN_3917_F	RETA1-2	33.28
H2_MPIMM19_C1_MPIAQ_ABN_4745_G	RETA1-2	100.00
H2_MPIMM19_C3_MPIAQ_ABN_4745_I	RETA1-2	100.00
H2_MPIMM19_C4_MPIAQ_ABN_3917_G	Fabibacter1	58.67
H2_MPIMM19_C4_MPIAQ_ABN_3917_G	Pseudomonadales	26.30
H2_MPIMM19_C4_MPIAQ_ABN_3917_G	Rhodobacteraceae	15.03
H2_MPIMM19_C7_MPIAQ_ABN_3917_J	RETA1-2	100.00
H2_MPIMM19_E3_MPIAQ_ABN_3917_L	Limnobacter1	100.00

Library	Symbiont clade	Relative abundance (%)
H2_MPIMM19_E5_MPIAQ_ABN_3917_M	Pseudomonadales	89.28
H2_MPIMM19_E5_MPIAQ_ABN_3917_M	Rhodobacteraceae	10.72
H2_MPIMM19_E7_MPIAQ_ABN_3917_O	Fabibacter1	2.22
H2_MPIMM19_E7_MPIAQ_ABN_3917_O	Enterobacterales	4.70
H2_MPIMM19_E7_MPIAQ_ABN_3917_O	Pseudomonadales	89.21
H2_MPIMM19_E7_MPIAQ_ABN_3917_O	Rhodobacteraceae	3.87
H2_MPIMM19_F4_MPIAQ_ABN_3917_S	Ruthmannia	100.00
H2_MPIMM19_G2_MPIAQ_ABN_3917_V	RETA1-2	100.00
H2_MPIMM19_K2_MPIAQ_ABN_4745_T	RETA1-2	100.00
H3_HW15MH_A26_KWWS_ABY_1717_A	Teraskiella1	7.20
H3_HW15MH_A26_KWWS_ABY_1717_A	Ruthmannia	1.29
H3_HW15MH_A26_KWWS_ABY_1717_A	Flavobacteriales	88.92
H3_HW15MH_A26_KWWS_ABY_1717_A	Pseudomonadales	1.10
H3_HW15MH_A26_KWWS_ABY_1717_A	RETA1-2	1.49
H4_BZ15_009_TCMC-B14_ABN_1717_B	RETA1-2	100.00
H4_BZ15_015_TCMC-B04_ABN_1717_C	Schlegelella1	5.72
H4_BZ15_015_TCMC-B04_ABN_1717_C	Coxiella1	17.00
H4_BZ15_015_TCMC-B04_ABN_1717_C	Vulcaniibacterium1	5.94
H4_BZ15_015_TCMC-B04_ABN_1717_C	Enterobacterales	12.14
H4_BZ15_015_TCMC-B04_ABN_1717_C	Flavobacteriales	4.81
H4_BZ15_015_TCMC-B04_ABN_1717_C	Pseudomonadales	13.38
H4_BZ15_015_TCMC-B04_ABN_1717_C	Rhodobacteraceae	28.75
H4_BZ15_015_TCMC-B04_ABN_1717_C	RETA1-2	12.26
H4_BZ15_024_TCMC-B025_ABN_1717_D	Schlegelella1	4.41
H4_BZ15_024_TCMC-B025_ABN_1717_D	Limnobacter1	2.82
H4_BZ15_024_TCMC-B025_ABN_1717_D	Limnobacter2	6.09
H4_BZ15_024_TCMC-B025_ABN_1717_D	Vulcaniibacterium1	11.76
H4_BZ15_024_TCMC-B025_ABN_1717_D	Sutterella1	2.11
H4_BZ15_024_TCMC-B025_ABN_1717_D	Enterobacterales	32.00
H4_BZ15_024_TCMC-B025_ABN_1717_D	Pseudomonadales	17.45
H4_BZ15_024_TCMC-B025_ABN_1717_D	RETA1-2	23.35
H4_BZ15_055_TCFI-B087_ABN_1717_E	Coxiella1	32.73
H4_BZ15_055_TCFI-B087_ABN_1717_E	Ricketsia1	7.46
H4_BZ15_055_TCFI-B087_ABN_1717_E	RETA1-2	59.81
H4_BZ15_102_TCFI-B028_ABN_1717_F	Coxiella1	31.13
H4_BZ15_102_TCFI-B028_ABN_1717_F	RETA1-2	68.87
H4_BZ15_106_TCFI-B010_ABN_1717_G	Coxiella1	59.74
H4_BZ15_106_TCFI-B010_ABN_1717_G	RETA1-2	40.26
H4_BZ17NL_351_TCFI-83_ABN_4127_I	Pseudomonadales	26.33
H4_BZ17NL_351_TCFI-83_ABN_4127_I	RETA1-2	73.67
H4_HW15MH_004_KWWS_ABN_1272_A	Coxiella1	5.17
H4_HW15MH_004_KWWS_ABN_1272_A	Enterobacterales	69.28
H4_HW15MH_004_KWWS_ABN_1272_A	Pseudomonadales	18.36
H4_HW15MH_004_KWWS_ABN_1272_A	RETA1-2	7.20
H4_HW15MH_014_KWWS_ABN_1272_D	Enterobacterales	72.83
H4_HW15MH_014_KWWS_ABN_1272_D	Pseudomonadales	24.12
H4_HW15MH_014_KWWS_ABN_1272_D	RETA1-2	3.05
H4_HW15MH_015_KWWS_ABN_1272_C	Coxiella1	2.81
H4_HW15MH_015_KWWS_ABN_1272_C	Enterobacterales	77.01
H4_HW15MH_015_KWWS_ABN_1272_C	Pseudomonadales	14.74
H4_HW15MH_015_KWWS_ABN_1272_C	RETA1-2	5.45

Library	Symbiont clade	Relative abundance (%)
H4_HW15MH_A01_KWWS_ABY_1272_G	Coxiella1	13.58
H4_HW15MH_A01_KWWS_ABY_1272_G	Flavobacteriales	56.46
H4_HW15MH_A01_KWWS_ABY_1272_G	RETA1-2	29.95
H4_HW15MH_A07_KWWS_ABY_1272_L	Coxiella1	11.99
H4_HW15MH_A07_KWWS_ABY_1272_L	Limnobacter1	0.84
H4_HW15MH_A07_KWWS_ABY_1272_L	Flavobacteriales	70.15
H4_HW15MH_A07_KWWS_ABY_1272_L	RETA1-2	17.02
H4_HW15MH_A12_KWWS_ABY_1272_I	Coxiella1	6.04
H4_HW15MH_A12_KWWS_ABY_1272_I	Flavobacteriales	87.57
H4_HW15MH_A12_KWWS_ABY_1272_I	RETA1-2	6.39
H4_HW15MH_A16_KWWS_ABY_1272_J	Coxiella1	20.96
H4_HW15MH_A16_KWWS_ABY_1272_J	Limnobacter1	1.91
H4_HW15MH_A16_KWWS_ABY_1272_J	Pseudarciella1	0.78
H4_HW15MH_A16_KWWS_ABY_1272_J	Flavobacteriales	51.42
H4_HW15MH_A16_KWWS_ABY_1272_J	RETA1-2	24.94
H4_HW15MH_MG4_KWRR_ABN_1992_D	Enterobacterales	62.05
H4_HW15MH_MG4_KWRR_ABN_1992_D	Pseudomonadales	37.95
H4_HW15MH_MG6_KWRR_ABN_1992_F	Rhodobacteraceae	100.00
H6_HW15MH_007_KWWS_ABN_1272_B	Coxiella1	0.30
H6_HW15MH_007_KWWS_ABN_1272_B	Teraskiella1	1.21
H6_HW15MH_007_KWWS_ABN_1272_B	Chlamydia2	2.29
H6_HW15MH_007_KWWS_ABN_1272_B	Limnobacter1	2.33
H6_HW15MH_007_KWWS_ABN_1272_B	Limnobacter2	0.23
H6_HW15MH_007_KWWS_ABN_1272_B	Pseudarciella1	1.14
H6_HW15MH_007_KWWS_ABN_1272_B	UBA10353	0.75
H6_HW15MH_007_KWWS_ABN_1272_B	Sutterella1	0.47
H6_HW15MH_007_KWWS_ABN_1272_B	RETA3	0.74
H6_HW15MH_007_KWWS_ABN_1272_B	Enterobacterales	61.73
H6_HW15MH_007_KWWS_ABN_1272_B	Flavobacteriales	3.92
H6_HW15MH_007_KWWS_ABN_1272_B	Pseudomonadales	24.32
H6_HW15MH_007_KWWS_ABN_1272_B	Rhodobacteraceae	0.56
H6_HW15MH_016_KWWS_ABN_1272_E	Teraskiella1	12.88
H6_HW15MH_016_KWWS_ABN_1272_E	Chlamydia2	4.31
H6_HW15MH_016_KWWS_ABN_1272_E	Enterobacterales	56.77
H6_HW15MH_016_KWWS_ABN_1272_E	Pseudomonadales	26.05
H6_HW15MH_020_KWWS_ABN_1272_F	Teraskiella1	3.92
H6_HW15MH_020_KWWS_ABN_1272_F	Chlamydia2	10.01
H6_HW15MH_020_KWWS_ABN_1272_F	RETA3	1.45
H6_HW15MH_020_KWWS_ABN_1272_F	Enterobacterales	59.18
H6_HW15MH_020_KWWS_ABN_1272_F	Pseudomonadales	25.43
H6_HW15MH_A03_KWWS_ABY_1272_K	Teraskiella1	32.66
H6_HW15MH_A03_KWWS_ABY_1272_K	Chlamydia2	12.33
H6_HW15MH_A03_KWWS_ABY_1272_K	Limnobacter1	0.96
H6_HW15MH_A03_KWWS_ABY_1272_K	Pseudarciella1	0.59
H6_HW15MH_A03_KWWS_ABY_1272_K	RETA3	2.33
H6_HW15MH_A03_KWWS_ABY_1272_K	Flavobacteriales	48.30
H6_HW15MH_A03_KWWS_ABY_1272_K	Pseudomonadales	2.84
H6_HW15MH_A11_KWWS_ABY_1272_H	Teraskiella1	3.96
H6_HW15MH_A11_KWWS_ABY_1272_H	Chlamydia2	6.08
H6_HW15MH_A11_KWWS_ABY_1272_H	Limnobacter2	0.63
H6_HW15MH_A11_KWWS_ABY_1272_H	RETA3	1.86

Library	Symbiont clade	Relative abundance (%)
H6_HW15MH_A11_KWWS_ABY_1272_H	Enterobacterales	2.69
H6_HW15MH_A11_KWWS_ABY_1272_H	Flavobacteriales	80.13
H6_HW15MH_A11_KWWS_ABY_1272_H	Pseudomonadales	4.66
H7_BZ15_011_CBOH-B067_ABN_1717_M	Schlegelella1	2.20
H7_BZ15_011_CBOH-B067_ABN_1717_M	Teraskiella1	29.47
H7_BZ15_011_CBOH-B067_ABN_1717_M	Limnobacter2	2.22
H7_BZ15_011_CBOH-B067_ABN_1717_M	Vulcaniibacterium1	2.87
H7_BZ15_011_CBOH-B067_ABN_1717_M	Teraskiella2	2.81
H7_BZ15_011_CBOH-B067_ABN_1717_M	Sutterella1	0.90
H7_BZ15_011_CBOH-B067_ABN_1717_M	RETA3	1.24
H7_BZ15_011_CBOH-B067_ABN_1717_M	Enterobacterales	28.85
H7_BZ15_011_CBOH-B067_ABN_1717_M	Flavobacteriales	3.24
H7_BZ15_011_CBOH-B067_ABN_1717_M	Pseudomonadales	18.03
H7_BZ15_011_CBOH-B067_ABN_1717_M	Rhodobacteraceae	8.18
H7_BZ15_012_CBOH-B067_ABN_1717_N	Chlamydia1	21.77
H7_BZ15_012_CBOH-B067_ABN_1717_N	Schlegelella1	5.92
H7_BZ15_012_CBOH-B067_ABN_1717_N	Coxiella1	17.53
H7_BZ15_012_CBOH-B067_ABN_1717_N	Teraskiella1	15.36
H7_BZ15_012_CBOH-B067_ABN_1717_N	Limnobacter2	3.92
H7_BZ15_012_CBOH-B067_ABN_1717_N	Vulcaniibacterium1	5.37
H7_BZ15_012_CBOH-B067_ABN_1717_N	Sutterella1	2.66
H7_BZ15_012_CBOH-B067_ABN_1717_N	RETA3	9.14
H7_BZ15_012_CBOH-B067_ABN_1717_N	Enterobacterales	18.33
H7_BZ15_020_TCMC-B096_ABN_1717_O	Schlegelella1	1.57
H7_BZ15_020_TCMC-B096_ABN_1717_O	Teraskiella1	74.85
H7_BZ15_020_TCMC-B096_ABN_1717_O	Vulcaniibacterium1	1.66
H7_BZ15_020_TCMC-B096_ABN_1717_O	RETA3	6.28
H7_BZ15_020_TCMC-B096_ABN_1717_O	Enterobacterales	4.25
H7_BZ15_020_TCMC-B096_ABN_1717_O	Pseudomonadales	11.39
H7_BZ15_037_TCMC-B096_ABC2_1717_P	Teraskiella1	79.29
H7_BZ15_037_TCMC-B096_ABC2_1717_P	Vulcaniibacterium1	1.62
H7_BZ15_037_TCMC-B096_ABC2_1717_P	Ricketsia1	1.82
H7_BZ15_037_TCMC-B096_ABC2_1717_P	RETA3	7.44
H7_BZ15_037_TCMC-B096_ABC2_1717_P	Pseudomonadales	9.84
H7_BZ15_040_TCMC-B096_ABC3_1717_Q	Schlegelella1	4.59
H7_BZ15_040_TCMC-B096_ABC3_1717_Q	Teraskiella1	53.51
H7_BZ15_040_TCMC-B096_ABC3_1717_Q	Vulcaniibacterium1	4.40
H7_BZ15_040_TCMC-B096_ABC3_1717_Q	RETA3	17.67
H7_BZ15_040_TCMC-B096_ABC3_1717_Q	Pseudomonadales	19.83
H7_BZ15_041_TCMC-B096_ABC2_1717_R	Teraskiella1	78.00
H7_BZ15_041_TCMC-B096_ABC2_1717_R	RETA3	10.98
H7_BZ15_041_TCMC-B096_ABC2_1717_R	Pseudomonadales	11.02
H7_BZ15_045_TCMC-B096_ABT2_1717_S	Teraskiella1	78.61
H7_BZ15_045_TCMC-B096_ABT2_1717_S	RETA3	10.03
H7_BZ15_045_TCMC-B096_ABT2_1717_S	Pseudomonadales	11.36
H7_BZ15_048_TCMC-B096_ABT2_1717_T	Teraskiella1	59.18
H7_BZ15_048_TCMC-B096_ABT2_1717_T	RETA3	18.35
H7_BZ15_048_TCMC-B096_ABT2_1717_T	Pseudomonadales	22.47
H7_BZ15_049_TCMC-B096_ABT2_1717_L	Teraskiella1	72.85
H7_BZ15_049_TCMC-B096_ABT2_1717_L	RETA3	11.98
H7_BZ15_049_TCMC-B096_ABT2_1717_L	Pseudomonadales	15.17

Library	Symbiont clade	Relative abundance (%)
H8_BZ15_013_CBOH-B061_ABN_1717_H	Schlegelella1	1.02
H8_BZ15_013_CBOH-B061_ABN_1717_H	Coxiella1	3.71
H8_BZ15_013_CBOH-B061_ABN_1717_H	Teraskiella1	4.88
H8_BZ15_013_CBOH-B061_ABN_1717_H	Limnobacter1	1.92
H8_BZ15_013_CBOH-B061_ABN_1717_H	Ruthmannia	0.60
H8_BZ15_013_CBOH-B061_ABN_1717_H	Limnobacter2	0.42
H8_BZ15_013_CBOH-B061_ABN_1717_H	Vulcaniibacterium1	2.31
H8_BZ15_013_CBOH-B061_ABN_1717_H	Sutterella1	0.42
H8_BZ15_013_CBOH-B061_ABN_1717_H	RETA3	2.15
H8_BZ15_013_CBOH-B061_ABN_1717_H	Enterobacterales	45.03
H8_BZ15_013_CBOH-B061_ABN_1717_H	Flavobacteriales	3.98
H8_BZ15_013_CBOH-B061_ABN_1717_H	Pseudomonadales	29.39
H8_BZ15_013_CBOH-B061_ABN_1717_H	Rhodobacteraceae	4.16
H8_BZ15_014_CBOH-B061_ABN_1717_I	Coxiella1	11.06
H8_BZ15_014_CBOH-B061_ABN_1717_I	Teraskiella1	36.93
H8_BZ15_014_CBOH-B061_ABN_1717_I	Ruthmannia	4.90
H8_BZ15_014_CBOH-B061_ABN_1717_I	Vulcaniibacterium1	4.21
H8_BZ15_014_CBOH-B061_ABN_1717_I	Teraskiella2	1.50
H8_BZ15_014_CBOH-B061_ABN_1717_I	RETA3	5.23
H8_BZ15_014_CBOH-B061_ABN_1717_I	Enterobacterales	20.60
H8_BZ15_014_CBOH-B061_ABN_1717_I	Flavobacteriales	2.02
H8_BZ15_014_CBOH-B061_ABN_1717_I	Pseudomonadales	9.36
H8_BZ15_014_CBOH-B061_ABN_1717_I	Rhodobacteraceae	4.20
H8_BZ15_101_TCFI-B028_ABN_1717_J	Coxiella1	13.92
H8_BZ15_101_TCFI-B028_ABN_1717_J	Teraskiella1	58.71
H8_BZ15_101_TCFI-B028_ABN_1717_J	Ruthmannia	8.03
H8_BZ15_101_TCFI-B028_ABN_1717_J	Vulcaniibacterium1	3.26
H8_BZ15_101_TCFI-B028_ABN_1717_J	RETA3	16.08
H8_BZ15_103_TCFI-B028_ABN_1717_K	Schlegelella1	1.77
H8_BZ15_103_TCFI-B028_ABN_1717_K	Coxiella1	7.55
H8_BZ15_103_TCFI-B028_ABN_1717_K	Teraskiella1	74.30
H8_BZ15_103_TCFI-B028_ABN_1717_K	Ruthmannia	4.15
H8_BZ15_103_TCFI-B028_ABN_1717_K	Limnobacter2	1.33
H8_BZ15_103_TCFI-B028_ABN_1717_K	Vulcaniibacterium1	2.40
H8_BZ15_103_TCFI-B028_ABN_1717_K	Sutterella1	1.56
H8_BZ15_103_TCFI-B028_ABN_1717_K	RETA3	6.94
SRR5311040	RETA1-2	100.00
SRR5934055	RETA1-2	100.00
SRR5934125	RETA1-2	100.00

Table S4 | List of symbiont clades that are unique to host haplotype, sampling location or environment.

	Unique to…		
Symbiont clade	**…host haplotype**	**…sampling site**	**…environment**
Babeliales1	H6	Bocas del Toro	Free-living
Chlamydia2	H6		
Fabibacter1	H2	Bremen	Aquarium
Pseudarciella1		Hawai‘i	Aquarium
RETA4	H11	Mallorca	Aquarium
Ricketsia1		Twin Cays	Free-living
Salinisphaera	H2	Hawai'i	Aquarium
Schlegelella1			Free-living
Teraskiella2		Carrie Bow Cay	Free-living
UBA10353		Hawai‘i	Aquarium
Vulcaniibacterium1			Free-living

Table S5 | Statistically significant co-occurrences and exclusions between symbiont clades.

Symbiont clade A	**Symbiont clade B**	**Co-occurrence probability**	**p-value for less frequent co-occurrence than expected by chance**	**p-value for more frequent co-occurrence than expected by chance**
Enterobacterales	RETA1-2	0.037	0.01383	0.99996
Pseduomnadales	RETA1-2	0.055	0.00167	1.00000
RETA1-2	Teraskiella1	0.043	0.00826	1.00000

Notes

Note S1: Host haplotype H7 from Belize

We detected a previously undescribed host haplotype among the placozoan individuals sampled from Belize. This host haplotype appears to form a sister clade to the previously described host haplotypes H23 and H7. Here, we refer to this host haplotype as H7 but suggest that future analysis of the relations between this potentially novel host haplotype and the described H7 and H23 host haplotypes are required to define whether this haplotype represents a novel host haplotype lineage or is considered a subtype of previously described host haplotype lineages.

Extended data

The host and symbiont phylogenies that were generated and used during this study are available as newick files under https://itol.embl.de/shared/skowjdt77213.

Chapter V | General discussion, preliminary results and future directions

Since their first recognition in 1879 by Anton de Bary, our knowledge on symbiotic associations and how they shape life on earth has tremendously increased[1,2]. Today, we understand that bacterial symbioses not only gave rise to the eukaryotic lineage but that potentially all organisms live in symbiotic associations which largely impact the hosts' biology, e.g. their ecology, evolution, development and health[2–8]. Despite the growing recognition of multimember mutualism in nature, we still lack a comprehensive understanding of the symbiont diversity in many associations as well as their evolution and the impact of symbiont diversity on these association.

In my PhD thesis, I analyzed the symbiont consortia of two very different marine invertebrates: gutless oligochaetes that solely rely symbiotic consortia for nutrition and waste recycling, and placozoans, the arguably most simple animals that are commonly associated with bacterial symbionts which however, have not been extensively studied yet[9–14]. The focus of my analyses ranged from symbiont genome evolution (Chapter II) to symbiont community composition (Chapter III and IV) and the evolution of symbiont communities (Chapter III). I showed that genome reduction differs even between closely related symbiont lineages. In addition, it appeared that the lifestyle of symbionts and the resulting degree of genetic isolation of symbiont populations, rather than their phylogenetic relations, determine the progression of genome reduction (Chapter II). Focusing on the evolution of symbiont community composition in the obligate symbiosis of gutless oligochaetes, I found that the same dynamics that could prevent genome reduction in different symbiont lineages, e.g. switching of symbionts between co-occurring hosts and *de novo* acquisition from the environment, also lead to the evolution of highly flexible symbiont communities (Chapter III). Placozoans are a well-established model for eukaryote evolution and development, yet their associations with diverse symbiont communities remains widely unstudied. Given the diversity of placozoan symbionts, I

hypothesize placozoan hosts could serve as a model to further understand how symbiotic associations influence their hosts biology and *vice versa*, how diverse microbiomes evolve in association with different hosts at changing environmental conditions (Chapter IV).

Taken together, my thesis offers a potential answer to one of the fundamental questions in mutualism evolution: *how can genetically diverse symbiont consortia evolve?* and it discusses how symbiotic associations could benefit from genetically variable symbionts, both on the level of symbiont genomes and communities[15]. In addition, this work provides ideas how to further improve the analyses of symbiont community variation and it introduces the placozoans as a potential model system to study microbiome evolution and to generate further understanding of the impact from diverse microbiomes on their hosts' biology.

5.1 Symbiont community composition is the result of several ecological and evolutionary processes

Likely all animals are associated with diverse symbiont communities that strongly influence their hosts' evolution, ecology and health[2–5]. Our understanding of how different symbiont community compositions can influence different aspects of their hosts' biology has constantly been increasing. For example, the probably best studied symbiont community, the human gut microbiome, has undeniable implications for the health status of its hosts[16]. Yet, it remains challenging to understand which factors drive the symbiont community composition.

In my thesis, I asked which ecological and evolutionary parameters drive the symbiont community composition in gutless oligochaetes (Chapter III) and placozoans (Chapter IV). For gutless oligochaetes, symbiont community composition is highly specific to a given host species, yet this specificity pattern is lost over longer evolutionary time scales, indicating that the symbiont community composition evolves independently of the host evolution. Other

ecological parameters that I tested for could also not explain the symbiont community composition across the diversity of gutless oligochaete hosts. In placozoans, I observed a similar pattern: the symbiont community composition of these animals was highly variable, across and even within closely related haplotype lineages and none of the evolutionary and ecological parameters that I tested for, such as host taxonomy and phylogeny, geography and habitat could explain the observed composition patterns. In fact, the symbiont community composition of both, gutless oligochaetes and placozoans appear to be influenced by a variety of parameters.

The observation that different parameters influence the composition of symbiont communities is not unusual. In many other symbiotic associations, diverse biotic, physical and chemical parameters appear to influence symbiont community composition. For example, the symbionts of some coral species appear to vary with water depth and thus, light availability; in other coral species, symbiont community composition appeared to be linked to the taxonomy of the hosts[17–19]. The symbiont communities of sponges were shown to vary with the hosts' taxonomy, lifecycle stage as well as the composition of the co-occurring, free-living bacterial communities[20]. The microbiomes of vertebrate hosts such as bats and humans appear to be influenced by a variety of host-related and environmental factors, such as host genetics and diet[16,21]. Despite growing indications that diverse factors shape the symbiont community composition in different hosts, it often remains unclear *how* different parameters act on the symbiont community composition and *why* the composition changes in response to a given parameter[22]. I postulate that in order to understand how symbiont community composition changes between host individuals, we need to tease apart the community into its individual members and ask how a certain parameter could influence the presence and abundance of each of them. By unraveling the evolutionary and ecological dynamics that influence each of the members, we can then interpret the overall community composition as a result of these

individual dynamics. For example, many studies on the composition of the human gut microbiome provide indications that indeed, certain members of the human gut microbiome respond differently to external factors such as the mode of delivery, diet, microbiome manipulation with anti-, pre- and probiotics, and fecal microbiota transplantation[23].

Given the complexity of many symbiont communities, such as the vertebrate gut microbiomes, teasing apart the communities and analyzing all their individual members is often not feasible. This limitation shows the power of low complexity systems to study the evolution of composition of symbiont communities. The gutless oligochaetes and the placozoans were associated with ~30 different symbiont clades which allowed me to not only analyze the symbiont community composition but also the presence/absence patterns of each symbiont clade in respect to the host phylogeny and environmental parameters. In addition, both the gutless oligochaetes and placozoans are phylogenetically diverse and occur at various locations and environments around the globe. Sampling a broad phylogenetic range of host species from different locations and environments is a key step if we want to start untangling the effects of host evolution and the environment on symbiont community composition. Another limitation that I encountered during my analyses is the availability of information on the environmental parameters from the different sampling sites. I argue that we would need much more comprehensive metadata for our study organisms to understand how environmental parameters effect symbiont community composition and evolution. I assume that e.g. the availability of certain nutrients could alter the symbiont community composition at a given location as it was shown for different symbionts strains of deep sea mussels[24]. Ansorge *et al.* showed that the prevalence of strains that were capable of hydrogen oxidation was higher at vent fields where higher hydrogen concentrations were measured, thus indicating that environmental parameters select for the prevalence of the best-adapted symbionts. The work of Ansorge *et al.* leads me to the last limitation that I would like to discuss in the context of the analyses of symbiont

community composition: the taxonomic level at which we analyze individual symbionts. Here, I focused on comparing the genus-level composition of symbiont communities based on 16S rRNA gene sequences as well as single 16S rRNA phylotypes of the different symbiont genera. I argue that the analysis of symbiont communities based on marker genes and broader taxonomic levels is an efficient and straight forward method to assess symbiont community composition. In addition, previous work on the gutless oligochaete species *O. algarvensis* at the metabolism of its symbionts showed that the different symbiont genera express different metabolic pathways and thus occupy unique niches in the symbiosis[9,11,25]. I decided to analyze symbiont community composition on the genus level as I hypothesized that the genus-level cutoff divides the symbiont communities into biologically meaningful units. However, this analysis level potentially masks genomic differences between symbiont species from the same genus or even between symbiont strains. The presence and/or abundance of a given symbiont in a certain host and at a given location is not necessarily linked to the symbiont's taxonomy but rather to the symbiont's ability to interact with the host animal or to thrive at the local conditions. Such traits could be conserved between closely related symbiont lineages, i.e. within a symbiont genus but they could also be highly variable between closely related symbiont lineages as observed for free-living bacterial communities[26]. At best, future analyses of symbiont communities would take the genetic variability of symbiont genera, species and strains into account. I am convinced that phylogenetically broad and global taxon sampling of host animals, extensive collection of metadata and analyzing symbiont diversity down to the strain level will improve our future understanding of symbiont community composition and its evolution.

5.2 Genetic diversity in symbiont communities

The central themes of my thesis are the genetic diversity in symbiont communities, which evolutionary paths lead to the observed diversity, and how this diversity impacts symbiotic

associations. Evolutionary theory predicts that symbiont diversity in mutualistic, and especially obligate associations should be low as they are assumed to most efficiently persist when stabilizing mechanisms such as high symbiont fidelity and/or high symbiont specificity ensure partner quality[27–29]. However, symbiont communities that consist of several, genetically diverse members, e.g. several strains, species and/or genera, are abundant in nature[24,30,31]. How such symbiont diversity evolves and persists in mutualistic associations remains largely elusive[15]. In the following paragraphs, I will discuss the levels of symbiont diversity that I observed in gutless oligochaete and placozoan hosts and how I imagine the evolution of the given diversity based on the data at hand. In addition, I will discuss potential benefits of the observed symbiont diversity and speculate which factors favored the evolution of the strikingly diverse symbiont communities of gutless oligochaetes.

5.2.1 What is the level of genetic variability of symbiont communities from gutless oligochaete and placozoan host?

In general, symbiont communities can exhibit genetic diversity on different levels ranging from harboring genetically different strains to harboring several taxonomically different symbionts.[24,30,31]. Gutless oligochaetes and placozoans were previously found to associated with a variety of symbiont genera[9,32,41–45,33–40]. Thus, and given the previously discussed metabolic differences between different symbiont genera in the gutless oligochaete *O. algarvensis*, I started my analyses of the genetic diversity of symbiont communities on the level of symbiont genera. My results from Chapter III and IV confirmed that individuals of both hosts, gutless oligochaetes and placozoans, were associated with a variety of symbiont genera and that the genus-level composition is highly variable between host individuals. In addition, the subsequent phylogenetic analyses of members from each symbiont genus revealed that several 16S rRNA phylotypes from the same symbiont genus were associated with different host individuals. Each host individual was usually associated with a single 16S rRNA phylotype

but not all host individuals shared the same phylotype. Thus, the genetic variability between different host individuals is not only the result of the association with different symbiont genera but also with different phylotypes from the same symbiont genus. In addition, initial analyses of the prevalence of SNPs in single-host symbiont populations of the Gamma4 symbiont from *I. exumae* indicates a high intraspecific variation for at least some gutless oligochaete symbionts (Chapter II).

5.2.2 How can genetic variation of symbiont communities evolve?

The evolution of genetic diversity in symbiont communities is mainly constrained by high symbiont fidelity and high symbiont specificity, i.e. strictly vertical symbiont transmission and/or the selection for certain symbionts. Strict vertical symbiont transmission from one host generation to the next imposes an extreme genetic bottleneck on the symbiont populations as only a subset of the symbiont community that is associated with the parental animals will be transmitted to the offspring[46]. A similar genetic bottleneck effect can result from stringent host selection for certain symbionts[47,48]. In contrast, low symbiont fidelity and low symbiont specificity, e.g. the repeated acquisition of symbionts from the environment and the switching of symbionts between hosts, can increase the genetic variability of symbiont communities[49–51]. In Chapter III, I analyzed the fidelity of gutless oligochaete symbionts. Gutless oligochaetes obligately depend on their symbiont consortia[9–11]. As obligate associations are commonly observed to display high symbiont fidelity in order to maintain their stability, such associations are expected to display low symbiont diversity[52]. However, the symbiont communities of gutless oligochaetes were strikingly diverse and the community composition appears to be the result of varying levels of symbiont fidelity. In concordance with morphological studies of the symbiont transmission in gutless oligochaetes, I indeed observed comparably high fidelity between symbiont and host mitochondrial lineages within host populations, indicating that maternal transmission is most common from one host generation to the next[53,54]. However, the

overall community composition of gutless oligochaete symbionts and the symbiont fidelity patterns across a broad host phylogeny indicated that over longer evolutionary time scales, other mechanisms than maternal transmission dominate. I showed that low symbiont fidelity in form of environmental acquisition, loss and host switching of symbionts are likely responsible for the observed genetic variability and the composition of symbiont communities (Chapter III, Figure 1). On the one hand, acquisition of new symbiont genera from the environment or other hosts increases the genetic variability in a given community. On the other hand, the switching of symbiont phylotypes from the same genus between hosts species enables genetic exchange between symbiont populations which in consequence, increases the genetic variability in a given symbiont community without altering its genus-level composition.

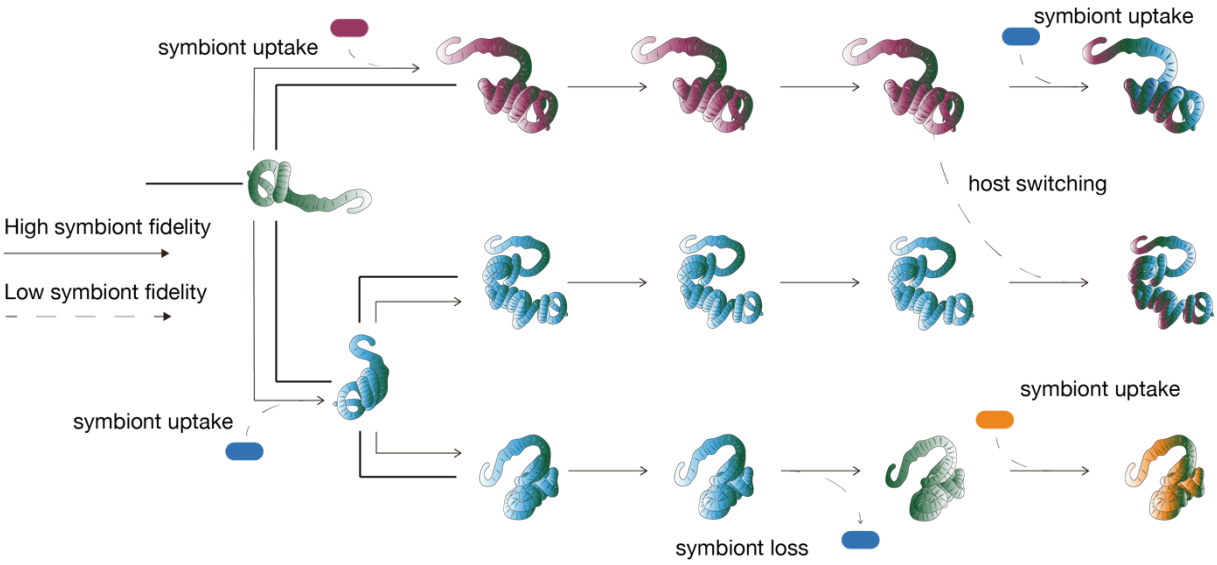


Figure 1 | The composition of gutless oligochaete symbiont communities is shaped by varying levels of symbiont fidelity. The scheme illustrates how varying levels of symbiont fidelity could theoretically influence the symbiont community composition in three distinct host lineages. The color of the worms reflects symbiont community composition. Solid arrows indicate high symbiont fidelity, dashed arrows indicate low symbiont fidelity.

In Chapter IV, I applied similar methods to understand the evolution of the symbiont communities from placozoan hosts. The placozoan symbiosis is quite different to the gutless oligochaete symbiosis. Whilst the gutless oligochaetes solely depend on their microbial partners

for nutrition and waste product recycling, Placozoa live from the external digestion of algae and biofilms and do not obligately rely their microbial consortia[42,55,56]. Potentially, many of the placozoan-associated bacteria are commensalists that feed on the nutrients that are generated during the external digestion and thus, benefit from attaching to the placozoan hosts (Chapter IV). In concordance with the apparent low dependence of the partners, I also observed low symbiont fidelity for most of the symbiont clades (Figure 2). Strikingly, this includes symbionts that likely occur intracellularly and even within the endoplasmatic reticulum, i.e. not only within a host cell but within a cell compartment[42]. I am fascinated by this finding as it indicates that intracellular symbionts of placozoans are invading host cells and compartments on a regular basis. Given that placozoans can be easily sampled and cultivated in the lab, they could serve as a model to understand fundamental molecular mechanisms on colonization and persistence of microbes that enter hosts' cells and compartments, such as Chlamydiae, Coxiellaceae and Rickettsiales[42,57–59].

Figure 2 | All placozoan symbionts show varying levels of symbiont fidelity across their host diversity. Pairwise nucleotide dissimilarity of the symbionts 16S rRNA gene are shown in relation to the pairwise nucleotide dissimilarity of the host m16S rRNA gene. R^2 values of each linear regression from testing correlation between the pairwise 16S rRNA nucleotide distance of each symbiont clade versus the pairwise m16S rRNA gene nucleotide distance are included in each panel.

I was intrigued to observe varying levels of symbiont fidelity in the gutless oligochaete and the placozoan hosts. Based on the observation that obligate symbionts commonly tend to show high symbiont fidelity whereas facultative symbionts commonly tend to show lower symbiont fidelity, I hypothesized that the variably degrees of symbiont fidelity could be a mechanism that enables the evolution of genetic diversity in obligate symbiont communities (Chapter III)[52]. The combination of different levels of symbiont fidelity could ensure the presence of obligate symbionts via high symbiont fidelity and at the same time genetically and metabolically diverse symbiont communities can be maintained via lower fidelity of accessory symbionts. The results

from the placozoan dataset indicate that this mechanism might not only be at work in obligate symbiosis but could fulfill a similar function in other microbiomes as well. I am convinced that linking a symbiont's function, e.g. through its genetic potential, to environmental parameters and its degree of fidelity will increase our understanding of why genetic variation arises and persists in complex symbiont communities.

5.2.3 What are the benefits of genetically diverse symbiont communities?

Whilst the association with several symbionts can destabilize the symbiotic association due to competition between different symbionts or the emergence of cheaters, i.e. bacterial lineages that exploit host resources without returning any goods, hosts can also benefit from genetically diverse symbiont communities. First, genetic diversity of the symbionts often translates into phenotypic heterogeneity, i.e. functional diversity, through the ability to use e.g. different carbon and energy source or to synthesize essential nutrients such as vitamins and amino acids[9,11,24,25,60]. Thus, a more diverse symbiont community could provide optimized nutrition their hosts', which could be crucial when the host solely relies on its symbionts for nutrition. The ability to use different sources of nutrition becomes especially important in the marine habitat where environmental conditions are highly fluctuating. Again, hosts that solely rely on their symbionts for nutrition are particularly reliant on their symbionts to be able to cope with these environmental changes. Second, the association of different host species with distinct symbiont communities could enable their co-existence in the same environment through niche separation. The probably most famous example for niche separation between species is the Darwin finches[61]. In this example, the beaks of different species have different shapes that allow them to feed on different food sources and thus, reduce or even eliminate the competition for certain resources between species[61,62]. Similarly, the association with distinct symbiont communities can reduce the competition and enable the coexistence of several host species[63]. Lastly, genetic diversity and the mechanisms enabling it, e.g. environmental acquisition or

symbiont switching can prevent massive genome reduction and decay of the symbiotic association and thus, provide long-term stability[50].

As relatively little is known about the metabolism of the different placozoan symbionts and their role within this symbiosis, it is yet impossible to tell why these animals are associated with the observed symbiont consortia and how they benefit from their diversity. For gutless oligochaetes however, it is well understood that the symbionts provide the entire nutrition for their hosts I hypothesize that their global success is the result of diverse symbiont communities which indeed could enable them to metabolize a variety of carbon and energy sources and thus thrive in the highly fluctuating environmental conditions that prevail in the marine sediments the worms inhabit as well as avoid competition between coexisting species[9,11,25]. As discussed in Chapter II and III, I also assume that the long-term stability of such a flexible symbiosis is the result of varying levels of symbiont fidelity and high symbiont diversity.

5.2.4 Why does the genetic variability of symbiont communities vary across different thiotrophic symbioses?

The highly variable symbiont communities of gutless oligochaetes are exceptionally diverse compared to the symbionts of other hosts that obtain their nutrition from thiotrophic symbionts. The majority of these hosts are associated with a single thiotrophic symbiont, e.g. *Solemya* and *Vesicomyid* clams, different genera of giant tube worms, *Paracatenula* flatworms, *Kentrophoros* ciliates and *Astomonema* and stilbonematine nematodes[64–71]. Bathymodioline mussels and lucinid clams diverge from this pattern, hosting several different symbiont types. Still, even when considering the extreme strain variability of the thiotrophic symbiont in *Bathymodiolus* mussels, the symbiont diversity observed in gutless oligochaetes remains unmatched[24,72,73]. Throughout my studies, I have been asking myself *what are the differences between gutless oligochaetes and other chemotrophic hosts that would explain this striking*

diversity of symbionts? With the data at hand, I cannot provide a final answer to the question but I would like to discuss a few of my thoughts in the following paragraph.

I hypothesize that the symbiont diversity of gutless oligochaetes is mainly influenced by three parameters: the dependence on the symbionts for nutrition, the transmission mode and potential infection sites. The gutless oligochaetes entirely depend on their symbiont consortia for nutrition. In contrast to other thiotrophic animals that have direct access to nutrients from the environment, such as stilbonematine nematodes, bathymodioline mussels as well as *Vesicomyid* and lucinid clams that retained at least parts of their digestive system, gutless oligochaete are restricted to the environmental resources that are provided by their symbionts[74–77]. Thus, metabolically diverse symbiont communities are likely to represent a selective advantage for their gutless hosts. In addition, the reproduction, the mixed-mode symbiont transmission, and the anatomy of gutless oligochaetes could also favor the evolution of diverse symbiont communities. First, gutless oligochaetes reproduce sexually and the physical contact of different individuals during copulation provides the opportunity to exchange symbionts between individuals, even of different host species during interspecies mating attempts. Second, the majority of symbionts appears to be vertically transmitted from one generation to the next, reducing the need for selecting the right bacteria from environmental populations anew in every generation[53,54]. Third, gutless oligochaetes potentially are at relatively low risk of getting infected with pathogens as the anatomy of gutless oligochaetes leaves very few infection sites for bacteria from the environment; they do not only lack a digestive and excretory systems but also any respiratory organs such as gills that could be infected by environmental bacteria. The only ways for environmental bacteria to infect gutless oligochaetes, besides the reproductive organs, would be the closely after the egg deposition before the egg cocoon is fully developed. The eggs are already colonized with bacterial symbionts that were vertically transmitted, this could imply that invading bacteria already would need to compete against these symbionts in

order to persist in the developing animal. Given the short time frame for colonization and the reduced chance to persist in the host tissue for potential pathogens, *de novo* infection with pathogens is likely rare, as also reflected in my analysis of the gutless oligochaete symbionts (Chapter III). The few animals that could become infected could be outcompeted by uninfected individuals on the level of host populations. Thus, I assume that diverse symbiont communities of gutless oligochaetes evolved as a result of the reduced need of selection, both for suitable symbiont lineages and against putative pathogens, as well as the exchange between symbiont communities during host reproduction. The combination of complete nutritional dependence, mixed-mode symbiont transmission as well as few infection sites appears to be unique to the gutless oligochaetes and I postulate that in fact this combination could have favored the evolution of exceptionally diverse thiotrophic symbiont consortia (Table 1).

Table 1 | The combination of nutritional dependence and mixed-mode transmission is unique to gutless oligochaetes among chemosynthetic symbiosis.

Host	Digestive system	Symbiont transmission	Infection sites
Gutless oligochaetes	None[10]	Mixed-mode	Freshly laid eggs
Bathymodioline mussels	Reduced[77]	Horizontal[24,78,79]	Gills, digestive system
Lucinid clams	Reduced[76]	Horizontal[80–83]	Gills, digestive system
Solemya clams	None[84]	Mixed-mode[51]	Gills
Vesicomyid clams	Reduced[75]	Mainly vertical[75]	Gills
Giant tubeworms	None[85,86]	Horizontal[87]	Larvae
Stilbonematine nematodes	Present[74]	Horizontal (?)[74]	Digestive system
Astomonema nematodes	None[74]	Horizontal (?)[74]	-
Paracatenula flatworms	None[88]	Vertical[70]	-
Kentrophoros ciliates	None	Vertical[69]	-

5.3 Which factors mediate multimember symbioses?

Despite their striking variability, the symbiont communities of gutless oligochaetes and placozoans consist of a limited number of bacterial symbionts, compared to the diversity of free-living bacterial populations. On the one hand, not all bacteria are adapted to live within a host animal, thus reducing the number of potential bacteria partners that could be acquired from the environment. On the other hand, selection for certain host-symbiont as well as for certain symbiont-symbiont combinations could reduce taxonomic diversity within a symbiont community.

5.3.1 Host selection could mediate genus-level composition of symbiont communities

Host selection appears to be less stringent in gutless oligochaetes compared to other thiotrophic symbioses. Still, host selection might still play a role in e.g. maintaining the stable association between nearly all gutless oligochaetes and their primary symbiont, *Cand.* Thiosymbion as observed for the association between *Cand.* Thiosymbion and stilbonematine nematodes[89,90]. In addition, the symbiont community composition of gutless oligochaetes was very stable within host species, indicating that on the host species level, host selection via e.g. the immune systems, could influence the overall community composition (Chapter III)[91]. In placozoans, host selection could determine the association between a certain host haplotype and the associated Rickettsiales symbiont type (Chapter IV).

5.3.2 Symbiont-symbiont interactions appear to play a minor role in shaping symbiont community composition

Symbiont-symbiont interactions appear to play a subordinated role for symbiont community composition in the gutless oligochaete and the placozoan symbioses. In gutless oligochaetes, only two symbiont pairs showed significant co-occurrence. However, these symbionts only occurred in small, monophyletic clades of closely related hosts. Thus, the apparent co-

occurrence could be the result of host selection. In the placozoans, I observed three pairs of symbionts that showed significant exclusion. Similar to the observations made for gutless oligochaetes, at least one of the symbionts was specific to a certain host lineage, which indicates that the host genotype rather than symbiont-symbiont interactions could select against the co-occurrence of the given symbiont pairs. In contrast to my studies on animal hosts, the composition of the root microbiome of plants is influenced by symbiont-symbiont interactions[92]. I obviously did not acquire enough data to claim that symbiont-symbiont interactions are more important in shaping plant microbiomes than animal microbiomes but it would be interesting to conduct a study that systematically reviews the importance of symbiont-symbiont interactions for shaping the microbiome composition across a broad diversity of hosts. I would like to point out that the low number of observed symbiont exclusion patterns does not imply that the symbionts of gutless oligochaetes and placozoans do not interact with other bacteria. Indeed, certain symbionts could be preventing the association with other bacterial clades, e.g. pathogens so efficiently that we were never able to capture them attempting to infect the host animals.

5.3.3 How can gutless oligochaete symbionts interact with their hosts and with each other?

I screened genome drafts of all gutless oligochaete symbiont genera for the presence of genes that encode secretion systems that are known to be involved in the interaction of bacteria with other cells (Figure 3)[93–95]. Several of the gutless oligochaete symbionts encode secretion systems. Strikingly, these features appear to be more wide-spread in symbionts that are commonly switched between host or repeatedly acquired from the environment indicating that they could play an important role in mediating the association between the symbiont and their hosts. In addition, the prevalence of genes that encode for flagella and thus, bacterial mobility indicates that these symbionts could still be capable of thriving outside of a host environment[96]. This includes symbionts which are phylogenetically intermixed with free-living bacteria, e.g.

the Delta1 and the Gamma3 symbionts (Chapter III). For these symbionts, I assumed that they are regularly acquired from the environment and thus, are closely related to members of the free-living populations. In contrast, some symbionts that were thought to only persist as gutless oligochaete symbionts retained genes that encode for flagella, e.g. the Spirochete and the alphaproteobacterial symbionts (Chapter III). These symbionts were not phylogenetically intermixed with any free-living bacteria that are represented in public databases, indicating that they have become genetically isolated from free-living populations. Yet, the presence of genes that encode for flagella indicate that these symbionts could have a free-living stage and thus, could switch between gutless oligochaete hosts not only during host mating but also through environmental transmission. However, the free-living symbiont stage is apparently not viable over longer time scale as it did not appear to persist in the free-living population yet.

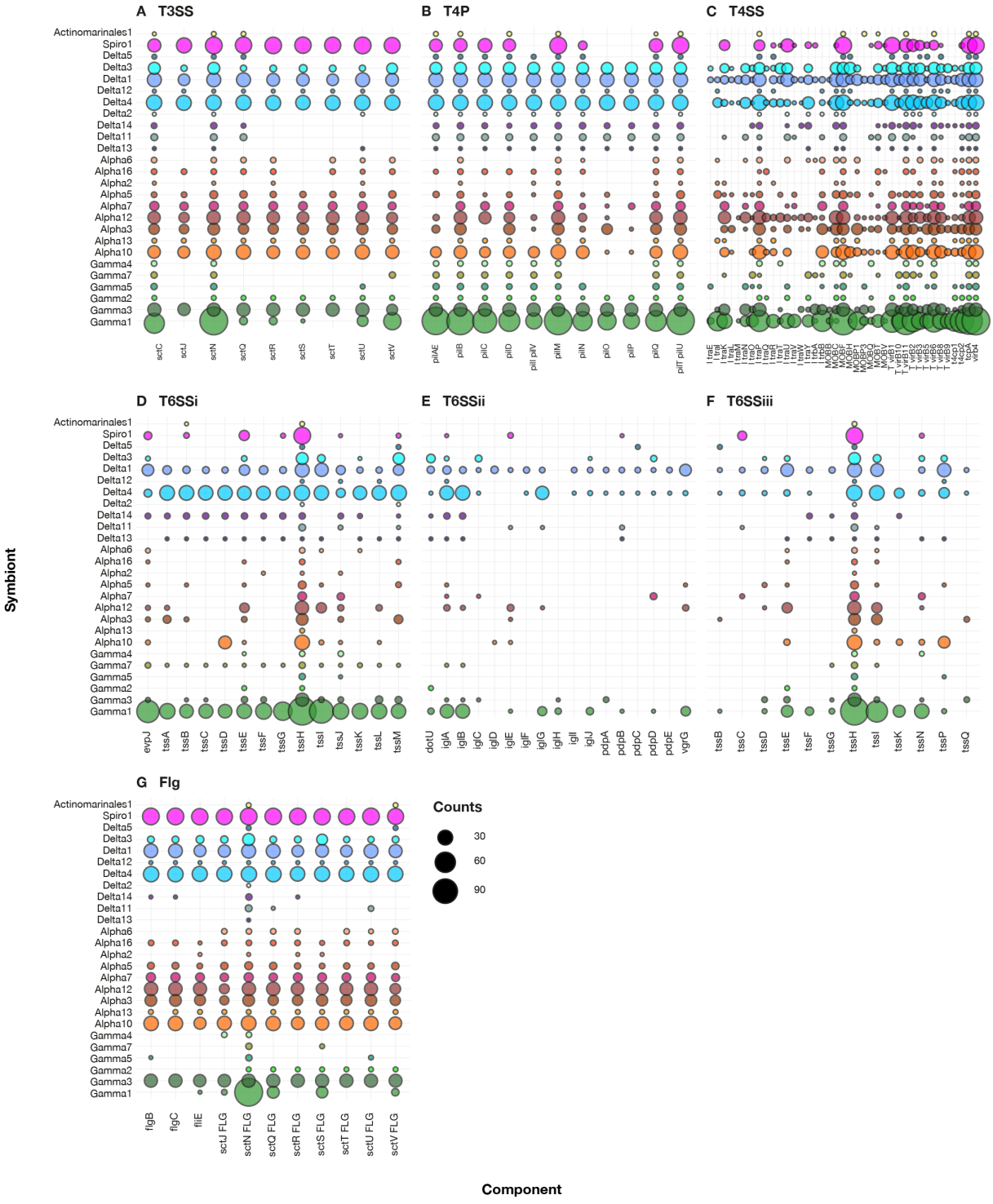


Figure 3 | Secretion systems as well as motility related genes are more wide-spread in symbionts that are commonly switched or environmentally acquired. Each panel shows the distribution and counts of genes that are assigned to certain secretion systems **(A-F)** or flagella **(G)** in symbiont MAGs. The dot size indicates the number of symbiont MAGs in which the respective gene was detected.

5.4 How do the diverse symbiont communities of gutless oligochaetes influence the metabolic capacities of individual hosts?

In Chapter III, I speculated that the phylogenetic diversity and composition of symbiont communities could indicate phenotypic differences between host individuals that harbor different symbiont communities. Based on our knowledge on the metabolism of the *O. algarvensis* symbionts, I assume differently composed communities of gutless oligochaete symbionts would encode for different metabolic pathways and thus, enable the hosts to thrive at least partly on different nutrient sources[9,11,25]. In order to test this hypothesis, I statistically compared the symbiont community composition as well as the metabolic pathway composition of each symbiont community. I found that differences in the pathway composition do not perfectly reflect the phylogenetic symbiont community composition. This could be the result of metabolic redundancy between certain symbiont genera, i.e. different genera of deltaproteobacterial symbionts that are all capable of sulfate reduction as well as differences in the metabolic pathways that are encoded by different phylotypes of the same genus; e.g. some representatives of the Gamma1 and the Alpha10 symbionts showed the potential to fix nitrogen but this respective pathway was not encoded by all members of the Gamma1 and Alpha10 genera. Yet, the link between phylogenetic and metabolic composition of symbiont communities was highly significant, emphasizing the importance of phylogenetic composition of symbiont communities for the encoded metabolic potential ($p=0$). As pointed out by Kleiner *et al.* (2018), the theoretical redundancy between two symbionts based on the pathways they encode and express does not always result in actual metabolic redundancy of the symbionts as they still could use the same pathways to metabolize different sources of e.g. organic carbon[25]. Thus, it would be interesting to test how the usage of different resources of a symbiont community relates to its phylogenetic composition.

During the analysis of the metabolic pathways that are encoded by gutless oligochaete symbionts, I became curious how wide-spread certain pathways are and whether there are certain core functions that are present in all gutless oligochaete individuals. I thus compiled a list of pathways that were encoded by symbionts of at least 95% of the analyzed host individuals in order to get a first idea whether our knowledge on the metabolism of the *O. algarvensis* symbionts is representative of the symbiont metabolism of gutless oligochaetes in general (Figure 4). The symbiont metabolism that was shared by at least 119 of 125 host individuals included the degradation of sugars via glycolysis and the TCA cycle, the biosynthesis and degradation of fatty acids, the incomplete 3-hydroxypropionate bicycle (3-HPB), the biosynthesis of amino acids and their precursors, the degradation of few amino acids, the biosynthesis of storage compounds such as glycogen and polyhydroxyalkanoates (PHA) and sulfur oxidation. Interestingly, the shared symbiont metabolism of the majority of gutless oligochaete symbionts is typical for other thiotrophic symbioses where it is commonly encoded by just a single thiotrophic symbiont, such as the *Cand.* Riegeria symbionts of *Paracatenula* flatworms[97]. Such a stable set of core functions, despite the striking flexibility of symbiont community composition, illustrates how strong the hosts depend on these services that provide their nutrition.

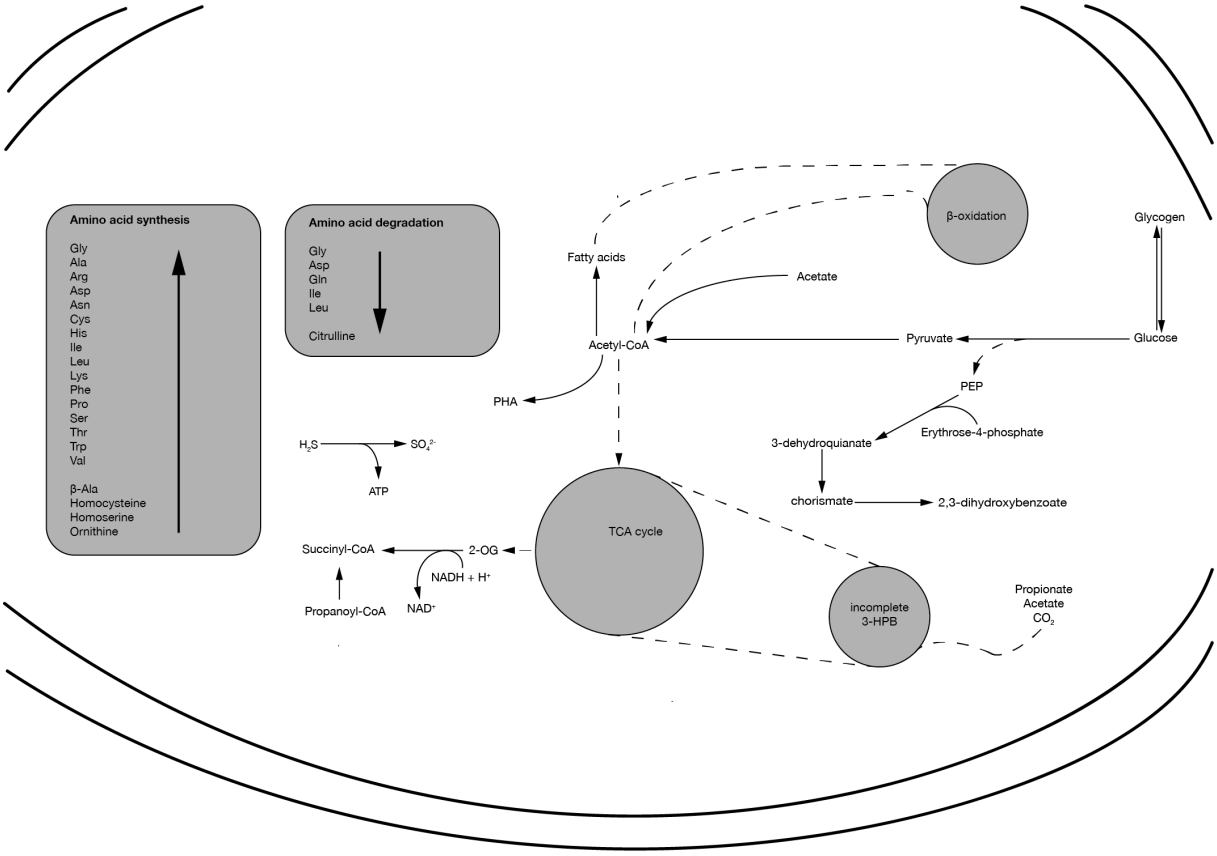


Figure 4 | The symbiont communities of different gutless oligochaete individuals share only few metabolic pathways. Scheme of the pathways involved in the central symbiont metabolism that were annotated in the symbiont communities of at least 95% of the analyzed individuals. Solid arrows represent detected reactions, dashed lines represent likely exchange of metabolites between pathways.

Strikingly, carbon fixation via the Calvin-Benson-Bassham cycle as well as sulfate reduction were encoded in less than 119 symbiont communities. The absence of genes encoding for the CBB cycle in not only one individual of *I. exumae* but in seven other gutless oligochaetes individuals indicates that the hosts are not necessarily dependent on autotrophic carbon fixation by their symbionts. Instead, heterotrophic assimilation or anaplerotic fixation of carbon appear to be sufficient to meet the demands for carbon of gutless oligochaete symbiosis (Chapter II). In addition, the syntrophic sulfur cycling that was observed between the delta- and gammaproteobacterial symbiont of *O. algarvensis* does not appear to be crucial for all gutless oligochaete individuals; especially in environments where hydrogen sulfide concentrations are

sufficient, the association with sulfate reducing bacteria might not be crucial to maintain the metabolism of the thiotrophic symbionts. In summary, our in-depth knowledge on the symbiont metabolism of *O. algarvensis* appears to be only partly representative for the metabolism that is shared by all gutless oligochaete individuals. Whilst thiotrophy, heterotrophic carbon degradation and the synthesis of amino acids and storage compounds appear to be wide-spread, the use of other carbon and energy sources such as carbon monoxide and hydrogen are only present in certain host individuals. In addition, other pathways such as the fixation of nitrogen that were known to be encoded by thiotrophic symbionts of other host groups such as stilbonematine nematodes and lucinid clams but that were not detected in the symbionts of *O. algarvensis,* are encoded by the symbionts of other host species[98]. These findings emphasize the flexibility of the gutless oligochaete symbiosis and show the need to tease apart the symbiont metabolism across the diversity of host species, symbiont genera and environments to understand the metabolic interactions in these associations.

5.5 Can the symbiont community composition affect symbiont genome evolution?

In Chapter II, I discussed potential factors that could influence the degree of genome reduction, e.g. the genetic isolation of a symbiont population and the function of the symbiont within the symbiosis. Preliminary results indicate that in addition, the symbiont community composition in *I. exumae* could influence the genome reduction of the Gamma4 symbionts. During my analyses of the *I. exumae* metagenomes and their symbiont community composition, I was surprised to see that one individual (IexuBAH1) was associated with a symbiont community that was largely different to the symbiont communities of the other two individuals (IexuBAH2 and IexuBAH3). Whilst IexuBAH2 and IexuBAH3 shared the Gamma4 symbiont as well as an alphaproteobacterial and a deltaproteobacterial symbiont, the Gamma4 symbiont was the only one that was also shared by IexuBAH1 (Figure 5). Interestingly, the genome of the Gamma4 symbiont of IexuBAH1 appears to be further reduced than the genomes of the other two *I.*

exumae individuals as seen by a lower number of pathways that were completely or partially annotated (Figure 5). It is tempting to speculate that symbiont community composition could influence the degree of genome reduction. Due to the very small sample size, I cannot draw any conclusions from this initial observation. However, it could be worthwhile to generate a bigger dataset containing more *I. exumae* specimens in order to investigate whether symbiont community composition and genome reduction are indeed linked and whether certain sets of secondary symbionts fulfill metabolic functions that were lost due to genome reduction.

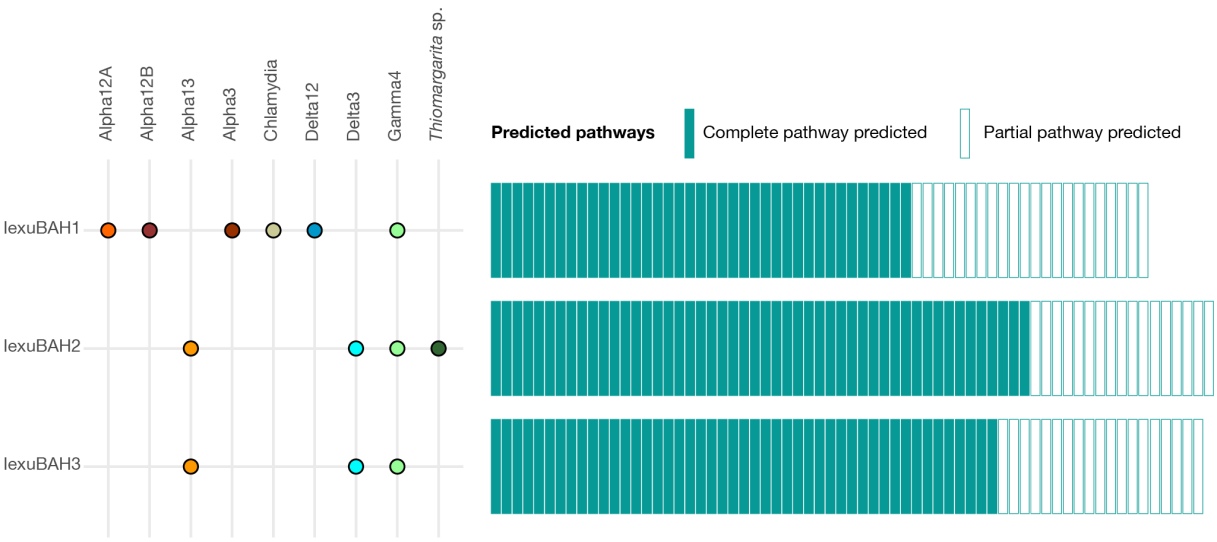


Figure 5 | The composition of symbiont communities could affect symbiont genome reduction. Left panel: presence/absence patterns of *I. exumae* symbionts in three individuals. Right panel: Number of pathways involved in the central symbiont metabolism that were completely or partially annotated. Boxes are ordered by annotation status, not by pathway annotation.

5.6 Concluding remarks

The data and analyses presented in my thesis provide new insights into the evolution of symbiotic associations and highlights the importance of studying non-model organisms such as marine invertebrates. Marine invertebrates are often easy to sample *in situ*, thus allowing us to understand the composition and role of symbiotic associations in nature. In addition, we can

relatively easily sample a broad phylogenetic diversity of a given host group from globally distributed sampling locations which enables us to understand the environmental and host evolutionary effects on the composition and evolution of symbiont communities.

By studying a broad diversity of hosts and thus symbionts, I was able to gain new insights that improve our understanding of symbiont genome evolution. I showed that genome reduction appears to be a highly variable process that proceeds differently across closely related symbiont lineages and can be largely influence by the association type. Furthermore, I could pinpoint processes, e.g. symbiont switching between distantly related host lineages, that appear to be important drivers for the phylogenetic and very likely, functional composition of symbiont communities; not only in the associations studied in this thesis but potentially in many other symbioses.

References

1. de Bary, A. Die Erscheinung der Symbiose . *Karl J. Trübner Verlag*. (1879).
2. McFall-Ngai, M. *et al.* Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad. Sci.* **110**, 3229 LP – 3236 (2013).
3. Sapp, J. The Symbiotic Self. *Evol. Biol.* **43**, 596–603 (2016).
4. Gilbert, S. F., Bosch, T. C. G. & Ledón-Rettig, C. Eco-Evo-Devo: developmental symbiosis and developmental plasticity as evolutionary agents. *Nat. Rev. Genet.* **16**, 611–622 (2015).
5. Henry, L. P., Bruijning, M., Forsberg, S. K. G. & Ayroles, J. F. The microbiome extends host evolutionary potential. *Nat. Commun.* **12**, 5141 (2021).
6. 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. Yale University Press.* (1970).
7. Sagan, L. On the origin of mitosing cells. *J. Theor. Biol.* **14**, 225-IN6 (1967).
8. Margulis, L. & Fester, R. Symbiosis assource of evolutionary innovation: speciation and morphogenesis. *The MIT Press.* 93–101 (1991).
9. Woyke, T. *et al.* Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**, 950–955 (2006).
10. Dubilier, N. *et al.* Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. *Nature* **411**, 298–302 (2001).
11. Kleiner, M. *et al.* Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E1173--82 (2012).
12. Laumer, C. E. *et al.* Support for a clade of Placozoa and Cnidaria in genes with minimal compositional bias. *Elife* **7**, e36278 (2018).
13. Eitel, M., Osigus, H.-J., DeSalle, R. & Schierwater, B. Global Diversity of the Placozoa. *PLoS One* **8**, e57131 (2013).
14. Simion, P. *et al.* A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals. *Curr. Biol.* **27**, 958–967 (2017).
15. Heath, K. D. & Stinchcombe, J. R. Explaining Mutualism Variation: A New Evolutionary Paradox? *Evolution (N. Y).* **68**, 309–317 (2014).
16. Gomaa, E. Z. Human gut microbiota/microbiome in health and diseases: a review. *Antonie Van Leeuwenhoek* **113**, 2019–2040 (2020).

17. Sampayo, E. M., Franceschinis, L., Hoegh-Guldberg, O. V. E. & Dove, S. Niche partitioning of closely related symbiotic dinoflagellates. *Mol. Ecol.* **16**, 3721–3733 (2007).
18. Cooper, T. F. *et al.* Niche specialization of reef-building corals in the mesophotic zone: metabolic trade-offs between divergent Symbiodinium types. *Proc. R. Soc. B Biol. Sci.* **278**, 1840–1850 (2011).
19. Evans, J. S., Erwin, P. M., Shenkar, N. & López-Legentil, S. A comparison of prokaryotic symbiont communities in nonnative and native ascidians from reef and harbor habitats. *FEMS Microbiol. Ecol.* **94**, (2018).
20. Sacristán-Soriano, O. *et al.* Ontogeny of symbiont community structure in two carotenoid-rich, viviparous marine sponges: comparison of microbiomes and analysis of culturable pigmented heterotrophic bacteria. *Environ. Microbiol. Rep.* **11**, 249–261 (2019).
21. Phillips, C. D. *et al.* Microbiome analysis among bats describes influences of host phylogeny, life history, physiology and geography. *Mol. Ecol.* **21**, 2617–2627 (2012).
22. Hacquard, S. *et al.* Microbiota and Host Nutrition across Plant and Animal Kingdoms. *Cell Host Microbe* **17**, 603–616 (2015).
23. Hasan, N. & Yang, H. Factors affecting the composition of the gut microbiota, and its modulation. *PeerJ* **7**, e7502 (2019).
24. Ansorge, R. *et al.* Functional diversity enables multiple symbiont strains to coexist in deep-sea mussels. *Nat. Microbiol.* **4**, 2487–2497 (2019).
25. Kleiner, M. *et al.* Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. *Proc. Natl. Acad. Sci.* **115**, E5576 LP-E5584 (2018).
26. Gibbons, S. M. Microbial community ecology: Function over phylogeny. *Nat. Ecol. Evol.* **1**, 32 (2017).
27. Leigh Jr, E. G. The evolution of mutualism. *J. Evol. Biol.* **23**, 2507–2528 (2010).
28. Archetti, M. *et al.* Economic game theory for mutualism and cooperation. *Ecol. Lett.* **14**, 1300–1312 (2011).
29. Sachs, J. L., Mueller, U. G., Wilcox, T. P. & Bull, J. J. The Evolution of Cooperation. *Q. Rev. Biol.* **79**, 135–160 (2004).
30. Ellegaard, K. M. & Engel, P. Genomic diversity landscape of the honey bee gut microbiota. *Nat. Commun.* **10**, 446 (2019).

31. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
32. Blazejak, A., Erseus, C. & Amann Rudolf & Dubilier, N. Coexistence of bacterial sulfide oxidizers, sulfate reducers, and spirochetes in a gutless worm (Oligochaeta) from the Peru margin. *Appl. Environ. Microbiol.* **71**, 1553–1561 (2005).
33. Dubilier, N. *et al.* Phylogenetic diversity of bacterial endosymbionts in the gutless marine oligochete *Olavius loisae* (Annelida). *Mar. Ecol. Prog. Ser.* **178**, 271–280 (1999).
34. Ruehland, C., Blazejak, A., Lott, C., Loy, A. & Erséus Christer & Dubilier, N. Multiple bacterial symbionts in two species of co-occurring gutless oligochaete worms from Mediterranean sea grass sediments. *Environ. Microbiol.* **10**, 3404–3416 (2008).
35. Blazejak, A., Kuever, J., Erséus, C., Amann, R. & Dubilier, N. Phylogeny of 16S rRNA, Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase, and Adenosine 5′-Phosphosulfate Reductase Genes from Gamma- and Alphaproteobacterial Symbionts in Gutless Marine Worms (Oligochaeta) from Bermuda and the Bahamas. *Appl. Environ. Microbiol.* **72**, 5527–5536 (2006).
36. Dubilier, N., Giere, O., Distel, D. L. & Cavanaugh, C. M. Characterization of chemoautotrophic bacterial symbionts in a gutless marine worm (Oligochaeta, Annelida) by phylogenetic 16S rRNA sequence analysis and in situ hybridization. *Appl. Environ. Microbiol.* **61**, 2346–2350 (1995).
37. Krieger, J., Giere, O. & Dubilier, N. Localization of RubisCO and sulfur in endosymbiotic bacteria of the gutless marine oligochaete *Inanidrilus leukodermatus* (Annelida). *Mar. Biol.* **137**, 239–244 (2000).
38. Bergin, C. *et al.* Acquisition of a Novel Sulfur-Oxidizing Symbiont in the Gutless Marine Worm *Inanidrilus exumae*. *Appl. Environ. Microbiol.* **84**, e02267-17 (2018).
39. Kamm, K., Osigus, H.-J., Stadler, P. F., DeSalle, R. & Schierwater, B. *Trichoplax* genomes reveal profound admixture and suggest stable wild populations without bisexual reproduction. *Sci. Rep.* **8**, 11168 (2018).
40. Driscoll, T., Gillespie, J. J., Nordberg, E. K., Azad, A. F. & Sobral, B. W. Bacterial DNA Sifted from the *Trichoplax adhaerens* (Animalia: Placozoa) Genome Project Reveals a Putative Rickettsial Endosymbiont. *Genome Biol. Evol.* **5**, 621–645 (2013).
41. Eitel, M. *et al.* Comparative genomics and the nature of placozoan species. *PLOS Biol.* **16**, e2005359 (2018).

42. Gruber-Vodicka, H. R. *et al.* Two intracellular and cell type-specific bacterial symbionts in the placozoan *Trichoplax* H2. *Nat. Microbiol.* **4**, 1465–1474 (2019).
43. Srivastava, M. *et al.* The *Trichoplax* genome and the nature of placozoans. *Nature* **454**, 955–960 (2008).
44. Guidi, L., Eitel, M., Cesarini, E., Schierwater, B. & Balsamo, M. Ultrastructural analyses support different morphological lineages in the phylum placozoa Grell, 1971. *J. Morphol.* **272**, 371–378 (2011).
45. Klinges, J. G. *et al.* Phylogenetic, genomic, and biogeographic characterization of a novel and ubiquitous marine invertebrate-associated Rickettsiales parasite, *Candidatus* Aquarickettsia rohweri, gen. nov., sp. nov. *ISME J.* **13**, 2938–2953 (2019).
46. Mira, A. & Moran, N. A. Estimating Population Size and Transmission Bottlenecks in Maternally Transmitted Endosymbiotic Bacteria . *Microb. Ecol.* **44**, 137–143 (2002).
47. McFall-Ngai, M. J. Consequences of Evolving With Bacterial Symbionts: Insights from the Squid-Vibrio Associations. *Annu. Rev. Ecol. Syst.* **30**, 235–256 (1999).
48. Visick, K. L. & McFall-Ngai, M. J. An Exclusive Contract: Specificity in the *Vibrio fischeri-Euprymna scolopes* Partnership. *J. Bacteriol.* **182**, 1779–1787 (2000).
49. Guyomar, C. *et al.* Multi-scale characterization of symbiont diversity in the pea aphid complex through metagenomic approaches. *Microbiome* **6**, 181 (2018).
50. Russell, S. L. *et al.* Horizontal transmission and recombination maintain forever young bacterial symbiont genomes. *PLOS Genet.* **16**, e1008935 (2020).
51. Russell, S. L., Corbett-Detig, R. B. & Cavanaugh, C. M. Mixed transmission modes and dynamic genome evolution in an obligate animal–bacterial symbiosis. *ISME J.* **11**, 1359–1371 (2017).
52. Fisher, R. M., Henry, L. M., Cornwallis, C. K., Kiers, E. T. & West, S. A. The evolution of host-symbiont dependence. *Nat. Commun.* **8**, 15973 (2017).
53. Giere, O. Ecology and Biology of Marine Oligochaeta – an Inventory Rather than another Review. *Hydrobiologia* **564**, 103–116 (2006).
54. Giere, O. & Langheld, C. Structural organisation, transfer and biological fate of endosymbiotic bacteria in gutless oligochaetes. *Mar. Biol.* **93**, 641–650 (1987).
55. Pearse, V. B. & Voigt, O. Field biology of placozoans (*Trichoplax*): distribution, diversity, biotic interactions. *Integr. Comp. Biol.* **47**, 677–692 (2007).
56. Smith, C. L., Pivovarova, N. & Reese, T. S. Coordinated Feeding Behavior in *Trichoplax*, an Animal without Synapses. *PLoS One* **10**, e0136098 (2015).

57. Horn, M. Chlamydiae as Symbionts in Eukaryotes. *Annu. Rev. Microbiol.* **62**, 113–131 (2008).
58. Duron, O., Doublet, P., Vavre, F. & Bouchon, D. The Importance of Revisiting Legionellales Diversity. *Trends Parasitol.* **34**, 1027–1037 (2018).
59. Thomas, S. Rickettsiales: Biology, Molecular Biology, Epidemiology, and Vaccine Development. *Springer International Publishing* (2016).
60. McCutcheon, J. P. & Moran, N. A. Functional Convergence in Reduced Genomes of Bacterial Symbionts Spanning 200 My of Evolution. *Genome Biol. Evol.* **2**, 708–718 (2010).
61. Darwin C., On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. *John Murray.* (1859).
62. Lamichhaney, S. *et al.* Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**, 371–375 (2015).
63. Joy, J. B. Symbiosis catalyses niche expansion and diversification. *Proc. R. Soc. B Biol. Sci.* **280**, 20122820 (2013).
64. Ott, J. A. *et al.* Tackling the Sulfide Gradient: A Novel Strategy Involving Marine Nematodes and Chemoautotrophic Ectosymbionts. *Mar. Ecol.* **12**, 261–279 (1991).
65. Polz, M. F. *et al.* Phylogenetic analysis of a highly specific association between ectosymbiotic, sulfur-oxidizing bacteria and a marine nematode. *Appl. Environ. Microbiol.* **60**, 4461–4467 (1994).
66. Polz, M. F., Felbeck, H., Novak, R., Nebelsick, M. & Ott, J. A. Chemoautotrophic, sulfur-oxidizing symbiotic bacteria on marine nematodes: Morphological and biochemical characterization. *Microb. Ecol.* **24**, 313–329 (1992).
67. Krueger, D. M., Dubilier, N. & Cavanaugh, C. M. Chemoautotrophic symbiosis in the tropical clam *Solemya occidentalis* (Bivalvia: Protobranchia): ultrastructural and phylogenetic analysis. *Mar. Biol.* **126**, 55–64 (1996).
68. Fisher, C. R. Chemoautotrophic and methanotrophic symbioses in marine invertebrates. *Rev. Aquat. Sci.* **2**, 399–436 (1990).
69. Seah, B. K. B. *et al.* Specificity in diversity: single origin of a widespread ciliate-bacteria symbiosis. *Proc. R. Soc. B Biol. Sci.* **284**, 20170764 (2017).
70. Gruber-Vodicka, H. R. *et al. Paracatenula* an ancient symbiosis between thiotrophic Alphaproteobacteria and catenulid flatworms. *Proc. Natl. Acad. Sci.* **108**, 12078 LP – 12083 (2011).

71. Robidart, J. C. *et al.* Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed through metagenomics. *Environ. Microbiol.* **10**, 727–737 (2008).
72. Dubilier, N. & Bergin Claudia & Lott, C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat Rev Micro* **6**, 725–740 (2008).
73. Lim, S. J. *et al.* Extensive Thioautotrophic Gill Endosymbiont Diversity within a Single *Ctena orbiculata* (Bivalvia: Lucinidae) Population and Implications for Defining Host-Symbiont Specificity and Species Recognition. *mSystems* **4**, e00280-19 (2021).
74. Ott, J. & Bright Monika & Bulgheresi, S. Symbioses between marine nematodes and sulfur-oxidizing chemoautotrophic bacteria. *Symbiosis* **36**, 103–126 (2004).
75. Ozawa, G. *et al.* Ancient Occasional Host Switching of Maternally Transmitted Bacterial Symbionts of Chemosynthetic Vesicomyid Clams. *Genome Biol. Evol.* **9**, 2226–2236 (2017).
76. Gros, O., Frenkiel, L. & Felbeck, H. Sulfur-oxidizing endosymbiosis in *Divaricella quadrisulcata* (Bivalvia: Lucinidae): morphological, ultrastructural, and phylogenetic analysis. *Symbiosis* (2000).
77. Franke, M., Geier, B., Hammel, J. U., Dubilier, N. & Leisch, N. Coming together—symbiont acquisition and early development in deep-sea bathymodioline mussels. *Proc. R. Soc. B Biol. Sci.* **288**, 20211044 (2021).
78. 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.* **16**, 3699–3713 (2014).
79. Won, Y.-J. *et al.* Environmental Acquisition of Thiotrophic Endosymbionts by Deep-Sea Mussels of the Genus *Bathymodiolus Appl. Environ. Microbiol.* **69**, 6785 LP – 6792 (2003).
80. Gros, O., Darrasse, A., Durand, P., Frenkiel, L. & Moueza, M. Environmental transmission of a sulfur-oxidizing bacterial gill endosymbiont in the tropical lucinid bivalve *Codakia orbicularis*. *Appl. Environ. Microbiol.* **62**, 2324–2330 (1996).
81. Gros, O., Frenkiel, L. & Moueza, M. Embryonic, larval, and post-larval development in the symbiotic clam *Codakia orbicularis* (Bivalvia: Lucinidae). *Invertebr. Biol.* 86–101 (1997).
82. Gros, O., Frenkiel, L. & Mouëza, M. Gill filament differentiation and experimental colonization by symbiotic bacteria in aposymbiotic juveniles of *Codakia orbicularis* (Bivalvia: Lucinidae). *Invertebr. Reprod. Dev.* **34**, 219–231 (1998).

83. Gros, O., Duplessis, M. R. & Felbeck, H. Embryonic development and endosymbiont transmission mode in the symbiotic clam *Lucinoma aequizonata* (Bivalvia: Lucinidae). *Invertebr. Reprod. Dev.* **36**, 93–103 (1999).
84. Reid, R. G. B. & Bernard, F. R. Gutless Bivalves. *Science (80-.).* **2008**, (1980).
85. Felbeck, H. Chemoautotrophic Potential of the Hydrothermal Vent Tube Worm, *Riftia pachyptila* Jones (Vestimentifera). *Science* **213**, 336–338 (1981).
86. Cavanaugh, C. M., Gardiner, S. L., Jones, M. L. & Jannasch H. W. & Waterbury, J. B. Prokaryotic Cells in the Hydrothermal Vent Tube Worm *Riftia pachyptila* Jones: Possible Chemoautotrophic Symbionts. *Science (80-.).* **213**, 340–342 (1981).
87. Nussbaumer, A. D., Fisher, C. R. & Bright, M. Horizontal endosymbiont transmission in hydrothermal vent tubeworms. *Nature* **441**, 345–348 (2006).
88. Sterrer, W. & Rieger, R. Retronectidae-a new cosmopolitan marine family of Catenulida (Turbellaria). *Biol. Turbellaria* 63–92 (1974).
89. Nussbaumer, A. D., Bright, M., Baranyi, C., Beisser, C. J. & Ott, J. A. Attachment mechanism in a highly specific association between ectosymbiotic bacteria and marine nematodes. *Aquat. Microb. Ecol.* **34**, 239–246 (2004).
90. Zimmermann, J. *et al.* Closely coupled evolutionary history of ecto- and endosymbionts from two distantly related animal phyla. *Mol. Ecol.* **25**, 3203–3223 (2016).
91. Wippler, J. *et al.* Transcriptomic and proteomic insights into innate immunity and adaptations to a symbiotic lifestyle in the gutless marine worm *Olavius algarvensis. BMC Genomics* **17**, 942 (2016).
92. Uroz, S., Courty, P. E. & Oger, P. Plant Symbionts Are Engineers of the Plant-Associated Microbiome. *Trends Plant Sci.* **24**, 905–916 (2019).
93. Russell, A. B., Peterson, S. B. & Mougous, J. D. Type VI secretion system effectors: poisons with a purpose. *Nat. Rev. Microbiol.* **12**, 137–148 (2014).
94. J., H. C. Type III Protein Secretion Systems in Bacterial Pathogens of Animals and Plants. *Microbiol. Mol. Biol. Rev.* **62**, 379–433 (1998).
95. Wallden, K., Rivera-Calzada, A. & Waksman, G. Microreview: Type IV secretion systems: versatility and diversity in function. *Cell. Microbiol.* **12**, 1203–1212 (2010).
96. Bardy, S. L., Ng, S. & Jarrell, K. F. Prokaryotic motility structures. Microbiology. *Microbiology* **149**, 295–304 (2003).
97. Jäckle, O. *et al.* Chemosynthetic symbiont with a drastically reduced genome serves as primary energy storage in the marine flatworm *Paracatenula. Proc. Natl. Acad. Sci.* **116**, 8505 LP – 8514 (2019).

98. Petersen, J. M. *et al.* Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation. *Nat. Microbiol.* **2**, 16195 (2016).

Personal contribution to each manuscript

Manuscript 1 | Chapter II

Conceptual design	60%
Data acquisition and experiments	50%
Analysis and interpretation of results	80%
Preparation of figures and tables	95%
Writing the manuscript	95%

Manuscript 2 | Chapter III

Conceptual design	50%
Data acquisition and experiments	70%
Analysis and interpretation of results	65%
Preparation of figures and tables	90%
Writing the manuscript	70%

Manuscript 3 | Chapter IV

Conceptual design	55%
Data acquisition and experiments	70%
Analysis and interpretation of results	85%
Preparation of figures and tables	95%
Writing the manuscript	95%

Acknowledgments

“Love is not a finite emotion. We don’t have only so much to share.
Our hearts create love as we need it.”
Dan Brown, Origin

Without constant support and advice throughout the last four and a half years, this thesis would not have been possible. I would like to thank everyone who was part of this incredible journey!

Prof. Dr. Nicole Dubilier, thank you for the opportunity to do my PhD in your department, for your support and advice throughout the years and of course for reviewing this thesis and being part of my examination committee.
Prof. Dr. Alexander J. Probst, not only for reviewing my thesis but also for your support and advice; thank you very much.
I'd also like to thank **Prof. Dr. Hinrich Schulenburg** for also reviewing this thesis.
Thank you **Prof. Dr. Christian Wild, Prof. Dr. Michael Friedrich, Dr. Luis Humberto Orellana Retamal** and **Anna Carlotta Kück** for agreeing to join my examination committee.

Dr. Harald R. Gruber-Vodicka, Harald, thank you very much for your supervision and your support during my time as a PhD student. Thank you for giving me the freedom to independently conduct my research whilst always being available for giving motivational speeches or advice.

Thank you, **Dr. Juliane Wippler,** without whom I might never started to work as a bioinformatician. Thank you for supervising my master thesis and the first part of my PhD thesis. Thank you for introducing me to bioinformatics, teaching me how to code and your endless patience in teaching me how to read and troubleshoot error messages. Thank you for your brutal honesty, your support and your friendship.

This PhD would not have been possible without many, many people that were involved in sample collection and processing, sequencing and providing the IT infrastructure that allowed me to analyze a total of ~500 metagenomes. This includes the crews and scientists of various research cruises and field expeditions. Special thanks to **NADYA,** for a magical field trip. I would jump on the next plain to Belize to spend another few weeks on Carrie Bow with you. 'Free beers tomorrow!'. Also, I would like to thank **Dr. Christer Erséus** for sharing his amazing collection of gutless oligochaete specimens, **Miriam Sadowski** for performing DNA extractions, **Dr. Tanja Woyke** and **Dr. Bruno Hüttel** for their efforts in sequencing our samples, **the IT department** here at the MPI and **Roland Dietrich** from the Max Planck genomes center for providing IT infrastructure and immediate trouble shooting.

Thank you to many colleagues at the MPI for making our everyday life easier, especially **Bernd Stickfort** for providing access to every publication and to **Susanne Krüger, Ulrike Tietjen, Tina Peters, Ralf Schwenke, Bettina Lahmers, Dr. Christiane Glöckner** and **Martina Patze** for your administrative support.

I would like to thank all my collaborators and co-authors for their scientific input. Thank you **Dr. Manuel Kleiner** and **Marlene Jensen** for your investment into my PhD projects. I am super thankful for discussing science with you and I really enjoy collaborating with the two of you, I hope this will continue. Thank you **Dr. Brandon Seah,** discussing my projects with you blows my mind away every single time and I always need a few days to digest all the input you provide but this is always 100% worth it and I learned so much from you. I would also like to thank my lab rotation student **Jan Brüwer** and my master student **Marius Poulain** for the chance to gain some supervision experience.

I never regretted joining the **Department of Symbiosis**; thank you to all past and current members for being great colleagues. I loved working with you on a day to day basis and your productive criticism helped me to push myself and made me a better scientist.
Thank you **Benedikt, Dolma, Max, Miguel and Niko!** Thank you for your friendship, for your advice on science and life, for being cheerful and uplifting. Thank you for board game and movie and pizza and gin tonic nights, for several rum tastings, for every BBQ at the Weser, sun downers and the Thirsty Thursdays. Thank you Dolma for a short but amazing pre field trip vacation. Bene, Max and Miguel, thank you so much for a great time at the ISME in Leipzig, I am not sure if I ever laughed that much again in just a few days.
Thank you **Rebecca** for being such an inspiration. I always appreciated you for your kick-ass science and for being one of the nicest persons I know. Your advice for the finishing phase of my PhD really made my life during the last months a lot easier.
Thank you **Merle** for bringing so much good mood into the department. Thank you for the helpful discussions on bioinformatics we had and for being a supportive colleague in all ups and downs during the life as a PhD student. I am so happy that we will be living in the same area again soon.
Thank you **Caro** for the great company in the office.
Tina! I once said that you would deserve a full page of acknowledgments and I 100% meant it but I will try to keep it short. Thank you for being the most direct and honest person I know. Thank you for always being critical of my science, my presentations and my manuscript drafts. Your feedback always helped me to improve. Thank you for going through the highest highs and the lowest lows of life and love together. Thank you for being such spontaneous and adventurous friend (I still think you are one of the most fearless persons I know). Thank you for endless discussions and making me feel understood. I cannot imagine the last four and a half years without you.

I would also like to thank **Miri** for an amazing inter-department flat share. Moving in with you was one of the best decisions I ever made. Thank you for these beautiful months and for our continuing friendship.

There is more to life than the PhD and I am incredible thankful to have great friends that made me forget about work from time to time. Thank you **Svenja** and **Marius** for your long lasting friendship**,** the two of you mean the world to me and I cannot wait to spend the next decades with you by my side (at least emotionally). Thank you **Svea** for making my life so much more exciting; spending time with you never gets boring and I am looking forward to future adventures. Thank you **Annika, Freia** and **Lucia** for bringing me down to earth, life feels so much easier when I hang out with you. Thank you **André, Lisa, Lukas** and **Matthias,** for being great training partners and more importantly friends. Our training sessions helped me to shut down my computer and get my head free nearly every day during the last three years. I hope to compete in many more triathlons and spend many more adventurous cycling vacations with you. Of course, I would also like to thank not only **Lisa**, but also **Eva** and **Katrin** for being amazing team mates; I cannot wait for our next triathlon seasons. Thank you **Jan** and **Alex** for fun cycling tours. Thank you **Theresa** for sharing the PhD experience and the mutual support for going through this part of our lives. Thank you **Margot, Lisa, Isi** and **Victor** for (market) coffees, climbing sessions, dinners, beers and star gazing.

Of course, I would also like to thank my family. **Mama, Papa,** ich weiß nicht wo ich anfangen und wo ich aufhören soll. Danke, für eine Kindheit am Mittelmeer, die Begeisterung für's Meer muss ja irgendwo herkommen. Danke, dass ihr meine Monologe über Biochemie und Zellbiologie am Frühstückstisch ertragen habt („Was redet unser Kind da nur schon wieder für wirres Zeug?"). Danke für eure anhaltenden Unterstützung. Besonderer Dank gilt außerdem **Michael** und **Pascal,** eure Unterstützung bedeutet mir eine Menge, ich bin froh euch als Teil der Familie zu wissen.

Last but not least, **Dennis.** Thank you so much for your support and for always believing in me, even or especially when I do not. I am looking forward to all upcoming adventures and challenges that we will master together.

Bremen, 28. September 2021

Versicherung an Eides Statt

Ich, **Anna Mankowski,**
Matrikelnummer 2987645

versichere an Eides Statt durch meine Unterschrift, dass ich die vorstehende Arbeit selbständig und ohne fremde Hilfe angefertigt und 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.

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.

Des Weiteren bestätige ich hiermit gemäß §7, Abs. 7, Punkt 4, dass die zu Prüfungszwecken beigelegte elektronische Version meiner Dissertation identisch ist mit der abgegebenen gedruckten Version. Ich bin mit der Überprüfung meiner Dissertation gemäß §6 Abs. 2, Punkt 5 mit qualifizierter Software im Rahmen der Untersuchung von Plagiatsvorwürfen einverstanden.

_____
Ort, Datum, Unterschrift