# Characterization of host-symbiont molecular interactions and evolutionary relationships in the gutless oligochaete *Olavius algarvensis*

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

Dem Fachbereich Biologie/Chemie der
Universität Bremen
vorgelegt von

Juliane Wippler

Bremen
April 2016

Die Untersuchungen zur vorliegenden Arbeit wurden in der Abteilung Symbiose am Max-Planck-Institut für Marine Mikrobiologie in Bremen durchgeführt.

Universität Bremen

1. Gutachterin: Prof. Dr. Nicole Dubilier

2. Gutachterin: Prof. Dr. Monika Bright

Tag des Promotionskolloquiums: 1. Juni 2016

Umschlag:

Vorderseite: Stereomikroskopisches Bild eines darmlosen Oligochaeten.

Hinterseite: Probennahme in der Bucht von Sant‘ Andrea, Elba

# Table of Contents

## Summary

Once thought to be a curiosity, it has now become increasingly recognized that beneficial symbioses between animals and microbes are common and wide-spread in nature. Animal-microbe interactions had mostly been studied from a medical perspective. However, thanks to major advances in sequencing technology that allowed the genomic study of non-cultivable microorganisms, it has become apparent that not only most (if not all) animals are colonized by microbes, but that the majority of these microbes is harmless or even beneficial to the host animal and often contributes integral functions to the biology of these animals. In some cases, animals form such highly intimate associations with bacteria that the association becomes obligate for their survival.

One such example is the obligate symbiosis between marine gutless oligochaete worms (Annelida, Phallodrilinae) and their chemosynthetic bacterial symbionts. Over the course of evolution, gutless oligochaetes have lost their entire digestive system, including mouth, gut, and anus, as well as their excretory organs, becoming entirely dependent on their symbionts to provide all necessary nutrients and to remove the waste products of the host. Each gutless oligochaete host harbors its own highly species-specific consortium of bacteria.

Among the gutless oligochaetes, the model species *Olavius algarvensis* is one of the best studied. This Mediterranean species lives in symbiotic association with multiple bacterial phylotypes, including two gammaproteobacterial sulfur-oxidizers (*OalgG1* and *OalgG3*), two deltaproteobacterial sulfate-reducers (*OalgD1*, *OalgD4*) and a spirochaete symbiont of unknown function (*OalgS1*). Although this species is one of the best studied gutless oligochaetes, many aspects of this symbiosis remain unresolved, in particular with regard to the evolutionary history of host and symbionts, their population level diversity, the transmission of symbionts from parent to offspring, the molecular mechanisms that enable the symbiosis to be functional and maintained across all life stages and across host generations, and the function of the spirochaete symbiont within the symbiosis.

In the first part of this thesis (chapter 2), I used direct COI and 16S rRNA gene sequencing together with high-throughput metagenomic sequencing to investigate the population structure of *O. algarvensis* and its symbionts in order to gain insights into the recent evolutionary history of this symbiosis and to study the diversity within the symbiosis on an intra-specific level. I show that the Sant' Andrea population of *O. algarvensis* consists of two haplotypes and that each haplotype is specifically associated with its own unique strain of *OalgG1*. By constructing phylogenetic trees from single nucleotide polymorphism (SNP) data, I could show that the phylogenies of the two host haplotypes and their *OalgG1* symbiont phylotypes were highly congruent, strongly suggesting maternal vertical transmission. The two *OalgG1* strains also showed divergent evolution in their gene content, since several genes were unique to either of the two phylotypes. With respect to the other symbionts I observed decreasing or absent congruence with host phylogeny, suggesting horizontal or mixed-mode transmission, and varying degrees of sequence divergence, suggesting different levels of specificity for these symbionts. Two novel deltaproteobacterial symbiont phylotypes were identified through metagenomic sequencing, and near-complete genomes of them, as well as the elusive spirochaete symbiont, were obtained.

In the second part of this thesis I investigated the genome of the spirochaetal symbiont with respect to the role it might play in this symbiosis, focusing on its metabolic capabilities and its repertoire of genes to interact with the host. I found that the spirochaete is likely a mutualistic symbiont, fermenting environmentally derived carbohydrates to different short chain fatty acids like acetate and to hydrogen. Since the fermentation end products of the spirochaete are known substrates for the deltaproteobacterial symbionts, I propose that the

interaction between these symbionts is syntrophic and positively contributes to the carbon and energy budget of the whole symbiosis.

In the third part of my thesis (chapter 4), using transcriptomic and proteomic analyses, I investigated the molecular mechanisms that allow the host to successfully live with symbionts of greatly differing metabolic demands (anoxic vs. oxic, sulfide producing, carbon monoxide requiring) and of very different phylogenetic origin. I found that the host expresses digestive enzymes, even in the absence of a gut, hemoglobin that is predicted to be able to bind symbiont-produced sulfide, and extremely high expression of hemerythrin, a protein insensitive to carbon monoxide. Both respiratory proteins aid the host in avoiding noxious gases that are required by the symbionts. In addition, I established an inventory of immune-related genes that could enable host-symbiont molecular interactions and symbiosis maintenance.

The work of this thesis provides insight into the recent evolution of the host and its symbionts at the population level, the likely transmission modes of each symbiont, and the first functional characterization of the spirochaete symbiont. It furthermore establishes a database of improved or completely new symbiont genomes and host genes for future research of symbiont functions and the molecular mechanisms that allow this symbiosis to be maintained.

## Zusammenfassung

Gegenseitig nützliche Symbiosen zwischen Tieren und Mikroorganismen wurden einst als sonderbare Einzelfälle betrachtet, da Interaktionen zwischen Tieren und Mikroben in erster Linie aus dem Blickwinkel der Medizin betrachtet wurden. Dank der Entwicklung neuer Sequenziertechniken, die es erstmals ermöglichten die Genome von Organismen zu untersuchen, die nicht kultiviert werden können, wurde schnell offenbar, dass die meisten (falls nicht sogar alle) Tiere von Mikroorganismen besiedelt werden, die harmlos oder sogar von Vorteil für das Wirtstier sind, und oft wesentliche Funktionen in der Biologie dieser Tiere erfüllen. In einigen Fällen sind die Assoziationen so eng, dass sie für das Wirtstier obligat, d.h. unverzichtbar, für das Überleben werden.

Ein solches Beispiel stellt die obligate Symbiose zwischen marinen darmlosen Oligochaeten (Annelida, Phallodrilinae) und ihren chemosynthetischen Bakterien dar. Im Laufe der Evolution haben diese Tiere ihren gesamten Verdauuungstrakt, inklusive Mund, Darm und Anus, sowie ihre Exktretionsorgane, die Nephridien, verloren, wodurch sie völlig abhängig von der Aktivität ihrer Symbionten wurden. Diese Symbionten stellen sämtliche benötigten Nährstoffe bereit, und entsorgen auch die Abfallprodukte des Wirtsstoffwechsels. Jede Spezies von darmlosen Oligochaeten besitzt ihr eigenes, arten-spezifisches Konsortium von bakteriellen Symbionten.

Von allen darmlosen Oligochaetenarten ist die Mittelmeer-Art *Olavius algarvensis* am besten untersucht. Die Art beherbergt zwei verschiedene Schwefel-oxidierende gammaproteobakterielle Symbionten (*OalgG1* und *OalgG3*), zwei Sulfat-reduzierende deltaproteobakterielle Symbionten (*OalgD1* und *OalgD4*), und einen Spirochaeten-Symbionten (*OalgS1*), dessen Funktion unbekannt ist. Obwohl diese Art zu den am besten untersuchten gehört, ist vieles über diese Symbiose noch unklar, vor allem in Bezug auf die Evolutionsgeschichte von Wirt und Symbionten, ihre Diversität auf Populationsebene, die Art der Weitergabe von Symbionten von Generation zu Generation, die molekularen Mechanismen, die den Fortbestand der Symbiose ermöglichen, und die Funktion des Spirochaten innerhalb der Symbiose.

Im ersten Teil meiner Arbeit (Kapitel 2), untersuchte ich die Populationsstruktur von *O. algarvensis* und seinen Symbionten mit PCR, Markergen Sequenzierung und metagenomischen „high-throughput“ Sequenzierungen um die jüngste Evolutionsgeschichte dieser Symbiose nachzuvollziehen und um die genetische Diversität innerhalb der Art näher zu untersuchen. Ich konnte zeigen, dass die Sant‘ Andrea *O. algarvensis* Population aus zwei unterschiedlichen Haplotypen besteht, die jeweils ihren eigenen *OalgG1* Phylotypen besitzen. Mit Hilfe von phylogenetischen SNP (*single nucleotide polymorphism*) Bäumen konnte ich zeigen, dass die Phylogenie zwischen diesen Symbionten und ihrem Wirt kongruent ist, und damit einen starken Hinweis darauf liefert, dass dieser Symbiont maternal und vertikal in die nächste Generation transmittiert wird. Zusätzlich unterschieden sich die *OalgG1* Genome auch in ihrer Genzusammensetzung, da einige Gene nur exclusiv in entweder dem einen, oder anderen *OalgG1* Phylotypen vorkamen. Die anderen Symbionten zeigten nur eine abgeschwächte oder gar keine Kongruenz mit der Wirtsphylogenie, was darauf hinweist, dass diese Symbionten horizontal, oder kombiniert mit vertikaler Transmission (*mixed-mode*) vererbt werden. Unterschiedliche Grade von Sequenzdiversität lieferten außerdem Hinweise darauf, dass diese Symbionten mit unterschiedlicher Spezifität aufgenommen werden. Zwei völlig neue Symbionten-Phylotypen wurden ebenfalls identifiziert, und sowohl ihre nahezu kompletten Genome, als auch das Genom des Spirochaeten, konnten assembliert werden.

Im zweiten Teil dieser Arbeit (Kapitel 3) untersuchte ich das Spirochaeten-Genom im Hinblick auf seine potenzielle Funktion innerhalb der Symbiose, wobei mein Fokus auf den

enkodierten Stoffwechselwegen und Genen, die eine Interaktion mit dem Wirt erlauben, lag. Die Analyse ergab, dass dieser Symbiont höchstwahrscheinlich nützlich für den Wirt ist, in dem er Kohlenhydrate aus der Umwelt aufnimmt und zu Produkten wie Acetat und molekularem Wasserstoff fermentiert, welche von den deltaproteobakteriellen Symbionten als Substrat verwendet werden können. D.h. der Spirochaet steht in einem mutualistischen, syntrophen Verhältnis zu den deltaproteobakteriellen Symbionten, und trägt insgesamt positiv zur Kohlenstoff- und Energiebilanz der Symbiose bei.

Im dritten Teil meiner Arbeit (Kapitel 4) widmete ich mich den molekularen Mechanismen, die es *O. algarvensis* erlauben, mit einer so metabolisch (oxisch vs. anoxisch, Sulfid-produzierend, und Kohlenstoffmonoxid-oxidierend) und phylgenetisch diversen Symbiontengemeinschaft zu leben. Ich fand heraus, dass der Wirt verschiedene Verdauungsenzyme produziert, obwohl er gar keinen Darmtrakt mehr besitzt, dass er ein Hemoglobin produziert, welches für den Wirt toxisches Sulfid vermutlich binden kann, und außerdem in großer Menge Hemerythrin synthetisiert, welches unempfindlich gegenüber Kohlenstoffmonoxid ist. Beide Atmungsproteine helfen dem Wirt die negativen Effekte beider toxischen Gase auf seinen Organismus zu mindern. Desweiteren habe ich die Proteine untersucht und katalogisiert, die Teil des Immunsystems des Wirts sind, und damit einen wichtigen Faktor in der Etablierung und im Fortbestand der Symbiose darstellen.

Diese Arbeit trägt zum Verständnis der jüngesten Evolution von Wirt und Symbionten, ihrer intraspezifischen Diversität und Transmission bei, und liefert die erste funktionelle Beschreibung und Interpretation des Spirochaeten-Genoms. Desweiteren wurde in dieser Arbeit eine Datenbank von verbesserten oder sogar komplett neuen Symbionten-Genomen erzeugt, sowie eine Katalogisierung von Wirtgenen vorgenommen, die die Basis von zukünftigen Untersuchung zur Funktion von diesen Symbionten und den molekularen Interaktionen mit ihrem Wirtstier sein werden.

## List of Abbreviations

| | |
|---|---|
| AMP | antimicrobial protein or peptide |
| CO | carbon monoxide |
| DAMP | damage-associated molecular pattern |
| DAP | diaminopimelic acid |
| FISH | fluorescence *in situ* hybridization |
| LPS | lipopolysaccharide |
| LTA | lipoteichoic acid |
| Lys | lysine |
| MAMP | microbe-associated molecular pattern |
| MHC | major histocompatibility complex |
| MOX | methane oxidizing bacterium |
| OMP | outer membrane protein |
| PHA | polyhydroxyalkanoate |
| RNS | reactive nitrogen species |
| ROS | reactive oxygen species |
| rRNA | ribosomal ribonucleic acid |
| PAMP | pathogen-associated molecular pattern |
| PGN | peptidoglycan |
| PGRP | peptidoglycan recognition protein |
| PHA | polyhydroxyalkanoate |
| PRR | pattern recognition receptor |
| SNP | single nucleotide polymorphism |
| SOX | sulfur-oxidizing bacterium |
| sp. | species |
| spp. | multiple species |
| SRB | sulfate-reducing bacterium |
| TEM | transmission electron microscopy |
| TLR | Toll-like receptor |

# Chapter 1: Introduction

## 1.1 Symbiosis – Definitions and Relevance

In natural environments, organisms do not live by themselves but are in constant contact with other types of organisms. When organisms of different species form intimate and long-lasting associations with each other, this is referred to as “symbiosis” (from the Greek words “syn” meaning “with” and “bios” meaning “life”).

The concept of symbiosis was developed in the 19$^{th}$ century as a result of the detailed study of lichens, which are highly intimate, mutually beneficial associations of fungi with algae or cyanobacteria. The idea that lichens weren’t self-contained, discrete entities, but instead composites of two different organisms, was revolutionary and met with much skepticism at the time [1]. At first, many believed that this association had to be detrimental, because the concept of two species merging to benefit each other, whilst simultaneously upsetting conventional systematics, was hard to accept. However, several scientists recognized the true nature of this association, and understood that interactions between different species are not limited to competitive, predatory or parasitic modes, but range from loose to highly intimate and from pathogenic or parasitic to mutually beneficial [1]. Albert Bernhard Frank (1839 – 1900) was the first who gave this phenomenon a name by referring to different species that live on or within one another as symbiosis (*symbiotism*, German *Symbiontismus*, [2]). Anton De Bary (1831 - 1888), who is often credited with inventing the term, conveyed the concept to a wider scientific audience at a meeting of the German Association of Naturalists and Physicians in Kassel in 1878, where he defined it as "the continuous living together of differently named organisms" (German: *Das fortwährende Zusammenleben ungleichnamiger Organismen*, [3, 1]).

The meaning of this term was initially not restricted to beneficial interactions, but encompassed neutral and harmful associations as well. Despite this, the term symbiosis has since then often

been used to exclusively describe beneficial associations, but is nowadays mostly used in the original, broader sense in the scientific literature [4]. In this thesis, I will use the term “symbiosis” in its original broad sense. Further, I will refer to the smaller (usually microbial) one of the symbiotic partners as “symbiont” and the larger (usually multicellular, macroscopic) one as “host”, or, in cases where this distinction is not relevant, simply as the “biont(s)”.

### 1.1.1 Definition and classification of microbial symbiotic interactions

As defined by Frank and De Bary, the term symbiosis covers all types of close and lasting associations between different species, which range from beneficial for both partners (mutualism), beneficial to only one partner, but without detrimental effect on the other (commensalism), to harmful associations where the fitness of one partner is negatively affected by the other (antagonism, pathogenicity or parasitism).

Often, a clear classification into one of these categories is difficult or impossible, either because the mutual fitness effects have not been demonstrated and are challenging to rigorously test, or because they are better described as a continuous spectrum where dynamic, environmental or genetic factors define the nature of the relationship at any given time [5, 6, 7, 8, 9]. Many such examples exist in nature. For instance, *Wolbachia*, a common bacterial symbiont in arthropods and nematodes, acts as a reproductive parasite in many insects species [10], but is a mutualistic symbiont essential for normal development and fertility in filarial nematodes [11]. But even in insect species, *Wolbachia* can be mutualistic by supplying essential vitamins to their host [12]. *Wolbachia* could therefore be characterized as either mutualistic or parasitic, depending on the host species it is associated with. As a further example, plant – fungal associations are often even more plastic, and can switch from mutualism to parasitism within the same host when environmental factors cause an imbalance in the reciprocal exchange of nutrients between plant and fungal bionts [13]. An example of conditional parasitism in the marine environment is the association of reef coral with the bacterium *Vibrio shiloi* [14], which was identified as the

causative agent of bleaching in Mediterranean coral [15, 16]. However, *V. shiloi* only becomes pathogenic at elevated temperatures, which, among other things, cause it to express a peptide toxin that inhibits the photosynthetic activity of the host's symbiotic zooxanthellae. However, it is harmless at temperatures under 25 °C [17, 14].

Symbiotic interactions are further classified by whether the association is highly specific or unspecific (permissible) and whether the association is optional (facultative) or essential (obligate) for the partners. Symbiotic interactions can be facultative for one biont, and obligate for another within the same symbiosis (example pathogenic symbiosis: *Pneumocystis* in mammals [18], example mutualistic symbiosis: sulfur-oxidizing symbionts in *Riftia* [19]). Lastly, symbioses are further defined by whether the symbionts are located on the outside (ectosymbiosis) or within (endosymbiosis) the host, and whether they occur intra- or extracellularly. Commensalistic ectosymbionts are often referred to as epibionts.

### 1.1.2 Significance and functions of beneficial microbial symbioses

Once regarded as a curiosity, it is now increasingly understood that beneficial symbioses between organisms are common and ubiquitous, fundamentally shape the evolutionary path of organisms and significantly influence nearly all biological aspects of life on earth [20, 21, 22, 23, 24]. Symbiotic interactions exist between many different lineages within all three domains of life and encompass a large variety of different lifestyles and functions [25, 26, 27]. Mutualistic partnerships are wide-spread in nature, because they allow the exploitation of resources and the occupation of ecological niches that would be inaccessible to the individual partners, but become available in concerted effort.

Symbioses dominate large and important ecosystems on this planet. For example, more than 90% of all land plants form symbiotic associations with fungi, called mycorrhizae, which allow the plants to mobilize nutritional minerals from soil [28], while virtually all herbivorous animals

rely on cellulose-degrading gut microbes to digest plant fiber [29, 30, 31]. Several plant groups, especially legumes, form endosymbiotic root-nodule symbioses with nitrogen-fixing bacteria that allow them to grow on nitrogen-deprived substrates [32]. Similar to the terrestrial environment, symbioses between nitrogen-fixing cyanobacteria and marine algae allow primary productivity in nitrogen-limited ocean waters [33]. Coral reefs, the “rainforests of the sea”, are built by mutualistic coral – algal symbioses [34], and enormous animal communities at deep sea hydrothermal vents, “oases of life” in an otherwise desolate environment, are supported by chemosynthetic associations (section 1.4). Microbial symbionts further provide nutritional benefits to a majority of animals by synthesizing essential vitamins, amino acids and co-factors that are lacking or low in their normal diet [12, 35, 36]. However, microbial symbionts not only confer nutritional benefits, but can also provide many other functions to their hosts, including waste product recycling, defense against pathogens and predators, attraction and killing of prey and resistance to abiotic stressors, like toxins and heat (Table 1, p. 14).

While in all these examples the host clearly benefits from the symbiosis, the fitness benefits for the microbial symbionts are often much less clear and sometimes debatable [8, 37]. Mostly they are hypothesized to lie in the provision of surfaces for colonization [38, 39, 40], a “sheltered environment” with reduced competition and protection from predators [41, 42], increased dispersal rates [43, 44], and increased accessibility to nutritional resources [45, 46, 47, 48, 40].

**Table 1: Benefits of microbial symbionts to eukaryote hosts**

| Symbiont function | Examples |
|---|---|
| Fixation of inorganic carbon into digestible biomass | Photosynthetic algae/cyanobacteria in lichen [49] |
| | Photosynthetic algae chloroplasts in sea slugs [50] |
| | Photosynthetic algae in corals [51] |
| | Chemosynthetic bacteria in invertebrates and ciliates [52, 53] |
| Fixation of atmospheric nitrogen | Plant root nodule symbioses with Rhizobia [54] |
| | Cyanobacteria in coral reef sponges [55] |
| | Gammaproteobacteria in shipworm symbioses [56] |
| | Hindgut bacteria of termites [57] |
| Synthesis of essential nutrients | Synthesis of essential amino acids by *Buchnera* symbionts in plant-sap feeding aphids [35]<br>Synthesis of B-vitamins by *Wigglesworthia* in blood-feeding tsetse flies [58] |
| Cellulose degradation | Rumen symbioses of mammalian herbivores [29] |
| | Wood digestion by hindgut microbiota in termites [30] |
| Recycling and conservation of metabolic waste products | Gutless oligochaete symbionts [59, 60] |
| | Algal symbionts in coral [61] |
| Detoxification of harmful substances | Sulfide detoxification in chemosynthetic symbioses [62, 39] |
| | Symbiont-mediated pesticide tolerance in insects [63] |
| | Bacterial breakdown of plant toxins in guts of herbivorous insects [31] |
| Stress tolerance | Thermal tolerance in aphis [64] |
| Defense against pathogens | Gut microbiota in vertebrates [65] |
| | Resistance to pathogenic fungi in ants [66] |
| Defense against predators | Counterillumination in bobtail squid [67] |
| | Resistance to parasitic wasps in aphids [68] |
| Attraction and killing of prey | Bioluminescence in deep-sea fish [69] |
| | Production of toxins in entomophagous nematodes [70] |
| Bridging of chemical gradients | Meiofaunal chemosynthetic symbioses [39, 40] |

## 1.2 Symbiosis as a driving force in evolution

Without a doubt, symbiotic interactions have profoundly shaped the evolution of life on earth. The evolution of eukaryotes themselves is a result of symbiosis: mitochondria are thought to have resulted from the endosymbiotic uptake of an alphaproteobacterial *Rickettsia*-like bacterium by the proto-eukaryotic cell [71], a theory that is well supported by morphological, biochemical and genetic evidence [71, 72, 73]. The eukaryotic nucleus is also hypothesized to be derived from endosymbiosis (engulfment of an archaeum by a eubacterium), although this is still highly debated [74]. Later, endosymbiosis of a cyanobacterium by a eukaryotic cell lead to the evolution of chloroplasts, and the rise of photosynthetic eukaryotes [72, 75]. Further (secondary and tertiary) endosymbiotic events lead to the evolution of many other photosynthetic eukaryotic lineages [76, 77, 75]. In the following sections, I will give an overview of i) how microbial symbionts have influenced the evolution of animals, and how they contribute to host speciation, and ii) how association with a eukaryotic host influences the evolution of symbionts.

### 1.2.1 Impact of mutualistic symbiosis on animal evolution and speciation

The evolution of animals has been mostly investigated leaving symbiosis out of the picture [78, 79, 80, 81, 82]. However, microbial symbionts undoubtedly made significant contributions to animal diversification. Speciation (i.e. the evolution of genetically distinct populations) requires the formation of reproductive barriers that prevent interbreeding between diverging insipient species (i.e. speciation requires reproductive isolation). Speciation is enabled through various processes, presented in Figure 1 (p. 16).

When complex multicellular eukaryotes arose, they did so in an environment that was already teeming with microbial life for at least two billion years [83]. Since the beginning, animals (and of course also plants) have evolved in the presence of microbes and have formed remarkable beneficial symbioses with many of them (section 1.1.2). Symbionts have greatly influenced their hosts' evolutionary trajectories by providing them with new traits that allowed them to exploit

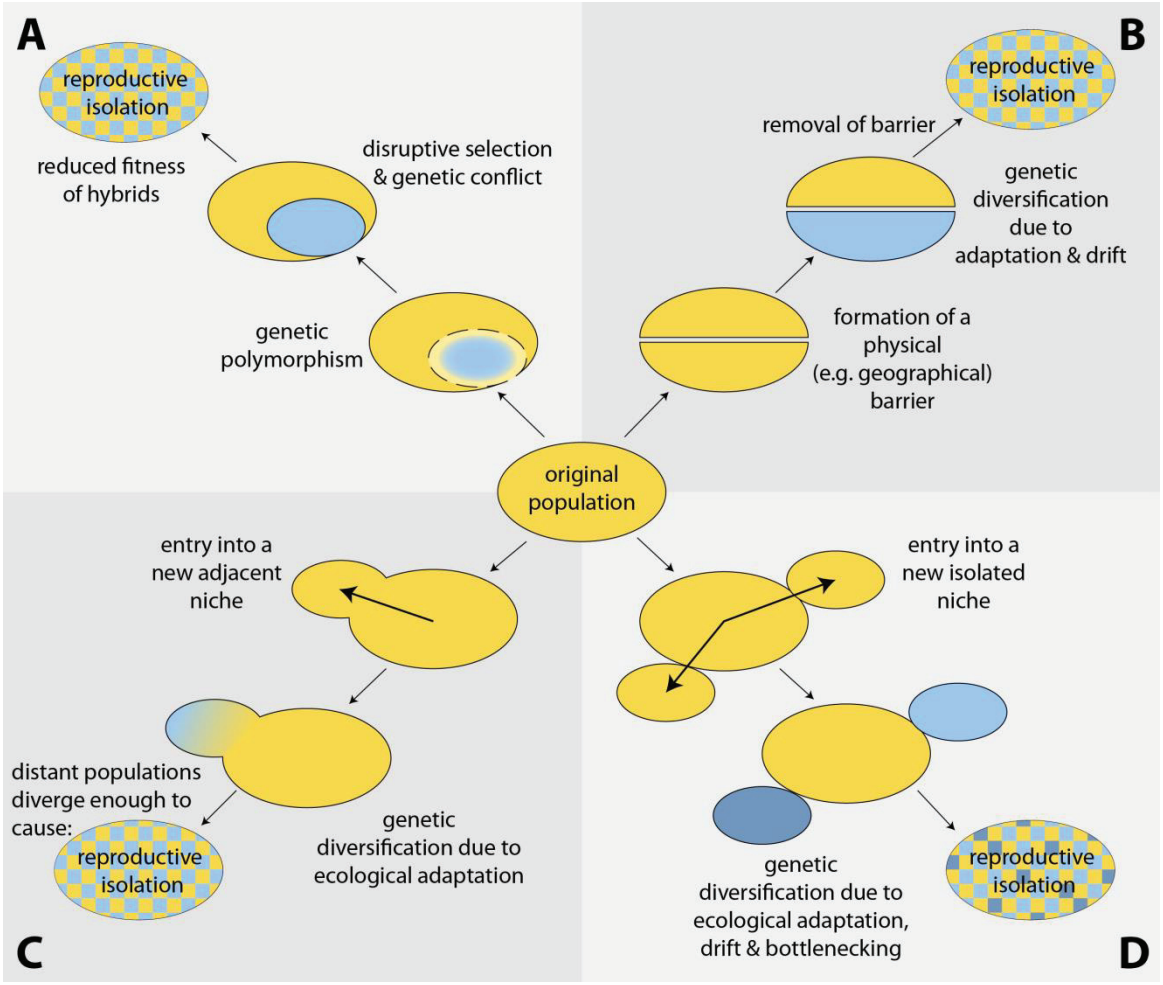



**Figure 1: Mechanisms that drive speciation. A)** *Sympatric speciation*: Divergence of populations without migratory barriers due to genetic polymorphisms that cause disruptive selection (extreme ends of a trait spectrum are favored by selection, while intermediates are not, including sexual conflict and assortative mating), and the creation of a "magic trait" (i.e. a trait that underlies disruptive selection and also pleiotropically promotes reproductive isolation) [78, 84]. **B)** *Allopatric speciation*: Divergence of populations due to migratory barriers which prevent gene flow. The so separated populations go different evolutionary paths, due to genetic drift and/or adaptation to different conditions across the physical barrier. If the populations have diverged sufficiently, removal of the migratory barrier will not reinstate interbreeding [85]. **C)** *Parapatric speciation*: Speciation through adaptation to geographically adjacent ecological niches in an environmentally continuous gradient. Hybridizations occur at a thin line of contact, but end populations are too diverged to interbreed successfully [86]. A typical example is the formation of "ring species" [87]. **D)** *Peripatric speciation*: a sub-form of allopatric speciation, in which a much smaller, peripherally isolated population diverges faster than in classical allopatric speciation due to selection bottlenecks [88]. Peripatric speciation allows the formation of more than one sister species from the same common ancestor, i.e. breaking the typical dichotomy of diverging species. It is often observed at the edges of large populations (e.g. brown bear -> polar bear [89]), or in species colonizing small islands from a large mainland population. *Yellow, original population; light and dark blue, new diverging populations; yellow-blue gradient, geographically overlapping diverging populations; checkerboard pattern, reproductive isolation prevents genetic mixing.*

new ecological niches that were previously inaccessible. For example, the evolution of herbivorous and xylophagous animals is tightly linked to the acquisition of mutualistic microbes that break down indigestible food components, like cellulose and lignin [90]. The association with nutritional endosymbionts has strongly influenced the evolution and diversification of many insect groups by allowing adaptation to new host plants (=host-shift) or other food sources, including herbivorous and plant-parasitic aphids [91, 92, 93, 94, 95], grain weevils [96], fruit flies [97], leafhoppers [98], stinkbugs [99], and blood-feeding tsetse flies [100]. As another example, the evolution of herbivorous ants from carnivorous ancestors was independently facilitated through the uptake of nutritional *Rhizobiales*-symbionts at least five times [101]. Further examples in other animal phyla include the evolution of rumen symbioses in herbivorous mammals [29] and chemosynthetic symbioses in ciliates, sponges, annelids, mollusks, and nematodes [52, 53]. These examples illustrate how microbial symbionts contribute to host diversification by facilitating the adaptation to ecological niches (=ecological speciation), a major driver for the evolution of new species (Figure 1).

It has recently been argued that microbial symbionts also cause reproductive isolation directly, without involving ecological isolation. This includes pre- and post-mating isolation mechanisms, like behavioral isolation (pre-mating), direct interference with host reproductive biology (pre- and post-mating), and isolation through immunological adaptations to the symbiotic microbes that cause immune incompatibilities in hybrids (post-mating, colloquially termed “The Large Immune Effect”) [102, 103]. Interestingly, all of these mechanisms allow the reduction of gene flow between populations that are not geographically isolated, i.e. they are ideal mechanisms for explaining strict sympatric speciation, the existence of which has been debated since Darwin and Wallace [78].

**Behavioral isolation.** Microbial symbionts may contribute to reproductive isolation by influencing host mating preferences or courtship behavior and thereby reducing gene flow

between populations that carry different microbiota [103]. It was shown that gut microbes influence mating behavior in *Drosophila melanogaster*, presumably through changing the levels of cuticular sex pheromones [104]. In this study, *Drosophila* were reared on two different media (molasses medium vs. starch medium) for one generation and exhibited a strong mating preference towards individuals that were reared on the same medium, which lasted for at least 37 generations (more were not tested). This behavior could be cured with antibiotic treatment or by artificially infecting hosts with microbes that are typical for flies reared on the other respective medium. Another study carried out on *D. melanogaster* showed that *Wolbachia* symbionts influence mate discrimination dependent on *Wolbachia* infection load [105]. *Wolbachia* was also shown to increase mate discrimination between incipient species of *Drosophila paulistorum* [106]. In grub beetles, sex pheromones produced by symbiotic bacteria located in special glands of the female reproductive organs influence mating behavior as well [107]. In vertebrates, proteins of the major histocompatibility complex (MHC), an important part of adaptive immunity, have been shown to play a role in mate preference [108, 109, 110, 111], and it has therefore been argued, that immunological adaptation to pathogens and the resulting changes in MHC diversity promote speciation [112].

**Influence on host reproductive biology.** Between 20-75 % of arthropod species harbor the intracellular reproductive parasite *Wolbachia*, an Alphaproteobacterium [113, 114, 115]. *Wolbachia* are vertically transmitted through the female germline, but host switches occur occasionally. *Wolbachia* influence the reproduction of their hosts in four major ways: i) killing of infected male embryos, ii) feminization of infected males, iii) induction of parthenogenesis in infected females, and iv) inducing cytoplasmic incompatibility (CI) in hybrids of infected males and uninfected females (unidirectional CI), or males and females that are infected with incompatible strains of *Wolbachia* (bidirectional CI) [10]. These mechanisms increase the number of infected females in the host population, and hence, *Wolbachia* fitness, but often

reduce host reproductive success in the process. These mechanisms also reduce gene flow between infected and uninfected parts of the population to varying degrees from blocking gene flow only in one direction (e.g. in unidirectional CI) or blocking successful interbreeding completely (e.g. between specimens infected with incompatible *Wolbachia* strains). In the latter case, reproductive isolation arises spontaneously even between animals that are genetically identical. Other bacteria than *Wolbachia* manipulate arthropod reproduction: *Cardinium* [116, 117], *Rickettsia* [118], and *Spiroplasma* [119] were all shown to interfere with host reproduction in similar ways.

**"The Large Immune Effect."** The term was coined in [103] and refers to immune-related incompatibilities in hybrids that arose from fast adaptation of immune genes to resident microbiota in the parent species. To illustrate, hybrids between very closely related species that each harbor different microbiota might not be viable (hybrid autoimmunity) or experience significant fitness defects (hybrid susceptibility), because their immune systems are not properly adapted to the new microbiota and immune responses are insufficient or get out of hand. Two recent publications from the Bordenstein lab demonstrate this immune breakdown in hybrids caused by i) *Wolbachia* [120] and ii) gut microbiota [121] in *Nasonia* wasps. The same phenomenon is also documented in plant hybrids [122, 123, 124].

### 1.2.2 Impact of obligate mutualism and transmission mode on symbiont evolution

Symbionts not only impact host evolution, but their own evolution is also influenced by a host-associated lifestyle. First, hosts provide new and unique ecological niches that drive the adaptive radiation of symbionts [125, 126, 127, 128], and promote the evolution of clades that are unique to a particular host species or host group [129, 130]. Second, obligate host restriction and strict vertical transmission has profound impacts on symbiont genome evolution.

**Symbiont transmission.** Symbionts can be transmitted from one host generation to the next via three routes: i) horizontal transmission, in which aposymbiotic (symbiont-free) offspring must acquire the symbionts from the environment anew in each generation (Figure 2a, p. 21), ii) vertical transmission, in which the symbionts are transferred from parent to offspring via direct transfer through the female germline (Figure 2b), and iii) mixed mode transmission, in which symbionts are mostly transferred via vertical transmission, but are occasionally also transmitted horizontally from other hosts (host-switching, Figure 2c).

**Influence of obligate host-association and vertical transmission on symbiont genome evolution.** The genomes of obligately host-associated, vertically transmitted symbionts often show the same trends in the evolution of their genomes: small genome size, low GC content, high coding density, accelerated rates of amino acid substitutions, loss of functions that are not necessary within the host environment, loss of functions involved in DNA repair, and loss of mobile genetic elements, like transposases and phages [131, 132, 133]. However, some vertically transmitted symbionts show high loads of transposable elements [134, 135, 136, 137]. This is hypothesized to occur in the early stages of host-restriction, as symbionts derived from free-living ancestors with large genomes and few mobile elements are subject to other evolutionary forces and selection pressures within the host [138]: i) host-restricted symbionts have smaller effective population sizes which reduces purifying selection and allows the inactivation of beneficial genes through genetic drift, and ii) the new host environment reduces purifying selection on genes that are no longer essential for survival [139, 140]. Genes that are commonly lost due to these processes include those involved in DNA repair and maintenance [131]. As a result, mobile genetic elements, which are usually present at low levels in free-living bacteria, proliferate without a check [138]. The spread of mobile elements promotes gene deletions and gene inactivation, accelerating the process of gene loss [141, 142]. Eventually, mobile elements and inactivated genes are deleted from the genomes, leading to highly reduced genomes free of

mobile DNA that keep slowly deteriorating over time [138, 143]. High loads of mobile elements have been reported from the intracellular symbionts *Wolbachia* and *Shigella flexneri* that did not recently become host-restricted [144, 145, 146], which seems to contradict the model of symbiont genome evolution proposed by Moran and Plague. However, these symbionts are prone to host-switches, and, although intracellular, often come into contact with other strains and bacteria, giving them the opportunity to pick up new genes, including mobile elements, via horizontal gene transfer [145].

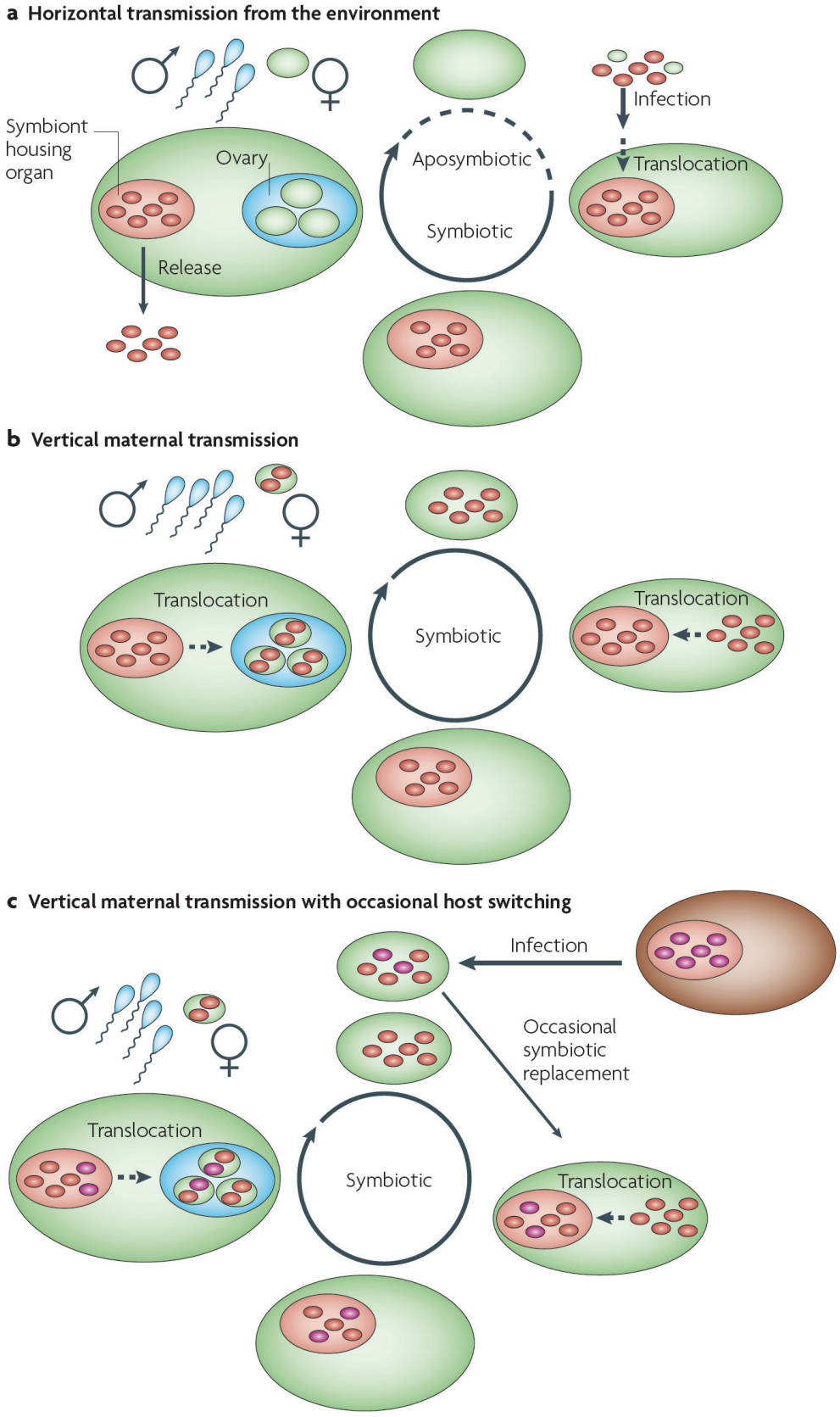



**Figure 2: Different transmission modes by which symbionts are transferred to offspring.** (Adapted from [147])

## 1.3 Role of the animal innate immune system in microbial symbiosis

Beneficial symbiotic interactions with microbes require specific recognition and tight regulation by the animal host during all stages of its life cycle. At the beginning of infection, the host must specifically recognize and respond to the correct symbiont phylotype(s), in order to avoid uptake of unwanted microbes. Then, symbionts must be guided and restricted to the intended locations for colonization in order to prevent misdirected, harmful interactions that can lead to disease. Over the course of the symbiotic relationship, constant molecular cross-talk between symbionts and host is required to safely establish and maintain a beneficial interaction. These functions are largely fulfilled by the host's immune system and its reciprocal interaction with molecular microbial cues [148, 149, 150].

Historically, the immune system has been regarded primarily as an arsenal of weapons intended to rapidly fight off any pathogenic intruders, and its involvement in inducing and maintaining beneficial interactions with microbes was long overlooked. With the realization that most animals are colonized by a diverse microbial community that is highly integrated with host physiology and immunity, often to the host's benefit, this view has recently changed considerably [26, 151, 152]. Molecular mechanisms that foster beneficial symbioses between animals and microbes turned out to be essentially the same as those that were initially seen simply as pathogen extermination strategies and microbial virulence factors causing disease [153, 154, 149, 26, 155, 151, 156].

In invertebrate animals, adaptive immunity does not exist, and all immune functions are carried out by components of the native innate immune system [157]. Jawed vertebrates, on the other hand, additionally possess adaptive (acquired) immunity, which allows a highly specific, amplified response to pathogenic encounters and confers immunological memory, i.e. the ability to immediately recognize and efficiently respond to specific microbes on repeated contact [158]. The innate immune system has long been regarded as a simple unspecific defensive barrier,

incapable of discriminating between microbes beyond basic categories such as viruses, gram-positive or gram-negative bacteria and fungi, by employing a rather limited set of microbial pattern recognition receptors [159, 160]. Again, fairly recent advances have brought a new perspective to invertebrate immunity, demonstrating its ability to recognize and respond to microbes with high specificity and nuance [161, 162, 163, 149, 164, 148, 165]. This should not come as a surprise, as many symbioses are highly specific [166, 167, 168, 148, 169, 130]. Although specificity can be achieved through various mechanisms (such as highly selective competition for resources in a particular niche within the host that automatically excludes certain strains), the immune system is clearly involved in many cases [170, 167, 168, 171]. Furthermore, new studies are beginning to uncover alternative mechanisms for high specificity, adaptive immunity and memory in some invertebrates [172, 173, 162, 174] (see [175] for a critical review of the more controversial findings). The following sections will give a brief introduction and overview of the innate immune mechanisms of invertebrates, and how they are employed in beneficial symbiotic interactions.

### 1.3.1 Components of the innate immune system in invertebrates

Mirroring invertebrate phylogenetic diversity, the diversity of immune mechanisms and molecules in these animals is high [176, 177]. However, basic concepts and broader categories of immune components have a long evolutionary history or evolved convergently several times, and are therefore still comparable between organismic groups [178].

As with adaptive immunity, the innate immune system can be conceptually divided into cellular and humoral components. Cellular immunity is conferred by mobile cells that are able to eliminate detrimental microbes by either engulfing them (phagocytosis) or by immobilizing and destroying them through various other mechanisms (Table 2, reviewed in [179]). These types of cells are differently named depending on the anatomy of the animal and their location, but fulfill similar functions in different animals. For example, acoelomates such as cnidarians possess

mesogleal wandering cells called amoebocytes; coelomate animals like annelids possess coelomocytes which patrol the coelomic fluid, while arthropods and mollusks possess hemocytes that circulate in the hemolymph. Typically, specialized sub-types of these cells exist that can differ in size, shape, behavior and/or specific function [158].

Humoral immunity refers to those components of the immune system which consist of molecules secreted into extracellular fluids, like blood, lymph, hemolymph and coelomic fluid. These include complement proteins, antimicrobial peptides and other cytotoxic compounds, soluble pattern recognition molecules and chemokines/cytokines [180, 158, 181]. Table 3 (p. 26) summarizes the different classes and functions of humoral immune molecules in invertebrates.

**Table 2: Functional roles of immune cells (cellular immunity)**

| Functional role | Explanation |
|---|---|
| Coagulation | Release of coagulation (clotting) factors that agglutinate in order to close open wounds and to trap microbes for subsequent elimination. |
| Encapsulation | Used to eliminate particles that are too large for direct phagocytosis; immune cells gather around the particle and destroy it with cytotoxic molecules and digestive enzymes. The cells form a tight sheath around the target through surface cell adhesion molecules. Encapsulation is usually followed by melanization. |
| Melanization | Production and deposition of melanin, which polymerizes and traps the target, and also produces cytotoxic reactive oxygen species as side-products. |
| Opsonisation | Release of proteins (called opsonins or agglutinins) such as lectins that coat and agglutinate the target and make it easier to be subsequently phagocytosed by host cells. |
| Phagocytosis | Engulfment of cells and other particles and subsequent intracellular digestion within phagosomes (phagocytosis also has a nutritional role in some animals like filter-feeders) |
| Production of AMPs | Release of antimicrobial proteins (AMPs) or other microbicidal compounds, such as reactive oxygen and nitrogen species (ROS/RNS), to kill a target extra- or intracellularly (following phagocytosis). |

**Table 3: Classes and functions of humoral components in innate immunity**

| Humoral immune component | Function |
|---|---|
| Antimicrobial proteins/defensins | A large group of microbicidal molecules of different composition and structure, unified by their ability to disintegrate microbial membranes. Examples include cecropins, cysteine-, proline- or glycine-rich peptides, lumbricins/fetidins, perforins, bactericidal permeability increasing protein (BPI) and lysozyme. |
| Chemokines/Cytokines | A large group of small proteins that mediate immune responses by activating and trafficking immune cells. They include macrophage inflammatory protein (MIP), interferons, interleukins, and tumor necrosis factor. Cytokines are also released by cells upon recognition of microbe-associated patterns (MAMPs). |
| Coagulation cascade | A protease-mediated activation cascade triggered by soluble pattern recognition proteins and resulting in the agglutination of soluble precursor-proteins to close open wounds and entrap microbes. |
| Complement proteins | The complement system consists of a proteolytic cascade that results in the recruitment of phagocytes via cytokines, and opsonisation or lysis of microbes. Different complement pathways exist. |
| Pattern recognition proteins | Proteins able to specifically bind conserved microbe-associated molecular patterns (MAMPs), such as lipopolysaccharide and peptidoglycan. These include peptidoglycan recognition proteins (PGRPs), fibrinogen-related proteins (FREPs), glucan binding proteins (GNBP) and various classes of lectins. Pattern recognition proteins usually trigger an antibacterial immune response via numerous possible pathways (further explored in chapter 1.3.2). |

### 1.3.2 Interactions of the innate immune system with beneficial symbionts

All immune functions described in the previous two sections require the reliable and specific recognition of microbes by the host in order to mount an appropriate response. It is important for the host to be able to distinguish between beneficial and harmful microbes and to modulate the immune response accordingly. Failure to recognize and combat pathogenic colonization results in disease and death. However, sustained inflammation in response to harmless

symbionts is highly damaging to the host as well [182, 183, 167]. Hosts are capable of differentiating between microbes and responding appropriately by employing a diverse array of recognition receptors and molecular signaling pathways, which I will summarize in the following.

Invertebrates sense the presence of microbes by binding microbe-derived molecular structures called microbe-associated molecular patterns (MAMPs) via pattern recognition receptors (PRRs) [184, 185]. MAMPs are conserved molecules that microbes release or carry on their cell surfaces and that are common and unique to a particular group of microorganisms. Typical MAMPs include peptidoglycan (PGN), lipopolysaccharide (LPS), lipoteichoic acid (LTA), flagellin and outer membrane proteins (OMPs), which are characteristic for different bacterial groups, and different surface carbohydrates like β-glucan and chitin which are characteristic of fungi. MAMPs, which are clearly not restricted to the pathogenic microbes, were originally called PAMPs (pathogen-associated molecular patterns), reflecting the pathogen-centric context in which these mechanisms were discovered [186, 187, 156].

Recognition of MAMPs by host PRRs initializes an intracellular signaling cascade that results in a change of gene expression and behavior of the host cell (Figure 3, p. 28). Typical host MAMP receptors include peptidoglycan recognition proteins (PGRPs), glucan-binding proteins (GNBPs), and Toll-like receptors (TLRs), which directly or indirectly activate immune signaling pathways, like Toll and IMD, upon MAMP binding (Table 4, p. 29). The activation of these pathways leads to proinflammatory, antibacterial responses, for example through the production of antimicrobial proteins (AMPs), release of reactive oxygen/nitrogen species (ROS/RNS) that damage cellular structures, release of chemokines to attract immune cells, activation of complement and induction of phagocytosis [185, 148].

While PRR signaling serves to initiate immune responses and eliminate pathogens, it also plays an important role in establishing and maintaining beneficial symbioses, through both, immune

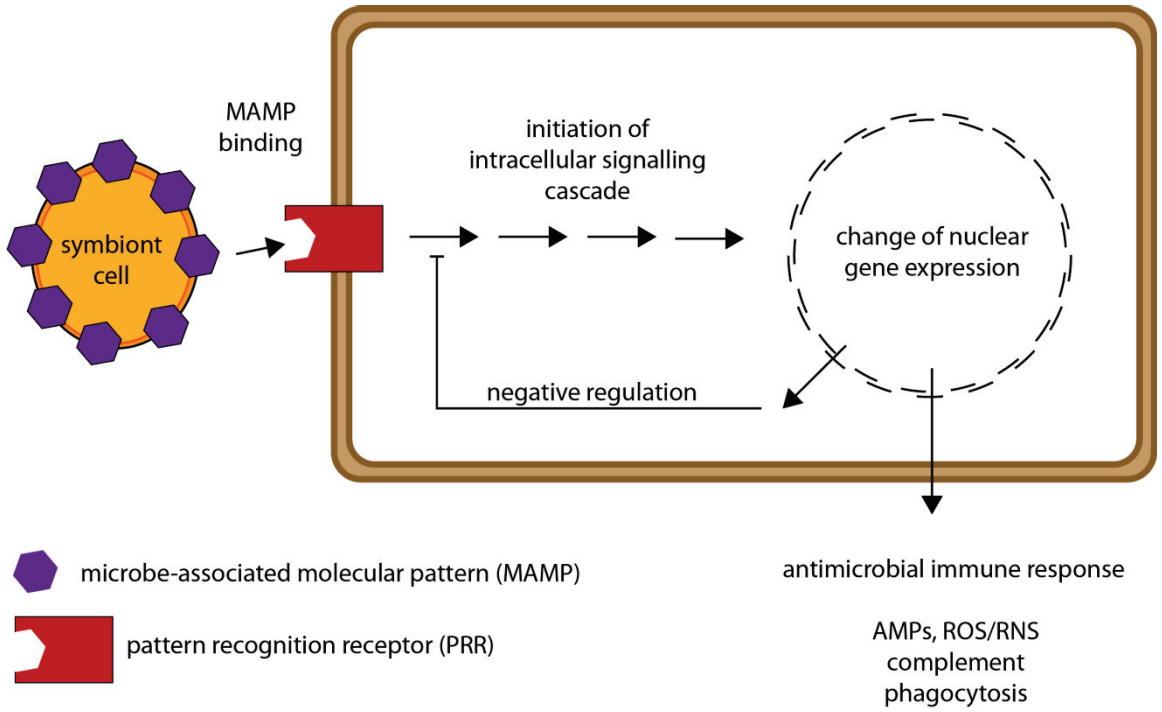



**Figure 3: Interaction of the host immune system with microbes.** In beneficial and harmful associations alike, the binding of MAMPs to host PRRs results in the activation of an antimicrobial immune response, mediated through intracellular signaling cascades and an alteration of gene expression. The antimicrobial immune response includes the production of antimicrobial proteins (AMPs) and reactive oxygen or nitrogen species (ROS/RNS), as well as the activation of complement and initiation of phagocytosis. In commensal/beneficial symbioses, the immune response is often down-regulated after the initial encounter to avoid damaging the symbionts and to avoid constant inflammation in the host.

activation and inhibition. For example, in the cnidarian *Hydra*, activation of Toll signaling by a MAMP binding TLR-like receptor results in the production of antimicrobial proteins, which prevent pathogenic colonization and promote the establishment of a host-specific symbiont community in embryos and adult polyps [188, 189, 171]. In the corn weevil *Sitophilus*, growth of the obligate endosymbiont SPE is controlled and restricted to the symbiotic tissue (the bacteriome) by the expression of coleoptericin-A, an antimicrobial peptide that inhibits bacterial cell division [190]. At the same time, the expression of a peptidoglycan-degrading PGRP and Tollip, an inhibitor of Toll signaling, prevents excessive immune responses towards the symbiont

**Table 4: Common intracellular immune signaling pathways in invertebrates**

| Pathway | Function |
|---|---|
| Toll pathway | Signaling cascade which is activated by ligand-binding of Toll-like receptors (TLRs). TLRs can bind MAMP ligands directly, or are activated indirectly by binding a mediator protein, which is previously activated after MAMP binding by recognition proteins like PGRP or GNBP [191, 192]. Activation of the Toll pathway induces the expression of genes that convey an antimicrobial and proinflammatory response, e.g. genes coding for cytokines and AMPs [193, 194]. Toll signaling is also specifically involved in mediating beneficial host-microbe associations [195, 196, 197]. TLRs are conserved throughout the Metazoa, albeit missing in the Platyhelminthes. The Toll pathway is furthermore involved in embryonic development in insects and nematodes [194, 198]. |
| IMD pathway | Signaling cascade which is activated by binding gram-negative DAP-type peptidoglycan to membrane-integral peptidoglycan recognition proteins (PGRPs) in insects [199]. It is named after *Drosophila* ***im***muno***d***eficiency mutants. Like the Toll pathway, IMD activates nuclear transcription factors which lead to the expression of antimicrobial proteins [200]. Unlike Toll, IMD has no additional functions in development. |
| p38 MAPK pathway | A phosphorylation cascade employing at least three core kinases (MAPK kinase kinase (MKKK) -> activates MAPK kinase (MKK) -> activates p38 mitogen-activated protein kinase (p38 MAPK)). Plays a role in many biological processes including immunity, apoptosis, cell cycle regulation and cell differentiation, often in cross-talk with other signaling pathways [201, 202]. A wide range of stimuli activate p38 MAPK signaling, including LPS, cytokines, heat and osmotic stress. |

in the bacteriome [203, 204]. In *Drosophila*, binding of DAP-type peptidoglycan (indicative of gram-negative bacteria) to a membrane-integral PGRP in the gut epithelium triggers an antimicrobial response via IMD and p38 MAPK signaling. This leads to the production of an AMP (via IMD) and ROS (via p38 MAPK), which were shown to be essential for resisting food-borne pathogenic bacteria [205, 206]. At the same time, IMD signaling is dampened in three major ways in the *Drosophila* gut to protect the resident microbiota from constant AMP expression: i) induction of Pirk (***p****oor* ***I****MD* ***r****esponse upon* ***k****nock-down*), a negative regulator of IMD signaling,

via IMD signaling itself [207], ii) action of amidase-active PGRPs in the gut lumen that degrade peptidoglycan (i.e. hampering IMD signaling by destroying the elicitor) [208] and iii) symbiont-induced nuclear translocation of the transcriptional inhibitor Caudal, which represses AMP expression directly [209].

Another extremely well-studied example of the involvement of MAMP-PRR interactions in symbiosis establishment and maintenance is the light-organ symbiosis between the bobtail squid *Euprymna scolopes* and bioluminescent *Vibrio fisheri*. The counterillumination provided by the symbionts through bioluminescence enhances host camouflage and helps the animal to avoid predation while foraging at night [67]. Freshly hatched squid are aposymbiotic and acquire their symbionts horizontally from the environment [210]. Selection of the specific symbiont from the highly diverse seawater community (in which it is not abundant) and subsequent colonization and maturation of the light organ are achieved by intricate molecular cross-talk of symbiont released MAMPs, and MAMP recognition and response by the host [211, 170, 212, 213, 155] (Figure 4). After successful colonization of the light organ by *V. fisheri*, MAMP degradation by a host-expressed PGRP and reduced binding of *V. fisheri* cells to hemocytes contributes to immune tolerance of this symbiont [167, 214].

Since MAMPs are not specific to pathogens or beneficial symbionts, and MAMP recognition is crucial in establishing and maintaining beneficial symbioses as well as fending off pathogenic intruders, the question remains how hosts are able to discriminate between beneficial and harmful encounters. (It should be noted that, although MAMPs are highly conserved, microbes do have the ability to modify details of the chemical structure of MAMPs, resulting in differential PRR stimulation. For example, several pathogens can produce different forms of LPS with PRR affinities ranging from very low to high [215, 216, 217].) Two main strategies for discriminating between beneficial symbionts and normal commensal microbiota or pathogenic infection have been proposed. First, mutualistic symbionts are usually restricted to specific tissues,

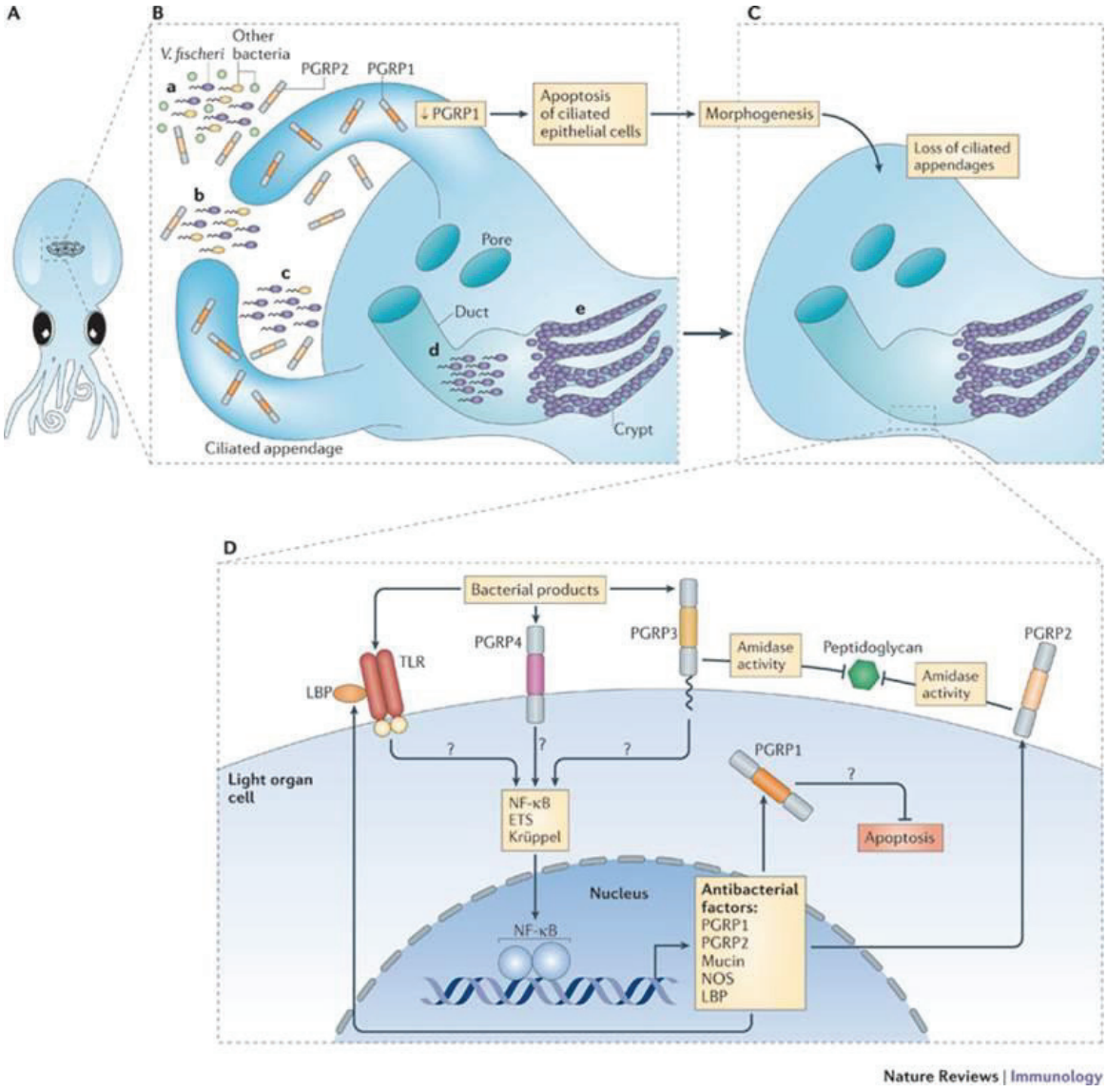



**Figure 4: Role of MAMP recognition in the establishment of the squid-*Vibrio* light organ symbiosis. A)** *Location of the light organ in the body cavity of E. scolopes*. **B)** *Colonization of the light organ crypts by V. fischeri.* The presence of bacteria stimulates the secretion of mucus, a peptidoglycan recognition protein (PGRP2) and other factors that promote *V. fisheri* growth and inhibit other bacteria (a, b), resulting in *V. fischeri* becoming the dominant bacterium (c). *V. fischeri* subsequently enters the light organ ducts (d) to colonize the light organ crypts (e). **C)** *Colonization of the crypts results in the loss of ciliated appendages*. **D)** *MAMP recognition and signaling in the light organ cells.* PGRP3, PGRP4 and TLR serve as PRRs which initiate the production of mucus, nitric oxide synthase (NOS), lipopolysaccharide binding protein (LBP), PGRP1 and PGRP2. PGRP2 and PGRP3 both possess amidase enzymatic activity, which cleaves immunogenic peptidoglycan fragments and reduces immune activation (promoting symbiont tolerance). After colonization of the crypts, peptidoglycan derived tracheal cytotoxin (TCT) from the symbionts induces the loss of PGRP1 from host nuclei in cells of the ciliated appendages. The loss of nuclear PGRP1 results in apoptosis of these cells and in loss of the ciliated appendages at the final stages of symbiosis establishment. Figure adapted from [185].

where their tolerance is achieved through limited expose to the immune system, or immune modulation and attenuation [170, 190, 218]. Second, pathogenic infection causes tissue damage, which results in the release of DAMPs (damage-associated molecular patterns), like DNA, ATP, uric acid, and DNA-binding proteins, into the extracellular space. DAMPs initiate an immune response and intensify the immune response in the presence of bacteria [219, 220, 221]. It was therefore proposed that the establishment and maintenance of beneficial symbioses is achieved through the interplay of i) co-evolved mechanisms that modulate MAMP recognition and signaling, ii) the sequestration of symbionts to designated tissues that allows for localized immune responses and physicochemical conditions that limit symbiont growth, and iii) the lack of DAMP signals that signify tissue damage in beneficial symbioses [148] (Figure 5).

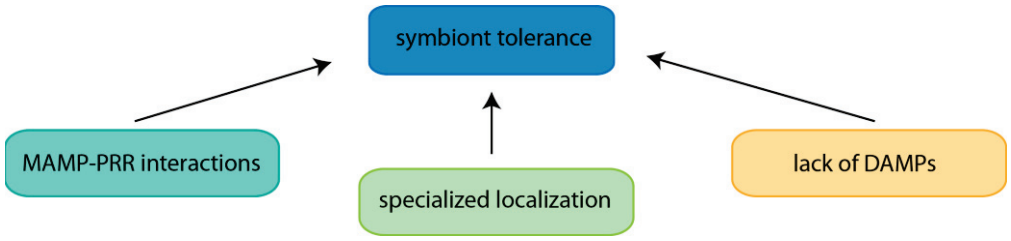



**Figure 5: Model for establishment and maintenance of beneficial symbiosis. (**Adapted from [148])

### 1.3.3 Annelid immunity and mutualistic symbioses

Within the annelids, the innate immune system has been studied intensively in earthworms and leeches (both belong to the Clitellata), since they are easily accessible experimental systems, are of medicinal and ecological value and are classical models for comparative immunology [222, 223, 224, 225, 226]. Earthworms have been studied with particular focus on tissue transplantation and short term immune memory [227], and leeches with respect to immune responses of the central nervous system [228]. Similar to other invertebrates, annelids employ a variety of cellular and humoral immune responses to fend off pathogens (reviewed in [229, 230,

228]), which include MAMP recognition through PRRs and the production of antimicrobial proteins (Table 5) [231, 232, 233]. While the immune system of these annelids is well studied from a classical pathogen-centric view, surprisingly little is known with respect to the molecular mechanisms that facilitate beneficial symbioses. In the following I will summarize what is currently known about the molecular interactions between hosts and mutualistic symbionts in annelids.

**Table 5: Components of the innate immune system in earthworms and leeches**

| | Leech | Earthworm | Examples/functions |
|---|---|---|---|
| Coelomocytes | yes | yes | Phagocytosis, opsonisation, encapsulation, wound healing |
| Microglia cells | yes | ? | Migratory immune cells of the CNS |
| Prophenol oxidase cascade | ? | yes | Melanization, potentially activating other antimicrobial responses as well [234, 235] |
| MAMP recognition | yes | yes | Toll-like receptors, mostly involved in immune defense [236, 237, 238], NOD-like receptor (only leech, [236]), uncharacterized lectins [239, 240, 241], CFF (coelomic cytolytic factor, in earthworms [242]) |
| Immune signaling via Toll | yes | yes | Activation of antimicrobial response |
| Antimicrobial proteins | yes | yes | **Earthworms:** lysozyme, lumbricin, PP-1, OEP3121, fetidin, lysenin, eiseniapore, hemolysins, CFF [230] **Leeches:** lumbricin, neurohemerythrin [243], neomacin, theromacin, theromycin [244] |

**Hydrothermal vent tube worm symbiosis: *Ridgeia piscesae* – SOX symbiont**. Chemosynthetic tube worm symbioses have been extensively studied in terms of ecology and metabolism, in particular with respect to the sulfur-oxidizing (SOX) symbionts and their contribution to host nutrition [245, 246, 247]. However, the molecular mechanisms which are responsible for symbiont acquisition (symbionts are horizontally transmitted), and symbiosis establishment and maintenance have so far received less attention. A single study has investigated the expression of immune genes that are potentially involved in host-symbiont interaction in the hydrothermal

tube worm *Ridgeia piscesae* [233]. This study examined the expression of immune genes in the symbiotic tissue (trophosome) compared to non-symbiotic tissue (plume) using EST and 454 pyro-sequencing of transcriptomes and qPCR. Several MAMP recognition molecules, components of intracellular immune signaling, and various immune effectors were identified and shown to be more abundantly expressed in the trophosome compared to the plume (Table 6). These results show that complex MAMP-PRR interactions likely play a very important role in symbiont maintenance and regulation in tube worms, similar to other microbial invertebrate symbioses.

**Table 6: Immune genes potentially involved in host-symbiont interaction in *R. piscesae* [148]**

| Immune genes over-expressed in the trophosome | Hypothesized role in symbiosis |
|---|---|
| LPS-induced tumor necrosis factor-alpha (LITAF) | Not discussed, but induction of immune response in other animals [248] |
| Peptidoglycan recognition proteins PGRP Rpi1, PGRP Rpi3, PGRP Rpi4, PGRP Rpi5 | Symbiont MAMP detection and mediation of immune responses |
| Peptidoglycan recognition protein PGRP Rpi2 | Amidase activity -> down-regulation of immune response through peptidoglycan degradation |
| Toll-like receptor (TLR) | Symbiont MAMP detection and mediation of immune responses |
| Alpha-2-macroglobulin receptor associated protein (A2M) | Activation of cytolytic activity |
| Bactericidal permeability increasing protein (BPI) | Not discussed, but antibacterial and LPS-detoxifying roles in other animals [249, 250] |
| NF-kappa-B inhibitor cactus | Not discussed, but down-regulation of Toll signaling and antimicrobial response in other animals [251, 252] |

**Earthworm nephridial symbiosis: *Eisenia fetida* - *Verminephrobacter***. Lumbricid earthworms like *Eisenia* harbor species-specific, extracellular, and vertically transmitted symbionts in their nephridia (excretory organs) [253]. The facultative symbionts have a beneficial effect on host reproduction, hypothesized to be related to the provision of vitamins by the symbionts [254,

255]. Although experimentally and genetically accessible, the earthworm nephridial symbiosis is a relatively new model for host-symbiont interaction, and therefore the molecular mechanisms that regulate this symbiosis are still unknown. However, the symbiont-expressed type IV pili and flagella were recently shown to be required for colonization of the nascent nephridia during embryogenesis [256]. The host immune mechanisms involved in this symbiosis remain to be studied.

**Leech-bacterial crop symbiosis: *Hirudo verbana* – *Aeromonas*/*Rikenella*.** The blood-feeding leech *Hirudo verbana* harbors a simple microbial community in its crop (the main compartment of its digestive tract) that is heavily dominated by two bacterial symbionts: *Aeromonas veronii* and a *Rikenella*-like bacterium [257]. The low diversity of gut microbes is unusual, even for an invertebrate. Several reasons that promote this low community complexity have been brought forth: i) the extremely alkaline conditions of the gut environment, ii) the antibacterial, prey-derived complement system of the ingested blood, which remains active for 1-2 days after feeding, iii) antibacterial peptides released by *Aeromonas*, iv) production of gut antimicrobial proteins by the host [258, 259]. *Aeromonas* and the *Rikenella*-like symbiont are hypothesized to benefit the host threefold: i) by supplementing essential B-vitamins, which are naturally lacking in the host's diet that consists exclusively of vertebrate blood, ii) aiding in blood digestion, especially in the lysis of erythrocytes, and iii) supporting the leech immune system with the production of antimicrobials [260, 258, 257, 259]. *Aeromonas veronii* is of particular interest to comparative immunologists, since it is not only a mutualistic symbiont in the leech crop, but also an opportunistic pathogen in mammals, fish and amphibians [261, 257]. It therefore lends itself to investigations focusing on the mechanisms that allow pathogenic colonization in vertebrates and beneficial colonization in leeches [262, 263, 264]. Recent studies showed that *Aeromonas veronii* requires a type 3 secretion system (T3SS), for successful host colonization as a pathogen and as a beneficial symbiont [262]. In both cases, the T3SS helped to escape the host immune

system, albeit in different ways. In mice that were injected with *A. veronii* T3SS was crucial for killing macrophages, while in the leech the T3SS allowed *A. veronii* to attach to coelomocytes without inducing phagocytosis and without killing the coelomocytes [262]. The *Rikenella*-like symbiont does not possess a T3SS, but might escape phagocytosis by being embedded in crop mucus, and/or a bacteria-derived polysaccharide matrix [265]. Further colonization mutants where identified in [263], which, based on the annotation of the inactivated genes, were hypothesized to be the result of altered bacterial cell wall features, gene regulation, reduced capacity to import nutrients, and loss of function in the type 2 secretion system (T2SS). T2SS-negative *A. veronii* mutants were further analyzed and revealed to be unable to export hemolysin, which is hypothesized to be involved in erythrocyte lysis and therefore heme acquisition [264]. Putative leech immune genes have been recently identified using transcriptomic sequencing [266], and next-generation sequencing is now used to investigate the transcriptomes of the leech microbial community [267], promising significant advances in the study of leech-symbiont interactions in the future.

## 1.4 Marine chemosynthetic symbioses

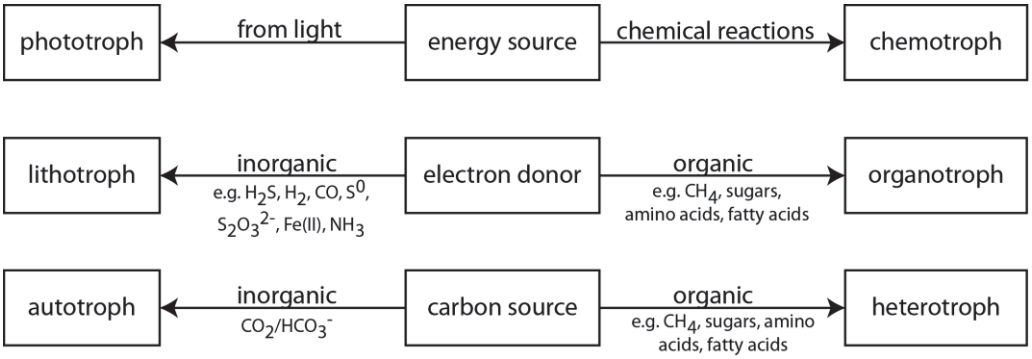
In marine chemosynthetic symbiosis, an invertebrate animal or ciliate protist lives in close, often obligate association with chemosynthetic bacteria. The chemosynthetic symbionts are able to synthesize complex organic molecules from simple, inorganic substrates, and thereby provide their host with nutrition from sources that are otherwise inaccessible to animals. There are numerous types of chemosynthetic symbioses in the marine environment, covering a wide range of habitats, host taxa, symbiont phylotypes and types of symbiotic interactions. The following sections will give a definition of important terms and a brief introduction into marine chemosynthetic symbioses.

### 1.4.1 Definition – chemosynthesis

Primary production, i.e. the production of biomass from inorganic carbon sources (carbon fixation), is achieved through two principal processes on earth: photosynthesis and chemosynthesis. Photosynthesis converts energy from sunlight into chemical energy, which is then used to synthesize organic molecules, such as sugars, from water and carbon dioxide [268]. Likewise, in chemosynthesis, energy from the oxidation of reduced inorganic molecules (e.g. hydrogen sulfide) is used instead of sunlight to convert inorganic carbon ($CO_2$/$HCO_3^-$) or organic one-carbon molecules ($CH_4$) into biomass. Many inorganic electron donor/electron acceptor redox couples could potentially yield sufficient energy to fuel carbon fixation, and some of these are realized in chemosynthetic symbioses (Table 7, p. 39).

Organisms are classified by the types of energy, electron sources and carbon sources they use to fuel their metabolism (Figure 6, p. 38). In chemosynthetic symbioses, the most common types of chemosynthesis are thiotrophy (fixation of inorganic carbon coupled to the oxidation of reduced sulfur compounds, i.e. a form of chemolithoautotrophy -> sulfur-oxidizing (SOX) symbionts), and methanotrophy (use of $CH_4$ as both electron donor and carbon source, i.e. a form of chemoorganoheterotrophy -> methane-oxidizing (MOX) symbionts). However, many organisms

defy strict classification into only one of these nutritional categories, because they are able to use different sources of energy, electrons and carbon [269]. For example, many pelagic bacteria are able to use sunlight to generate ATP with proteorhodopsin, while still gaining energy, as well as electrons and carbon, from the degradation of organic matter taken up from the environment [270]. Such organisms, which combine multiple trophic strategies in their metabolism, are termed mixotrophs. Mixotrophy is more costly because more biochemical machinery is needed to exploit multiple resources, but also allows for the use of a wider range of substrates and energy sources. In a dynamic and resource-limited environment, this provides a significant advantage over more cost-efficient but metabolically restricted organisms. Mixotrophs therefore often play a role in symbioses that are characterized by fluctuating conditions and limited resource availability, for example in some sponges [271], hydrothermal vent tube worms [247], and in the gutless oligochaetes [59, 60].




| | |
|---|---|
| photolithoautotroph | plants, algae, cyanobacteria |
| photolithoheterotroph | some purple sulfur bacteria, e.g. *Marichromatium fluminis* [272] |
| photoorganoautotroph | some green non-sulfur bacteria, e.g. *Chloroflexus aurantiacus* [273] |
| photoorganoheterotroph | some marine pelagic Gammaproteobacteria, e.g. *Pelagibacter ubique* [274] , *Congregibacter litoralis* [275] |
| chemolithoautotroph | sulfur-oxidizing bacteria, e.g. *Acidithiobacillus thiooxidans* [276] |
| chemolithoheterotroph | sulfur-oxidizing bacteria, e.g. *Thiobacillus* strain Q [277] |
| chemoorganoautotroph | methanotrophic prokaryotes, e.g. *Methylococcus spp.* |
| chemoorganoheterotroph | animals, fungi, methanotrophic symbionts |

**Figure 6: Classification of primary nutritional groups and prominent examples of each group**

**Table 7: Examples of chemosynthetic processes realized in animal symbiosis**

| Energy metabolism | e- donor | e- acceptor | Redox reaction | Host phylum | Host family | Host species | Reference |
|---|---|---|---|---|---|---|---|
| Aerobic methane oxidation | $CH_4$ | $O_2$ | $CH_4 + 2\ O_2 \rightarrow HCO_3^- + H^+ + H_2O$ | Porifera | Cladorhizidae | *Cladorhiza methanophila* | [278] |
| | | | | Mollusca | Bathymodiolinae | *Bathymodiolus childressi* | [279] |
| Aerobic sulfide oxidation | $H_2S$ | $O_2$ | $H_2S + 2\ O_2 \rightarrow SO_4^{2-} + 2\ H^+$ | Ciliophora | Zoothamniidae | *Zoothamnium niveum* | [280] |
| | | | | Nematoda | Stilbonematinae | *Laxus oneistus* | [281] |
| | | | | Nematoda | Siphonolaimidae | *Astomonema southwardorum* | [282] |
| | | | | Mollusca | Solemyidae | *Solemya reidi* | [283] |
| | | | | Mollusca | Lucinidae | *Lucinoma aff. kazani* | [284] |
| | | | | Mollusca | Thyasiridae | *Thyasira equalis* | [285] |
| | | | | Mollusca | Vesicomyidae | *Calyptogena magnifica* | [286] |
| | | | | Mollusca | Bathymodiolinae | *Bathymodiolus azoricus* | [287] |
| | | | | Mollusca | Provannidae | *Alviniconcha hessleri* | [288, 289] |
| | | | | Mollusca | Lepetodrilidae | *Lepetodrilus fucensis* | [290, 291] |
| | | | | Annelida | Alvinellidae | *Alvinella pompejana* | [292] |
| | | | | Annelida | Siboglinidae | *Riftia pachyptila* | [293, 294] |
| | | | | Annelida | Phallodrilidae | *Olavius algarvensis* | [295] |
| | | | | Arthropoda | Alvinocarididae | *Rimicaris exoculata* | [296] |
| | | | | Echinodermata | Loveniidae | *Echinocardium cordatum* | [297] |
| Aerobic hydrogen oxidation | $H_2$ | $O_2$ | $2\ H_2 + O_2 \rightarrow 2\ H_2O$ | Mollusca | Bathymodiolinae | *Bathymodiolus puteoserpentis* | [298] |
| | | | | Annelida | Siboglinidae | *Riftia pachyptila* | [298] |
| | | | | Arthropoda | Alvinocarididae | *Rimicaris exoculata* | [298] |
| Anaerobic sulfide oxidation with nitrate | $HS^-$ | $NO_3^-$ | $5\ HS^- + 8\ NO_3^{2-} + 3\ H^+ \rightarrow 5\ SO_4^{2-} + 4\ N_2 + 4\ H_2O$ | Nematoda | Stilbonematinae | *Laxus oneistus* | [281] |
| | | | | Mollusca | Lucinidae | *Lucinoma aequizonata* | [299] |
| | | | | Annelida | Siboglinidae | *Riftia pachyptila* | [300] |
| | | | | Annelida | Phallodrilidae | *Olavius algarvensis* | [59] |
| Anaerobic hydrogen oxidation with sulfate | $H_2$ | $SO_4^{2-}$ | $4\ H_2 + SO_4^{2-} + H^+ \rightarrow HS^- + 4\ H^2O$ | Annelida | Phallodrilidae | *Olavius algarvensis* | [60] |

### 1.4.2 Diversity of chemosynthetic symbioses and their habitats

Chemosynthetic symbioses were first discovered at deep sea hydrothermal vents along the Galapagos Rift in the Pacific Ocean in 1977 [301]. There, tube worms and mussels that live in symbiosis with chemosynthetic symbionts form vast communities [302], which, at the time, were unexpected to exist in the nutrient-poor deep sea. Soon, the presence of intracellular sulfur- and methane oxidizing symbionts, and their significance to the animals' nutrition was recognized (reviewed in [52, 53]).

Since then, chemosynthetic symbioses have been found in many diverse habitats, including cold seeps, whale and wood falls, shallow water sediments in association with coral reefs, mangroves and seagrasses, and muddy sediments along continental slopes (Figure 7). To date, hundreds of species from six different animal phyla and two groups of ciliates have been identified (Figure 7, a selection of prominent examples also listed in Table 7, p. 39).

Likewise, the phylogenetic diversity of chemosynthetic symbionts is very high, because many symbionts are host lineage- over even host species-specific [52, 130]. Some hydrothermal vent polychaetes (*Alvinella*), shrimp (*Rimicaris*) and snails (*Alviniconcha*) associate with chemosynthetic Epsilonproteobacteria [303, 304, 305], but most chemosynthetic symbionts (SOX as well as MOX) are from various clades within the Gammaproteobacteria (Figure 8, p. 42). Chemosynthetic symbionts form several well-separated clades that often contain free-living bacteria as well (Figure 8). The fact that many symbiont clades are more closely related to free-living bacteria than to other chemosynthetic symbionts is evidence that chemosynthetic symbioses independently arose numerous times [52]. Chemosynthetic symbionts are not only phylogenetically diverse, but also employ many different chemosynthetic pathways for $CO_2$ fixation, transmission strategies (from horizontal to strictly vertical, [147]) and modes of association (from epibionts and extracellular ectosymbionts to intracellular endosymbionts, [53]).

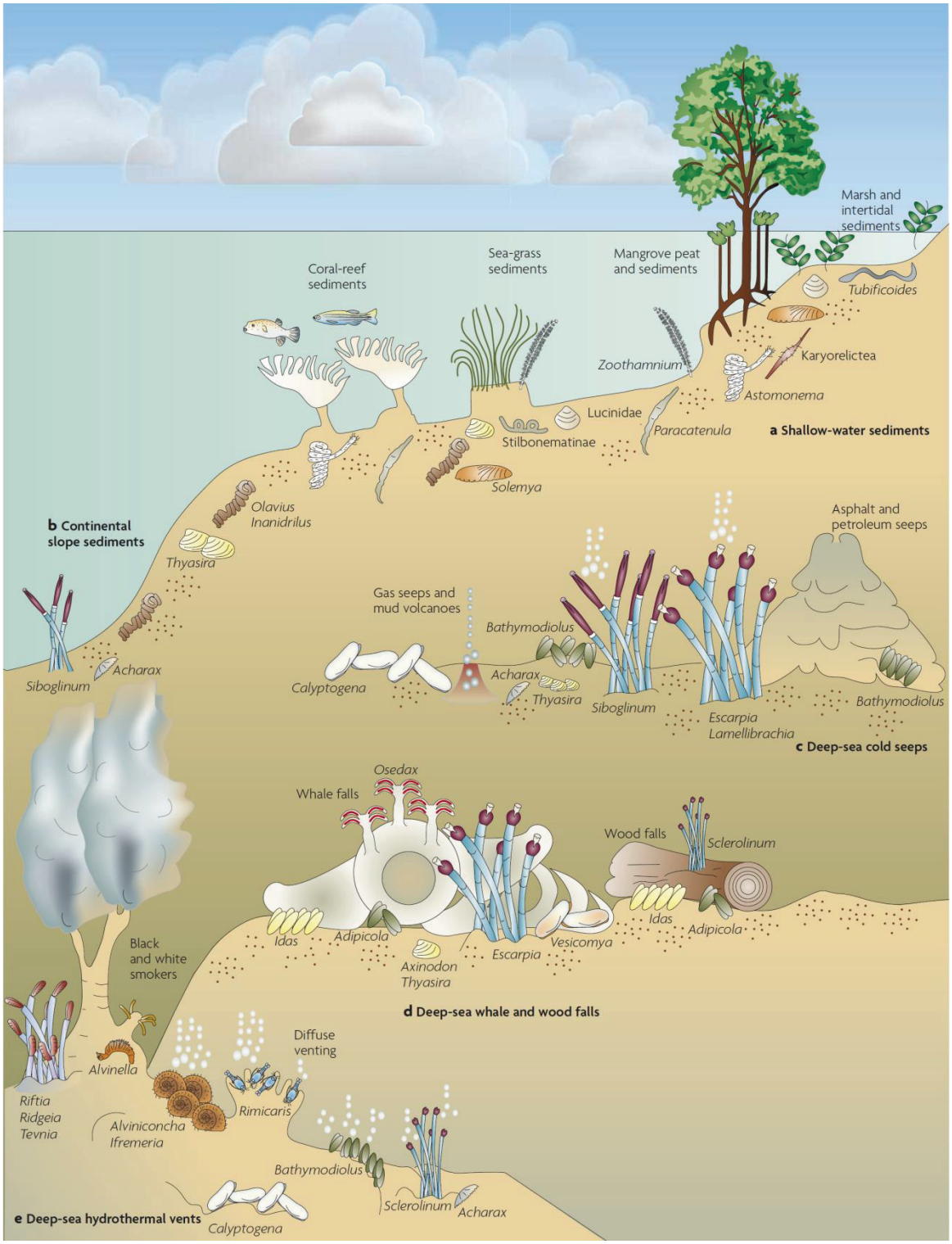
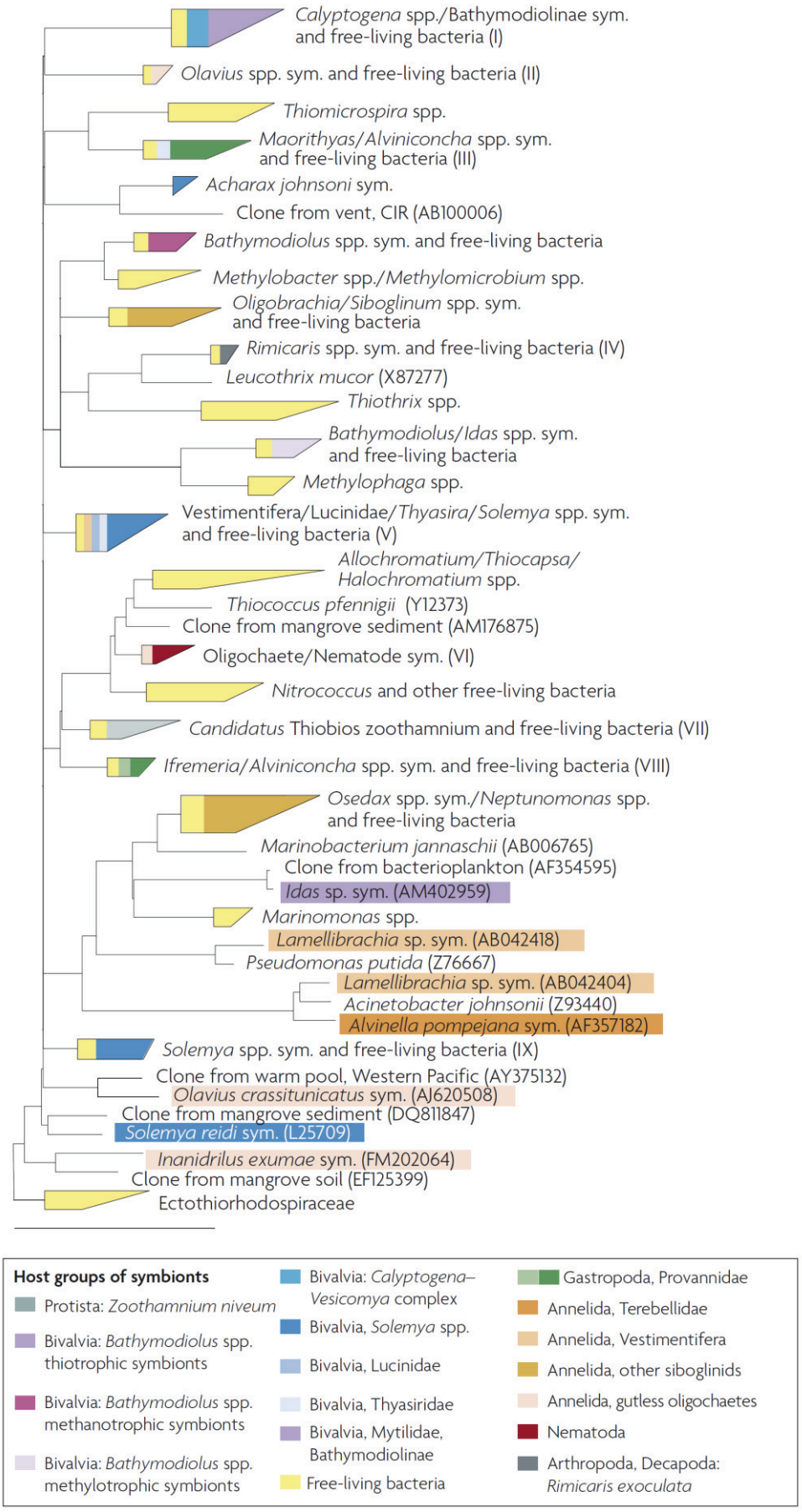



**Figure 7:** Overview of the diversity of host animals that form chemosynthetic symbioses and the range of marine habitats they occur in. Adapted from [52].

**Figure 8 (next page):** Diversity of gammaproteobacterial 16S rRNA phylotypes of chemosynthetic symbionts associated with animal and ciliate hosts. Distinct chemosymbiotic clades are numbered with roman numerals. Adapted from [52].


*Calyptogena* spp./Bathymodiolinae sym. and free-living bacteria (I)
*Olavius* spp. sym. and free-living bacteria (II)
*Thiomicrospira* spp.
*Maorithyas/Alviniconcha* spp. sym. and free-living bacteria (III)
*Acharax johnsoni* sym.
Clone from vent, CIR (AB100006)
*Bathymodiolus* spp. sym. and free-living bacteria
*Methylobacter* spp./*Methylomicrobium* spp.
*Oligobrachia/Siboglinum* spp. sym. and free-living bacteria
*Rimicaris* spp. sym. and free-living bacteria (IV)
*Leucothrix mucor* (X87277)
*Thiothrix* spp.
*Bathymodiolus/Idas* spp. sym. and free-living bacteria
*Methylophaga* spp.
Vestimentifera/Lucinidae/*Thyasira/Solemya* spp. sym. and free-living bacteria (V)
*Allochromatium/Thiocapsa/ Halochromatium* spp.
*Thiococcus pfennigii* (Y12373)
Clone from mangrove sediment (AM176875)
Oligochaete/Nematode sym. (VI)
*Nitrococcus* and other free-living bacteria
*Candidatus* Thiobios zoothamnium and free-living bacteria (VII)
*Ifremeria/Alviniconcha* spp. sym. and free-living bacteria (VIII)
*Osedax* spp. sym./*Neptunomonas* spp. and free-living bacteria
*Marinobacterium jannaschii* (AB006765)
Clone from bacterioplankton (AF354595)
*Idas* sp. sym. (AM402959)
*Marinomonas* spp.
*Lamellibrachia* sp. sym. (AB042418)
*Pseudomonas putida* (Z76667)
*Lamellibrachia* sp. sym. (AB042404)
*Acinetobacter johnsonii* (Z93440)
*Alvinella pompejana* sym. (AF357182)
*Solemya* spp. sym. and free-living bacteria (IX)
Clone from warm pool, Western Pacific (AY375132)
*Olavius crassitunicatus* sym. (AJ620508)
Clone from mangrove sediment (DQ811847)
*Solemya reidi* sym. (L25709)
*Inanidrilus exumae* sym. (FM202064)
Clone from mangrove soil (EF125399)
Ectothiorhodospiraceae
**Host groups of symbionts**
Protista: *Zoothamnium niveum*
Bivalvia: *Bathymodiolus* spp. thiotrophic symbionts
Bivalvia: *Bathymodiolus* spp. methanotrophic symbionts
Bivalvia: *Bathymodiolus* spp. methylotrophic symbionts
Bivalvia: *Calyptogena–Vesicomya* complex
Bivalvia, *Solemya* spp.
Bivalvia, Lucinidae
Bivalvia, Thyasiridae
Bivalvia, Mytilidae, Bathymodiolinae
Free-living bacteria
Gastropoda, Provannidae
Annelida, Terebellidae
Annelida, Vestimentifera
Annelida, other siboglinids
Annelida, gutless oligochaetes
Nematoda
Arthropoda, Decapoda: *Rimicaris exoculata*

## 1.5 Gutless oligochaete chemosynthetic symbioses

The first gutless oligochaete species, *Phallodrilus albidus* (now *Olavius albidus*), was described as early as 1977 by Jamieson [306]; however, the two most outstanding characteristics of this animal, i) the lack of a gut and ii) the presence of bacterial symbionts, were not detected at the time. The presence of symbiotic bacteria and the chemosynthetic and nutritional nature of the symbiosis were recognized just a few years later, around the same time the chemosynthetic nutritional symbiosis of the deep sea hydrothermal vent tube worm *Riftia pachyptila* was discovered [307, 308, 309, 310]. Since then, a wealth of morphological, ecophysiological, taxonomical, and, in recent years, molecular studies have uncovered a highly diverse and complex symbiotic system unlike any other in the world of marine chemosynthetic symbioses. In the following, I will review the current knowledge on gutless oligochaetes, with particular emphasis on the model species *Olavius algarvensis*, which is the focus of this thesis.

### 1.5.1 Morphological characteristics of the gutless oligochaete symbiosis

The gutless oligochaetes are small marine annelids within the family Phallodrilinae (Clitellata) that inhabit the interstitial pore water of marine sediments [311]. Their bodies are very thin and elongated (0.1-0.2 mm in diameter and 10-40 mm in length); a morphology that is well adapted to life in the interstitium and that is typical for interstitial meiofauna in general [312]. Gutless oligochaete species are morphologically very similar to each other and hard to distinguish even for experienced taxonomists. Species are distinguished morphologically mainly by features of the genital organs and the presence, shape and number of penial setae [313].

The gutless oligochaetes are highly unusual in that they not only lack a digestive tract (no mouth, gut, and anus), but also nephridia (excretory organs) [308]. Instead they harbor large amounts of bacterial symbionts (constituting ~25% of the worm's biomass [314]), which provide the host with nutrition through chemoautotrophic carbon fixation, and which also recycle host waste products [59]. Although the cuticle is permeable for compounds at least up to 70 kDa ([311], and

the worms do show some potential for the uptake and utilization of dissolved organic compounds if supplied externally [315, 316], it seems safe to assume that under most circumstances, the symbionts contribute most if not all of the nutrition, and that the symbiosis is thus obligate at least for the host [311].

Easily distinguished from other meiofauna by eye, the gutless oligochaetes are further characterized by their unusual bright, chalky-white color (Figure 9, A and B), which stems from large amounts of light-reflecting sulfur and PHA granules stored inside the sulfur-oxidizing symbionts [306, 308]. PHA (polyhydroxyalkanoate) is a polymeric carbon and energy storage compound in bacteria [317], and in gutless oligochaetes, PHA can contribute up to 10% of the total worm dry weight [318]. When these storage compounds get depleted, the worms lose their white coloring and turn beige-transparent (called “pale”, Figure 9D [319]).

All gutless oligochaete species harbor at least two morphological types of bacterial endosymbionts that are discernible in TEM images [308, 310]: a) a large (2-7 µm) oval morphotype containing large amounts of the aforementioned inclusions which give rise to the white color of the worms, and b) a much smaller (0.7-1.5 µm) rod or croissant shaped

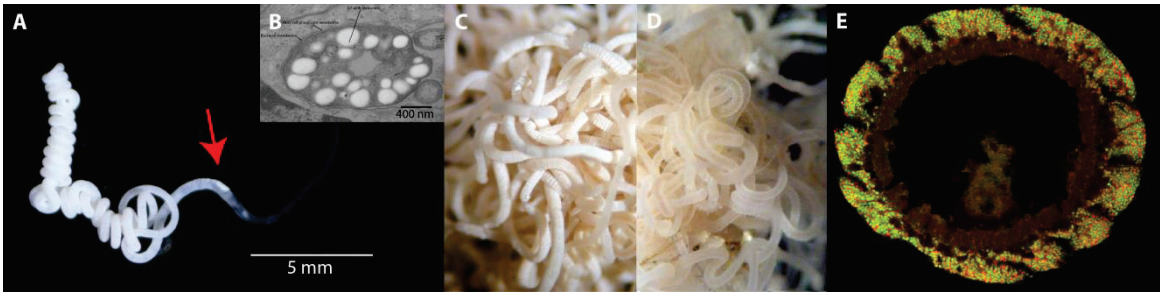



**Figure 9: Morphology of gutless oligochaetes. A)** Mature *Olavius algarvensis* worm, arrow indicates location of genital pad, image courtesy of Alexander Gruhl, **B)** TEM image of primary SOX symbiont in *O. algarvensis* showing storage granules, image courtesy of Nikolaus Leisch, **C)** Collection of “white” gutless oligochaetes, **D)** Collection of “pale” gutless oligochaetes. Images C) and D) are courtesy of Christian Lott. **E)** Cross section of *O. algarvensis* stained with symbiont-targeting FISH probes (green: Gammaproteobacteria, red: Deltaproteobacteria) showing the location of the symbionts between the epidermis and the cuticle. Adapted from [320].

morphotype with no or fewer inclusions [311]. The bacterial endosymbionts occur extracellularly along the entire length of the worm between the cuticle and the epidermis of the worm, but the two morphotypes are not evenly distributed. The first anterior segments of the worm up to the clitellar region which contains the genital organs (segments XI - XIII) only contain the smaller morphotype in low numbers, and as a result this part of the worm is always pale (Figure 9A). The postclitellar part of the body is colonized by both morphotypes and thus usually appears white.

On the ventral side of the genital segments is an area where the space between epidermis and cuticle is much wider than in other parts of the body. This area is called the genital pad. It appears bright white in sexually mature worms, because it is densely filled with symbiont cells of both morphotypes in reproductively active worms (Figure 9A). The genital pad is implicated in the transmission of symbionts from the parent worm to the egg (see section 1.5.4).

Gutless oligochaete symbionts are generally described as being extracellular endosymbionts. However, the accuracy of this description has been previously challenged for the following two reasons [321, 322]. First, depending on the host species, the symbionts can be engulfed and reside inside epidermal vacuoles, thus becoming *de facto* intracellular [321]. It is unknown if these symbionts are able to persist within epidermal vacuoles for longer periods of time before eventually being lysed, and this may vary considerably between host species. Second, when extracellular, the symbionts are i) subject to environmental influences due to the permeable nature of the host's cuticle and ii) have not invaded any host tissue or crossed tissue boundaries. Since both points are of biological significance, it needs to be considered when drawing general conclusions by comparing these symbionts to "true" endosymbionts which are permanently housed within host tissues (e.g. the intracellular endosymbionts of aphids [35] or extracellular endosymbionts of earthworms [253]).

### 1.5.2 Biogeography and ecology of gutless oligochaetes

Gutless oligochaetes are a significant part of the marine meiofauna in many tropical and subtropical sediments throughout the world (Figure 10) [311], and can reach densities of 100,000 individuals per $m^2$ in some locations [323]. Most species are described from warm shallow water coastal sediments (e.g. [324, 325, 326]), where they are easy to find and collect, but some species have also been found in colder continental shelf sediments up to a depth of 583 m [327, 328].

Typical habitats of gutless oligochaetes are subtidal calcareous sediments associated with coral reefs, where the highest numbers are found in heterogeneous sand of varying grain sizes that has collected in depressions between coral blocks ([329] and personal observation). There, anoxic and sulfidic conditions develop within the first few millimeters of the sediment surface due to microbial respiration of organic material. Most worms are found within 5 to 15 cm depth, avoiding both completely oxic and highly sulfidic zones [329].

Other environments include anoxic and sulfidic shelf sediments composed of soft, muddy material in depths of 100-400 m [327, 328] and coarse, oligotrophic, non-sulfidic siliceous or calcareous sediments associated with seagrass meadows ([321], discussed in more detail in the context of symbiont metabolism in section 1.5.5).

The biogeographical pattern of gutless oligochaetes is not uniform. Some species are highly endemic to a particular location [330], while others are cosmopolitan with wide but disjunct distributions (summarized in [331]). Often, several species co-occur at the same site (e.g. [323, 321]) suggesting a high degree of micro-niche partitioning between species [323]. In many cases co-occurring species are more closely related to species from other regions of the world than with each other (see section 1.5.3, [130]). This is particularly interesting in light of the fact that gutless oligochaetes have limited means of dispersal, as they do not form planktonic egg or

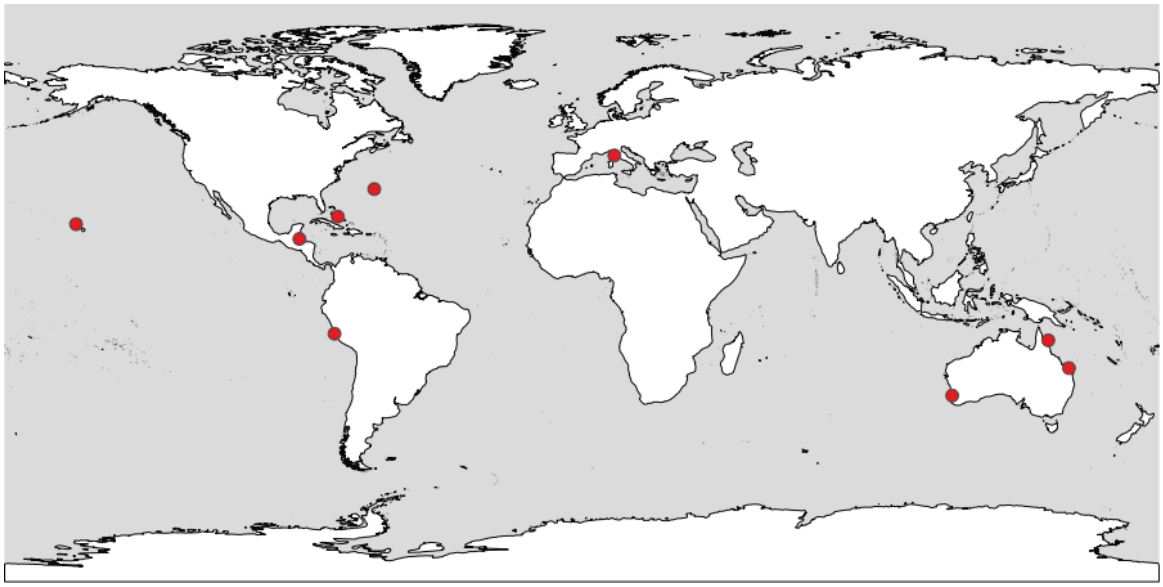


**Figure 10: Distribution of gutless oligochaetes across the world.** Map was generated with simplemappr (http://www.simplemappr.net). Gutless oligochaete species used to generate map: *O. albidoides* (Rottnest Island, West Australia, -31.999, 115.49), *I. leukodermatus* A (Harrington Sound, Bermuda, 32.324, -64.738), *I. leukodermatus* B (Carrie Bow Cay, Belize, 16.80243, -88.08213), *O. loisae* (Heron Island, East Australia, -23.443, 151.913), *O. imperfectus* (Lee Stocking Island, Bahamas, 23.767, -76.1), *O. algarvensis* (Elba, Italy, 42.80816, 10.14202), *O. crassitunicatus* (continental margin off Peru, -12.73217, -77.13267), *I. manae* (Lizard Island, East Australia, -14.787, 145.452), *Inanidrilus* 'Hawaii sp. 1' (Oahu, Hawaii, 21.394446, -157.714627).

larval stages [330]. It is entirely possible that there are gutless oligochaete species that live in much deeper and more diverse habitats than currently known, and that these take part in bridging this gap.

Since sulfide is almost always available in gutless oligochaete habitats, it was assumed and shown early on that the large symbiont morphotype is thiotrophic and oxidizes the environmental sulfide present in its habitat for energy conservation and subsequent inorganic carbon fixation [329, 309, 319].

Most worms inhabit the anoxic and sulfidic layers of the sediment, and are only rarely found in the top oxygen-containing layer [329, 45, 331]. Like other marine invertebrates, the worms are capable of switching to an aerobic fermentative metabolism for short periods of time [332] but

die if exposed to anoxia for several days (unpublished observation in [311]). It is hypothesized that the worms migrate back and forth between anoxic and oxygenated layers, taking up oxygen and storing it to keep up aerobic respiration for some time under hypoxic conditions ("whale hypothesis"). However, the respiratory pigments in these animals and their oxygen-binding/-storing properties remain to be investigated. The work presented in this thesis significantly expands our knowledge on the oxygen-binding proteins present in the model gutless oligochaete *Olavius algarvensis* and gives fresh impetus to the whale hypothesis (chapter 4).

As stated before, gutless oligochaetes are able to take up dissolved organic compounds from the surrounding medium, and such compounds might contribute to the overall nutrition of the worms [329, 315]. Previous studies report the presence of dissolved organic compounds in the habitats of gutless oligochaetes wherever this has been investigated ([329, 333], Manuel Liebeke and Erik Puskas, unpublished results). However, for which species, under which circumstances, and to what extent this plays a role remains to be determined. This thesis provides genomic evidence that simple carbohydrates from the environment could be used to additionally fuel the symbiosis in *O. algarvensis* via its spirochaete symbiont (chapter 3).

### 1.5.3 Diversity and phylogeny of gutless oligochaete hosts and their symbionts

The diversity of gutless oligochaetes is large. To date, 88 species have been formally described (summarized in [334]), but there are many more that still await taxonomic description. A recent phylogenetic study by Zimmermann et al. showed the existence of two cryptic species in the nominal *Olavius imperfectus* (Figure 11, p. 51, [130]). As mentioned in section 1.5.1, identifying species morphologically is challenging and therefore more such cases might surface as new molecular data becomes available. More species are found with virtually every field excursion to a new region, suggesting that the true diversity is heavily undersampled. The highest diversity of gutless oligochaetes is recorded from coral reef sediments in the Caribbean and the Australian Great Barrier Reef, with as many as 18 species described from a single collection site [323]. The

gutless oligochaetes are a monophyletic group [313, 335, 336, 337] composed of the two sister genera *Olavius* and *Inanidrilus*. While the genus *Inanidrilus* is monophyletic, the genus *Olavius* is paraphyletic and requires revision (Figure 11, p. 51, [130]). Two observations indicate that the gutless oligochaetes have recently experienced or are experiencing evolutionary radiation: a) the large number of very closely related species comprised within only two genera, and b) extant species that represent morphologies ranging from “primitive” to highly derived.

On the symbiont side, the diversity is even larger, since each gutless oligochaete species harbors a species-specific consortium of at least three, and up to six, symbiont phylotypes [311]. Gutless oligochaete symbionts fall within three clades of Alphaproteobacteria, four clades of Gammaproteobacteria, ten clades of Deltaproteobacteria, and one clade of Spirochaeta (Figure 12, p. 53, not all clades are shown). Attempts to cultivate of any of the symbionts have so far been unsuccessful [338].

All host species analyzed to date, with the sole exception of *Inanidrilus exumae* [334], possess a primary sulfur-oxidizing symbiont of the Gamma1 clade, which corresponds to the large morphotype described in section 1.5.1 (soon to be named *Candidatus* Thiosymbion (Gruber-Vodicka et al., in prep.). In *I. exumae*, this symbiont is replaced by a novel sulfur-oxidizing gammaproteobacterial type (Gamma4) which looks morphologically very similar to Gamma1 and has so far not been found in any other species [334]. *Olavius crassitunicatus* possess a Gamma2 symbiont in addition to Gamma1 [339], and *Olavius algarvensis* and *Olavius ilvae* contain a Gamma3 symbiont in addition to Gamma1 [340]. The Gamma1 symbionts form a closely related monophyletic group related to the Chromatiaceae, together with the sulfur-oxidizing ectosymbionts of stilbonematine and intracellular endosymbionts of astomonematine nematodes (Figure 12, [130]). It is remarkable that this group of bacteria is able to associate with animals from two unrelated animal phyla with likely no environmental intermediates [130].

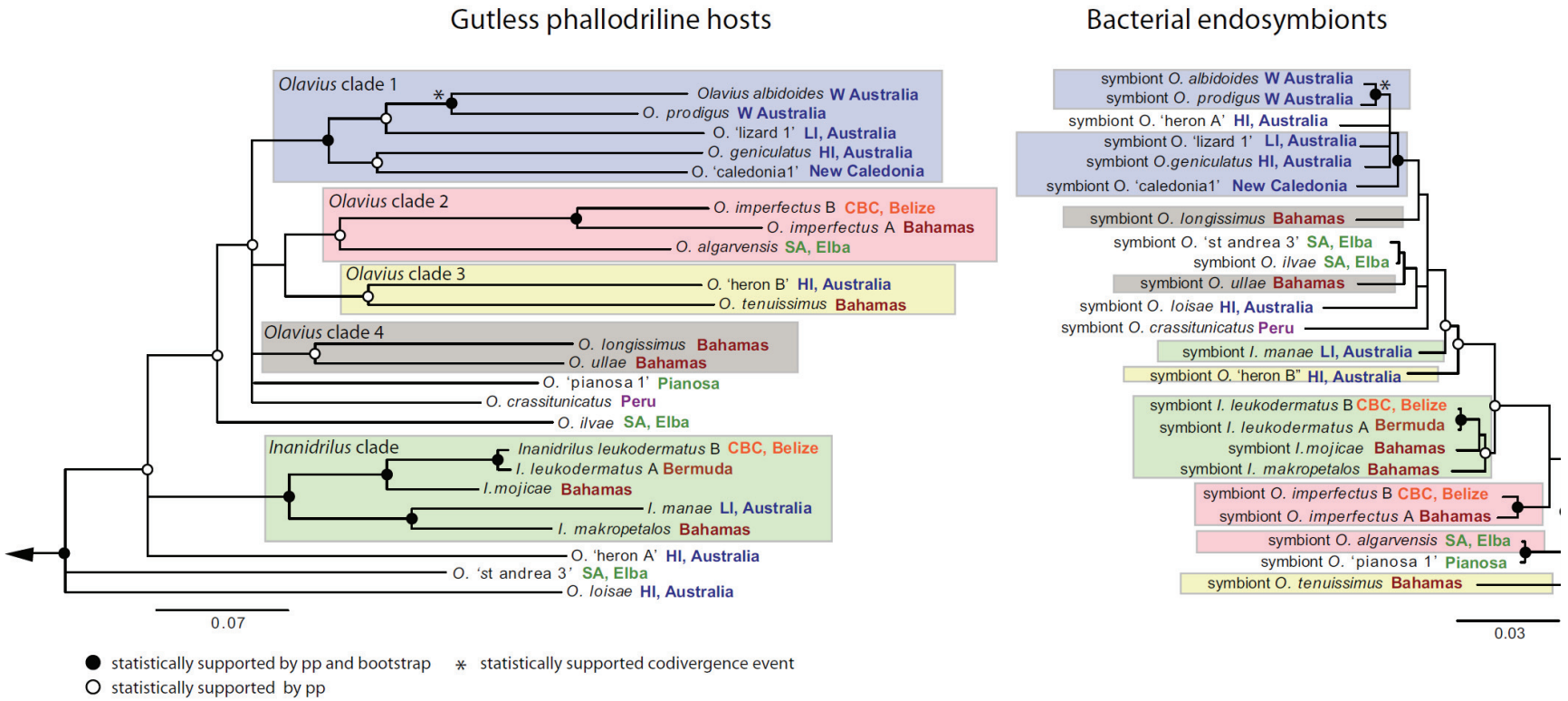



**Figure 11: Phylogeny of gutless oligochaetes and their corresponding primary symbiont Gamma1/Candidatus *Thiosymbion*.** Host tree was co from concatenated marker genes (mitochondrial (mt)12S, mt16S, 18S, 28S rRNA genes, internal transcribed spacer between 18S and 28S rRNA (IT (cytochrome oxidase I)), symbiont tree was constructed from 16S rRNA genes. Colored boxes indicate monophyletic groups. Stars indicate st supported co-diversification events. Black nodes indicate >80% bootstrap support in both Bayesian inference and maximum-likelihood calculatic nodes are only supported in Bayesian inference. Figure adapted from [130].

Host-switches between gutless oligochaetes and stilbonematine nematodes seem to have occurred at least three times [130], raising the question by which molecular mechanism(s) this clade of symbionts is able to colonize both host groups, while retaining this high degree of host species specificity. As outlined in section 1.3, molecular host-symbiont interactions are largely a result of the interplay between microbial molecular factors (MAMPs) and recognition molecules of the host immune system, such as lectins. In the stilbonematine nematodes *Laxus oneistus* and *Stilbonema majum* highly specific c-type lectins called Mermaid proteins seem to be responsible for recognition and attachment of their respective species-specific Gamma1 ectosymbionts [341, 168]. The first insight into the innate immune system of a gutless oligochaete, and its suggested role in symbiont recognition and interaction is provided in chapter 4.

The alpha- and deltabacterial symbionts correspond to the small morphotypes described in section 1.5.1. In contrast to the Gamma1/*Thiosymbion* clade, most of these do not form symbiont-exclusive clades, but contain closely related environmental bacteria as well. Within the Delta1 and Gamma3 clade, different symbiont phylotypes share their direct common ancestor not with each other, but with environmental bacteria (Figure 12), indicating that they were repeatedly taken up from the environment by different gutless oligochaete species. In chapter 2, I show that the diversity of deltaproteobacterial symbionts in *O. algarvensis* is even greater than previously known, and I present the genomes of two novel deltaproteobacterial symbionts. In the case of alphaproteobacterial symbionts, multiple very closely related phylotypes can co-occur in the same species (e.g. *Inanidrilus leukodermatus*, *Olavius loisae*, Figure 12). The existence of such closely related symbionts with presumably very similar metabolisms within the same host is puzzling because it is expected to cause strong competition for resources. On the other hand, if different symbionts are adapted to slightly different micro-niches, they might be able to exploit a wider range of resources. So far, three gutless oligochaete species have been found that also contain a spirochaete symbiont (Figure 12, [342, 339, 340]). The spirochaete

symbionts form a monophyletic clade distinct from any other spirochetes (>95.4% sequence identity). Since these spirochetes are found in gutless oligochaetes from diverse geographic regions and habitats, they appear to be regular symbionts, independent of geography or niche. However, their function is completely unknown. Chapter 3 provides first functional insights into the spirochaete symbiont of *Olavius algarvensis*.

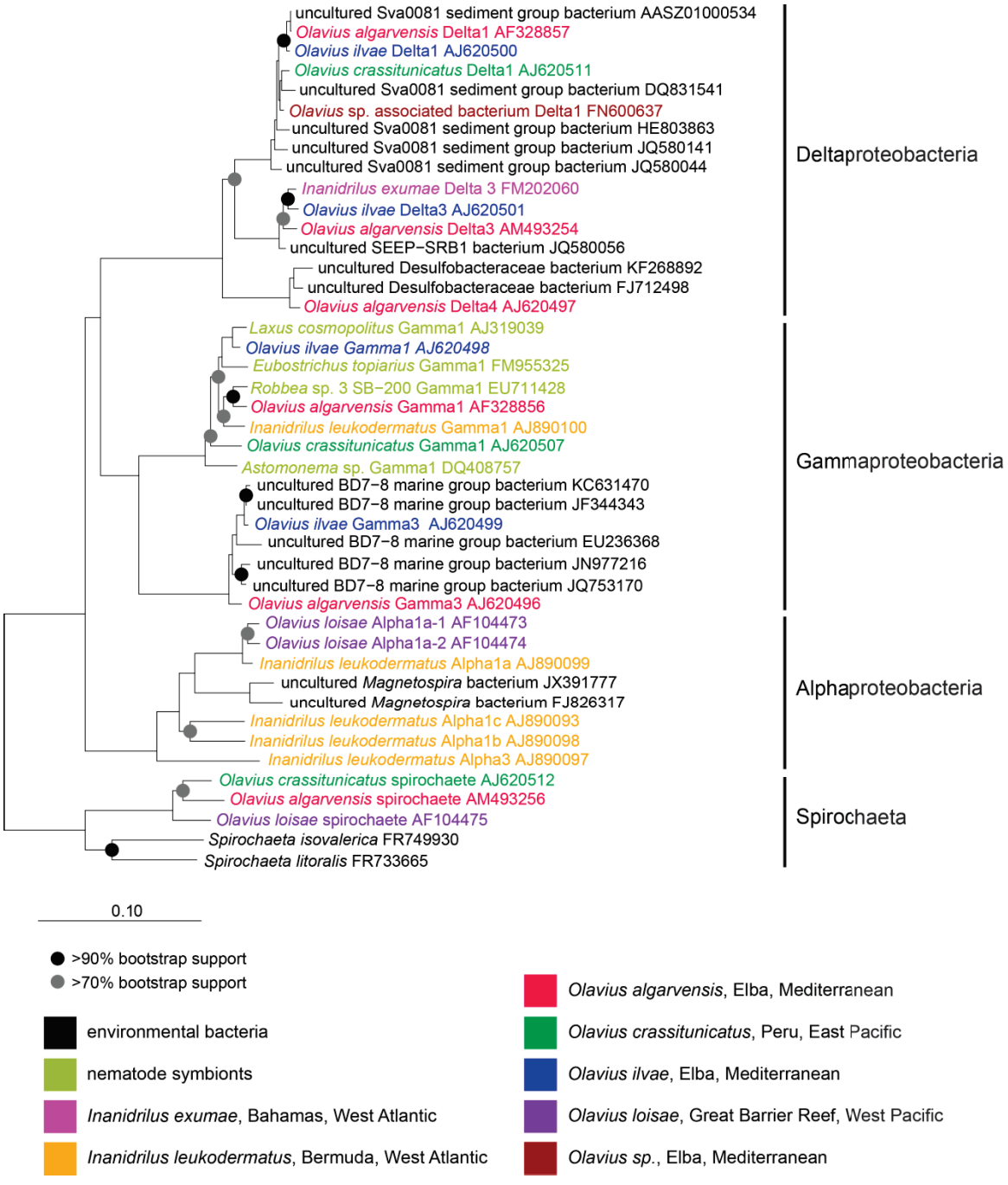


**Figure 12 (opposite page): Phylogenetic 16S rRNA Neighbor Joining gene tree of primary and secondary gutless oligochaete symbionts.** Gutless oligochaetes associate with secondary symbionts from clades within the Alpha-, Gamma-, and Deltaproteobacteria and Spirochaeta. Not all clades are shown (mostly those not yet confirmed by FISH).

### 1.5.4 Transmission of gutless oligochaete symbionts

Since the symbionts are functionally indispensable to the host, their reliable transmission from parent worm to offspring is of paramount importance to ensure survival of the next generation. This can be achieved through either strict vertical transmission via the germline or through highly specific uptake of symbionts from free-living populations or resting stages [147].

Like other oligochaetes, the gutless oligochaetes are hermaphrodites, i.e. each worm contains the full set of male and female genital organs; however self-fertilization is generally not possible due to the anatomical arrangement of genitalia [343]. Unusual for oligochaetes, the gutless Phallodrilinae only develop a single egg at a time, and the egg is not enclosed inside a cocoon after oviposition [343]. Instead, the egg is coated by a sticky mucus sheath during oviposition, which later hardens to form a more rigid egg integument [330].

It was determined through ultrastructural analysis [318, 330] and fluorescence *in situ* hybridizations (FISH, [344]) that the male and female gonads as well as the egg maturing inside the worm are free of any bacteria, but that juveniles already contain all symbiont morphotypes [318]. It was proposed early on that symbionts are transferred onto the egg surface during oviposition: the egg is squeezed through the oviduct and oviporus, rupturing the thin cuticle of the genital pad, which releases a large number of symbionts to the environment. It is thought that some of these released symbionts infect the freshly lain egg by sticking to the mucus sheath that surrounds the egg and subsequently invading it [318, 330, 322]. Since the eggs are deposited directly into the sediment and lack a rigid outer shell initially, environmental bacteria and symbionts released from other worms could also potentially infect the egg.

Schimak et al. recently traced the vertical transmission of $^{15}$N-labelled symbionts in *O. algarvensis* from parent worms to eggs and developing embryos [322]. The eggs were deposited into native sediment, allowing for potential infection by natural environmental strains. All bacterial cells later found in the egg mucus and the developing egg carried the stable N-isotope label. *O. algarvensis* harbors a Gamma1, Gamma3, Delta1, Delta4, and spirochete symbiont (although the true diversity is in fact higher, see chapter 2). Using specific FISH probes, $^{15}$N-labelled Gamma1, Gamma3 and Delta1 symbionts were clearly identified in various developmental stages. These results suggest that at least the Gamma1, Gamma3 and Delta1 symbionts are vertically transmitted, but do not preclude occasional horizontal transmission events. In chapter 2 of this thesis, using genomic approaches, I provide evidence that the Gamma1 and Gamma3 are indeed strictly vertically transmitted, but that all other symbionts, including the spirochaete, at least occasionally experience horizontal transmission.

### 1.5.5 Symbiont metabolism in the gutless oligochaete *Olavius algarvensis*

The Mediterranean species *Olavius algarvensis* is the best studied gutless oligochaete on the physiological and molecular level. As is true for all gutless oligochaete symbioses, *O. algarvensis* and its symbionts currently cannot be cultivated in the lab. Therefore, all studies on gutless oligochaetes are carried out with specimens collected from the wild, and the methods employed are culture-independent (short-term physiological experiments, molecular PCR-based approaches, metagenomic and transcriptomic sequencing, metaproteomics, metabolomics, and microscopic imaging methods).

The symbiont community has been intensively characterized using 16S rRNA gene clone libraries and FISH [295, 345, 340], and consists of the primary Gamma1 symbiont (*OalgG1*), a secondary Gamma3 symbiont (*OalgG3*), a Delta1 and Delta4 deltaproteobacterial symbiont (*OalgD1*, *OalgD4*), and a spirochaete symbiont (*OalgS1*). Two new symbiont phylotypes are established by

the work presented in this thesis (see chapter 2): a second Delta1-related phylotype (*OalgD1b*) and a Delta3 phylotype (*OalgD3*).

*O. algarvensis* was first described from shallow-water sediments off the Algarve coast of Portugal [346], and has since been found at several sites in the Tuscan archipelago off the islands of Elba and Pianosa [321]. There, the worms occur in a water depth of 6-8 m close to seagrass beds composed of *Posidonia oceanica* [295, 347] (Figure 14, p. 59). Most worms are found in a sediment depth of 12 cm, which is usually anoxic (C. Lott, unpublished data in [347]). *P. oceanica* forms large, dense meadows which are anchored in the sediment by an extensive root and rhizome system [348]. Underneath, dead, decaying rhizome material builds an impenetrable mat of ligneous peat. This peat is often found buried under the sediment where the worms occur (Figure 14). It is currently unknown if the sea grass meadow or peat also harbor gutless oligochaetes, due to their impervious nature.

The Elba sediments are unusual habitats for gutless oligochaetes, because they are oligotrophic, i.e. poor in nutrients and inorganic energy sources such as sulfide that fuel chemosynthesis (see section 1.4). Reduced sulfur compounds, phosphate, ammonium and nitrate are present only at nanomolar concentrations, if at all (summarized in [347]). It was shown that in *O. algarvensis*, the deltaproteobacterial *OalgD1* symbionts provide reduced sulfur compounds to the chemosynthetic sulfur-oxidizing *OalgG1* symbionts internally, eliminating the need to take up reduced sulfur compounds from the environment [295].

In 2006, draft genomes of *OalgG1*, *OalgG3*, *OalgD1* and *OalgD4* were obtained in one of the first large-scale metagenomic studies, at the time still performed through massive parallel end sequencing of fosmid clones using Sanger technology [59]. This study provided fundamental new insights into the metabolism of the *O. algarvensis* symbionts, showing that the *OalgG3* symbiont is a sulfur-oxidizing chemoautotroph like *OalgG1*, but uses nitrate instead of oxygen as a

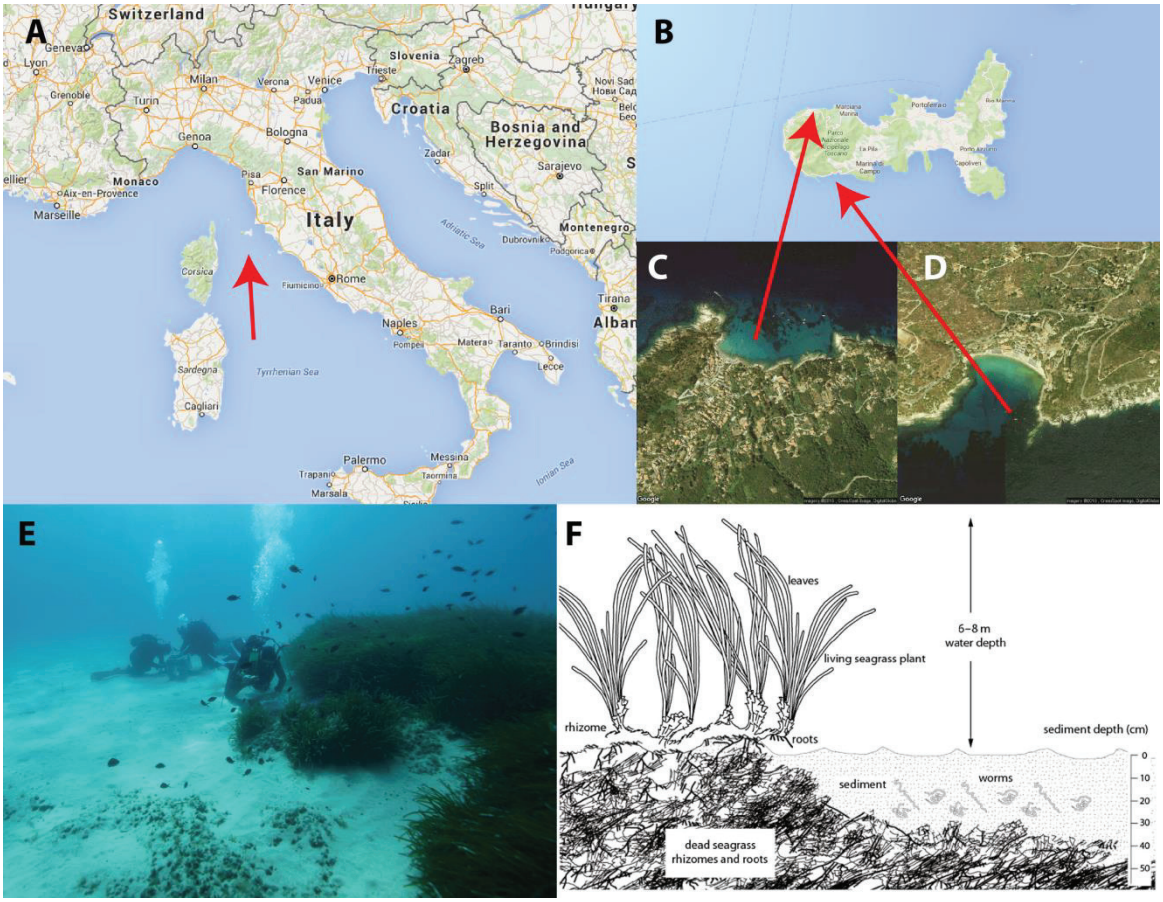



**Figure 14: Geographic location of *Olavius algarvensis* worms and their habitat. A)** The Island of Elba is located in the in the Tyrrhenian Sea, west of Italy, **B)** *O. algarvensis* specimens are found near seagrass beds close to the coast of Elba and Pianosa; shown here are sampling sites on Elba in the Bay of **(C)** Sant' Andrea and **(D)** Cavoli, **E)** Gutless oligochaetes occur in depths of ~8 m water depth; shown here is a usual sampling site in Sant' Andrea Bay, **F)** Schematic representation of the habitat structure. Worms typically occur in medium- to coarse grained sands close to beds of the seagrass *Posidonia oceanica*. This seagrass forms large peat-like structures composed of living root and rhizome material covering older rhizomes that have died off. Throughout the bay, mats of dead seagrass peat can be found buried underneath the sediment inhabited by the worms. A), B), C), D) Google Maps, E) Photo courtesy of the HYDRA Institute, F) Adapted from [349].

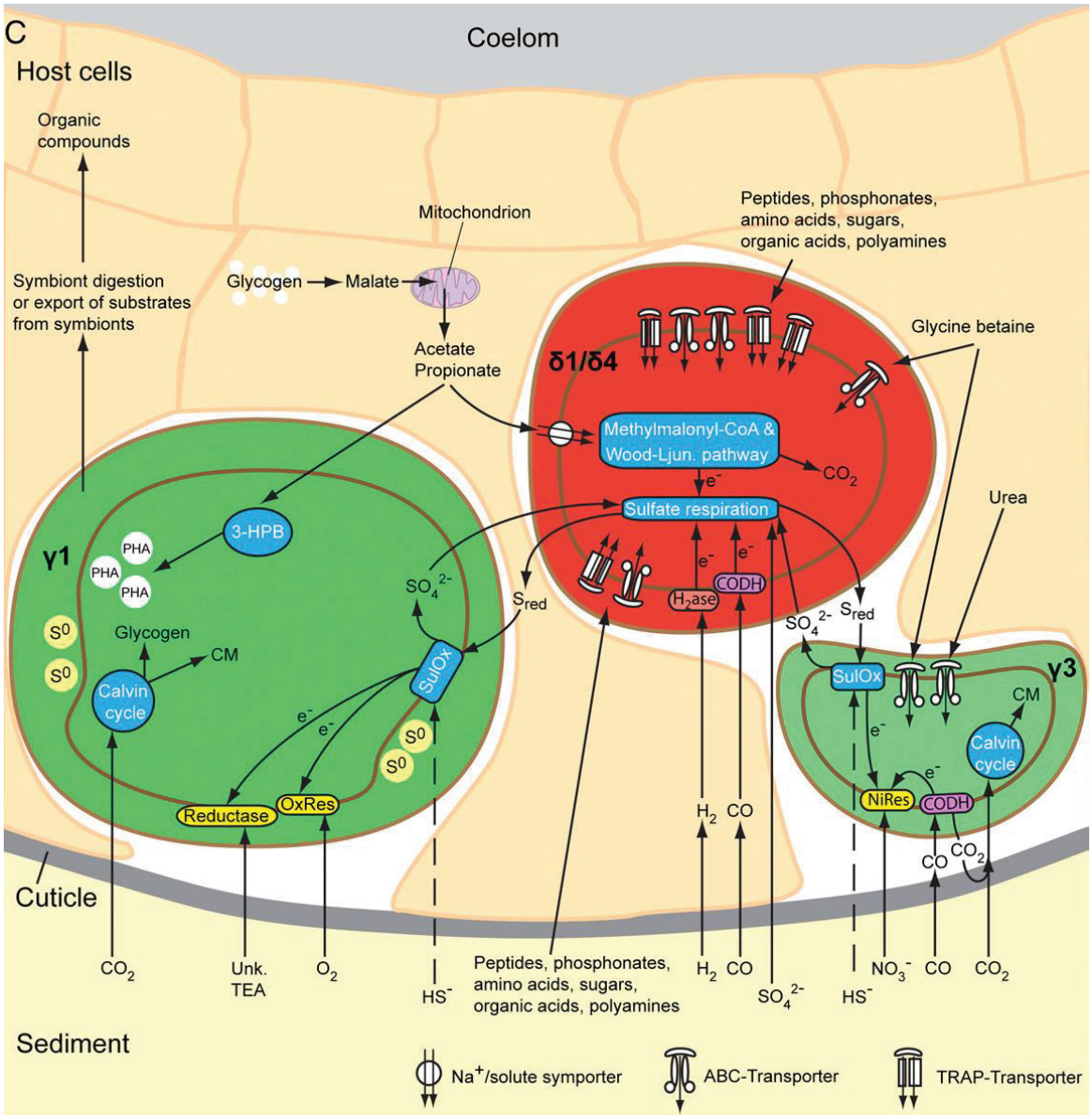



**Figure 15: Metabolic model of the *O. algarvensis* symbiotic consortium**.
Schematic is based on the metagenomic study by [59] and the metaproteomic study by [60]. *OalgD1* and *OalgD4* are shown as single entity, because of their functional similarity. 3-HPB, partial 3-hydroxypropionate bicycle; CM, cell material; CODH, aerobic or anaerobic carbon monoxide dehydrogenase; NiRes, nitrate respiration; OxRes, oxygen respiration; PHA, polyhydroxyalkanoate granules; $S^0$, elemental sulfur; $S_{red}$, reduced sulfur compounds; SulOx, sulfur oxidation; Unk. TEA, unknown terminal electron acceptor. Adapted from [60].

terminal electron acceptor. It confirmed that the deltaproteobacterial symbionts *OalgD1* and *OalgD4* are sulfate-reducers that engage in an internal syntrophic sulfur cycle with the gammaproteobacterial SOX symbionts. It further showed that both *OalgG1* and *OalgG3* fix $CO_2$ via the Calvin cycle using energy derived from sulfide oxidation [59]. The study also revealed that *OalgG3* and *OalgD1* are able to recycle energy- and nitrogen-rich waste products of the host, such as fermentation end products and nitrogenous waste compounds (Figure 15). It further suggested that *OalgD1* can use $H_2$ as an energy source.

The expression of these metabolic pathways was confirmed to in a metaproteomic study by Kleiner et al. in 2012 [60]. In addition, this study discovered that the gammaproteobacterial symbionts employ a modified, more energy efficient version of the Calvin cycle, and that *OalgG1* is able to assimilate short chain fatty acids derived from host fermentation into PHA as an energy and carbon store. Both deltaproteobacterial symbionts as well as *OalgG3* highly expressed carbon monoxide dehydrogenase, suggesting that they use CO as an energy source. Indeed, it was shown that high concentrations of carbon monoxide (and also $H_2$), present in the habitat of the worms, are sufficient to fuel symbiont metabolism [349].

The metabolism of the symbionts makes use of two gases that are considered toxic to the invertebrate host. Hydrogen sulfide blocks cytochrome oxidase c of the mitochondrial respiratory chain and also has a detrimental effect on other enzymes [350]. Carbon monoxide is an important energy source for most of the symbionts, but carbon monoxide strongly binds to hemoglobin, the suspected respiratory pigment in the worms, severely cutting oxygen availability. In chapter 4 of this thesis, I show that the host might be adapted to the presence of sulfide and carbon monoxide though the expression of respiratory proteins that bind or are insensitive to these gases.

Kleiner et al. showed, that the seeming functional redundancy of the gammaproteobacterial symbionts only exists at the surface, and that in reality, they fulfill similar, but different roles, making full use of the available resources. In this thesis, I obtained near complete genomes of *OalgD1* and *OalgD4*, as well as two novel deltaproteobacterial symbionts in *O. algarvensis*. Combinations of these deltaproteobacterial symbionts often coexist within the same worm, and they also appear to be functionally redundant on the surface. Further studies will show if these too are able to make use of subtle differences in available resources.

The metabolism of the spirochaete has so far remained completely unknown due to the difficulty of obtaining genomic information from this relatively low abundant symbiont. In this thesis, I obtained the first draft genome of this symbiont and describe its potential metabolic capacity in chapter 3.

## 1.6 Aims of the thesis

This thesis focuses on the symbiosis between the gutless oligochaete species *Olavius algarvensis* and its chemosynthetic bacterial consortium. It investigates the relationship between host and symbionts on a population genetics, evolutionary, metabolic and physiological level. The first part examines the population structure of *O. algarvensis* in its Sant' Andrea habitat, and deals with the diversity and flexibility of the associated symbiotic consortium, and the evolutionary relationships between host and symbionts (chapter 2). The second part describes the metabolic capabilities of the spirochaete symbiont, for which genomic information has so far been lacking (chapter 3). The last part of the thesis focuses on the physiological adaptations and immunological mechanisms that enable the host to live in such close association with a highly diverse symbiont consortium (chapter 4).

### 1.6.1 Host-symbiont population structure and evolutionary relationships

As outlined in section 1.2, the symbiotic lifestyle has profound effects on the evolution of hosts as well as symbionts. When I started out with my PhD, an in depth examination of the large scale phylogenetic relationships between gutless oligochaete hosts, stilbonematine/astomonematine nematode hosts and their shared primary Gamma1-clade symbionts by Zimmermann et al. was underway [130]. This study sought to resolve the phylogeny of all three groups, with the aim to uncover potential co-diversification patterns between hosts and symbionts, to identify possible host-switching events of the Gamma1 symbionts between and within oligochaete and nematode hosts, and to establish the species-specificity of Gamma1-animal associations. With respect to the gutless oligochaete symbiosis, the study confirmed the high species-specificity and showed that despite this specificity, host and symbiont phylogenies showed only weak congruence, indicating that host-switches are frequent in gutless oligochaetes over long evolutionary timeframes (see Figure 11 in section 1.5.3). Co-divergence patterns where only found in gutless oligochaete sister species or sub-species that were relatively young, indicating that different patterns of co-evolution and host-switching exist on smaller evolutionary timescales. Also, this study was exclusively concerned with the Gamma1/*Thiosymbion* clade, since only this clade of symbionts is shared between the two host groups.

A major aim of this thesis was to examine the evolutionary patterns of symbionts within a single gutless oligochaetes on a short-term, i.e. population genomics scale, in order to better understand i) the population structure of the host and the within-worm population structure and diversity of the symbionts, ii) the presence of co-evolutionary tendencies or the lack thereof, depending on symbiont species, and iii) the implications of the observed patterns for symbiont transmission fidelity and host-symbiont (co-)evolution, (co-diversification) and speciation. The research concerning this aim is in the process of being published and is therefore presented in the form of a self-contained manuscript in chapter 2.

### 1.6.2 Characterization of the metabolic capabilities of the spirochaete symbiont

Metagenomic Sanger sequencing of *Olavius algarvensis* fosmid clones by [59] in 2006 lead to the assembly of draft genome sequences for four out of the five symbionts that were known at the time. No genome sequence could be obtained of the spirochaete symbiont, leaving its metabolic potential and putative role within the symbiosis completely unknown. The metagenomic data generated in this thesis allowed the assembly of an essentially complete genome draft of the spirochaetal symbiont. I functionally analyzed its genome to gain insight on the metabolic potential and function of this symbiont in the symbiosis. This research is in the process of being published and therefore presented in the form of a manuscript in chapter 3.

### 1.6.3 Molecular mechanisms that enable the *O. algarvensis* symbiosis

The metagenomic analyses carried out in this thesis revealed that *O. algarvensis* not only harbors an even higher diversity of secondary symbionts as previously thought, but that the symbiont community is also highly variable between *O. algarvensis* individuals. The secondary symbionts which were flexibly associated and showed no co-diversification pattern with *O. algarvensis* are likely horizontally transmitted, at least occasionally. Still, a high specificity is maintained and no strain variability could be identified on the 16S level within each symbiont group. This raises the question which molecular mechanisms are responsible for selecting and taking up these symbionts, and how the host avoids wrongful uptake of closely related environmental strains or closely-related symbiont phylotypes of other, co-occurring gutless oligochaete species. Previous to this thesis, no information was available on the molecular mechanisms that might enable the host to initiate, establish and maintain these highly specific, yet flexible associations. Therefore, another aim of this thesis was to identify and characterize

genes of the host immune system that might play a role in these processes, using metatranscriptomics and metaproteomics.

In addition, previous studies on the metabolism of the symbionts established that the deltaproteobacterial symbionts produce endogenous sulfide [295] and that the deltaproteobacterial symbionts and *OalgG3* use environmental carbon monoxide as an energy source. Both gases are toxic to animals [351, 352, 350]. Further, the host must endure frequent periods of anoxia to accommodate the anaerobic metabolism of its secondary symbionts. Another aim was therefore to analyze the transcriptomes and proteomes with respect to the physiological adaptations that allow the host to life in association with these symbionts. The research of this aim has been submitted for publication and is therefore presented in this thesis in the form of a manuscript in chapter 4.

# Chapter 2: Metagenomic sequencing reveals host-linked divergent evolution and potential for niche differentiation in chemosynthetic symbionts of a gutless marine worm

**Authors:** Juliane Wippler[1], Cecilia Wentrup[2], Amandine Nunes-Jorge[1], Miriam Sadowski[1], Harald Gruber-Vodicka[1], Tanja Woyke[4], Nicole Dubilier[1], Manuel Kleiner[3]

[1]Max Planck Institute for Marine Microbiology, Celsiusstr. 1, D-28359 Bremen, Germany

[2]University of Vienna, Department of Microbiology and Ecosystem Science, Althanstr. 14, A-1090 Vienna, Austria

[3]University of Calgary, Department of Geoscience, Calgary T2N 1N4, AB, Canada

[4]DOE Joint Genome Institute, Walnut Creek, California 94598, USA

**Publication status:** Manuscript in preparation

**Author's contributions:**

**JW**: Conceived the project and manuscript with significant conceptual contributions, provided ideas and wrote the manuscript, collected most worms, extracted DNA, did PCRs and COI sequencing of those worms, did all metagenomic DNA extractions, metagenomic data processing, genome assembly, binning, annotation and functional analyses, prepared all figures and tables. **CW:** Conceived the project and manuscript, provided ideas and conceptional input**, ANJ:** Collected worms, and performed DNA extractions, PCR and COI sequencing, **MS:** Performed DNA extractions, PCR and COI sequencing of some of the worms, **HGV:** Provided ideas and conceptual input, **MK:** Conceived the project and manuscript, edited the manuscript, was involved in logistics and project management for all JGI sequenced metagenomes, provided significant conceptual input and ideas, supervised the project**, TW:** Help with project organisation, project and sample logistics on JGI side, coordination of library preparation & metagenomic sequencing at the JGI (6/22 worms), **ND:** Conceived the project and manuscript, provided ideas and was involved in project organization and coordination.

# Titlepage

**Metagenomic sequencing reveals host-linked divergent evolution and potential for niche differentiation in symbionts of marine chemosynthetic worm**

Juliane Wippler[1], Cecilia Wentrup[4], Tanja Woyke[2], Nicole Dubilier[1], Manuel Kleiner[1,3]

[1]Max Planck Institute for Marine Microbiology, Celsiusstr. 1, D-28359 Bremen, Germany

[2]DOE Joint Genome Institute, Walnut Creek, California 94598, USA

[3]University of Calgary, Department of Geoscience, Calgary T2N 1N4, AB, Canada

[4]University of Vienna, Department of Microbiology and Ecosystem Science, Althanstr. 14, A-1090 Vienna, Austria

Corresponding authors:

Juliane Wippler: jwippler@mpi-bremen.de, phone: +49 (421) 2028-907, fax: +49 (421) 2028-580

Manuel Kleiner: manuel.kleiner@ucalgary.ca, phone: +1 (403) 210-6617

## Abstract

Gutless oligochaetes are a monophyletic group of small marine annelids comprising more than a hundred described species. They lack digestive and excretory organs and instead, harbor a species-specific consortium of chemoautotrophic primary and secondary symbionts which provide nutrition and recycle waste products. Often, several morphologically uniform species co-occur, thriving in the same sediment interstitial habitat. What drives the diversification of this group and what allows many of the species to share what appears to be the same ecological niche is unknown. In this study, we extensively characterized the symbiotic community in the model species *Olavius algarvensis* using PCR screening and high throughput metagenomic and -transcriptomic sequencing to investigate how its obligate symbiont community might influence gutless oligochaete evolution and *vice versa*.

We found that the community of secondary symbionts in *O. algarvensis* is highly diverse, flexible and shows little to no congruence with host mitochondrial genome evolution, suggesting frequent host switching events and metabolic versatility. In contrast, the primary symbiont has clearly co-diverged together with its host into two main haplogroups/phylogroups, suggesting strict vertical transmission through the maternal line. Divergent and reductive genome evolution in the primary symbiont is apparent in lineage-specific loss of multiple functional genes involved in carbon and energy metabolism as well as molecular interaction with the host. These observations demonstrate how symbiont genome evolution is influenced by host association and different modes of transmission, and how, in turn, they can shape and differentiate the ecological niche of their host, and, on a larger evolutionary scale, could provide the genetic foundation for host diversification, speciation and species co-existence.

## Introduction

The majority of animals form intimate and beneficial associations with microbial symbionts that significantly contribute to their health, nutrition, and their development [1, 2]. Many of these associations are stable over time and are transmitted between host generations with confidence [3], which is often reflected in strong congruence between host and symbiont marker gene phylogenies [4]. Two of the many examples of high host specificity and transmission fidelity include the chemosynthetic ectosymbionts of stilbonematine nematodes [5], which are acquired anew in each host generation and each host molting event, and the obligate intracellular endosymbionts of aphids, the latter almost perfectly mirroring the phylogenetic history of their hosts [6].

As animals and their associated microbes often share a long history together, they have profoundly shaped each other's biology over the course of evolution. For one, animals had to evolve molecular mechanisms to specifically recognize and interact with their symbionts and to reliably transmit them from parent to offspring [7, 3]. In addition, they often underwent substantial morphological, physiological and/or behavioral changes in adaptation to the symbiotic lifestyle, e.g. developing tissues to specifically house the symbionts [8, 9, 10], modifying the biochemical properties of proteins to accommodate symbiont metabolism [11], or developing certain behaviors that ensure symbiont transmission to the offspring [12].

Likewise, the evolution of symbionts is often heavily influenced by selective forces imposed by the host animal as well [13, 14]. For example, symbiont genome evolution is greatly influenced by the fidelity with which symbionts are transmitted from one host generation to the next, and whether or not genetic recombination is possible between subpopulations of symbionts that live either inside other host animals or in the environment. Symbionts with a free-living stage usually have relatively large genomes that allow them to thrive inside the host as well as in the environment, while strictly host associated bacteria tend to have much smaller genomes than their free-living relatives [13, 3]. It has been postulated that as symbionts shift their lifestyle from free-living to strictly host-associated, their genomes undergo a phase of rapid genome deterioration due to reduced purifying selection in

the host environment [15]. The process is characterized by a significant increase in transposable elements, the formation of pseudogenes, chromosomal rearrangements and small- to large-scale deletions in the early stages of strict host association. Eventually, symbionts which live exclusively within a host and do not experience genetic recombination with other populations develop highly reduced, AT-rich genomes without transposable elements [16]. Examples of this include the highly reduced genomes of obligate intracellular insect symbionts [17] and the chemosynthetic endosymbionts of clams [18].

While it is clear that animals and their symbionts have fundamentally influenced each other's evolution, there is considerable debate on how much symbionts contribute to the diversification and speciation of their animal hosts [19, 20]. Speciation requires, in essence, that barriers to gene flow are established, which prevent or at least significantly reduce successful interbreeding between populations [21]. Such barriers can be created through divergent ecological or sexual selection, which render hybrids unfit to their environment or unattractive to putative mates, or through genetic incompatibilities resulting from genetic drift or genomic conflict that leave hybrids unviable independent of environmental or sexual interactions [22]. Since beneficial symbionts significantly influence the ecology, health, development and behavior of most animals [1, 2] they could potentially play an important role in host speciation as well, by influencing any of the mechanisms that cause reproductive isolation.

Evidence that microbial symbionts influence host speciation in many different animal groups is slowly accumulating. Some bacterial symbionts of arthropods, like *Wolbachia*, *Cardinium* and *Spiroplasma*, significantly reduce gene flow between infected and uninfected host populations by interfering with host reproduction in various ways, and are thus heavily implicated in the speciation of many species [23, 24, 25]. In insects and vertebrates, it was shown that bacterial symbionts might also promote host speciation by influencing mate choice behavior [26, 27] or by highly reducing the fitness of hybrids due to immune conflicts [26, 28]. Many symbionts confer important phenotypic traits that allow a host to exploit resources that would otherwise be inaccessible, e.g. by synthesizing essential

nutrients lacking from the host's diet [29] or allowing hosts to better endure environmental stressors [30]. Symbiont-facilitated adaptation to a new habitat or resource may therefore result in ecological speciation [31, 19].

Gutless oligochaetes (Annelida; Clitellata; Phallodrilinae) are a highly diverse group of marine annelid worms that live in obligate nutritional association with multiple bacterial symbiont species [32]. The influence of the symbionts on the evolution of their hosts is apparent, as these worms completely lack a digestive and excretory system, and instead rely on their symbionts for nutrition and waste removal [33, 34]. Each host species carries a stable, species-specific set of symbiont phylotypes that are closely related to, but not shared with any other species [32, 5]. The gutless oligochaetes are very speciose, but morphologically uniform. Several superficially identical species often co-occur at the same site [35], raising the question of how such diversity has evolved and how it is maintained within the same habitat. It is tempting to speculate that it is the symbionts that play a critical role in metabolic resource partitioning and diversification of this oligochaete group.

To better understand the evolutionary relationships between a gutless oligochaete host and its symbionts and to examine how the symbiotic community might influence gutless oligochaete diversification, we chose the model species *Olavius algarvensis* for our investigations [36, 33, 34]. This Mediterranean species harbors a bacterial consortium of two gammaproteobacterial, two deltaproteobacterial and one spirochaetal symbiont. We used host mitochondrial cytochrome c oxidase (COI) and symbiont 16S rRNA genotyping to extensively characterize the host and primary symbiont population at one collection site on Elba where they abundantly occur. Furthermore, we sequenced the metagenomes of 22 individual *O. algarvensis* worms collected from this and one other site to study the recent evolutionary history of this species with its symbionts, to gain insights into the transmission fidelity of the symbionts, and to trace divergent evolutionary tendencies within the symbiont population that might promote host niche differentiation and diversification.

## Materials and Methods

### Specimen collection

Sediment containing *O. algarvensis* worms was collected at 7 meters water depth from the Bay of Sant' Andrea, Elba, Italy (42°48'26"N, 010°08'28"E), in November 2010, November 2011, March and June 2012, June 2013, August 2014, and August 2015. The worms were extracted from the sediment by decantation with seawater, rinsed in sterile-filtrated seawater (2 µm pore size, Millipore, Darmstadt, Germany), briefly tapped on blotting paper to remove excess liquid, and either flash-frozen in liquid nitrogen, or fixed in RNAlater Stabilization Solution (Ambion, Thermo Fisher Scientific, Waltham, USA) at 4 °C overnight, then kept frozen at -80 °C until further use. Live worms for the incubation experiment were collected from Sant' Andrea sediment in the same manner, and were kept in washed Elba sediment for up to two weeks until further use.

### DNA extraction

Two gutless oligochaete species, *O. algarvensis* and *O. ilvae*, co-occur in Sant' Andrea. They are morphologically indistinguishable unless reproductively active. To determine species affiliation of individual worm specimens and for genotyping the target species *O. algarvensis*, worms used for metagenomic and metatranscriptomic sequencing were PCR screened prior to whole worm DNA/RNA extractions as follows. A small piece was removed with a sterile scalpel from the anterior tip of each RNAlater-fixed specimen. DNA for PCR screening was obtained from each piece through homogenization with a sterile disposable plastic pistil (VWR International, Darmstadt, Germany) in a 1.5 ml reaction tube and heating to 70 °C for 10 minutes and then used directly as template for amplification.

DNA for metagenomic sequencing and COI/16S rRNA gene sequencing was extracted from individual whole worms using the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany). Frozen worms were thawed in 180 µl buffer ATL with Proteinase K @ 55 °C for 10 minutes, and then incubated for up to six days at 37°C until the worms were completely dissolved, to maximize DNA yield. Subsequent

extraction steps were carried out according to the manufacturer's instructions. DNA for metagenomic sequencing of worms used in the incubation experiment was co-extracted together with RNA using Qiagen's Allprep RNA/DNA kit (see below).

**COI/16S rRNA gene amplification, sequencing and analysis**

Host mitochondrial cytochrome oxidase I (COI) sequences were amplified from DNA extracts using general primers COI-1490F (5'-GGT-CAA-CAA-ATC-ATA-AAG-ATA-TTG-G-3', [37]) and COI-2189R (5'-TAA-ACT-TCA-GGG-TGA-CCA-AAA-AAT-CA-3', [37]). *OalgG1* symbiont 16S rRNA gene sequences were amplified using primers OalgG1_644F (5'-TGT-CCG-GCT-AGA-GTG-TGG-TA-3', which specifically targets *OalgG1*) and GM4R (3'-TAC-CTT-GTT-ACG-ACT-T-5', [38]). Target DNA was amplified using Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific, Braunschweig, Germany). 1 µl of DNA extract was used as template for the PCR reaction. The following thermocycler conditions were used: 95°C for 5 min, followed by 35 cycles of 95°C for 1 min, 55 °C (COI) or 58°C (16S) for 1 min, 72°C for 1 min, 50 sec and an extension at 72°C for 10 min. PCR products were purified on a Sephadex G-50 Superfine column (Amersham Pharmacia Biotech, Freiburg, Germany), then directly sequenced using the BigDye Sanger sequencing kit (Life Technologies, Darmstadt, Germany) on an Applied Biosystems Hitachi capillary sequencer (Applied Biosystems, Waltham, USA). Resulting sequences were automatically filtered and quality end-trimmed with Sequencher 4.6 (Gene Codes Corp., Ann Arbor, USA) software and aligned with ClustalW [39]. Aligned sequences were manually inspected in BioEdit version 7.2.5 [40] to determine host species, host haplotypes and *OalgG1* symbiont phylotypes.

**Illumina metagenomic sequencing**

22 *O. algarvensis* specimens were selected for metagenomic sequencing based on their genotype. Illumina TruSeq compatible libraries were constructed with Illumina TruSeq library preparation kit according to the manufacturer's instructions and fragmented to DNA inserts 300-500 bp length. Libraries were paired-end sequenced (2x 100 bp) on an Illumina HiSeq2500 (Supplementary Table 1).

**Bioinformatic analysis**

Reads were quality filtered and adapter-trimmed using nesoni version 0.114 [41] and corrected for sequencing errors using BayesHammer [42], as implemented in Spades version 2.5.1 [43]. Metagenomic reads were assembled de novo using idba_ud version 1.1.1 [44]. Symbiont genomes were binned using a combined approach of differential coverage binning as described in [45], Metawatt version 1.7 [46], and targeted reassembly. Completeness estimates of binned symbiont drafts were performed with CheckM version 1.0.1 [47]. Symbiont genome draft assembly metrics were determined with QUAST version 2.3 [48]. Reads were mapped to reference sequences with bowtie2 version 2.1.0 [49]. 16S rRNA gene sequences were assembled and relative symbiont abundances estimated with EMIRGE version 0.6 [50, 51]. Genome bins were automatically annotated in RAST version 2.0 [52, 53]. Sequence similarity searches were performed with BLAST 2.2.28+ [54]. SNP (single nucleotide polymorphism) analysis was performed with wombac version 1.2 (http://www.vicbioinformatics.com/software. wombac.shtml) using the draft host mitochondrial and symbiont genomes obtained in this study. Neighbor Joining distance transformations and bootstrap replications were calculated with SplitsTree4 version 4.13.1, [55]. We used the binned draft genomes as mapping references to confirm that particular symbionts are really missing from single worms. Mappings were manually inspected in Tablet version 1.15.09.1 [56] to detect the presence of low-abundance symbionts. Orthologous *OalgG1* proteins used for gene set comparisons of symbionts from different host haplotypes were identified using ProteinOrtho version 5.11 [57]. Proteins that were predicted to be restricted to the symbionts of a particular host haplotype were confirmed to be

unique by cross-mapping the metagenomic reads of all specimens onto the gene sequences of the respective proteins. Genes were only considered unique to a particular haplotype they did not recover reads from metagenomes of the other host haplotype. Average amino acid identity (AAI) between closely related symbiont species was determined with the AAI calculator tool implemented at http://enve-omics.ce.gatech.edu/aai/ [58], using the protein coding sequences obtained in the respective RAST annotations of each symbiont genome bin. Average nucleotide identities (ANIs) between symbiont genomes were calculated using the respective symbiont draft genomes with the ANI calculator implemented at http://enve-omics.ce.gatech.edu/ani/ [58].

## Results

### *The O. algarvensis population on Elba consists of three host haplotypes with specific primary symbiont phylotypes OalgG1*

*Olavius algarvensis* is a Mediterranean species of gutless oligochaete which is found abundantly in the Bay of Sant' Andrea (North) and in lower abundance in the Bay of Cavoli (South) of the island of Elba, Italy (Figure 1, [59]). This species of gutless oligochaete harbors five symbiont phylotypes – two Gammaproteobacteria, two Deltaproteobacteria, and one Spirochaeta (Table 1), whose presence has been previously confirmed by fluorescence *in situ* hybridizations (FISH) [59].

In order to characterize the population structure of *O. algarvensis* in Sant' Andrea, we sequenced the mitochondrial cytochrome c oxidase I (COI) gene of 380 *O. algarvensis* individuals collected from the Bay of Sant' Andrea throughout the years 2010 to 2015. We found two COI haplotypes, which differed by six consistent SNPs (single nucleotide polymorphisms). We designated them haplotype "A" and haplotype "B" (Figure 2A, see Supplementary File 1 for full multiple sequence alignment of all 380 sequences). Haplotype A always dominated the population in numbers, being approximately three times more abundant than haplotype B (295x type A, 85x type B).

To investigate if the primary *OalgG1* symbiont shows any sequence variation that reflects the host's population structure, we sequenced the 16S rRNA gene (16S) from 60 worms. This revealed that each of the two haplotypes is exclusively associated with a specific *OalgG1* phylotype (Figure 2A-B, Supplementary Figure 1). We named the phylotype specific to host haplotype A "*OalgG1-A"* and the phylotype specific to haplotype B "*OalgG1-B"*. The *OalgG1-A* and *OalgG1-B* phylotypes differed only by a single base transition (A<->G) across the 1497 bp long 16S sequence. The same two (and no additional) *OalgG*1 phylotypes were also detected in a 16S clone library generated in 2005 (Supplementary File 2 [33], 158 sequenced *OalgG1* 16S clones), which was prepared from a batch of 600 worms, providing further evidence that only two *OalgG1* phylotypes occur in this worm population.

***Metagenomic sequencing reveals variability in the secondary symbionts***

In order to investigate the 16S gene diversity and relative abundance of the secondary symbionts and how they might be linked to host haplotypes we selected eighteen *O. algarvensis* specimens collected from Sant' Andrea in the North of Elba (haplotype A: A1-A9, haplotype B: B1-B9) for high-throughput metagenomic sequencing. We also sequenced four specimens of *O. algarvensis* obtained from the Bay of Cavoli in the South of Elba, of which three were assigned to haplotype A (A10 - A12) and one was assigned to the new haplotype C (C1). This new phylotype differed at four single base positions in the COI sequence compared to haplotype B from Sant' Andrea (Figure 2A), and showed substantial sequence divergence in the mitochondrial genome sequence (Figure 4). Supplementary Table 1 summarizes the number of raw and processed Illumina reads generated from each metagenomic library.

In order to assess the diversity and symbiont phylotype frequency associated with each host specimen, we reconstructed full-length 16S rRNA gene sequences and estimated relative symbiont abundance in each metagenome using EMIRGE [50] (Figure 3). Very low-abundant symbionts that

could not be detected by EMIRGE were independently identified by mapping the metagenomic reads back onto the individual symbiont draft genomes, which were also obtained in this study (see below).

We recovered 16S sequences of both primary SOX symbiont phylotypes, *OalgG1-A* and *OalgG1-B*, and they were exclusively associated with either host haplotypes A or B, as expected (Figure 3). The *OalgG1* 16S sequences of Cavoli haplotype A worms were identical to the 16S sequence of *OalgG1-A* from Sant' Andrea. However, the 16S sequence from the Cavoli haplotype C worm differed by four single base substitutions from the 16S of *OalgG1-B* and by three substitutions from *OalgG1-A* from Sant' Andrea (thus designated *OalgG1-C*). This 16S divergence reflects the sequence divergence pattern observed in the host COI sequences (Figure 2A-B). In contrast, the 16S sequences recovered from the secondary SOX symbiont, *OalgG3*, were identical in all specimens (Supplementary file 3). Surprisingly, no *OalgG3* sequences were recovered from specimen A9 (Figure 3).

We observed a lot of variability in the presence-absence pattern and individual abundances of different deltaproteobacterial symbionts across worm specimens. In eight out of eighteen metagenomes we detected a deltaproteobacterial phylotype, not previously recognized as a symbiont (Figure 2C, Figure 3), which was closely related to the known *O. algarvensis* symbiont *OalgD1* (Table 1). The same phylotype had been found in a large clone library containing a mix of host species, and was at the time dismissed as contamination [60]. Although the 16S sequence identity between this new phylotype and *OalgD1* was high (98.79%), the genomes were considerably divergent with 81.92 – 82.10% average nucleotide identity (two-way ANI) and 77.88 – 78.63% average amino acid identity (two-way AAI) between their protein coding sequences, justifying the separation of *OalgD1* into two distinct symbiont species. Due to its close relatedness to *OalgD1*, we designated this novel phylotype *OalgD1b*, and renamed *OalgD1* to *OalgD1a*. Two worm specimens (A10 from Cavoli and B2 from Sant' Andrea) each contained another deltaproteobacterial phylotype, which was previously found in a clone library, but also dismissed as a contaminating sequence. This phylotype was highly similar to *OilvD3* (98.22% 16S sequence identity), a deltaproteobacterial

symbiont of the co-occurring gutless oligochaete species *Olavius ilvae* [59]. The two *OalgD3* 16S sequences differed in two single nucleotide positions from each other, indicating that worms from Cavoli might harbor a different strain than worms from Sant' Andrea. However, with only one worm carrying this symbiont per site, this remains to be tested. These phylotypes, which we both named *OalgD3*, contributed significantly (12-17%) to the symbiont community in these two worms (Figure 3). All 16S sequences recovered for the *OalgD4* symbiont were identical (Supplementary file 4).

We found no haplotype restriction of any of the deltaproteobacterial 16S variants (apart from possibly *OalgD3*), and thus no haplotype-related 16S pattern. However, we did observe a haplotype-biased occurrence and abundance pattern in the *OalgD1a* and *OalgD1b* symbionts. *OalgD1a* occurred more frequently and with significantly higher abundance in haplotype A worms, while *OalgD1b* was significantly more frequent and abundant in haplotype B worms (Figure 3)(t-test *OalgD1*: P=0.045688; t-test *OalgD1b*: 0.015941, null hypothesis: symbiont abundance is equal in both host haplotypes). Moreover, in worm specimens that harbored both symbionts, either *OalgD1a* or *OalgD1b* was dominating in abundance, while the other symbiont was heavily marginalized (Figure 3). This suggests that growth of *OalgD1b* is hampered in haplotype A worms in the presence of *OalgD1a*, while growth of *Delta1a* is hampered in haplotype B worms in the presence of *OalgD1b*, and that these two symbionts cannot both grow successfully within the same worm. When only either *OalgD1a* or *OalgD1b* were present in the same worm, they reached normal abundance levels, independent of the host haplotype they were associated with (Figure 3, see worm A7, B2, B8).

Individual worms showed high variability with respect to which combination of deltaproteobacterial phylotypes they harbored (Figure 3). While most worms contained either *OalgD1a*, or *OalgD1b*, or both, worm B1 had no *OalgD1*-like symbiont at all. Worms B6 and B9 did not contain an *OalgD4* symbiont, and only two worms had an *OalgD3* symbiont. Overall, it seems that while the presence of at least one deltaproteobacterial symbiont type is required, none of the different types are *per se* essential in the symbiosis.

All deltaproteobacterial symbiont species were found in at least one worm from each collection site, indicating that, in principle, there is no geographical influence on symbiont sets in northern (Sant' Andrea) compared to southern (Cavoli) populations of *O. algarvensis* (an exception might be *OalgD3*, as mentioned above). All symbiont species also occurred in at least one specimen of each host haplotype, indicating that, with the exception of *OalgD1a*/*OalgD1b*, there is no conflict or incompatibility between haplotypes and symbiont species.

We recovered 10 phylotypes of the spirochaetal symbiont *OalgS1* that differed at the same exact four single base positions from each other (Figure 2D). *OalgS1* phylotypes that were found in multiple specimens were not restricted to a particular host haplotype or location. The average nucleotide sequence identity (two-way ANI) between the draft genomes of these phylotypes was high (99.74 – 99.85%), and therefore all spirochaete phylotypes recovered in this study will remain under the designated name *OalgS1*.

In summary, we observed no 16S sequence variability in *OalgG3*, *OalgD1a*, *OalgD1b*, and *OalgD4* symbionts from different worm specimens. Instead, we found high variability in the presence-absence pattern of these symbionts (in the case of *OalgD1a*/*OalgD1b* linked to host haplotype). We observed slight 16S sequence variation in *OalgG1* that was clearly linked to host haplotype, but not geographic location, and some sequence variation in *OalgD3* that may or may not be linked to host haplotype or geographic location. The *OalgS1* symbiont showed the highest 16S sequence variation, but without evidence for haplotype or geographic linkage.

***Phylogenetic SNP trees reveal that both sulfur-oxidizing symbionts are host-linked***

From the metagenomic data, we assembled and binned the host mitochondrial genome and draft genomes for each symbiont phylotype. Supplementary Table 2 provides information on the assembly quality and completeness of each symbiont draft genome used as reference in downstream analyses. Most genomes were assembled to at least 90% completeness.

In order to elucidate the recent evolutionary history of the symbionts and their host, and to trace possible linkage disequilibria between host mitochondrial genome and symbiont genomes, we used genomic sequence divergence in the form of SNPs (single nucleotide polymorphisms). We assembled and binned symbiont draft genomes from several metagenomes, and used the draft genomes with the highest completeness and best assembly metrics as references for read mapping, SNP calling and construction of phylogenetic trees from high quality core SNPs.

Based on the phylogenetic SNP tree, the mitochondrial host genomes of Sant' Andrea separate well into haplogroups corresponding to the COI-derived haplotypes A and B (Figure 4A). Two Cavoli haplotype A specimens formed a well-supported, but not much divergent, sub-group separate from the other haplotype A specimens. The placement of the third haplotype A specimen from Cavoli within haplogroup A is not resolved due to low support of its basal node. Overall, haplogroup A showed much higher in-group SNP divergence compared to haplogroup B. Within haplogroup B, mitochondrial sequences were often so similar that they failed to produce individual branches in the tree.

The phylogenetic SNP tree constructed from *OalgG1* SNP data mirrors the host mitochondrial phylogeny (Figure 4B). As predicted from *OalgG1* 16S sequences, *OalgG1* separated into three distinct phylogroups (A, B, and C) according to the haplotypes of their respective hosts. In contrast to the host, *OalgG1-A* from Cavoli formed a subgroup within phylogroup A that was well-supported (bootstrap value >99%). While the mitochondrial haplogroup A showed higher SNP divergence than

haplogroup B, *OalgG1* showed the opposite pattern, with phylogroup B being more divergent than phylogroup A.

While all 16S rRNA sequences of the *OalgG3* symbiont were identical, the SNP based phylogeny of the *OalgG3* symbiont revealed that it separated into distinct clades, which partially mirrored the host mitochondrial phylogeny (Figure 4C). All *OalgG3* from Sant' Andrea haplotype B worms fell into a well-defined clade that was distinct from all other *OalgG3*. The other *Oalg3* fell into several clades without a clear pattern emerging (Figure 4C). Interestingly, the *OalgG3* from the Cavoli haplotype C worm fell into a well-supported clade with two of the Cavoli haplotype A *OalgG3*, possibly showing a recent haplotype switch of *OalgG3*.

The SNP based phylogenetic trees of the deltaproteobacterial and spirochaete symbionts were incongruent with host mitochondrial phylogeny (Figure 4D to G). Together with the presence-absence pattern described above, this indicates that these symbionts are not strictly linked to a particular host haplotype.

***The genomes of the primary symbiont phylotypes OalgG1-A and OalgG1-B differ in gene content***

Since the primary *OalgG1* symbiont showed clear phylogenetic separation according to host haplotype on the SNP level, we analyzed the gene content of each Sant' Andrea *OalgG1* genome to investigate if these two phylotypes show functional divergence, as well. For this, we first assembled, binned and annotated the genomes of *OalgG1* from six metagenomes (A1-A3, B1-B3), and identified orthologous genes that were present in all *OalgG1* genomes of one phylotype, but absent in all genomes of the other *OalgG1* phylotype. To verify that the absence of genes in one phylotype was not simply due to incompletely assembled or binned genomes, and to include all available metagenomes in the analysis, the presence or absence of a particular gene was confirmed by cross-mapping the reads from each metagenome back onto each gene (summarized in Table 2, see Supplementary File 6 for individual mapping results).

We identified nine genes that were unique to *OalgG1-A*. Most of these genes coded for hypothetical proteins of unknown function (Table 2). Unique to *OalgG1-B* were 40 genes, which included genes involved in DNA modification (methylation and recombination) and energy metabolism (hydrogen metabolism, fumarate/nitrate reduction), acetate uptake and two proteins that may play a role in adhesion and attachment to host surfaces.

## Discussion

### *O. algarvensis* host haplotype diversity, phylogeny and geographic distribution

We found in total three host haplotypes, of which two appeared to be restricted to either one of the two locations were worms had been sampled, although more specimens from Cavoli need to be examined for a conclusive statement. In both localities, haplotype A seems to be the most abundant one, suggesting that this haplotype has a higher fitness compared to the other two. The higher abundance of haplotype A in Sant' Andrea has been observed for more than 10 years without exceptions, indicating that this haplotype frequency distribution is relatively stable and not due to a sampling artifact.

On the whole mitochondrial genomes, all haplotypes were well separated into distinct sequence groups. Haplotype A also showed more mitochondrial sequence divergence compared to haplotype B, suggesting that this haplotype has had the opportunity to accumulate more genetic changes over time either because it diverged before haplotype B, or because is experiences less purifying selection. The mitochondrial sequences were still too conserved to allow distinct separation of haplotype A sequences from Sant' Andrea and Cavoli. However, since we observe a clear linkage disequilibrium pattern with the primary SOX symbiont *OalgG1*, and since the split between *OalgG1-A* from Sant' Andrea and *OalgG1-A* from Cavoli is well supported, one could speculate that haplotype A actually also partitions into a Cavoli clade and a Sant' Andrea clade, and that migration of hosts between the north and south of Elba is very low. This is in line with the fact that gutless oligochaetes do not form

planktonic dispersal or larval stages, but lay single eggs that are tightly attached to sand grains with a sticky, mucous substance from which fully developed juvenile worms are hatched [61, 62].

***Linkage disequilibrium in two symbionts suggests vertical transmission***

Since the host does not have a digestive tract or excretory organs (nephridia), it is entirely dependent on its species-specific symbionts for survival. The transmission of symbionts from one host generation to the next is therefore of paramount importance for the continued existence of the species. The results of this study show that each host haplotype is exclusively associated with its own *OalgG1* 16S phylotype, and that switches between host haplotypes have not been observed. This is consolidated by the phylogenetic SNP trees which show perfect congruence between symbiont and mitochondrial genome phylogeny. One explanation for the observed pattern is that *OalgG1* is strictly vertically transmitted from one host generation to the next via the maternal line, i.e. together with the mitochondria. This is further supported by anatomical features of the genital region which ensure that the egg comes into contact with the parent's symbionts during egg laying [61, 62]. All gutless oligochaetes possess a pouch-like bulge of the cuticle, in an area close to the oviporus, called the genital pad. The genital pad is densely packed with symbiont cells and ruptures during egg laying, as the egg is squeezed through the very narrow oviduct and oviporus, releasing symbionts into the environment. As a result, the freshly lain egg comes into direct contact with parental symbionts, which can subsequently invade the egg. Moreover, in a recent study Schimak et al. were able to trace the transmission of symbionts labeled with $^{15}$N-ammonium from parent worm to offspring, suggesting vertical transmission [62]. Exceptionally high numbers of transposase genes were found in the symbionts draft genomes of the gutless oligochaete *Olavius algarvensis* [33] and many were abundantly expressed [63], suggesting that these symbiont might have recently become strictly host-associated [15], and lending further evidence that the symbionts are strictly vertically transmitted.

The other explanation for the observed pattern would be extremely specific uptake of the correct strain from the environment. In theory this would be possible, because, unlike other oligochaetes,

the gutless oligochaetes do not produce egg cocoons, but lay their eggs directly into the sediment. The freshly lain eggs are initially surrounded by a soft, sticky mucus layer which only hardens into a more rigid egg wall later on. Freshly lain eggs could therefore be infected by environmental bacteria as long as they are able to traverse the mucus layer and soft egg integument and are not eliminated by the innate immune system of the developing embryo. However, no *OalgG1*-like sequences have even been found in the environment, despite extensive sequencing efforts by us and others ([64, 65, 66], and Wippler, unpublished results). Environmental *G1*-like sequences have been reported previously [67]; however their origin is not clear and could be contamination from host-associated *G1* sequences. On the other hand, *O. algarvensis* reproduces seasonally, with a large proportion of worms laying eggs at the same time. Symbionts cells released during egg laying may remain infectious for a certain period of time and infect eggs from different parent worms, constituting a mixed transmission mode.

***Linkage disequilibrium in OalgG3***

We observed a similar pattern of host linkage in the genomes of *OalgG3* that was less prominent than in *OalgG1*. *OalgG3* genomes from Sant' Andrea appeared to group according to host haplotype, although the cluster was much less defined and showed much more divergence compared to *OalgG1*. However, three of the four *OalgG3* symbionts from Cavoli formed a cluster independent of host haplotype (the fourth one could not be placed in the tree with confidence; therefore its phylogenetic position remains unresolved). This might be explained by a recent host switch. Interestingly, *OalgG3* was missing from one worm, which further suggests that vertical transmission may not be perfect and that symbionts may be lost due to bottlenecks in the transmission process.

***Co-diversification with host on a local, but not on a larger evolutionary scale***

It is tempting to assume that, on a larger evolutionary scale, co-diversification of the *OalgG1* symbionts and its host should lead to co-speciation. At this point, we do not have sufficient data to assess how far the haplotypes have diverged on the level of the nuclear genome (i.e. if and how far

they have come along the path of speciation). A recent study found only two supported instances of co-speciation between a gutless oligochaete and its *G1* symbiont [5], while all other gutless oligochaete species had *G1* symbionts that were more closely related to *G1* symbionts from more distantly related hosts. Overall, geography seemed to have a higher influence on *G1* phylogeny than the phylogeny of the host, i.e. *G1* from distantly related, but syntopic host species were more closely related to each other than to *G1* symbionts from more closely related hosts. However, the phylogenetic trees presented in the same study show a co-divergence pattern between very closely related, cryptic host species and their primary *G1* symbiont in two gutless oligochaete species from the Caribbean (*Inanidrilus leukodermatus* and *Olavius imperfectus*). Taken together, this suggests that co-speciation could happen on a local scale, but that ultimately patterns of co-diversification and co-speciation are broken up by migration to new sites

**SNP diversity in *OalgG1* – phylotype A vs B**

*OalgG1-B* of Sant' Andrea shows more diversity on the SNP level than *OalgG1-B*. This could be an indication that the *OalgG1-B* symbiont has had more time to accumulate SNPs than the A type, suggesting that its origin lays further back in time. The fact that *OalgG1-A* symbionts show relatively little SNP diversity could be an indication that they experience more purifying selection than *OalgG1-B*.

***Novel deltaproteobacterial symbiont phylotypes***

In this study, the number of identified symbiont phylotypes is considerably higher than previously reported. We found two phylotypes, *OalgD1b* and *OalgD3*, which were previously detected in clone library, but were dismissed as contamination, rather than than seen as actual symbionts [60]. The new phylotype *OalgD1b* was never identified in previous FISH studies, because the FISH probe used targets a 16S rRNA region that is identical in both phylotypes, making differentiation impossible ([59], Figure 2C). The fact that these sequences had been found previously indicates that these are real symbionts that simply occur in only part of the *O. algarvensis* population. In addition, as near

complete draft genomes have been obtained for these symbionts in this study, future genome analysis and comparative genomics with the other deltaproteobacterial symbionts of *O. algarvensis* is now possible.

***Mix-and-match assembly of a non-neutral symbiont community phenotype suggests specific function requirements and functional redundancy***

The variability in secondary symbiont presence and abundance suggests a) functional redundancy of deltaproteobacterial symbionts to some degree, and b) horizontal transmission or frequent host switching. As mentioned above, a recent study by Schimak et al. [62] investigated the transmission mode of these symbionts using a labeling experiment. Adult worms, containing maturing eggs, were incubated in medium containing $^{15}$N-labeled ammonium. Directly after egg deposition, the freshly layn eggs were transferred to new incubation vials and incubated for several days in medium that did not contain $^{15}$N-labeled ammonium. The so incubated eggs were analyzed using FISH and NanoSIMS, showing that all symbionts present in the newly hatched and developing eggs carried the $^{15}$N label. Taken together, the findings of this and our study indicate that these symbionts are neither strictly vertically nor horizontally transmitted, but that they are most likely transferred to the next generation via both routes, in a mixed-mode transmission. Although horizontal transmission events might not be possible to be detected in a laboratory setting (because they do not happen often enough) the phylogenetic pattern obtained through SNP analyses reveals that horizontal transmission must occur occasionally in order to break the linkage pattern.

The spirochaete symbiont shows high levels of SNP diversity compared to the other symbionts, but in contrast to them, shows no increased distance between samples from Cavoli compared to Sant' Andrea. Both of these findings suggest that the spirochaete symbiont is able to infect worms independent from host reproduction, and that there is regular gene flow between the populations from Cavoli and Sant' Andrea. The fact that the *OalgD1* SNP tree has some strong outliers could be evidence that there are switches with a large population of free-living strains. With respect to the *D1*

clade, closely related environmental strains have been recently sequenced (>99% similarity on 16S, Marc Mußmann, personal communication).

In our study, we found that the secondary symbiont community was more diverse and variable between host specimens than previously thought. These results demonstrate the power of metagenomics over approaches like 16S rRNA clone libraries and 16S tag sequencing in identifying new phylotypes and discovering the true diversity within a microbial community.

***OalgD1* symbionts show host preference without host-linkage**

Although not restricted to one host haplotype, the *OalgD1a*/*OalgD1b* secondary symbionts still showed a statistically significant host preference, indicating that they grow better within the environment of their preferred host type. In worms that contain both symbionts at the same time, the "wrong" phylotype was heavily marginalized within the symbiont community and had a much lower abundance compared to the other respective phylotype. This sort of incompatibility could be the result of increased competition between two symbionts that are phenotypically too similar to each, marginalizing the symbiont that is less competitive. Alternatively (or in addition), host factors that are better at targeting one of the symbionts over the other (e.g. in phagocytosis efficiency, or effectiveness of population-controlling antimicrobials) could play a role.

**The *O. algarvensis* symbiosis covers a large range of symbiont-specificity**

Even though all symbionts are housed within the same symbiotic region between the cuticle of the host and its epidermis, and although the physical route for transmission is essentially the same for all symbionts (during egg laying via rupturing of the genital pad and release of symbiont cells to the outside), the symbionts display a wide range of host specificity and transmission mode.

On the one hand, specificity is extremely high in *OalgG1*, which shows a low amount of SNP divergence, very few, well defined phylotype groups, very high congruence with the host mitochondrial phylogeny, and has never been observed to switch between host haplotypes. In the case of the *OalgG3* symbiont, an intermediate pattern is observed, in which some linkage with the

host mitochondria is observed, and sequences from host haplotype B show a well-defined, tight cluster of sequences, while *OalgG3* associated with haplotypes A of Sant' Andrea and Cavoli form a diffuse cluster with higher sequence variation and possibly host switching events.

In the case of *OalgD4*, we observed a cluster without any clear diversification that reflects host haplotype, however, with a large outlier genome that had diverged substantially from the "cloud" of more similar *OalgD4* sequences. *OalgD1b* showed two defined clades, although they were not associated with a particular host haplotype, and only a small amount of sequence divergence. Assuming that this symbiont can be horizontally obtained from a free-living population of symbionts, the specificity with which this symbiont is taken up is quite high, but allows for some variation. In contrast, *OalgD1a* is more diverged on the SNP level, suggesting that its specificity is lower.

Several mechanisms might explain these observations. Strict vertical transmission, in which symbionts go through a transmission bottleneck each generation and in which no new phylotypes are introduced from the environment or other hosts, strain variability is expected to decrease [68]. Horizontal transmission, on the other hand, allows for new strains to enter the symbiosis and would increase genetic variability. Strong outliers in SNP divergence, as observed in *OalgD4* might also point to the recent uptake of a novel symbiont strain, while little genetic variation might also be a result of a recent selective sweep (and have less to do with transmission mode). Highly specific association with a particular phylotype might also be achieved through highly selective recognition and uptake mechanisms employed by the host or by imposing selective forces onto the symbiont that not all strains are able to handle equally well (e.g. antimicrobials produced by the host).

***Are the two phylotypes OalgG1-A and OalgG1-B functionally diverging?***

We compared the genomes of the primary symbionts from the two host haplotypes found in Sant' Andrea and found that divergent evolution has resulted in exclusive gene content restricted to the symbionts of one particular haplotype. Several genes involved in carbon- and energy metabolism, host interaction and DNA modification were absent in the primary symbiont of haplotype A, but not

haplotype B. Our study shows that the primary symbiont is a hot spot for the evolution of haplotype-specific gene sets. The absence of certain genes that we identified in this analysis could be the result of gene loss and genome deterioration that is predicted to be the result of a lifestyle which restricts a symbiont to be exclusively associated with a host [15]. These changes are typical for symbionts that are strictly vertically transmitted (appears to be the case for *OalgG1*) and genetically isolated. The latter does not apply to *OalgG1*, since it occurs together with several symbionts that, according to the SNP data, experience occasional exchange with other populations (be they free-living or associated with other hosts). Genome deterioration could possibly be further enhanced through the extremely high activity of transposase genes in these symbionts [63]. Transposases are enzymes that are able to move genetic elements within genomes, and contribute to genome erosion and gene loss by i) destroying the function of genes by inserting into them and ii) by inserting regions of high sequence similarity into the genome, which are anchors for homologous recombination and the deletion of regions between such anchors [69]. Host-linked genome diversification in the primary symbiont could lead to differentiation into metabolic niches, which could provide a molecular basis for host isolation and ultimately speciation and would allow similar host species to co-exist in the same habitat. The fast evolution of symbiont genomes, high flexibility of secondary symbionts, together with the low dispersal rate of the host might create local hot-spots for the diversification of both hosts and symbionts.

***Outlook***

While the results obtained in this study point to clear co-diversification between the host and the primary symbiont *OalgG1*, even on an intraspecific, population level scale, and divergent evolution is apparent in the symbiont genome, it is not clear how far the hosts have diverged in terms of their nuclear genome, and whether they are in the process of speciation. Future studies should therefore focus on investigating divergent evolution within the nuclear genome of the host, in order to assess if and how far these two haplotypes have diverged and how this might be influenced by the symbionts.

## Acknowledgements

Part of the sequencing was conducted by the U.S. Department of Energy Joint Genome Institute and is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Part of the sequencing was conducted by the Max Planck Society Genome Center at the Max Planck Institute for Plant Breeding in Cologne, Germany and was funded by the Max Planck Society and the Gordon and Betty Moore Foundation through Grant GBMF3811 to ND. MK was supported by a NSERC Banting Postdoctoral Fellowship. The authors have declared no conflict of interest.

## References

[1] McFall-Ngai, M. (2008) Are biologists in 'future shock'? Symbiosis integrates biology across domains. *Nature Reviews Microbiology* 6(10), 789–792

[2] McFall-Ngai, M., Hadfield, M. G., Bosch, T. C., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S. F., et al. (2013) Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences U S A* 110(9), 3229–3236

[3] Bright, M. and Bulgheresi, S. (2010) A complex journey: Transmission of microbial symbionts. *Nature Reviews Microbiology* 8(3), 218–230

[4] Hafner, M. S. and Nadler, S. A. (1988) Phylogenetic trees support the coevolution of parasites and their hosts. *Nature* 332(6161), 258–259

[5] Zimmermann, J., Wentrup, C., Sadowski, M., Blazejak, A., Gruber-Vodicka, H., Kleiner, M., Ott, J., Cronholm, B., De Wit, P., Erséus, C., and Dubilier, N. (2016) Closely coupled evolutionary history of ecto- and endosymbionts from two distantly-related animal phyla. *Molecular Ecology*, doi: 10.1111/mec.13554

[6] Clark, M. A., Moran, N. A., Baumann, P., and Wernegreen, J. J. (2000) Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. *Evolution* 54(2), 517–525

[7] Ruby, E. G. (2008) Symbiotic conversations are revealed under genetic interrogation. *Nature Reviews Microbiology* 6(10), 752–762

[8] Jones, M. L. and Gardiner, S. L. (1989) On the early development of the vestimentiferan tube worm *Ridgeia sp.* and observations on the nervous system and trophosome of *Ridgeia sp.* and *Riftia pachyptila*. *The Biological Bulletin* 177(2), 254–276

[9] Nyholm, S. V. and McFall-Ngai, M. J. (2004) The winnowing: Establishing the squid-*Vibrio* symbiosis. *Nature Reviews Microbiology* 2(8), 632–642

[10] Baumann, P. (2005) Biology bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annual Review of Microbiology* 59, 155–189

[11] Childress, J. J. and Girguis, P. R. (2011) The metabolic demands of endosymbiotic chemoautotrophic metabolism on host physiological capacities. *Journal of Experimental Biology* 214(Pt 2), 312–325

[12] Fukatsu, T. and Hosokawa, T. (2002) Capsule-transmitted gut symbiotic bacterium of the Japanese common plataspid stinkbug, *Megacopta punctatissima*. *Applied and Environmental Microbiology* 68(1), 389–396

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

[14] Oh, P. L., Benson, A. K., Peterson, D. A., Patil, P. B., Moriyama, E. N., Roos, S., and Walter, J. (2010) Diversification of the gut symbiont *Lactobacillus reuteri* as a result of host-driven evolution. *The ISME Journal* 4(3), 377–387

[15] Moran, N. A. and Plague, G. R. (2004) Genomic changes following host restriction in bacteria. *Current Opinion in Genetics and Development* 14(6), 627–633

[16] McCutcheon, J. P. and Moran, N. A. (2012) Extreme genome reduction in symbiotic bacteria. *Nature Reviwes Microbiology* 10(1), 13–26

[17] Baumann, P., Moran, N., and Baumann, L. (2013) *Bacteriocyte-Associated Endosymbionts of Insects*. Springer Berlin Heidelberg

[18] Newton, I. L. G., Woyke, T., Auchtung, T. A., Dilly, G. F., Dutton, R. J., Fisher, M. C., Fontanez, K. M., Lau, E., Stewart, F. J., Richardson, P. M., Barry, K. W., Saunders, E., Detter, J. C., Wu, D., Eisen, J. A., and Cavanaugh, C. M. (2007) The *Calyptogena magnifica* chemoautotrophic symbiont genome. *Science* 315(5814), 998–1000

[19] Brucker, R. M. and Bordenstein, S. R. (2012) Speciation by symbiosis. *Trends in Ecology and Evolution* 27(8), 443–451

[20] O'Malley, M. A. (2015) Endosymbiosis and its implications for evolutionary theory. *Proceedings of the National Academy of Sciences U S A* 112(33), 10270–10277

[21] Presgraves, D. C. (2010) The molecular evolutionary basis of species formation. *Nature Reviews Genetics* 11(3), 175–180

[22] Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, Å., et al. (2014) Genomics and the origin of species. *Nature Reviews Genetics* 15(3), 176–192

[23] Charlat, S., Hurst, G. D., and Merçot, H. (2003) Evolutionary consequences of *Wolbachia* infections. *Trends in Genetics* 19(4), 217–223

[24] Gotoh, T., Noda, H., and Ito, S. (2007) *Cardinium* symbionts cause cytoplasmic incompatibility in spider mites. *Heredity (Edinburgh)* 98(1), 13–20

[25] Gordon, I. J., Ireri, P., and Smith, D. A. (2014) Hologenomic speciation: synergy between a male-killing bacterium and sex-linkage creates a "magic trait" in a butterfly hybrid zone. *Biological Journal of the Linnean Society* 111(1), 92–109

[26] Eizaguirre, C., Lenz, T. L., Traulsen, A., and Milinski, M. (2009) Speciation accelerated and stabilized by pleiotropic major histocompatibility complex immunogenes. *Ecology Letters* 12(1), 5–12

[27] Sharon, G., Segal, D., Ringo, J. M., Hefetz, A., Zilber-Rosenberg, I., and Rosenberg, E. (2010) Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences* 107(46), 20051–20056

[28] Brucker, R. M. and Bordenstein, S. R. (2013) The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus *Nasonia*. *Science* 341(6146), 667–669

[29] Janson, E. M., Stireman, J. O., 3rd, Singer, M. S., and Abbot, P. (2008) Phytophagous insect-microbe mutualisms and adaptive evolutionary diversification. *Evolution* 62(5), 997–1012

[30] Dunbar, H. E., Wilson, A. C., Ferguson, N. R., and Moran, N. A. (2007) Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS Biology* 5(5), e96

[31] Rundle, H. D. and Nosil, P. (2005) Ecological speciation. *Ecology Letters* 8(3), 336–352

[32] Dubilier, N., Blazejak, A., and Rühland, C. (2006) Symbioses between bacteria and gutless marine oligochaetes. *Progress in Molecular and Subcellular Biology* 41, 251–275

[33] Woyke, T., Teeling, H., Ivanova, N. N., Huntemann, M., Richter, M., Gloeckner, F. O., Boffelli, D., Anderson, I. J., Barry, K. W., Shapiro, H. J., Szeto, E., Kyrpides, N. C., Mussmann, M., Amann, R., Bergin, C., Ruehland, C., Rubin, E. M., and Dubilier, N. (2006) Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* 443(7114), 950–955

[34] Kleiner, M., Wentrup, C., Lott, C., Teeling, H., Wetzel, S., Young, J., Chang, Y.-J., Shah, M., VerBerkmoes, N. C., Zarzycki, J., Fuchs, G., Markert, S., Hempel, K., Voigt, B., Becher, D., Liebeke, M., Lalk, M., Albrecht, D., Hecker, M., Schweder, T., and Dubilier, N. (2012) Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proceedings of the National Academy of Sciences U S A* 109(19), E1173–E1182

[35] Erseus, C. (2003) The gutless Tubificidae (Annelida: Oligochaeta) of the Bahamas. *Meiofauna Marina* 12, 59–84

[36] Dubilier, N., Mülders, C., Ferdelman, T., de Beer, D., Pernthaler, A., Klein, M., Wagner, M., Erséus, C., Thiermann, F., Krieger, J., Giere, O., and Amann, R. (2001) Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. *Nature* 411(6835), 298–302

[37] Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* 3(5), 294–299

[38] Kane, M. D., Poulsen, L. K., and Stahl, D. A. (1993) Monitoring the enrichment and isolation of sulfate-reducing bacteria by using oligonucleotide hybridization probes designed from environmentally derived 16S rRNA sequences. *Applied and Environmental Microbiology* 59(3), 682–686

[39] Larkin, M. A., Blackshields, G., Brown, N., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., et al. (2007) Clustal W and clustal X version 2.0. *Bioinformatics* 23(21), 2947–2948

[40] Hall, T. A. (1999) Bioedit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. In: *Nucleic Acids Symposium Series*, vol. 41, pp. 95-98

[41] Harrison, P. and Seemann, T. (2010) From high-throughput sequencing read alignments to confident, biologically relevant conclusions with nesoni. Poster presentation, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.509.7046&rep=rep1&type=pdf

[42] Nikolenko, S. I., Korobeynikov, A. I., and Alekseyev, M. A. (2013) Bayeshammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics* 14 Suppl 1, S7

[43] Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. (2012) SPADES: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19(5), 455–477

[44] Peng, Y., Leung, H. C. M., Yiu, S. M., and Chin, F. Y. L. (2012) IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28(11), 1420–1428

[45] Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G. W., and Nielsen, P. H. (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnology* 31(6), 533–538

[46] Strous, M., Kraft, B., Bisdorf, R., and Tegetmeyer, H. E. (2012) The binning of metagenomic contigs for microbial physiology of mixed cultures. *Frontiers in Microbiology* 3, 410

[47] Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research 25*(7), 1043-1055

[48] Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29(8), 1072–1075

[49] Langmead, B. and Salzberg, S. L. (2012) Fast gapped-read alignment with bowtie 2. *Nature Methods* 9(4), 357–359

[50] Miller, C. S., Baker, B. J., Thomas, B. C., Singer, S. W., and Banfield, J. F. (2011) EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biology* 12(5), R44

[51] Miller, C. S., Handley, K. M., Wrighton, K. C., Frischkorn, K. R., Thomas, B. C., and Banfield, J. F. (2013) Short-read assembly of full-length 16S amplicons reveals bacterial diversity in subsurface sediments. *PLoS One* 8(2), e56018

[52] Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. (2008) The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9, 75

[53] Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., and Stevens, R. (2014) The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Research* 42(Database issue), D206–D214

[54] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215(3), 403–410

[55] Huson, D. H. (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14(1), 68–73

[56] Milne, I., Bayer, M., Cardle, L., Shaw, P., Stephen, G., Wright, F., and Marshall, D. (2010) Tablet–next generation sequence assembly visualization. *Bioinformatics* 26(3), 401–402

[57] Lechner, M., Findeiss, S., Steiner, L., Marz, M., Stadler, P. F., and Prohaska, S. J. (2011) Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* 12, 124

[58] Rodriguez-R, L. M. and Konstantinidis, K. T. (2010) Bypassing cultivation to identify bacterial species. *Microbe*, *9*(3), 111-118

[59] Ruehland, C., Blazejak, A., Lott, C., Loy, A., Erséus, C., and Dubilier, N. (2008) Multiple bacterial symbionts in two species of co-occurring gutless oligochaete worms from Mediterranean sea grass sediments. *Environmental Microbiology* 10(12), 3404–3416

[60] Kleiner, M., Woyke, T., Ruehland, C., and Dubilier, N. (2011) The *Olavius algarvensis* metagenome revisited: Lessons learned from the analysis of the low-diversity microbial consortium of a gutless marine worm. In: *Handbook of Molecular Microbial Ecology II: Metagenomics in Different Habitats*, pp. 319-333

[61] Krieger, J. (2000) *Transmission in Inanidrilus leukodermatus*. Ph.D. thesis, University of Hamburg, Germany

[62] Schimak, M. P. (2015) *Transmission of bacterial symbionts in the gutless oligochaete Olavius algarvensis*. Ph.D. thesis, University of Bremen, Germany

[63] Kleiner, M., Young, J. C., Shah, M., Verberkmoes, N. C., and Dubilier, N. (2013) Metaproteomics reveals abundant transposase expression in mutualistic endosymbionts. *MBio* 4(3)

[64] Perner, M. (2003) *Biogeochemische und mikrobiologische Charakterisierung mariner Sedimente vor Elba - ein Beitrag zur ökosystemaren Analyse bakteriensymbiontischer Oligochaeten*. Master's thesis, University of Bremen, Germany

[65] Musat, N., Werner, U., Knittel, K., Kolb, S., Dodenhof, T., van Beusekom, J. E. E., de Beer, D., Dubilier, N., and Amann, R. (2006) Microbial community structure of sandy intertidal sediments in the North Sea, Sylt-Rømø Basin, Wadden Sea. *Systematic and Applied Microbiology* 29(4), 333–348

[66] Aravindraja, C., Viszwapriya, D., and Pandian, S. K. (2013) Ultradeep 16S rRNA sequencing analysis of geographically similar but diverse unexplored marine samples reveal varied bacterial community composition. *PLOS One* 8(10), e76724

[67] Heindl, N. R., Gruber-Vodicka, H. R., Bayer, C., Lücker, S., Ott, J. A., and Bulgheresi, S. (2011) First detection of thiotrophic symbiont phylotypes in the pelagic marine environment. *FEMS Microbiology Ecology* 77(1), 223–227

[68] Sachs, J. L., Skophammer, R. G., Bansal, N., and Stajich, J. E. (2014) Evolutionary origins and diversification of proteobacterial mutualists. *Proceedings of the Royal Society B* 281(1775), 20132146

[69] Mahillon, J. and Chandler, M. (1998) Insertion sequences. *Microbiology and Molecular Biology Reviews* 62(3), 725–774

[70] Ruehland, C., Bergin, C., Lott, C., and Dubilier, N. (2006) Symbiosen mit mikrobiellen konsortien. *BIOSPEKTRUM-HEIDELBERG* 12(6), 600

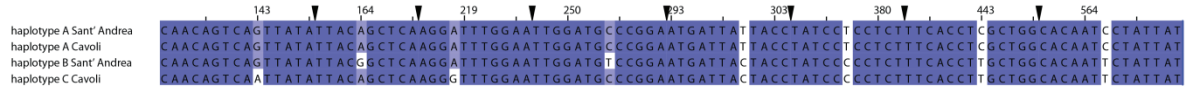
## Tables and Figures

**Figure 1: Sampling sites**

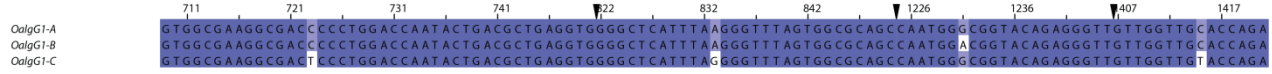
Worms in this study were collected from two sites off the Island of Elba, the Bay of Sant' Andrea (North) and the Bay of Cavoli (South).

**Figure 2: Multiple alignments of host COI and symbiont 16S rRNA gene sequences**
Shown are multiple sequence alignments of the host and symbiont phylogenetic marker genes. For full sequence alignments containing all sequences obtained in this study, refer to Supplementary Files 1 – 5.

A) Host cytochrome c oxidase COI haplotypes




B) *OalgG1* host-specific symbiont phylotypes




C) *OalgD1a*/*OalgD1b* symbiont phylotypes and FISH probes


D) *OalgS1* symbiont phylotypes

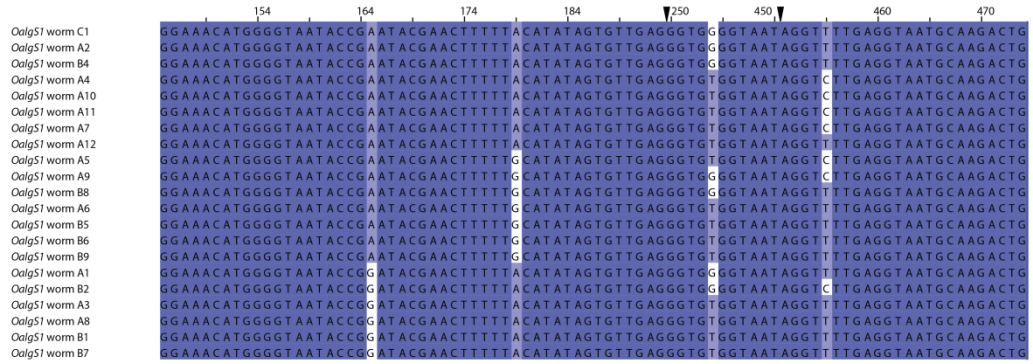


**Figure 3: Relative abundance of symbionts in each worm metagenome**

Relative estimated abundances of symbionts in each metagenome. Estimated abundances were calculated with EMIRGE [50]. HT-A: host COI haplotype A; HT-B: host COI haplotype B, HT-C host COI haplotype C. Top grey bar indicates presence of low-abundance symbionts that were not detected by EMIRGE and that were manually identified by mapping the metagenomic reads onto the respective symbiont draft genomes.

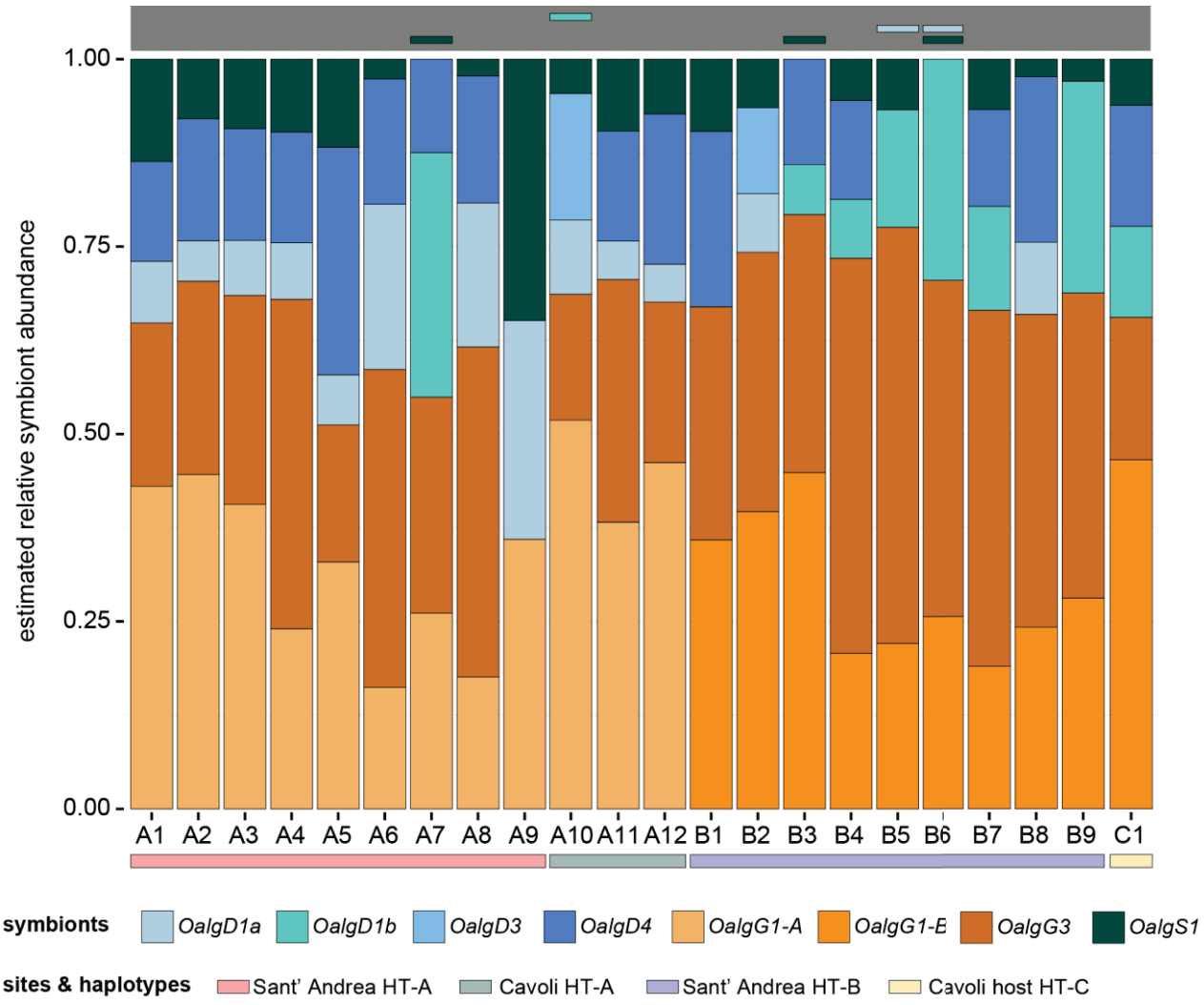


**Figure 4: SNP distance trees of host mitochondrial and symbiont genomes**

Neighbor Joining (NJ) transformed splits trees based on core SNPs called from read alignments to the respective genomes. Asterisks mark edges with >99% bootstrap support from 1,000 replications. Scale bar represents 10% SNP divergence, inset scale bars represent 1% SNP divergence. A) host mitochondrial genome, tree based on 162 core SNPs, B) *OalgG1* symbiont genome, tree based on 12,779 core SNPs, C) *OalgG3* symbiont genome, tree based on 9502 core SNPs, D) *OalgS1* symbiont genome, tree based on 3421 core SNPs, E) *OalgD1a* symbiont genome, tree based on 8960 core SNPs, F) OalgD1b symbiont genome, tree based on 240,062 core SNPs, G) *OalgD4* symbiont genome, tree based on 6210 core SNPs.

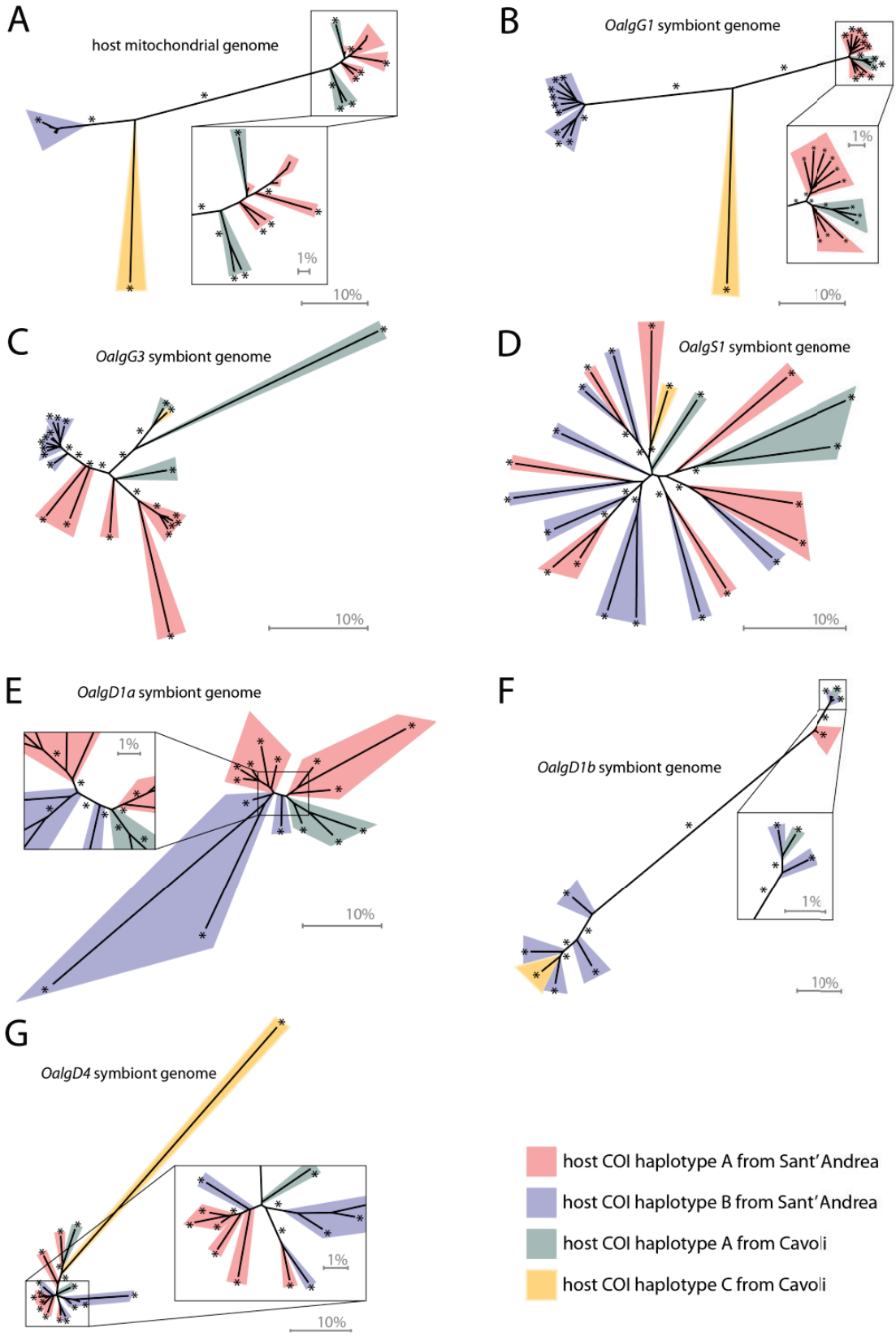

A
host mitochondrial genome
1%
10%
B
OalgG1 symbiont genome
1%
10%
C
OalgG3 symbiont genome
10%
D
OalgS1 symbiont genome
10%
E
OalgD1a symbiont genome
1%
10%
F
OalgD1b symbiont genome
1%
10%
G
OalgD4 symbiont genome
1%
10%
host COI haplotype A from Sant'Andrea
host COI haplotype B from Sant'Andrea
host COI haplotype A from Cavoli
host COI haplotype C from Cavoli

**Table 1: *Olavius algarvensis* symbionts**

| Phylogenetic group | Clade | Symbiont species | Synonym/previously used name | Subspecies phylotype name | Evidence | References |
|---|---|---|---|---|---|---|
| Gammaproteobacteria | 1 | *OalgG1* | Gamma1, Candidatus *Thiosymbion*, γ1 | *OalgG1-A* | FISH, clone library, metagenome | [36, 33, 59], this study |
| | | *OalgG1* | Gamma1, Candidatus *Thiosymbion*, γ1 | *OalgG1-B* | clone library, metagenome | [33], this study |
| | | *OalgG1* | Gamma1, Candidatus *Thiosymbion*, γ1 | *OalgG1-C* | metagenome | This study |
| | 3 | *OalgG3* | Gamma3, γ3 | - | FISH, clone library, metagenome | [33, 59] |
| Deltaproteobacteria | 1 | *OalgD1a* | Delta1, δ1 | - | FISH, clone library, metagenome | [36, 33, 59] |
| | | *OalgD1b* | Olavius sp.-associated clone | - | Clone library, metagenome | [33], this study |
| | | *OalgD3* | *Olavius algarvensis* Delta3-associated clone | - | Clone library, metagenome | [33, 60], this study |
| | 4 | *OalgD4* | Delta4, δ4 | - | FISH, clone library, metagenome | [36, 33, 70] |
| Spirochaeta | 1 | *OalgS1* | Spirochaete | - | FISH, clone library, metagenome | [59], this study |

**Table 2: Genes unique to A- or G-type *OalgG1* symbionts**

| Unique to [a] | Gene ID | Gene annotation [b] | # Reads A [c] | # Reads B [c] |
|---|---|---|---|---|
| A | 69204.peg.697 | hypothetical protein | 618 | 0 |
| A | 69204.peg.656 | hypothetical transmembrane protein | 267 | 0 |
| A | 69204.peg.367 | hypothetical protein | 584 | 0 |
| A | 69204.peg.250 | hypothetical protein | 216 | 0 |
| A | 69204.peg.2285 | HigA protein (antitoxin to HigB) | 399 | 0 |
| A | 69204.peg.1913 | hypothetical protein | 373 | 0 |
| A | 69204.peg.1334 | hypothetical protein | 173 | 0 |
| A | 69204.peg.1176 | Asparagine synthetase | 628 | 0 |
| A | 69204.peg.1 | hypothetical protein with DUF4160 (PF13711) | 426 | 0 |
| B | 69208.peg.73 | hypothetical protein | 0 | 228 |
| B | 69208.peg.1123 | hypothetical protein | 0 | 4452 |
| B | 69207.peg.980 | DNA-cytosine methyltransferase | 0 | 1134 |
| B | 69207.peg.912 | hypothetical protein with DUF820 (PF05685), restriction endonuclease type II-like domain (SSF52980) | 0 | 1419 |
| B | 69207.peg.883 | hypothetical transmembrane protein | 0 | 6930 |
| B | 69207.peg.882 | hypothetical transmembrane protein with signal peptide | 0 | 1015 |
| B | 69207.peg.726 | hypothetical protein | 0 | 1156 |
| B | 69207.peg.2390 | hypothetical protein | 0 | 218 |
| B | 69207.peg.2343 | hypothetical protein | 0 | 379 |
| B | 69207.peg.2299 | hypothetical protein | 0 | 665 |
| B | 69207.peg.2226 | hypothetical protein | 0 | 512 |
| B | 69207.peg.2225 | DNA-invertase | 0 | 2256 |
| B | 69207.peg.2201 | Hydrogenase maturation factor hoxX | 0 | 363 |
| B | 69207.peg.2186 | Rhodanese-related sulfurtransferase | 0 | 1214 |
| B | 69207.peg.2150 | Large exoproteins involved in heme utilization or adhesion | 0 | 4229 |
| B | 69207.peg.2114 | Fumarate and nitrate reduction regulatory protein | 0 | 2785 |
| B | 69207.peg.1999 | hypothetical protein | 0 | 767 |
| B | 69207.peg.1984 | hypothetical protein | 0 | 1833 |
| B | 69207.peg.1870 | hypothetical protein | 0 | 896 |
| B | 69207.peg.1838 | hypothetical protein | 0 | 950 |
| B | 69207.peg.1755 | Glycosyl transferase, group 1 | 0 | 1478 |
| B | 69207.peg.1731 | Polyferredoxin NapH (periplasmic nitrate reductase) | 0 | 2580 |
| B | 69207.peg.1666 | hypothetical protein | 0 | 631 |
| B | 69207.peg.1647 | hypothetical protein with Beta-lactamase/transpeptidase-like domain (PF13354, SSF56601) | 0 | 2868 |
| B | 69207.peg.1635 | hypothetical protein | 0 | 387 |
| B | 69207.peg.1634 | hypothetical protein | 0 | 245 |
| B | 69207.peg.1611 | Acetate permease ActP (cation/acetate symporter) | 0 | 4705 |
| B | 69207.peg.1579 | hypothetical protein | 0 | 626 |
| B | 69207.peg.1556 | hypothetical protein | 0 | 1881 |
| B | 69207.peg.1499 | FIG00637371: hypothetical protein | 0 | 783 |
| B | 69207.peg.1498 | Carbohydrate kinase, PfkB | 0 | 4084 |
| B | 69207.peg.1469 | hypothetical transmembrane protein | 0 | 1292 |

| | | | | |
|---|---|---|---|---|
| B | 69207.peg.1468 | hypothetical protein with transposase IS200-like domain (PF01797, SSF143422) | 0 | 123 |
| B | 69207.peg.1467 | hypothetical transmembrane protein with signal peptide | 0 | 2253 |
| B | 69207.peg.1374 | Dolichol-phosphate mannosyltransferase | 0 | 2628 |
| B | 69207.peg.1242 | hypothetical protein | 0 | 548 |
| B | 69207.peg.1239 | hypothetical protein with DUF2326 (PF10088) | 0 | 4925 |
| B | 69207.peg.1238 | hypothetical transmembrane protein | 0 | 482 |
| B | 69207.peg.1220 | Type IV pilus biogenesis protein PilM | 0 | 450 |
| B | 69207.peg.1159 | hypothetical protein | 0 | 343 |

[a)] Unique to *OalgG1* from host COI haplotype A or B, respectively

[b)] RAST annotation, augmented with blast2go domain searches (PFAM, SUPERFAMILY, TMHMM, SignalP4, Phobius)

[c)] Number of total mapped reads from all metagenomes of Sant' Andrea COI haplotype A hosts, or haplotype B hosts respectively, using *OalgG1* strain-unique genes as target

**Supplementary Table 1: Metagenome sequencing statistics**

| Number of sequenced paired reads in millions | | | | | |
|---|---|---|---|---|---|
| **Worm** | **Site** | **Raw** | **Processed** | **Sequencing center** | **Accession** |
| A1 | Sant' Andrea | 111 | 55 | JGI DOE | 1021950[a)] |
| A2 | Sant' Andrea | 77 | 38 | JGI DOE | 1021953[a)] |
| A3 | Sant' Andrea | 170 | 84 | JGI DOE | 1021956[a)] |
| A4 | Sant' Andrea | 2 | 18 | MPIPZ GC | Unpubl. |
| A5 | Sant' Andrea | 33 | 18 | MPIPZ GC | Unpubl. |
| A6 | Sant' Andrea | 28 | 16 | MPIPZ GC | Unpubl. |
| A7 | Sant' Andrea | 33 | 23 | MPIPZ GC | Unpubl. |
| A8 | Sant' Andrea | 35 | 25 | MPIPZ GC | Unpubl. |
| A9 | Sant' Andrea | 30 | 17 | MPIPZ GC | Unpubl. |
| A10 | Cavoli | 36 | 34 | MPIPZ GC | Unpubl. |
| A11 | Cavoli | 85 | 84 | MPIPZ GC | Unpubl. |
| A12 | Cavoli | 38 | 38 | MPIPZ GC | Unpubl. |
| B1 | Sant' Andrea | 84 | 42 | JGI DOE | 1021959[a)] |
| B2 | Sant' Andrea | 107 | 53 | JGI DOE | 1021962[a)] |
| B3 | Sant' Andrea | 90 | 44 | JGI DOE | 1021965[a)] |
| B4 | Sant' Andrea | 33 | 22 | MPIPZ GC | Unpubl. |
| B5 | Sant' Andrea | 32 | 16 | MPIPZ GC | Unpubl. |
| B6 | Sant' Andrea | 28 | 16 | MPIPZ GC | Unpubl. |
| B7 | Sant' Andrea | 32 | 17 | MPIPZ GC | Unpubl. |
| B8 | Sant' Andrea | 33 | 23 | MPIPZ GC | Unpubl. |
| B9 | Sant' Andrea | 34 | 22 | MPIPZ GC | Unpubl. |
| C1 | Cavoli | 34 | 32 | MPIPZ GC | Unpubl. |

**JGI DOE:** Joint Genome Institute Department of Energy
**MPIPZ GC:** Max Planck Institute für Pflanzenzüchtung Genome Center
[a)] JGI DOE project ID for metagenome retrieval from IMG database

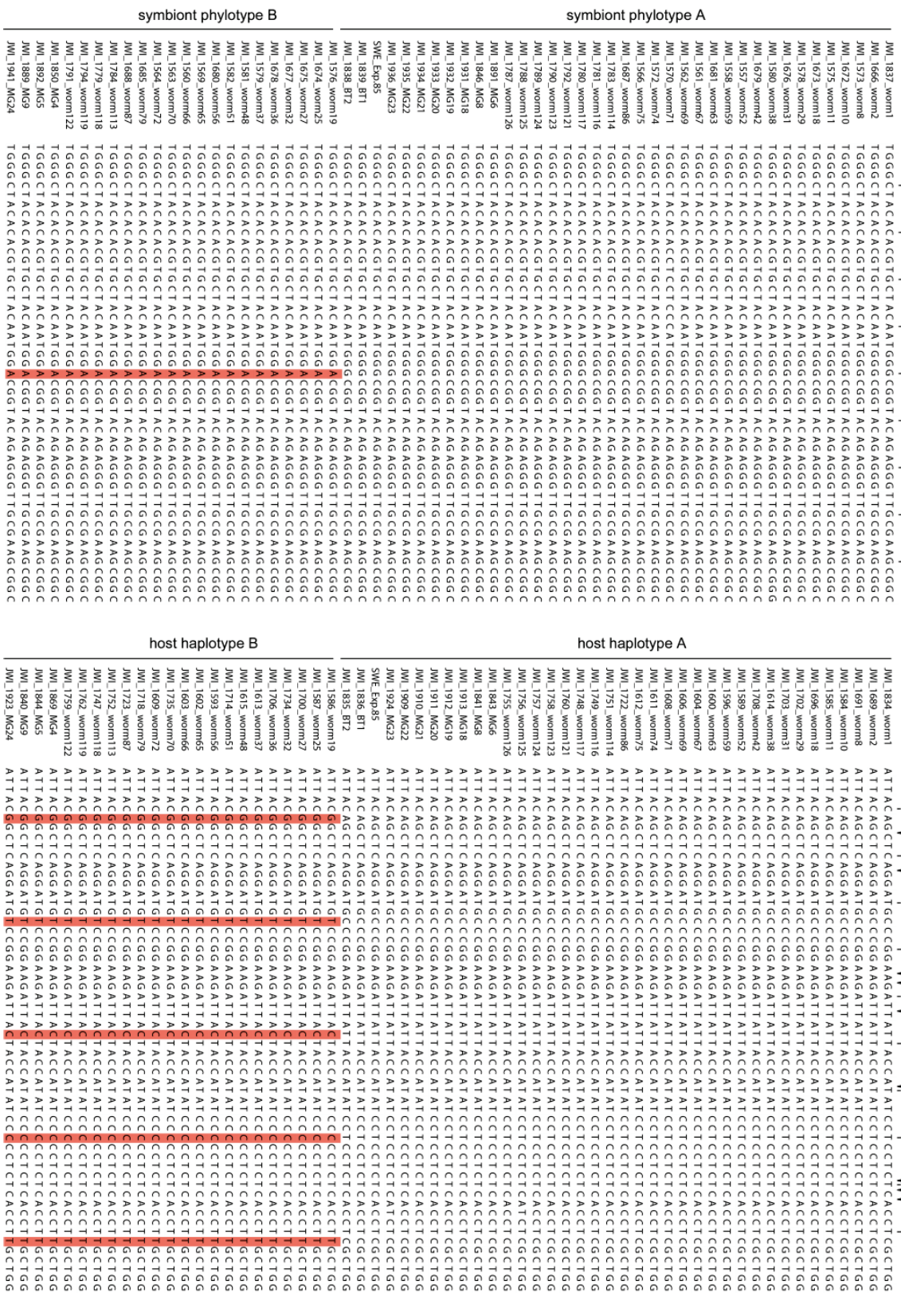
**Supplementary Table 2: Draft genome completeness and assembly quality**

Completeness estimate of symbiont genome assemblies, based on single copy genes as calculated using CheckM [47]. Previously published genome drafts are from [33]; n.a., not available.

| | Symbiont genome bins used as mapping reference for SNP analyses | | | | | | | Host mitochondrion |
|---|---|---|---|---|---|---|---|---|
| Genome | *OalgG1* | *OalgG3* | *OalgD1a* | *OalgD1b* | *OalgD3* | *OalgD4* | *OalgS1* | Haplotype A |
| Binned from sample | A3 | A1 | A3 | B3 | B2 | A1 | B2 | A2 |
| JGI DOE project ID for metagenome | 1021956 | 1021950 | 1021956 | 1021965 | 1021962 | 1021950 | 1021962 | 1021953 |
| RAST genome project ID | 6666666.69206 | 6666666.76579 | 6666666.83790 | 6666666.83916 | 6666666.84585 | 6666666.76586 | 6666666.76598 | n.a. |
| Completeness estimate | | | | | | | | |
| previous draft [33] | 26.00 | 89.76 | 90.22 | n.a. | n.a. | 38.62 | n.a. | n.a. |
| current draft | 89.18 | 98.24 | 92.37 | 85.14 | 91.61 | 99.35 | 44.55 | 100.00 (circular) |
| Genome bin metrics | | | | | | | | |
| # contigs | 2349 | 414 | 834 | 272 | 36 | 87 | 18 | 1 |
| total length | 2361199 | 4,194,833 | 8102240 | 6,555,209 | 3,589,405 | 5,440,777 | 963,155 | 14525 |
| largest contig (bp) | 7508 | 213,227 | 45438 | 99,901 | 469,213 | 549,397 | 230,242 | 14525 |
| GC (%) | 61.01 | 55.84 | 50.42 | 48.37 | 54.38 | 54.04 | 46.23 | 26.42 |
| N50 | 2539 | 31,529 | 13830 | 42,416 | 131,204 | 214,140 | 80,079 | 14525 |
| # cds | 1832 | 4141 | 7667 | 6546 | 4872 | 5045 | 1137 | 16 |
| # tRNAs | 41 | 36 | 44 | 31 | 47 | 47 | 31 | 22 |

**Supplementary Figure 1: Haplotype-specific association of *OalgG1* phylotypes**

Figure shows multiple sequence alignments of *OalgG1* symbiont 16S sequences (left) and of the host COI sequences from the same worm specimens (right). Black arrows indicate hidden sequence portions.



**Supplementary file 1: Multiple alignment of host COI sequences from Sant' Andrea**

-> Contains all host COI sequences obtained in this study through PCR and direct sequencing

**Supplementary file 2: Multiple alignment of *OalgG1* 16S sequences from clone library**

-> contains all *OalgG1* 16S sequences from [33] clone library

**Supplementary file 3: Multiple alignment of *OalgG3* 16S sequences reconstructed from metagenomes**

-> contains all 16S sequences of *OalgG3* symbionts reconstructed from metagenomes with EMIRGE

**Supplementary file 4: Multiple alignment of deltaproteobacterial symbiont 16S sequences reconstructed from metagenomes**

-> contains all 16S sequences of deltaproteobacterial symbionts reconstructed from metagenomes with EMIRGE

**Supplementary file 5: Multiple alignment of *OalgS1* 16S sequences reconstructed from metagenomes**

-> contains all 16S sequences of *OalgS1* symbiont reconstructed from metagenomes with EMIRGE

**Supplementary file 6: Confirmation of *OalgG1* phylotype-specific genes by read mapping analysis**

-> contains a list of putative phylotype-specific genes and the number of mapped reads from each metagenome for each gene

# Chapter 3: Functional genomic characterization of the spirochaetal endosymbiont of the gutless oligochaete *Olavius algarvensis*

**Authors:** Juliane Wippler[1], Cecilia Wentrup[2], Manuel Kleiner[3], Tanja Woyke[4], Nicole Dubilier[1]

[1]Max Planck Institute for Marine Microbiology, Celsiusstr. 1, D-28359 Bremen, Germany

[2]University of Vienna, Department of Microbiology and Ecosystem Science, Althanstr. 14, A-1090 Vienna, Austria

[3]University of Calgary, Department of Geoscience, Calgary T2N 1N4, AB, Canada

[4]DOE Joint Genome Institute, Walnut Creek, California 94598, USA

**Publication status:** Manuscript in preparation

**Author's contributions:**

**JW**: Conceived the project and manuscript, provided ideas and wrote the manuscript, collected worms, extracted DNA, did all data processing, genome assembly, binning, annotation and functional and phylogenetic analyses, prepared all figures and tables. **CW, MK, TW, ND:** Conceived the project and were involved in project organization and coordination.

## Abstract

The family of *Spirochaetaceae* is mostly known for its members that cause disease in humans and other mammals, and to a lesser extent, for the mutualistic symbionts in wood-eating termites. Here, we present the near complete genome sequence of the spirochaetal endosymbiont *OalgS1* of the gutless oligochaete *Olavius algarvensis*, which is most closely related to free-living, non-pathogenic and non-host associated *Spirochaeta* isolates from marine sediments and microbial mats. *OalgS1* is an extracellular endosymbiont that lives in a symbiotic consortium together with two chemosynthetic sulfur-oxidizing Gammaproteobacteria and sulfate-reducing Deltaproteobacteria. *OalgS1* is a chemoorganoheterotrophic organism capable of fermenting a large number of carbohydrates to acetate and other short chain fatty acids (SCFAs), and $CO_2$ + $H_2$. *OalgS1*, which is located directly beneath the highly permeable cuticle of the worm, encodes in its genome a large number of transporters for a wide range of sugars, some of which were recently shown to be present in the pore water that the host inhabits. It is therefore hypothesized to scavenge simple sugars from the environment and engage in a syntrophic relationship with deltaproteobacterial symbionts of *O. algarvensis*, which are known to use SCFAs as an energy and carbon source and $H_2$ as an energy source. We found no indication that *OalgS1* is pathogenic in healthy worms; in fact, it lacks several of the molecular components that typically make pathogenic spirochaetes cytotoxic to host epithelia.

## Introduction

The Spirochaetes comprise a group of morphologically unique bacteria that include a number of important human pathogens, like the causative agents of Lyme disease, syphilis, epidemic and endemic relapsing fever, leptospirosis and periodontal disease [1]. The family Spirochaetaceae contains pathogens like *Treponema* and *Borrelia*, as well as mutualistic treponeme symbionts of wood-eating termites and free-living *Spirochaeta* that inhabit marine or freshwater sediments and microbial mats [2, 3, 4]. Other spirochaetal symbionts have been reported from the crystalline styles of bivalve and gastropod digestive tracts [5], from the sponge genus *Clathrina* [6], from the body surface of the marine polychaete *Alvinella pompeijana* [7], and from three species of gutless marine oligochaetes [8, 9, 10]. The *Alvinella* and gutless oligochaete spirochaetes are closely related [10]. However, nothing is known about the physiology of any of these spirochaete symbionts and the possible role they might play in their hosts.

The gutless oligochaetes are a species-rich, monophyletic group of small annelids that inhabit the interstitial pore water of marine sediments [11]. All gutless oligochaetes form beneficial, obligate associations with chemosynthetic sulfur-oxidizing symbionts that compensate for the host's lack of a mouth, anus and gut by providing most, if not all, of its nutrition though autotrophic carbon fixation. In addition to the primary obligate sulfur-oxidizing symbiont (the Gammaproteobacterium *OalgG1*/Gamma1/*Candidatus* Thiosymbion), which is shared among all gutless oligochaete species, the gutless oligochaetes also associate with a number of secondary symbionts.

In the Mediterranean gutless oligochaete species *Olavius algarvensis*, these secondary symbionts comprise a number of different deltaproteobacterial sulfate-reducing bacteria, another sulfur-oxidizing Gammaproteobacterium in addition to the Gamma1 symbiont, and the spirochaete symbiont *OalgS1* mentioned above. While the functions of the gamma- and deltaproteobacterial symbionts have been extensively studied and mostly resolved through

metagenomic and metaproteomic approaches [12, 13], the role of the spirochaete symbiont remains elusive because of its relatively low abundance compared to the other symbionts (< 10% of the symbiotic community, [10, 12], Wippler et al. 2016 submitted). Using next-generation sequencing technology and the high read coverage that comes with Illumina sequencing, we were able to assemble a near complete genome of the *O. algarvensis* spirochaete *OalgS1* from the metagenomes generated from six *O. algarvensis* specimens. In this study, we present a characterization of the functional repertoire encoded in the *OalgS1* genome, compare it to the genomic features of other spirochaete genomes and speculate on the putative role of this symbiont in the gutless oligochaete symbiosis.

## Material and Methods

### Sample collection

Sediment containing *O. algarvensis* worms was collected at 7 meters water depth from the Bay of Sant' Andrea, Elba, Italy (42°48'26"N, 010°08'28"E), June 2012. Worms were extracted from the sediment by decantation with seawater, and six mature *Olavius algarvensis* specimens were morphologically identified under a dissection microscope. The live worms were subsequently rinsed in sterile-filtrated seawater (2 µm pore size, Millipore, Darmstadt, Germany), briefly tapped on blotting paper to remove excess liquid, then flash-frozen in liquid nitrogen, and stored at -80°C until DNA extraction.

### DNA extraction

DNA for metagenomic sequencing was extracted from individual whole worms using the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany). Frozen worms were thawed in 180 µl buffer ATL with Proteinase K @ 55 °C for 10 minutes, and then incubated for up to six days at 37°C until the

worms were completely dissolved, to maximize DNA yield. Subsequent extraction steps were carried out according to the manufacturer's instructions.

**Illumina metagenome sequencing**

Illumina libraries were constructed using the TruSeq Illumina library preparation kit (Illumina, San Diego, USA) according to the manufacturer's instructions and fragmented to 300-500 bp insert size. Libraries were paired-end sequenced (2x 100 bp) on an Illumina HiSeq2500 machine. Supplementary Table 1 summarizes the number of raw reads generated from each library, the number of processed reads used in subsequent bioinformatic analyses. The metagenomic sequences have been published under the accession numbers given in Supplementary Table 1.

**Bioinformatic analysis**

Reads were quality filtered and adapter-trimmed using nesoni version 0.114 [14] and corrected for sequencing errors using BayesHammer [15], as implemented in Spades version 2.5.1 [16]. Metagenomic reads were assembled de novo using idba_ud version 1.1.1 [17]. Spirochaeta symbiont genomes were binned using a combined approach of differential coverage binning as described in [18] and Wippler et al. 2016 (in prep.). Briefly, an initial spirochaete bin was created by identifying contigs that contained conserved spirochaetal markers genes. Further spirochaetal contigs were binned from the whole metagenome assembly by retrieving any contigs that were connected to the initial bin via paired-end read linkage as described in [18]. This procedure was carried out using single metagenomes and a combination of all six metagenomes (co-assembly). The assemblies obtained from using only single metagenomic data sets were inferior to the co-assembly (Supplementary Table 2), and therefore, the co-assembly was used for all subsequent functional analyses. Completeness estimates of binned spirochaete symbiont drafts were performed with CheckM version 1.0.1 [19]. Symbiont genome draft assembly metrics were determined with QUAST version 2.3 [20]. Reads were mapped to reference sequences with bowtie2 version 2.1.0 [21]. Read mappings for Figure 4 were visualized in artemis [22]. The draft

genome was automatically annotated in RAST version 2.0 [23, 24], which includes the detection of CRISPR repeats and spacers. Manual sequence similarity searches were performed with BLAST 2.2.28+ [25]. Average nucleotide identities (ANIs) between symbiont genomes were calculated using the respective symbiont draft genomes with the ANI calculator implemented at http://enve-omics.ce.gatech.edu/ani/ [26]. The phylogenetic 16S rRNA Neighbor Joining tree was generated with the software package ARB [27].

## Results and Discussion

### Genome properties of *O. algarvensis OalgS1* symbiont

A near complete (96.06%) genome draft sequence of the *Olavius algarvensis* spirochaete symbiont *OalgS1* was obtained combining the metagenomic data sequenced from six *O. algarvensis* specimens (Figure 1). Using the individual metagenomic data sets, we assembled and binned draft genomes of varying completeness from each worm specimen (Supplementary Table 2). A pairwise comparison of ANI (average nucleotide identity) values between these six draft genomes, showed that the genomes obtained from the six specimens differed only marginally (ANI 99.74 – 99.85%), indicating that the worms are all infected by the same species (Supplementary Table 3). Analysis of strain heterogeneity, based on conserved Spirochaetae-specific maker genes, also suggested that strain variability is very low (Figure 1). Therefore, we used the near-complete co-assembly of *OalgS1* for subsequent functional analyses.

The *OalgS1* genome is 2.18 Mb large, with a DNA G+C content of 46%. Of the 2022 predicted genes, 1975 were protein coding genes, 44 were tRNAs, and 2 were rRNAs. No pseudogenes were identified. 54.74% of all coding genes could be annotated with a putative function, of which 34% could be integrated into the curated annotation subsystems of the annotation pipeline RAST. Table 1 summarizes the genome characteristics of *OalgS1* and Table 2

summarizes the distribution of genes into RAST subsystem categories. Note that no flagella or motility genes are listed in subsystems; however, *OalgS1* does encode the regular genes for flagella synthesis (Supplementary Table 4).

**Phylogeny and comparison of genomic features to related Spirochaetaceae**

Using the 16S rRNA sequence from the draft assembly as query, highly scoring 16S rRNA sequences were retrieved from NCBI with blastn. The top scoring sequence hits were Spirochaetaceae of the genera *Spirochaeta*, *Treponema* and *Borrelia*. The 16S rRNA sequence of *OalgS1* was aligned to spirochaete 16S rRNA sequences representative of the diversity within this group, and used to construct a Neighbor Joining phylogenetic tree (Figure 2). *OalgS1* formed a well-supported monophyletic group with other gutless oligochaete spirochete symbionts and was closely related to free-living *Spirochaeta* members. Strains most closely related to the gutless oligochaete spirochetes had mainly been isolated from marine sediments. See Tables 1 and 2 for genomic and functional comparisons of *OalgS1* with other available Spirochaetaceae genomes.

In comparison to other free-living *Spirochaeta* strains from marine sediments, *OalgS1* had a much smaller genome, only about half the average size (Table 1). The reduced genome size is also noticeable in the reduced number of coding sequences (cds) in the *OalgS1* genome. The *OalgS1* genome was also significantly smaller than that of the termite hindgut symbiont *T. azotonutricium* ZAS-9, and was more similar in size to facultative or opportunistic spirochete pathogens (0.9 – 2.8 Mb), including the human dental isolate *Treponema denticola*, *Treponema pallidum*, the causative agent of syphilis, and *Borrelia spp*., which cause borreliosis (Lyme disease) in humans. In contrast, the DNA GC content was more similar to free-living mesophilic *Spirochaeta* and symbiotic termite isolates, than to the pathogenic spirochaetes. The properties and function of *OalgS1* within the gutless worm symbiosis were so far not known; however, based on the observation that spirochaete-infected worms showed no signs of disease it was

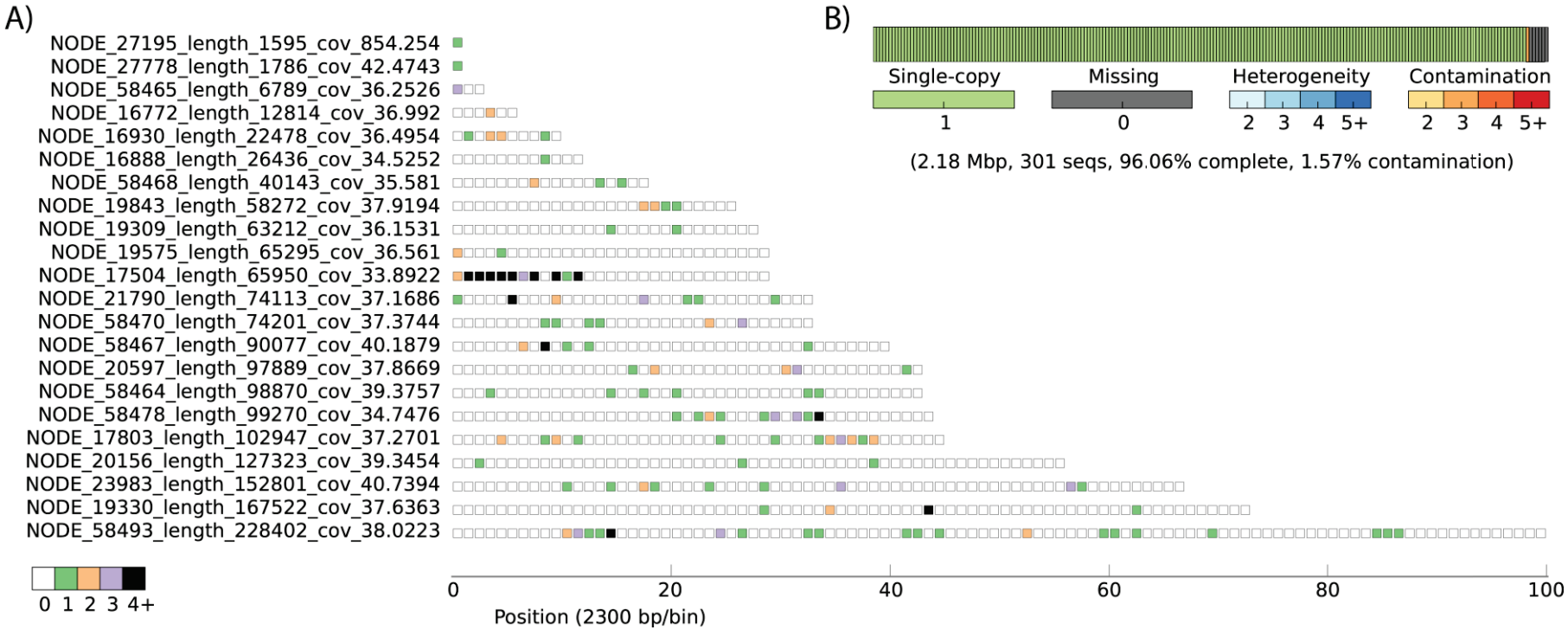



**Figure 1: Distribution of Spirochaetae phylogenetic marker genes across the *OalgS1* genome and estimation of genome completeness, strain heterogeneity, and contamination. A)** Displayed are all contigs containing conserved marker genes characteristic for the phylum Spirochaetae. Each box represents a fixed window size of 2300 bp. Box color represents the number of marker genes identified within a sequence window. Conserved markers are distributed evenly across the genome except ribosomal proteins, which typically occur in clusters (contig NODE_17504_length_65950_cov_33.8922). **B)** Estimation of completeness, strain heterogeneity, and contamination in *OalgS1* genome. Green bars: marker genes found in genome, grey bars: marker genes not found in genome. Strain heterogeneity and contamination is based on marker genes of which multiple copies were found in the genome. Colors represent amino acid similarity between these multiple copies. If they are at least 90% identical, the marker gene is expected to be of the same species, but a different strain (blue shades). If they are less than 90% identical they are assumed to be derived from a different species (yellow-red shades). Genome is estimated to be almost complete (96.06%), contain no strain heterogeneity and only minor contamination (1.57%).

**Table 1: Genome characteristics of *OalgS1* and related Spirochaetaceae bacteria**

| Genome | # contigs | Largest contig | Total length | GC (%) | # cds | # RNAs | % hypothetical proteins |
|---|---|---|---|---|---|---|---|
| *OalgS1* draft genome | 301 | 228402 | 2182993 | 45.97 | 1975 | 48 | 45.26 |
| *Spirochaeta bajacaliforniensis* DSM16054 | 80 | 375186 | 4574849 | 49.17 | 4303 | 51 | 32.35 |
| *Spirochaeta cellobiosiphila* DSM17781 | 37 | 511479 | 3945304 | 37.01 | 3741 | 51 | 38.25 |
| *Spirochaeta smaragdinae* DSM11293 | 1 | 4653970 | 4653970 | 48.97 | 4411 | 55 | 33.03 |
| *Spirochaeta thermophila* DSM6578 | 1 | 2560222 | 2560222 | 60.92 | 2329 | 52 | 29.49 |
| *Spirochaeta africana* DSM8902 | 1 | 3285855 | 3285855 | 57.77 | 2846 | 53 | 35.45 |
| *Spirochaeta alkalica* DSM8900 | 87 | 232938 | 3347706 | 60.55 | 2966 | 50 | 34.38 |
| *Treponema azotonutricium* ZAS-9 | 1 | 3855671 | 3855671 | 49.80 | 3524 | 49 | 38.82 |
| *Treponema denticola* ATCC 35405 | 1 | 2843201 | 2843201 | 37.87 | 2769 | 52 | 40.20 |
| *Treponema pallidum subsp. pallidum str.* Nichols | 1 | 1138011 | 1138011 | 52.78 | 1057 | 53 | 9.37 |
| *Borrelia afzelii* PKo | 1 | 905471 | 905471 | 28.32 | 1143 | 3 | 19.16 |
| *Borrelia burgdorferi* B31 | 1 | 910724 | 910724 | 28.59 | 1690 | 26 | 23.43 |

human dental pathogen
termite hindgut symbiont
marine sediment or microbial mat
mammalian pathogen
gutless oligochaete symbiont
halophile/thermophile/alkaliphile

**(Legend valid for Tables 1 and 2)**

**Table 2: Distribution of annotated genes into RAST subsystem categories and comparison with related Spirochaetae bacteria;** *OalgS1*, *O. algarvensis* spirochaete draft genome (co-assembly); *Sbaj*, *Spirochaeta bajacaliforniensis* DSM16054; *Scel*, *Spirochaeta cellobiosiphila* DSM17781; Ssma, *Spirochaeta smaragdinae* DSM11293; Sthe, *Spirochaeta thermophila* DSM6578; Safr, *Spirochaeta africana* DSM8902; Salk, *Spirochaeta alkalica* DSM8900; Tazo, *Treponema azotonutricium* ZAS-9; Tden, *Treponema denticola* ATCC 35405; Tpal, *Treponema pallidum subsp. pallidum str.* Nichols; Bafz, *Borrelia afzelii* PKo; Bbur, *Borrelia burgdorferi* B31

| RAST Subsystem features | *OalgS1* | *Sbaj* | *Scel* | *Ssma* | *Sthe* | *Safr* | *Salk* | *Tazo* | *Tden* | *Tpal* | *Bafz* | *Bbur* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cofactors, Vitamins, Prosthetic Groups, Pigments | 69 | 145 | 123 | 143 | 103 | 126 | 121 | 138 | 109 | 39 | 15 | 21 |
| Cell Wall and Capsule | 24 | 91 | 79 | 88 | 47 | 64 | 72 | 55 | 40 | 32 | 43 | 41 |
| Virulence, Disease and Defense | 22 | 76 | 65 | 75 | 51 | 58 | 57 | 70 | 31 | 17 | 21 | 23 |
| Potassium metabolism | 7 | 7 | 11 | 7 | 10 | 14 | 15 | 11 | 21 | 14 | 5 | 4 |
| Phages, Prophages, Transposable elements, Plasmids | 2 | 17 | 2 | 15 | 0 | 8 | 1 | 13 | 0 | 0 | 0 | 0 |
| Membrane Transport | 35 | 120 | 59 | 125 | 55 | 37 | 99 | 97 | 66 | 13 | 4 | 15 |
| Iron acquisition and metabolism | 1 | 18 | 1 | 7 | 5 | 1 | 13 | 7 | 6 | 0 | 0 | 0 |
| RNA Metabolism | 87 | 142 | 138 | 142 | 109 | 108 | 108 | 99 | 114 | 74 | 78 | 82 |
| Nucleosides and Nucleotides | 64 | 123 | 105 | 135 | 60 | 79 | 99 | 85 | 63 | 27 | 33 | 34 |
| Protein Metabolism | 119 | 218 | 222 | 224 | 197 | 207 | 221 | 229 | 221 | 160 | 167 | 164 |
| Cell Division and Cell Cycle | 14 | 29 | 31 | 30 | 31 | 31 | 31 | 36 | 27 | 25 | 26 | 26 |
| Motility and Chemotaxis | 0 | 146 | 136 | 152 | 108 | 123 | 140 | 68 | 64 | 60 | 58 | 58 |
| Regulation and Cell signaling | 14 | 26 | 44 | 32 | 25 | 40 | 25 | 48 | 32 | 9 | 10 | 12 |
| DNA Metabolism | 85 | 83 | 79 | 92 | 66 | 68 | 75 | 89 | 59 | 53 | 34 | 55 |
| Fatty Acids, Lipids, and Isoprenoids | 45 | 94 | 46 | 95 | 53 | 94 | 60 | 49 | 39 | 18 | 37 | 35 |
| Nitrogen Metabolism | 5 | 15 | 21 | 26 | 12 | 13 | 13 | 12 | 2 | 0 | 0 | 0 |
| Dormancy and Sporulation | 2 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 1 | 2 | 2 |
| Respiration | 19 | 68 | 46 | 71 | 36 | 21 | 24 | 32 | 8 | 2 | 9 | 9 |
| Stress Response | 19 | 84 | 89 | 86 | 55 | 65 | 58 | 38 | 28 | 20 | 17 | 15 |
| Metabolism of Aromatic Compounds | 0 | 3 | 2 | 4 | 3 | 0 | 2 | 0 | 1 | 0 | 0 | 0 |
| Amino Acids and Derivatives | 144 | 337 | 265 | 374 | 303 | 204 | 209 | 249 | 70 | 19 | 35 | 32 |
| Sulfur Metabolism | 2 | 20 | 17 | 30 | 22 | 13 | 20 | 23 | 1 | 2 | 1 | 1 |
| Phosphorus Metabolism | 7 | 54 | 45 | 47 | 28 | 46 | 12 | 32 | 10 | 0 | 14 | 14 |
| Carbohydrates | 116 | 601 | 298 | 638 | 228 | 230 | 328 | 292 | 93 | 55 | 55 | 49 |

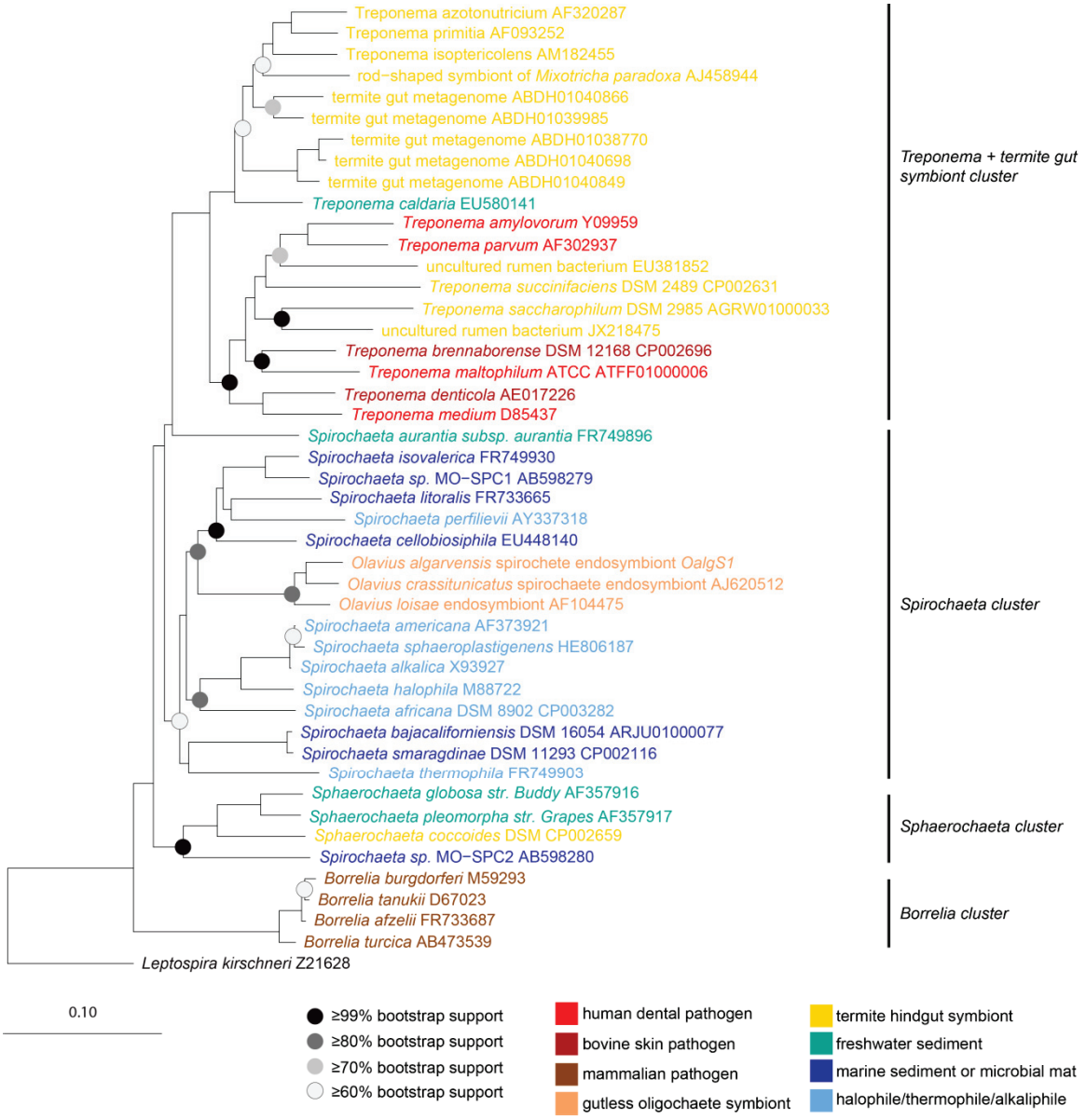



**Figure 2: Phylogeny of *OalgS1* and closely related members of the family Spirochaetaceae.** The phylogenetic tree was constructed using 16S rRNA sequences of Spirochaetaceae from isolates or important host-associated spirochetes (Neighbor Joining, 100 bootstrap replications). Sequences are colored by the habitat they were obtained from.

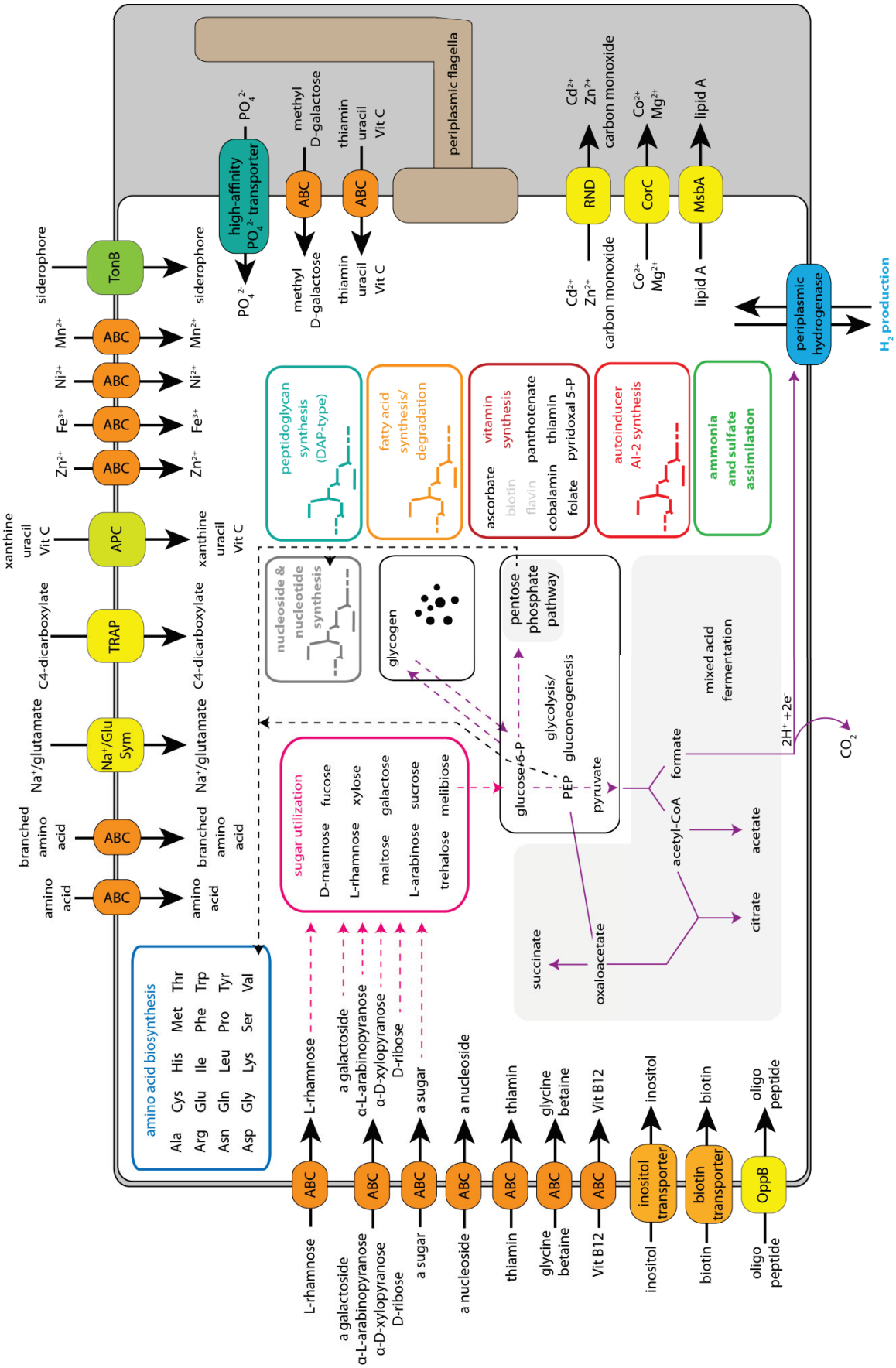
assumed that *OalgS1* is either a mutualist or a commensal, not a pathogen. The placement of *OalgS1* within non-pathogenic sediment bacteria provides further evidence that it is not pathogenic to its host.

**Metabolic insights from the *O. algarvensis OalgS1* symbiont genome**

According to the metabolic pathways encoded in its genome, *OalgS1* is a heterotrophic mixed acid fermenter, capable of taking up multiple carbohydrate substrates from the environment and fermenting them to various short chain fatty acids, like succinate and acetate (Figure 3). Compared to parasites like *Borrelia*, which generally can only utilize fructose and maltose (e.g. *B. burgdorferi* B31, *B. afzelii* Pko, *B. garinii* PBi), *OalgS1* possesses uptake transporters, and the biochemical pathways necessary, to make use of a multitude of different sugars, including mannose, maltose, fucose, xylose, melibiose, sucrose and arabinose (Figure 3). The habitat of *O. algarvensis* is oligotrophic and considered to be depleted in dissolved organic carbon compounds. As a result, hydrogen sulfide, which normally fuels chemosynthetic sulfur-oxidizing symbionts, is rarely detected in the sediment, raising the question of which energy sources feed this symbiosis. It was shown that the deltaproteobacterial symbionts are sulfate reducers which provide the sulfur-oxidizing symbionts with reduced sulfur compounds in an internal syntrophic sulfur cycle [28]. In addition, the gamma- and deltaproteobacterial symbionts are able to assimilate and recycle energy-rich metabolic waste products of the host, reducing the energy demand of the symbiosis [12]. Although this a perfect adaptation to the oligotrophic environment of the host, the symbiosis still ultimately requires external energy sources. Recently, environmental hydrogen and carbon monoxide derived from the degradation of seagrass rhizome were shown to serve as significant alternative energy sources for some of the symbionts. If *OalgS1* is able to take up energy rich dissolved carbon compounds like sugars from the environment, it might provide an additional source of energy to the symbiosis and reduce the need for costly inorganic carbon fixation. Although the sediments are considered to be

oligotrophic, it was recently shown that the pore water inhabited by the worms does contain surprising amounts of dissolved organic compounds in the µM range, in particular sucrose and inositol (Manuel Liebeke, Erik Puschkas, unpublished data). In addition, the cuticle of *O. algarvensis* is permeable for compounds with a molecular weight of at least up to 70 kDa [11]. Since the *OalgS1* genome encodes many different uptake transporters for various energy-rich compounds, in particular sugars, it might be able to scavenge the breakdown products derived from the degradation of seagrass material by xylophagous environmental bacteria. The spirochaete symbiont would support the energetic needs of the symbiosis by fermenting these compounds to short chain fatty acids, like acetate and succinate, which were shown to be used as carbon and electron sources by the deltaproteobacterial symbionts. Another observation that supports this hypothesis is that the spirochaete symbiont is mostly located directly underneath the cuticle of the host, i.e. in a position with the most direct access to pore water metabolites [10]. The *OalgS1* genome also encodes a periplasmic [Fe] hydrogenase, most similar to those found in other spirochaetes, suggesting that it is maintaining redox balance within the cell by releasing molecular hydrogen. The fermentation of carbohydrates to mixed acids and $CO_2$ + $H_2$ is a common metabolic strategy among the closely related environmental spirochaetes isolated from marine sediments [29, 30, 3]. This hydrogen could in turn also be used as energy source by the deltaproteobacterial symbionts. The $CO_2$ released during hydrogen formation might help to increase the concentration of inorganic carbon for autotrophic fixation by the sulfur-oxidizing

**Figure 3: Schematic overview of main metabolic functions in *OalgS1* (next page).** *OalgS1* is capable of synthesizing almost all of its amino acids and vitamins and co-factors (no shown) itself and possesses full pathways for the synthesis of nucleotides and nucleosides, fatty acids, and peptidoglycan. *OalgS1* is able to take up plenty of different sugars from the environment (multiple copies of transporters of the same type shown as one) and ferments them to mixed acids. *OalgS1* can release excess reducing power as hydrogen, or oxidize available environmental hydrogen via a periplasmic [Fe]-hydrogenase shared with other spirochetes. *OalgS1* can build up glycogen as a carbon and energy storage molecule and can use quorum sensing. Dashed lines indicate multiple reaction steps that are not shown.


periplasmic flagella
high-affinity PO4 2- transporter
PO4 2-
methyl D-galactose
thiamin uracil Vit C
Cd2+ Zn2+ carbon monoxide
Co2+ Mg2+
lipid A
RND
CorC
MsbA
periplasmic hydrogenase
H2 production
siderophore
TonB
Mn2+
Ni2+
Fe3+
Zn2+
ABC
xanthine uracil Vit C
APC
C4-dicarboxylate
TRAP
Na+/glutamate
Na+/Glu Sym
branched amino acid
amino acid
peptidoglycan synthesis (DAP-type)
fatty acid synthesis/ degradation
vitamin synthesis
ascorbate biotin flavin cobalamin folate panthotenate thiamin pyridoxal 5-P
autoinducer AI-2 synthesis
ammonia and sulfate assimilation
nucleoside & nucleotide synthesis
glycogen
pentose phosphate pathway
glucose-6-P
glycolysis/ gluconeogenesis
PEP
pyruvate
sugar utilization
D-mannose fucose
L-rhamnose xylose
maltose galactose
L-arabinose sucrose
trehalose melibiose
mixed acid fermentation
formate
acetyl-CoA
acetate
citrate
succinate
oxaloacetate
2H+ +2e-
CO2
amino acid biosynthesis
Ala Cys His Met Thr
Arg Glu Ile Phe Trp
Asn Gln Leu Pro Tyr
Asp Gly Lys Ser Val
L-rhamnose
a galactoside
α-L-arabinopyranose
α-D-xylopyranose
D-ribose
a sugar
a nucleoside
thiamin
glycine betaine
Vit B12
inositol
biotin
oligo peptide
inositol transporter
biotin transporter
OppB

symbionts. Since molecular hydrogen occurs naturally in the host environment, and is used by the deltaproteobacterial symbionts to gain energy, hydrogen might also be used to generate a proton motive force and synthesize ATP via a V-type ATPase (see below) in the spirochaete symbiont. In summary, the spirochaetal symbiont might contribute to the energy budget of the symbiosis by fermenting environmental carbohydrates and sugar alcohols and forming a syntrophic relationship with the deltaproteobacterial symbionts, which consume its fermentation end products (cross-feeding).

In contrast to other symbionts with a reduced genome among the Spirochaetaceae (mainly reduced *Borrelia* spp.), *OalgS1* has retained all of its capability to synthesize amino acids from precursors derived from glycolysis, the pentose phosphate pathways, and ammonia assimilation (Figure 3). Still, the *OalgS1* genome encodes several high-affinity uptake transporters for amino acids and glycine betaine, similar to the deltaproteobacterial symbionts of *O. algarvensis*. Both symbiont groups might therefore be able to recycle nitrogenous wastes of the host. No indication for $N_2$ fixation, which many free-living and symbiotic spirochetes are capable of, was found. This is quite surprising, as the pore water in the worm's habitat is very low in nitrogenous compounds [31], and the ability to fix atmospheric nitrogen should be advantageous in this context. On the other hand, nitrogen fixation is highly energy demanding and might not be sustainable in an energy-limited environment. The low amounts of dissolved nitrogen compounds found in the pore water in combination with the ability of the symbionts to efficiently recycle nitrogenous waste of the host, might be sufficient sources of nitrogen for the symbiosis. The *OalgS1* genome encodes a high-affinity phosphate uptake transporter, which would allow sequestering inorganic phosphate from the phosphate-poor sediment pore water. Like other spirochetes, *OalgS1* lacks a respiratory electron transport chain and gains energy via substrate level phosphorylation. Membrane potential is achieved by the activity of a V-type ATPase, which is unusual for eubacteria (V-type ATPases are typical of eukaryotes). This kind of

V-type ATPase was also identified in other spirochetes (*Treponema pallidum, Borrelia burgdorferi*), and it has been hypothesized that this kind of ATPase is a particular spirochaetal feature [32].

Interestingly, the genome of *OalgS1* encodes a rubredoxin and rubrerythrin (Supplementary Table 5), which are both proteins that contribute to oxidative stress tolerance in anaerobic bacteria [33]. Environmental *Spirochaeta* isolates are mostly anaerobes, but some, like *Spirochaeta perfilievii* are aerotolerant [34, 35]. The spirochaetes that tolerated microoxic conditions were isolated from bacterial sulfur mats, an environment that could, in many ways, be considered similar to the chemosynthetic consortium in *O. algarvensis*. Since the symbiont community of *O. algarvensis* encompasses both aerobic sulfur-oxidizers and anaerobic sulfur-oxidizers and sulfate-reducers, the worm is thought to shuttle between oxic and anoxic layers of the sediment to provide its aerobic symbionts and itself with oxygen. Therefore, these proteins might protect the *OalgS1* symbiont from oxidative stress during oxic phases.

**Molecular interactions between *OalgS1* and its host *O. algarvensis***

*OalgS1* is able to produce the quorum sensing autoinducer AI-2, indicating that it is capable of cell-density regulated responses, which often play a role in host colonization in pathogenic as well as mutualistic symbioses [36, 37, 38]. It also encodes a number of two-component regulatory and other response signaling proteins (Supplementary Table 5), indicating that it is able to respond to a variety of different stimuli within the host environment. Apart from genes encoding for classical MAMPs (microbe-associated molecular patterns), like flagellin and peptidoglycan, we identified a large number of genes that are implied or were shown to be important for host colonization and host-microbe interaction in pathogenic, commensal and mutualistic symbiosis in the *OalgS1* genome. The identified genes are summarized in Supplementary Table 5, and their putative function in *OalgS1* is discussed below. Interestingly,

like other host-associated spirochaetes, e.g. *Treponema* and *Borrelia*, *OalgS1* is not able to synthesize LPS.

**Leucine-rich repeat proteins**

The *OalgS1* genome encodes a leucine-rich repeat (LRR) protein which was highly similar to LRR proteins of other spirochaetes (Supplementary Table 5). LRR proteins were shown to play a role in mediating binding of *Treponema denticola* to other bacteria and epithelial cells of the host [39]. Likewise, *OalgS1* LRR protein might contribute to host and symbiont binding in *O. algarvensis*, which might play a role in host colonization and syntrophic metabolic interactions between the spirochaete symbiont and the deltaproteobacterial symbionts, which are able to metabolize its fermentation end products.

**Lipoproteins**

Lipoproteins are extremely abundant in pathogenic spirochaetes, ranging from 36 (*Borrelia garinii*) to 217 (*Leptospira interrogans* serovar Copenhageni) different lipoproteins encoded per genome [40]. The function of most of these lipoproteins is unclear, but a few were shown to play a role in pathogenesis, while others have roles in cell physiology [41]. Only three lipoproteins were identified in *OalgS1*, all of them most similar to other spirochaetal lipoproteins of unknown function. Although a more extensive search might reveal more candidates, it seems clear that the number of lipoproteins in *OalgS1* is much lower than is common in pathogenic spirochaetes. A reduced number of lipoproteins might be related to reduced virulence in spirochaetes; however, this hypothesis requires proper analysis of all available spirochaete genomes in order to reliably detect trends between pathogenic and non-pathogenic spirochetes. *OalgS1* encodes most of the necessary components for the transport of lipoproteins to the periplasmic space and subsequent translocation to the outer membrane (Apolipoprotein N-acyltransferase, lipoprotein ABC transporter ATP-binding protein, lipoprotein ABC transporter permease, lipoprotein signal

peptidase II, prolipoprotein diacylglyceryl transferase, SecA, SecD, SecF, SecY), indicating that lipoproteins do fulfill an important (but unknown) role in *OalgS1*.

**Tetratricopeptide repeat proteins (TPR), Ankyrin repeat protein and lectin**

The *OalgS1* genome encodes 14 different TPR proteins, one Ankyrin repeat protein, and a lectin-domain containing protein. All these proteins mediate protein-protein or carbohydrate-protein interaction and are implicated or shown to play important roles in bacteria-host interactions in mutualistic symbiotic and pathogenic systems [42, 6, 43]. In *OalgS1*, these proteins are promising candidates for elucidating the mechanisms that aid in host colonization and interaction.

**Outer membrane/surface proteins and antigens**

Several genes were present in the *OalgS1* genome that encode surface structures shown or hypothesized to be involved in recognition by the host's immune system. These include *Borrelia* P83/P100 antigen protein, outer membrane protein OmpA and an unspecified outer membrane surface antigen. P83/P100 is a family of major *Borrelia* antigen proteins of unknown function that is highly specific to *Borrelia* infections and is therefore used as diagnostic markers to detect Lyme disease [44]. The amino acid sequences of these proteins are conserved but also display regions of variability that allow discrimination between different species [45]. Since p83/p100 elicits the humoral production of antibodies in mammals, it must be accessible to the host's immune system, either by being attached to the bacterial surface, or by being released from the cell. As an invertebrate, *O. algarvensis* is not capable of producing antibodies, which is a feature of adaptive immunity; however, it does produce a large amount of lectins and other pattern recognition molecules (Wippler et al. 2016, submitted) which might allow for specific recognition of this symbiont. The actual expression and effect of this protein in the *O. algarvensis* symbiosis remains to be shown experimentally, and warrants further exploration.

The second surface antigen encoded by *OalgS1* contained the conserved bacterial surface antigen domain D15 (IPR000184), which has been shown to be involved in host-pathogen interactions before [46, 47]. Based on sequence similarity, it was most closely related to proteins from other Spirochaetaceae, including free-living environmental, termite-symbiotic and commensal/pathogenic species (data not shown), indicating that this protein is wide-spread among this group and that it is not restricted to pathogenic interactions. Like p83/p100, this protein could play a role in symbiont recognition in *O. algarvensis*.

Outer membrane proteins, like OmpA, are major proteins of the bacterial cell envelope, and often play a role in host-microbe interactions because they are microbe-asssociated molecular patterns that are recognized by the host immune system [48]. OmpA was shown to be a major factor of virulence in many pathogenic bacteria, including spirochaetes [49, 48]. However, it is also an important factor in mutualistic host-microbe associations, like the squid – *Vibrio* light organ symbiosis [50], and the tsetse fly – *Sodalis* nutritional symbiosis [51]. OmpA is also a surface receptor for bacteriophages and might serve as entry point for *OalgS1* phages [52], in addition to being involved in host recognition and interaction.

**Cytoplasmic filament**

*OalgS1* encodes cytoplasmic filament protein CfpA, a protein exclusively found in spirochetes [53, 54]. The concrete function of cytoplasmic filaments is unknown, but is hypothesized to play roles in cell motility, maintenance of cell structure, or cell division [53, 54]. In addition, cytoplasmic filament was shown to be necessary to successfully form multi-species biofilms and colonize the host in *T. denticola* [55] and might therefore play a role in host colonization in *OalgS1* as well.

**The lack of typical major host cell cytotoxic factors of spirochaetes in the *OalgS1* symbiont**

Two major factors have been indentified in pathogenic spirochaetes that contribute to host disease and host cell death: i) Dentilisin, a cell-surface located protease that is able to interefere

with host immune signaling, degrades host cell matrix proteins and is implicated in the penetration of host epithelial cells in *T. denticola* [56, 57], and ii) the Major Sheath Protein Msp which is an abundant outer membrane protein in pathogenic spirochaetes like *T. pallidum* and *T. denticola*, and was shown to have cytotoxic activity against epithelial cells [58, 59]. Neither Msp, nor dentilisin (which is encoded by the three genes prcB, prcA, and prtP), could be identified in the genome of *OalgS1* using blastp searches and the canonical protein sequences as queries. This marked difference might reflect that *OalgS1* is not pathogenic in *O. algarvensis*, and has not retained proteins that would be directly harmful to host cells.

**Mobile genetic elements**

*OalgS1* encoded a single IS1-type transposase in its genome. Transposases are exceptionally abundant in the genomes of the other *O. algarvensis* symbionts, and many of them are also extremely abundantly expressed [12, 60]. These symbionts possess the highest number of transposases ever reported in any bacteria. In this context, it is surprising to find only a single copy of an IS element transposase in *OalgS1*. The high numbers of transposases in the genomes of the other *O. algarvensis* symbionts are thought to be a reflection of recent host restriction [61]. According to this model, free-living bacteria or symbionts with a free-living stage that recently have become obligately host restricted experience rampant transposase multiplication due to factors such as reduced purifying selection and increased genetic drift within the host environment. The model also predicts that these transposases are eventually lost from ancient endosymbionts as a result of genome reduction and lack of intergenomic exchange with other bacteria. The lack of such transposase proliferation in *OalgS1* could either mean that this symbiont has been strictly associated with *O. algarvensis* for a very long time, or that this symbiont also has a free-living stage.

The *OalgS1* genome encoded a prophage (Figure 4), a CRISPR Cas protein (Supplemental Table 4), a CRISPR array (Supplementary Table 6) with 18 CRISPR repeats (two different repeats) and 17

(phage-derived) CRISPR spacers, and other isolated genes encoding phage proteins (Supplementary Table 5). This confirms that infections by bacteriophages have happened multiple times in the evolutionary history of this symbiont. One of the spacer sequences perfectly matches segments of the prophage shown in Figure 4, indicating that this phage has activated the symbiont's CRISPR-Cas defense system at least once. The metagenomic read coverage of the phage-encoding region is much higher compared to the rest of the surrounding genomic region in one out of the six sequenced worms (Figure 4). The likely explanation for this observation is that there were many more physical DNA copies of this phage present in that worm specimen, indicating that this phage was active and had entered its lytic phase. This also means that the phage's genome is made of dsDNA, since the method for preparing our sequencing libraries excludes ssDNA and RNA molecules. Interestingly, no reads mapped to a part of the gene coding for the phage tail protein in the other five specimens, indicating that this gene is missing or truncated. It would explain why the phage is not active in those specimens. The presence of mobile genetic elements like phages and transposases in the *OalgS1* genome indicates that *OalgS1* is open to infection with new genetic elements from the environment and supports the hypothesis that *OalgS1* is not obligately restricted to the host. Both types of mobile elements are known to serve as mechanisms for the acquisition and integration of foreign DNA (horizontal gene transfer) into bacterial genomes and might allow new genes to be brought into the symbiosis [62, 63].

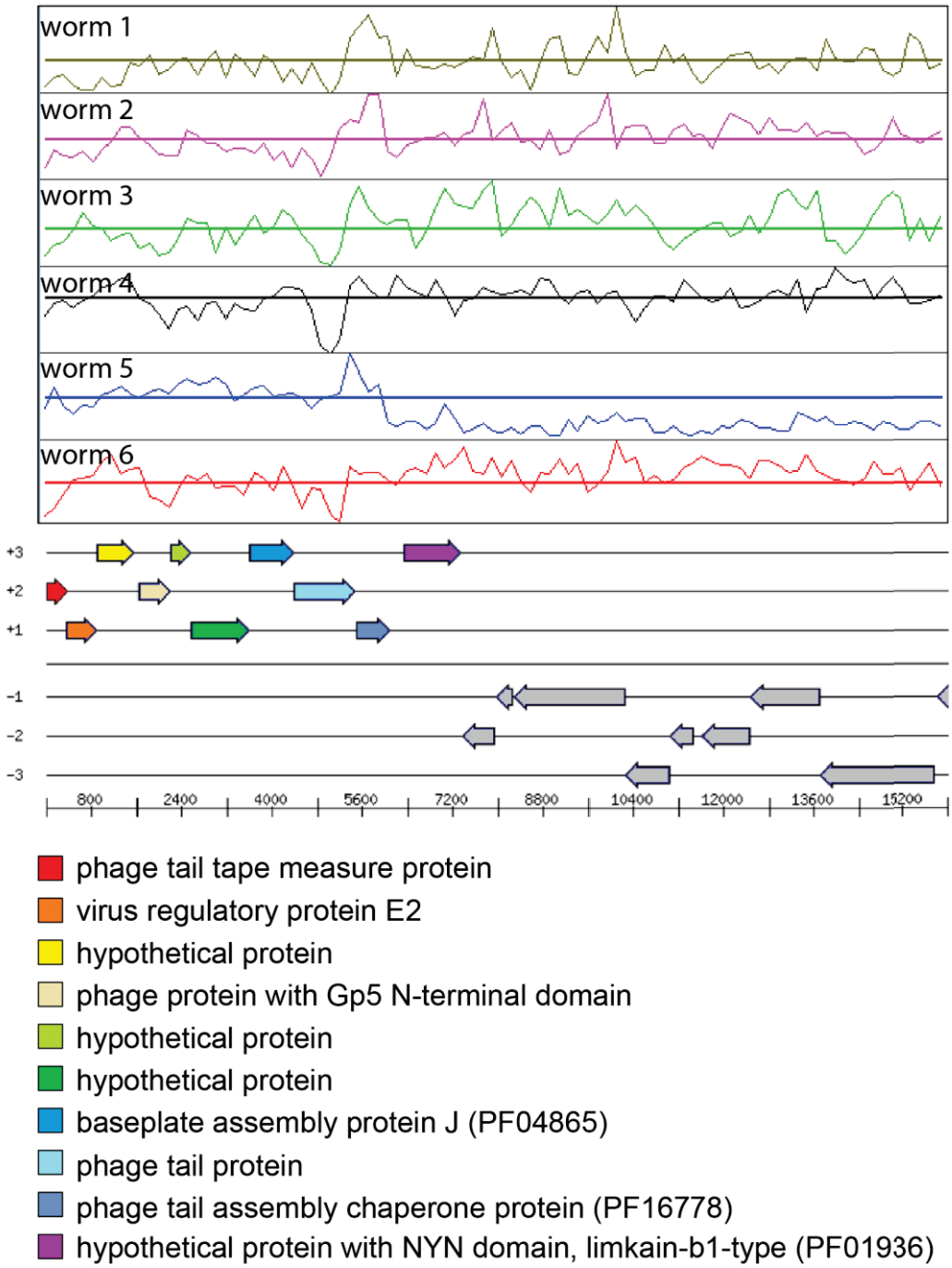



**Figure 4: High read coverage of the prophage encoded in the *OalgS1* genome in one worm specimen indicates active replication.** Top panels show read coverage of the genomic region containing the phage in each metagenome (metagenomes from worm specimens 1 – 6). Note that the coverage of the phage sequences is in the same range as the rest of the genomic region, with the exception of worm 5, where the read coverage of the phage is considerably higher, indicating that the phage was replicating in this specimen. Part of the phage tail protein appears to be missing in all other metagenomes, as its coverage is zero in all metagenomes except the one in which it was lilkey replicating. Truncation of the phage tail gene might explain why the phage was not active in these specimens. Genes econding phage proteins are colored as indicated, spirochaete genes are shown in gray.

**Conclusion**

In this study, we describe the first sequenced genome of a symbiotic spirochaete of the otherwise free-living *Spirochaeta* group. The metabolic functions of this symbiont are highly similar to those of other free-living *Spirochaeta*, and include the fermentation of various carbohydrates, but not amino acids and alcohols, to mixed short chain fatty acids and $CO_2$ + $H_2$. Due to the lack of cytotoxic proteins typical for tissue invading pathogenic spirochaetes, we propose that the spirochaete symbiont of *O. algarvensis* is a mutualistic symbiont that engages in a syntrophic cross-feeding relationship with the hydrogen and acetate/succinate oxidizing sulfate-reducing symbionts. The presence of uptake transporters for a variety of sugars, some of which are found in the surrounding pore water, and the symbiont's localization directly underneath the host's permeable cuticle, suggest that the spirochaete uses environmentally derived degradation products of abundant decaying seagrass rhizomes. Future transcriptomic, proteomic and physiological incubation experiments should further elucidate the role and contribution of this spirochaete to the symbiosis.

**Acknowledgements**

Sequencing was conducted by the U.S. Department of Energy Joint Genome Institute and is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. MK was supported by a NSERC Banting Postdoctoral Fellowship. We thank the HYDRA Institute and staff for their support in sample collection and Silke Wetzel for help with filling out customs forms.

The authors have declared no conflict of interest.

**References**

[1] Pavia, C. S. (1994) Overview of the pathogenic spirochetes. *Journal of Spirochetal and Tick-borne Disease* 1, 3–11

[2] Aksenova, H. Y., Rainey, F. A., Janssen, P. H., Zavarzin, G. A., and Morgan, H. W. (1992) *Spirochaeta thermophila sp. nov*., an obligately anaerobic, polysaccharolytic, extremely thermophilic bacterium. *International Journal of Systematic and Evolutionary Microbiology* 42(1), 175–177

[3] Breznak, J. A. and Warnecke, F. (2008) *Spirochaeta cellobiosiphila sp. nov*., a facultatively anaerobic, marine spirochaete. *International Journal of Systematic and Evolutionary Microbiology* 58(12), 2762–2768

[4] Ohkuma, M., Noda, S., Hattori, S., Iida, T., Yuki, M., Starns, D., Inoue, J.-I., Darby, A. C., and Hongoh, Y. (2015) Acetogenesis from $H_2$ plus $CO_2$ and nitrogen fixation by an endosymbiotic spirochete of a termite-gut cellulolytic protist. *Proceedings of the National Academy of Sciences U S A* 112(33), 10224–10230

[5] Margulis, L. and Hinkle, G. (2006) Large symbiotic spirochetes: *Clevelandina*, *Cristispira*, *Diplocalyx*, *Hollandina* and *Pillotina*. In: *The Prokaryotes*, Springer, pp. 971–982

[6] Thomas, T., Rusch, D., DeMaere, M. Z., Yung, P. Y., Lewis, M., Halpern, A., Heidelberg, K. B., Egan, S., Steinberg, P. D., and Kjelleberg, S. (2010) Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *The ISME Journal* 4(12), 1557–1567

[7] Campbell, B. J. and Cary, S. C. (2001) Characterization of a novel spirochete associated with the hydrothermal vent polychaete annelid, *Alvinella pompejana*. *Applied and Environmental Microbiology* 67(1), 110–117

[8] Dubilier, N., Amann, R., Erseus, C., Muyzer, G., Park, S., Giere, O., and Cavanaugh, C. M. (1999) Phylogenetic diversity of bacterial endosymbionts in the gutless marine oligochete *Olavius loisae* (Annelida). *Marine Ecology Progress Series* 178, 271–280

[9] Blazejak, A., Erséus, C., Amann, R., and Dubilier, N. (2005) Coexistence of bacterial sulfide oxidizers, sulfate reducers, and spirochetes in a gutless worm (Oligochaeta) from the Peru Margin. *Applied and Environmental Microbiology* 71(3), 1553–1561

[10] Ruehland, C., Blazejak, A., Lott, C., Loy, A., Erséus, C., and Dubilier, N. (2008) Multiple bacterial symbionts in two species of co-occurring gutless oligochaete worms from Mediterranean sea grass sediments. *Environmental Microbiology* 10(12), 3404–3416

[11] Dubilier, N., Blazejak, A., and Rühland, C. (2006) Symbioses between bacteria and gutless marine oligochaetes. *Progress in Molecular and Subcellular Biology* 41, 251–275

[12] Woyke, T., Teeling, H., Ivanova, N. N., Huntemann, M., Richter, M., Gloeckner, F. O., Boffelli, D., Anderson, I. J., Barry, K. W., Shapiro, H. J., Szeto, E., Kyrpides, N. C., Mussmann, M., Amann, R., Bergin, C., Ruehland, C., Rubin, E. M., and Dubilier, N. (2006) Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* 443(7114), 950–955

[13] Kleiner, M., Wentrup, C., Lott, C., Teeling, H., Wetzel, S., Young, J., Chang, Y.-J., Shah, M., VerBerkmoes, N. C., Zarzycki, J., Fuchs, G., Markert, S., Hempel, K., Voigt, B., Becher, D., Liebeke, M., Lalk, M., Albrecht, D., Hecker, M., Schweder, T., and Dubilier, N. (2012) Metaproteomics of a

gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proceedings of the National Academy of Sciences U S A* 109(19), E1173–E1182

[14] Harrison, P. and Seemann, T. (2010) From high-throughput sequencing read alignments to confident, biologically relevant conclusions with nesoni. Poster presentation, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.509.7046&rep=rep1&type=pdf

[15] Nikolenko, S. I., Korobeynikov, A. I., and Alekseyev, M. A. (2013) Bayeshammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics* 14 Suppl 1, S7

[16] Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. (2012) SPADES: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19(5), 455–477

[17] Peng, Y., Leung, H. C. M., Yiu, S. M., and Chin, F. Y. L. (2012) IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28(11), 1420–1428

[18] Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G. W., and Nielsen, P. H. (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnology* 31(6), 533–538

[19] Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* 25(7), 1043–1055

[20] Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29(8), 1072–1075

[21] Langmead, B. and Salzberg, S. L. (2012) Fast gapped-read alignment with bowtie 2. *Nature Methods* 9(4), 357–359

[22] Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* 16(10), 944–945

[23] Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. (2008) The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9, 75

[24] Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., and Stevens, R. (2014) The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Research* 42(Database issue), D206–D214

[25] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215(3), 403–410

[26] Rodriguez-R, L. M. and Konstantinidis, K. T. (2010) Bypassing cultivation to identify bacterial species. *Microbe*, *9*(3), 111-118

[27] Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Buchner, A., Lai, T., Steppi, S., Jobb, G., Förster, W., et al. (2004) ARB: a software environment for sequence data. *Nucleic Acids Research* 32(4), 1363–1371

[28] Dubilier, N., MÃ¼lders, C., Ferdelman, T., de Beer, D., Pernthaler, A., Klein, M., Wagner, M., Erséus, C., Thiermann, F., Krieger, J., Giere, O., and Amann, R. (2001) Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. *Nature* 411(6835), 298–302

[29] Hespell, R. B. and Canale-Parola, E. (1970) *Spirochaeta litoralis sp. n.*, a strictly anaerobic marine spirochete. *Archiv für Mikrobiologie* 74(1), 1–18

[30] Harwood, C. and Canale-Parola, E. (1983) *Spirochaeta isovalerica sp. nov*., a marine anaerobe that forms branched-chain fatty acids as fermentation products. *International Journal of Systematic and Evolutionary Microbiology* 33(3), 573–579

[31] Nemecky, S. (2008) Benthic degradation rates in shallow subtidal carbonate and silicate sands. Diploma thesis, University of Bremen, Germany

[32] Fraser, C. M., Casjens, S., Huang, W. M., Sutton, G. G., Clayton, R., Lathigra, R., White, O., Ketchum, K. A., Dodson, R., Hickey, E. K., et al. (1997) Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* 390(6660), 580–586

[33] Lumppio, H. L., Shenvi, N. V., Summers, A. O., Voordouw, G., and Kurtz, D. M. (2001) Rubrerythrin and rubredoxin oxidoreductase in *Desulfovibrio vulgaris*: a novel oxidative stress protection system. *Journal of Bacteriology* 183(1), 101–108

[34] Dubinina, G. A., Grabovich, M. I., and Chernyshova, I. I. (2004) The role of oxygen in the regulation of the metabolism of aerotolerant spirochetes, a major component of "thiodendron" bacterial sulfur mats. *Mikrobiologiia* 73(6), 725–733

[35] Dubinina, G., Grabovich, M., Leshcheva, N., Rainey, F. A., and Gavrish, E. (2011) *Spirochaeta perfilievii sp. nov*., an oxygen-tolerant, sulfide-oxidizing, sulfur-and thiosulfate-reducing spirochaete isolated from a saline spring. *International Journal of Systematic and Evolutionary Microbiology* 61(1), 110–117

[36] Winzer, K. and Williams, P. (2001) Quorum sensing and the regulation of virulence gene expression in pathogenic bacteria. *International Journal of Medical Microbiology* 291(2), 131–143

[37] Lupp, C. and Ruby, E. G. (2005) *Vibrio fischeri* uses two quorum-sensing systems for the regulation of early and late colonization factors. *Journal of Bacteriology* 187(11), 3620–3629

[38] Bassler, B. and Miller, M. (2013) *Quorum Sensing*. Springer Berlin Heidelberg

[39] Ikegami, A., Honma, K., Sharma, A., and Kuramitsu, H. K. (2004) Multiple functions of the leucine-rich repeat protein LRRA of *Treponema denticola*. *Infection and Immunity* 72(8), 4619–4627

[40] Setubal, J. C., Reis, M., Matsunaga, J., and Haake, D. A. (2006) Lipoprotein computational prediction in spirochaetal genomes. *Microbiology* 152(1), 113–121

[41] Haake, D. A. (2000) Spirochaetal lipoproteins and pathogenesis. *Microbiology* 146(7), 1491–1504

[42] Al-Khodor, S., Price, C. T., Kalia, A., and Kwaik, Y. A. (2010) Functional diversity of ankyrin repeats in microbial proteins. *Trends in Microbiology* 18(3), 132–139

[43] Chou, L.-F., Chen, Y.-T., Lu, C.-W., Ko, Y.-C., Tang, C.-Y., Pan, M.-J., Tian, Y.-C., Chiu, C.-H., Hung, C.-C., and Yang, C.-W. (2012) Sequence of *Leptospira santarosai serovar shermani* genome and prediction of virulence-associated genes. *Gene* 511(2), 364–370

[44] Ditton, H.-J., Neuss, M., and Zöller, L. (1992) Evidence that *Borrelia burgdorferi* immunodominant proteins p100, p94 and p83 are identical. *FEMS Microbiology Letters* 94(3), 217–220

[45] Fomenko, N., Sabitova, Y. V., Khasnatinov, M., Goltsova, N., Danchinova, G., Bataa, J., Ambed, D., and Stronin, O. (2007) Heterogeneity of the p83/100 gene of *Borrelia* from the *Borrelia burgdorferi sensu lato* complex. *Molecular Genetics, Microbiology and Virology* 22(4), 172–179

[46] Ruffolo, C. G. and Adler, B. (1996) Cloning, sequencing, expression, and protective capacity of the OMA87 gene encoding the *Pasteurella multocida* 87-kilodalton outer membrane antigen. *Infection and Immunity* 64(8), 3161–3167

[47] Loosmore, S. M., Yang, Y., Coleman, D. C., Shortreed, J. M., England, D. M., and Klein, M. H. (1997) Outer membrane protein D15 is conserved among *Haemophilus influenzae* species and may represent a universal protective antigen against invasive disease. *Infection and Immunity* 65(9), 3701–3707

[48] Jeannin, P., Magistrelli, G., Goetsch, L., Haeuw, J.-F., Thieblemont, N., Bonnefoy, J.-Y., and Delneste, Y. (2002) Outer membrane protein a (OmpA): a new pathogen-associated molecular pattern that interacts with antigen presenting cells - impact on vaccine strategies. *Vaccine* 20, A23–A27

[49] Ristow, P., Bourhy, P., da Cruz McBride, F. W., Figueira, C. P., Huerre, M., Ave, P., Saint Girons, I., Ko, A. I., and Picardeau, M. (2007) The OmpA-like protein LOA22 is essential for leptospiral virulence. *PLoS Pathogens* 3(7), e97

[50] Aeckersberg, F., Lupp, C., Feliciano, B., and Ruby, E. G. (2001) *Vibrio fischeri* outer membrane protein OmpU plays a role in normal symbiotic colonization. *Journal of Bacteriology* 183(22), 6590–6597

[51] Weiss, B. L., Wu, Y., Schwank, J. J., Tolwinski, N. S., and Aksoy, S. (2008) An insect symbiosis is influenced by bacterium-specific polymorphisms in outer-membrane protein A. *Proceedings of the National Academy of Sciences U S A* 105(39), 15088–15093

[52] Morona, R., Klose, M., and Henning, U. (1984) *Escherichia coli* K-12 outer membrane protein (OmpA) as a bacteriophage receptor: analysis of mutant genes expressing altered proteins. *Journal of Bacteriology* 159(2), 570–578

[53] You, Y., Elmore, S., Colton, L. L., Mackenzie, C., Stoops, J. K., Weinstock, G. M., and Norris, S. J. (1996) Characterization of the cytoplasmic filament protein gene (cfpA) of *Treponema pallidum subsp. pallidum. Journal of Bacteriology* 178(11), 3177–3187

[54] Izard, J., Samsonoff, W. A., and Limberger, R. J. (2001) Cytoplasmic filament-deficient mutant of *Treponema denticola* has pleiotropic defects. *Journal of Bacteriology* 183(3), 1078–1084

[55] Yamada, M., Ikegami, A., and Kuramitsu, H. K. (2005) Synergistic biofilm formation by *Treponema denticola* and *Porphyromonas gingivalis*. *FEMS Microbiology Letters* 250(2), 271–277

[56] Chi, B., Qi, M., and Kuramitsu, H. K. (2003) Role of dentilisin in *Treponema denticola* epithelial cell layer penetration. *Research in Microbiology* 154(9), 637–643

[57] Miyamoto, M., Ishihara, K., and Okuda, K. (2006) The *Treponema denticola* surface protease dentilisin degrades interleukin-1β (IL-1β), IL-6, and tumor necrosis factor alpha. *Infection and Immunity* 74(4), 2462–2467

[58] Egli, C., Leung, W., Müller, K., Hancock, R., and McBride, B. (1993) Pore-forming properties of the major 53-kilodalton surface antigen from the outer sheath of *Treponema denticola*. *Infection and Immunity* 61(5), 1694–1699

[59] Mathers, D. A., Leung, W. K., Fenno, J. C., Hong, Y., and McBride, B. C. (1996) The major surface protein complex of *Treponema denticola* depolarizes and induces ion channels in HeLa cell membranes. *Infection and Immunity* 64(8), 2904–2910

[60] Kleiner, M., Young, J. C., Shah, M., Verberkmoes, N. C., and Dubilier, N. (2013) Metaproteomics reveals abundant transposase expression in mutualistic endosymbionts. *MBio* 4(3)

[61] Moran, N. A. and Plague, G. R. (2004) Genomic changes following host restriction in bacteria. *Current Opinion in Genetics and Development* 14(6), 627–633

[62] Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.-L., and Brüssow, H. (2003) Phage as agents of lateral gene transfer. *Current Opinion in Microbiology* 6(4), 417–424

[63] Thomas, C. M. and Nielsen, K. M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology* 3(9), 711–721

**Supplementary Table 1: Sequencing statistics**

| Worm | Site | Raw reads ( million pairs) | Processed reads ( million pairs) | Sequencing center | Accession |
|---|---|---|---|---|---|
| A1 | Sant' Andrea | 111 | 55 | JGI DOE | 1021950[a)] |
| A2 | Sant' Andrea | 77 | 38 | JGI DOE | 1021953[a)] |
| A3 | Sant' Andrea | 170 | 84 | JGI DOE | 1021956[a)] |
| B1 | Sant' Andrea | 84 | 42 | JGI DOE | 1021959[a)] |
| B2 | Sant' Andrea | 107 | 53 | JGI DOE | 1021962[a)] |
| B3 | Sant' Andrea | 90 | 44 | JGI DOE | 1021965[a)] |

**Supplementary Table 2:** Assembly statistics for *OalgS1* genomes assembled from individual *O. algarvensis* worm specimens (A1-3, G1-3) and co-assembly from six *O. algarvensis* specimens (*OalgS1*).

| Assembly | A1 | A2 | A3 | G1 | G2 | G3 | *OalgS1* |
|---|---|---|---|---|---|---|---|
| # contigs (all) | 18 | 32 | 21 | 29 | 18 | 35 | 301 |
| # contigs (>=1 kb) | 17 | 29 | 19 | 28 | 17 | 28 | 301 |
| Total length (all) | 963155 | 180912 | 914391 | 549950 | 1138222 | 117016 | 2182993 |
| Total length (>=1 kb) | 962660 | 179172 | 913665 | 549454 | 1137727 | 111884 | 2182993 |
| Largest contig | 230242 | 19411 | 177514 | 46209 | 229499 | 15007 | 228402 |
| GC (%) | 46.23 | 46.26 | 46.35 | 46.39 | 46.44 | 46.9 | 45.97 |
| N50 | 80079 | 10270 | 134743 | 29651 | 117451 | 5719 | 90077 |
| N75 | 60010 | 4433 | 50034 | 19384 | 75935 | 2631 | 26436 |
| L50 | 4 | 6 | 3 | 8 | 4 | 7 | 9 |
| L75 | 7 | 13 | 7 | 14 | 7 | 15 | 18 |
| # Ns per 100 kbp | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Supplementary Table 3:** ANI (average nucleotide identity) between *OalgS1* draft assemblies from individual worm specimens.

| specimen | A1 | A2 | A3 | G1 | G2 | G3 |
|---|---|---|---|---|---|---|
| **A1** | X | 99.75 | 99.74 | 99.75 | 99.83 | 99.82 |
| **A2** | X | X | 99.81 | 99.77 | 99.75 | 99.77 |
| **A3** | X | X | X | 99.84 | 99.75 | 99.83 |
| **G1** | X | X | X | X | 99.77 | 99.85 |
| **G2** | X | X | X | X | X | 99.80 |
| **G3** | X | X | X | X | X | X |

**Supplementary Table 4: *OalgS1* flagellum synthesis proteins, not in RAST subsystems.**

| Feature ID | Start | Stop | Length (bp) | Function |
|---|---|---|---|---|
| fig\|260710.3.peg.375 | 90072 | 89725 | 348 | Flagellar protein FlgJ [peptidoglycan hydrolase] (EC 3.2.1.-) |
| fig\|260710.3.peg.376 | 90878 | 90084 | 795 | Flagellar basal-body rod protein FlgG |
| fig\|260710.3.peg.377 | 91763 | 90957 | 807 | Flagellar basal-body rod protein FlgF |
| fig\|260710.3.peg.573 | 103807 | 103382 | 426 | Flagellar biosynthesis protein FliS |
| fig\|260710.3.peg.632 | 3763 | 2795 | 969 | flagellar filament outer layer protein |
| fig\|260710.3.peg.713 | 12475 | 11603 | 873 | Flagellar synthesis regulator FleN |
| fig\|260710.3.peg.714 | 13710 | 12478 | 1233 | Flagellar biosynthesis protein FlhF |
| fig\|260710.3.peg.715 | 15737 | 13710 | 2028 | Flagellar biosynthesis protein FlhA |
| fig\|260710.3.peg.716 | 16947 | 15790 | 1158 | Flagellar biosynthesis protein FlhB |
| fig\|260710.3.peg.717 | 17753 | 16956 | 798 | Flagellar biosynthesis protein FliR |
| fig\|260710.3.peg.718 | 18004 | 17750 | 255 | Flagellar biosynthesis protein FliQ |
| fig\|260710.3.peg.719 | 18826 | 18029 | 798 | Flagellar biosynthesis protein FliP |
| fig\|260710.3.peg.721 | 20621 | 19503 | 1119 | Flagellar motor switch protein FliN |
| fig\|260710.3.peg.722 | 21646 | 20621 | 1026 | Flagellar motor switch protein FliM |
| fig\|260710.3.peg.723 | 22231 | 21665 | 567 | Flagellar biosynthesis protein FliL |
| fig\|260710.3.peg.725 | 23080 | 22334 | 747 | Flagellar motor rotation protein MotB |
| fig\|260710.3.peg.726 | 23863 | 23084 | 780 | Flagellar motor rotation protein MotA |
| fig\|260710.3.peg.728 | 25536 | 24124 | 1413 | Flagellar hook protein FlgE |
| fig\|260710.3.peg.729 | 25986 | 25552 | 435 | Flagellar basal-body rod modification protein FlgD |
| fig\|260710.3.peg.731 | 27841 | 27203 | 639 | Flagellar protein FlbB |
| fig\|260710.3.peg.734 | 30514 | 29612 | 903 | Flagellar assembly protein FliH |
| fig\|260710.3.peg.735 | 31603 | 30524 | 1080 | Flagellar motor switch protein FliG |
| fig\|260710.3.peg.736 | 33317 | 31605 | 1713 | Flagellar M-ring protein FliF |
| fig\|260710.3.peg.738 | 34166 | 33708 | 459 | Flagellar basal-body rod protein FlgC |
| fig\|260710.3.peg.739 | 34583 | 34173 | 411 | Flagellar basal-body rod protein FlgB |
| fig\|260710.3.peg.907 | 16411 | 14873 | 1539 | Flagellar synthesis regulator FleN |
| fig\|260710.3.peg.1007 | 17731 | 19722 | 1992 | Flagellar hook-associated protein FliD |
| fig\|260710.3.peg.1013 | 23480 | 23025 | 456 | Flagellar assembly factor FliW |
| fig\|260710.3.peg.1014 | 24666 | 23488 | 1179 | Flagellar hook-associated protein FlgL |
| fig\|260710.3.peg.1015 | 26611 | 24734 | 1878 | Flagellar hook-associated protein FlgK |
| fig\|260710.3.peg.1873 | 117407 | 116715 | 693 | flagellar filament outer layer protein FlaA, putative |
| fig\|260710.3.peg.1874 | 118189 | 117410 | 780 | flagellar filament outer layer protein FlaA, putative |
| fig\|260710.3.peg.1900 | 147393 | 146194 | 1200 | Flagellar motor switch protein FliG |

**Supplementary Table 5: Genes in *OalgS1* potentially involved in symbiotic interactions.**

| ID | Annotation | blast hit description | blast hits accession | organism | e-value | aa% | conserved domains |
|---|---|---|---|---|---|---|---|
| **Lipoproteins** | | | | | | | |
| peg.1352 | treponemal lipoprotein | lipoprotein | WP_010257069.1 | *Treponema primitia* | 3.51E-72 | 50.90% | no conserved domain identified |
| peg.954 | treponemal lipoprotein | putative lipoprotein | WP_013968214.1 | *Treponema caldarium* | 6.20E-50 | 48.30% | SH3-like domain, bacterial-type: PF08239 (PFAM) |
| peg.1028 | spirochaetal lipoprotein | lipoprotein | WP_024267541.1 | *Salinispira pacifica* | 1.45E-38 | 49.10% | Six-bladed beta-propeller, TolB-like: SSF101898 (SUPERFAMILY) |
| **Tetratricopeptide repeat proteins** | | | | | | | |
| peg.673 | Tetratricopeptide repeat protein | tetratricopeptide repeat protein,cyclic nucleotide-binding protein | AFG37963.1 | *Spirochaeta africana* | 1.03E | 55.90% | Tetratricopeptide repeat: PS50005 (PROSITE_PROFILES) |
| peg.943 | tetratricopeptide repeat protein | putative tetratricopeptide TPR_2 repeat protein | AEF86907.1 | *Treponema primitia ZAS-2* | 1.05E-43 | 49.30% | Tetratricopeptide repeat: PS50005 (PROSITE_PROFILES) |
| peg.957 | tetratricopeptide repeat protein | tetratricopeptide repeat protein | EEV19456.1 | *Treponema vincentii* | 1.70E-57 | 54.60% | Tetratricopeptide repeat: PS50005 (PROSITE_PROFILES) |
| peg.19 | tetratricopeptide repeat domain | tetratricopeptide repeat domain protein | AEF82438.1 | *Treponema azotonutricium ZAS-9* | 9.23E-70 | 55.80% | Tetratricopeptide repeat: PS50005 (PROSITE_PROFILES) |
| peg.313 | tetratricopeptide repeat protein | Tetratricopeptide TPR_2 repeat-containing protein | AEJ61275.1 | *Spirochaeta thermophila* | 4.68E-58 | 46.80% | Tetratricopeptide repeat: PS50005 (PROSITE_PROFILES) |
| peg.993 | tetratricopeptide repeat protein | TPR repeat-containing protein | ADK80642.1 | *Spirochaeta smaragdinae* | 2.35E-53 | 64.20% | Tetratricopeptide repeat: PS50005 (PROSITE_PROFILES) |
| peg.1092 | tetratricopeptide repeat protein | tetratricopeptide repeat protein | EEV20958.1 | *Treponema vincentii* | 7.56E-50 | 65.00% | Tetratricopeptide repeat: PS50005 (PROSITE_PROFILES) |
| peg.1837 | tetratricopeptide repeat protein | Tetratricopeptide TPR_2 repeat protein | ADK81198.1 | *Spirochaeta smaragdinae* | 4.86E-42 | 50.10% | Tetratricopeptide repeat: PS50005 (PROSITE_PROFILES) |
| peg.1846 | tetratricopeptide repeat protein | TPR repeat-containing protein | ADK81180.1 | *Spirochaeta smaragdinae* | 1.37E-31 | 61.10% | Tetratricopeptide repeat: PS50005 (PROSITE_PROFILES) |
| peg.1859 | tetratricopeptide repeat domain | Tetratricopeptide TPR_2 repeat-containing protein | AEJ19512.1 | *Treponema caldarium* | 3.18E-57 | 58.30% | Tetratricopeptide repeat: PS50005 (PROSITE_PROFILES) |
| peg.1308 | tetratricopeptide repeat protein | TPR repeat | AAZ71514.1 | *Methanosarcina barkeri* | 3.51E-14 | 48.40% | Tetratricopeptide repeat: PS50005 (PROSITE_PROFILES) |
| peg.97 | tetratricopeptide repeat protein | Tetratricopeptide TPR_1 repeat-containing protein | ADL55129.1 | *Gallionella capsiferriformans* | 8.09E-13 | 47.90% | Tetratricopeptide repeat: PS50005 (PROSITE_PROFILES) |
| peg.988 | Tetratricopeptide repeat protein | tetratricopeptide repeat family protein | AAX17048.1 | *Borrelia hermsii* | 1.49E-05 | 49.70% | Tetratricopeptide repeat: PS50005 (PROSITE_PROFILES) |
| peg.1040 | tetratricopeptide repeat protein | tetratricopeptide repeat family | GAK49871.1 | *bacterium UASB14* | 1.81E-15 | 38.00% | Tetratricopeptide repeat: PS50005 (PROSITE_PROFILES) |
| **Ankyrin repeat proteins** | | | | | | | |
| peg.1806 | ankyrin repeat protein | Ankyrin | AEJ19232.1 | *Treponema caldarium* | 5.45E-108 | 51.20% | Ankyrin repeat: PF00023 (PFAM), 15x PS50088 (PROSITE_PROFILES) |

| ID | Annotation | blast hit description | blast hits accession | organism | e-value | aa% | conserved domains |
|---|---|---|---|---|---|---|---|
| **Leucine-rich repeat proteins** | | | | | | | |
| peg.1429 | leucine-rich repeat protein | hypothetical protein | WP_020602485.1 | *Spirosoma spitsbergense* | 2.82E-16 | 56.00% | Leucine-rich repeats: PF13855 (PFAM), 5x PS51450 (PROSITE_PROFILES) |
| **Lectins** | | | | | | | |
| peg.1385 | Concanavalin A-like lectin domain containing protein | fibronectin type III domain-containing protein | WP_013970102.1 | *Treponema caldarium* | 1.44E-23 | 41.60% | Concanavalin A-like lectin/glucanase domain: PF13385 (PFAM) |
| **Outer membrane/surface proteins and antigens** | | | | | | | |
| peg.667 | Borrelia P83/P100 antigen protein | P83/100 family protein | WP_013758634.1 | *Treponema brennaborense* | 1.91E-76 | 51.20% | Borrelia P83/100: PF05262 (PFAM) |
| peg.524 | outer membrane protein OmpA | cell envelope biogenesis protein OmpA | WP_052299776.1 | *Treponema primitia* | 5.92E-71 | 54.30% | Outer membrane protein, OmpA/MotB, C-terminal: PF00691 (PFAM) |
| peg.1721 | outer membrane surface antigen | surface antigen | WP_013313911.1 | *Spirochaeta thermophila* | 0 | 55.70% | Bacterial surface antigen (D15): PF01103 (PFAM) |
| **sensory and signaling proteins** | | | | | | | |
| peg.218 | adenylate cyclase | adenylate cyclase | WP_028972611.1 | *Spirochaeta cellobiosiphila* | 7.91E-129 | 59.50% | Adenylyl cyclase class-3/4/guanylyl cyclase: PF00211 (PFAM) |
| peg.1614 | adenylate guanylate cyclase catalytic domain | adenylate/guanylate cyclase catalytic domain protein | WP_020776043.1 | *Leptospira meyeri* | 1.09E-112 | 58.50% | Adenylyl cyclase class-3/4/guanylyl cyclase: PF00211 (PFAM), HAMP domain: PF00672 (PFAM) |
| peg.1845 | adenylate cyclase with CHASE2 sensor | adenylate/guanylate cyclase with Chase sensor | ADK81181.1 | *Spirochaeta smaragdinae* | 0 | 61.00% | Adenylyl cyclase class-3/4/guanylyl cyclase: PF00211 (PFAM), CHASE2: PF05226 (PFAM) |
| peg.16 | PAS domain-containing sensor histidine kinase CheY | PAS domain-containing sensor histidine kinase | WP_053335577.1 | *Salinispira pacifica* | 8.25E-115 | 69.10% | Signal transduction histidine kinase, dimerisation/phosphoacceptor domain: PF00512 (PFAM) |
| peg.540 | response regulator with CheY-like receiver, AAA-type ATPase and DNA binding domain | response regulator with CheY-like receiver, AAA-type ATPase, and DNA-binding domains | AEV29832.1 | *Sphaerochaeta pleomorpha* | 4.06E-117 | 70.90% | RNA polymerase sigma factor 54 interaction domain: PF00158 (PFAM), AAA+ ATPase domain: SM00382 (SMART) |
| peg.908 | two component transcriptional regulator LuxR | LuxR family transcriptional regulator | KPK90994.1 | *Anaerolineae bacterium SM23_ 63* | 3.92E-66 | 69.10% | Signal transduction response regulator, receiver domain: PF00072 (PFAM), CheY-like superfamily: SSF52172 (SUPERFAMILY) |
| peg.450 | two-component sensor histidine kinase | two-component sensor histidine kinase | WP_014271234.1 | *Sphaerochaeta pleomorpha* | 2.98E-58 | 53.30% | Signal transduction histidine kinase, dimerisation/phosphoacceptor domain: PF00512 (PFAM) |
| peg.909 | two-component sensor histidine kinase | two-component sensor histidine kinase | WP_054532601.1 | *Herpetosiphon geysericola* | 1.31E-53 | 54.80% | Signal transduction histidine kinase, subgroup 3, dimerisation and phosphoacceptor domain: PF07730 (PFAM) |
| peg.804 | two component sensor histidine kinase | two-component sensor histidine kinase | WP_013253648.1 | *Spirochaeta smaragdinae* | 1.05E-87 | 59.00% | Signal transduction histidine kinase, dimerisation/phosphoacceptor domain: PF00512 (PFAM) |
| peg.602 | 7TM-DISM receptor diguanylate cyclase | 7TM domain sensor diguanylate cyclase | ABR50492.1 | *Alkaliphilus metalliredigens* | 4.90E-55 | 44.60% | 7TM-DISM receptor, extracellular domain, type 1: PF07695 (PFAM) |
| peg.805 | two component system transcriptional regulator | two component transcriptional regulator, winged helix family | ADK80183.1 | *Spirochaeta smaragdinae* | 4.36E-65 | 64.60% | Signal transduction response regulator, receiver domain: PF00072 (PFAM) |

| ID | Annotation | blast hit description | blast hits accession | organism | e-value | aa% | conserved domains |
|---|---|---|---|---|---|---|---|
| peg.1098 | Two component response regulator | Two-component response regulator SA14-24 | AHC14775.1 | *Salinispira pacifica* | 1.15E-93 | 78.90% | Signal transduction response regulator, receiver domain: PF00072 (PFAM) |
| peg.545 | Two component transcriptional regulator LuxR | two-component system regulator | WP_014932028.1 | *Brachyspira pilosicoli* | 3.86E-11 | 47.60% | Transcription regulator LuxR, C-terminal: PF00196 (PFAM) |
| **Drug efflux transporters** | | | | | | | |
| peg.1561 | outer membrane efflux protein | outer membrane efflux protein | WP_013253496.1 | *Spirochaeta smaragdinae* | 1.93E-64 | 57.00% | Outer membrane efflux protein: PF02321 (PFAM) |
| peg.1078 | Multi antimicrobial extrusion protein MATE family efflux transporter | MATE family efflux transporter | WP_037548891.1 | *Spirochaeta sp. JC230* | 5.39E-136 | 67.60% | Multi antimicrobial extrusion protein: PF01554 (PFAM) |
| peg.911 | acriflavin resistance protein efflux transporter | acriflavin resistance protein | WP_013969439.1 | *Treponema caldarium* | 0 | 58.00% | Acriflavin resistance protein: PF00873 (PFAM) |
| peg.1563 | acriflavin resistance protein efflux transporter | acriflavin resistance protein | WP_013252662.1 | *Spirochaeta smaragdinae* | 0 | 71.80% | Acriflavin resistance protein: PF00873 (PFAM) |
| peg.910 | RND efflux pump | RND family efflux transporter MFP subunit | WP_015709472.1 | *Treponema primitia* | 1.68E-30 | 50.30% | RND efflux pump, membrane fusion protein: PF13533 (PFAM), PF13437 (PFAM) |
| **Type II and III secretion system proteins** | | | | | | | |
| peg.1748 | type II secretion protein F | type II secretion system F domain-containing protein | WP_014625169.1 | *Spirochaeta thermophila* | 1.95E-36 | 55.30% | Type II secretion system F domain: PF00482 (PFAM) |
| peg.1747 | type II secretion system protein E | type II secretion system protein E | AEJ61840.1 | *Spirochaeta thermophila* | 1.10E-113 | 62.80% | Type II secretion system protein E: PF00437 (PFAM) |
| peg.1741 | type II secretion system protein G | type II secretion system protein G | EPF47858.1 | *Treponema vincentii* | 2.59E-44 | 66.40% | Type II secretion system protein G: PF08334 (PFAM) |
| peg.1746 | type II and III secretion system protein | type II and III secretion system protein | WP_014625171.1 | *Spirochaeta thermophila* | 3.11E-126 | 56.80% | Type II/III secretion system:PF00263 (PFAM) |
| **Phage-related proteins** | | | | | | | |
| peg.703 | prophage baseplate J protein | baseplate assembly protein | ALJ90069.1 | *Thermus aquaticus* | 8.70E-19 | 46.00% | Baseplate protein J-like: PF04865 (PFAM) |
| peg.657 | CRISPR-associated Cas4 endonuclease | CRISPR-associated endonuclease Cas4/Cas1 | WP_020717016.1 | *Acidobacteriaceae bacterium KBS 96* | 2.23E-132 | 55.70% | CRISPR-associated protein Cas1: PF01867 (PFAM) |
| peg.264 | bacteriophage lambda repressor-like regulator | hypothetical protein | WP_011811261.1 | *Verminephrobacter eiseniae* | 3.46E-126 | 66.30% | Lambda repressor-like, DNA-binding domain: SSF47413 (SUPERFAMILY) |
| peg.1099 | integration host factor subunit beta | integration host factor subunit beta | WP_028975167.1 | *Spirochaeta cellobiosiphila* | 1.75E-40 | 80.60% | Integration host factor (IHF)-like DNA-binding domain: SSF47729 (SUPERFAMILY) |
| peg.1668 | integrase | integron integrase | ABL64322.1 | *Chlorobium phaeobacteroides* | 4.43E-150 | 85.40% | Integrase, catalytic domain: PF00589 (PFAM) |
| peg.705 | phage tail protein | phage tail protein | WP_014486594.1 | *Brachyspira intermedia* | 3.02E-12 | 52.30% | no conserved domain identified |

| ID | Annotation | blast hit description | blast hits accession | organism | e-value | aa% | conserved domains |
|---|---|---|---|---|---|---|---|
| peg.237 | phage tail sheath protein | Phage tail sheath protein FI-like protein | ACY13287.1 | *Haliangium ochraceum DSM 14365* | 6.40E-154 | 88.30% | Tail sheath protein: PF04984 (PFAM) |
| peg.1695 | phage transcriptional regulator AlpA | phage transcriptional regulator, AlpA | WP_015334928.1 | *Desulfovibrio hydrothermalis* | 7.80E-34 | 96.80% | SinI-like, DNA-binding domain: PF12728 (PFAM) |
| **Spirochaetal cytoplasmic filament** | | | | | | | |
| peg.827 | cytoplasmic filament protein A | cytoplasmic filament protein A (cfpA) | AHC15697.1 | *Salinispira pacifica* | 0 | 80.50% | no conserved domain identified |
| **Transposases** | | | | | | | |
| peg.1301 | transposase, IS1 family | IS1 transposase | EFH87042.1 | *Ktedonobacter racemifer* | 7.33E-85 | 73.50% | Transposase, IS1: PF03400 (PFAM) |
| **Aerotolerance** | | | | | | | |
| peg.590 | rubredoxin | rubredoxin | WP_018527119.1 | *Spirochaeta alkalica* | 3.66E-19 | 85.20% | Rubredoxin domain: PF00301 (PFAM) |
| peg.1340 | rubrerythrin | rubrerythrin | WP_013969266.1 | *Treponema caldarium* | 2.26E-94 | 81.90% | Rubrerythrin: PF02915 (PFAM) |

**Supplementary Table 6: Spacer and repeat sequences of the *OalgS1* CRISPR array.** Spacer 7 (bold) matches prophage in Figure 4.

| Identifier | contig | start | stop | Repeat/spacer sequence |
|---|---|---|---|---|
| fig\|260710.4.crispr_repeat.9 | NODE_19575_length_65295_cov_36.561 | 24262 | 24297 | atcgcaccgcttgaaattgagttaattgactgacac |
| fig\|260710.4.crispr_repeat.10 | NODE_19575_length_65295_cov_36.561 | 24332 | 24367 | atcgcaccgcttgaaattgagttaattgactgacac |
| fig\|260710.4.crispr_repeat.11 | NODE_19575_length_65295_cov_36.561 | 24403 | 24438 | atcgcaccgcttgaaattgagttaattgactgacac |
| fig\|260710.4.crispr_repeat.12 | NODE_19575_length_65295_cov_36.561 | 24473 | 24508 | atcgcaccgcttgaaattgagttaattgactgacac |
| fig\|260710.4.crispr_repeat.13 | NODE_19575_length_65295_cov_36.561 | 24544 | 24579 | atcgcaccgcttgaaattgagttaattgactgacac |
| fig\|260710.4.crispr_repeat.14 | NODE_19575_length_65295_cov_36.561 | 24614 | 24649 | atcgcaccgcttgaaattgagttaattgactgacac |
| fig\|260710.4.crispr_repeat.15 | NODE_19575_length_65295_cov_36.561 | 24686 | 24721 | atcgcaccgcttgaaattgagttaattgactgacac |
| fig\|260710.4.crispr_repeat.16 | NODE_19575_length_65295_cov_36.561 | 24757 | 24792 | atcgcaccgcttgaaattgagttaattgactgacac |
| fig\|260710.4.crispr_repeat.17 | NODE_19575_length_65295_cov_36.561 | 24828 | 24863 | gtcgcaccgcttgaaattgagttaattgactgacac |
| fig\|260710.4.crispr_repeat.18 | NODE_19575_length_65295_cov_36.561 | 24899 | 24934 | gtcgcaccgcttgagaaggaaatcgcagattgatca |
| fig\|260710.4.crispr_spacer.1 | NODE_19575_length_65295_cov_36.561 | 23726 | 23759 | aaagattcttccactaatattgatgggtctgtca |
| fig\|260710.4.crispr_spacer.2 | NODE_19575_length_65295_cov_36.561 | 23796 | 23830 | ttcaataacttaccccagacacaggccgatagtcg |
| fig\|260710.4.crispr_spacer.3 | NODE_19575_length_65295_cov_36.561 | 23867 | 23901 | gacagcgtggccatcagcgcccgcctcttgatctt |
| fig\|260710.4.crispr_spacer.4 | NODE_19575_length_65295_cov_36.561 | 23938 | 23973 | tgttgcgtcctattcaccctcttgtcggtttatctc |
| fig\|260710.4.crispr_spacer.5 | NODE_19575_length_65295_cov_36.561 | 24010 | 24044 | attttgatgttcgacaaaaagttccgtcgtgtcct |
| fig\|260710.4.crispr_spacer.6 | NODE_19575_length_65295_cov_36.561 | 24081 | 24117 | ttgacataggtgtccaccgcatttttggcggtggtgc |
| **fig\|260710.4.crispr_spacer.7** | **NODE_19575_length_65295_cov_36.561** | **24154** | **24188** | **tgctgctgaatacgattgaatgtgattacgataaa** |
| fig\|260710.4.crispr_spacer.8 | NODE_19575_length_65295_cov_36.561 | 24225 | 24261 | tactgtcgaccgcaacgaccgtatcaatgatgttgcg |
| fig\|260710.4.crispr_spacer.9 | NODE_19575_length_65295_cov_36.561 | 24298 | 24331 | agcgccagttccaccatggcgagttttgccagat |
| fig\|260710.4.crispr_spacer.10 | NODE_19575_length_65295_cov_36.561 | 24368 | 24402 | cttataactgaaggtcaactgcgggcaaacatggc |
| fig\|260710.4.crispr_spacer.11 | NODE_19575_length_65295_cov_36.561 | 24439 | 24472 | tgccggcgcgcggaggctagattggcgctctctc |
| fig\|260710.4.crispr_spacer.12 | NODE_19575_length_65295_cov_36.561 | 24509 | 24543 | gtaatttgacgggacgttacattgcctctagtatc |
| fig\|260710.4.crispr_spacer.13 | NODE_19575_length_65295_cov_36.561 | 24580 | 24613 | tcgatgagctctccgaagtccgccttgaggagtt |
| fig\|260710.4.crispr_spacer.14 | NODE_19575_length_65295_cov_36.561 | 24650 | 24685 | ctagtggttattttggtaccgatctcgcggagccgg |
| fig\|260710.4.crispr_spacer.15 | NODE_19575_length_65295_cov_36.561 | 24722 | 24756 | atttccgttacacgtctaacaaccttgccggagga |
| fig\|260710.4.crispr_spacer.16 | NODE_19575_length_65295_cov_36.561 | 24793 | 24827 | tgccgaatccacaaccggcaggacagtcctaccgc |
| fig\|260710.4.crispr_spacer.17 | NODE_19575_length_65295_cov_36.561 | 24864 | 24898 | gacaaatttttctgtgttgatatttgagttttctt |

# Chapter 4: Transcriptomic and proteomic insights into innate immunity and adaptations to a symbiotic lifestyle in the gutless marine worm *Olavius algarvensis*

**Authors:** Juliane Wippler, Manuel Kleiner, Christian Lott, Alexander Gruhl, Paul E. Abraham, Richard J. Giannone, Jacque C. Young, Robert L. Hettich, Nicole Dubilier

**Publication status:** Submitted to BMC Genomics

**Author's contributions:**

**JW** and **MK** contributed equally to this manuscript. **JW**: Conceived and wrote the manuscript, analyzed and interpreted the proteomic data and provided ideas, did in-depth bioinformatic analyses on proteins of interest, prepared all figures and tables, collected all worms for transcriptomics experiments, conceived and performed all transcriptomics experiments, assembled and annotated the transcriptomes, predicted cds from transcriptome data and did all transcriptome statistics and analyses. **MK**: Conceived the study and the manuscript, edited the manuscript and provided ideas, conceived proteomics experiments, collected worms for proteomics experiments and performed two of the starvation experiments, compiled the protein reference database, did statistical analyses and processed and analyzed the proteomics data. **CL**: Provided ideas and collected worms for proteomics and transcriptomics experiments and performed one of the starvation experiments. **AG**: Provided all microscopic images used in Figure 1, and commented on the manuscript. **PEA**: Processed proteomics data. **RJG**: Processed proteomics data. **JCY**: Provided ideas and ran 2D-LC-MS/MS experiments. **RLH**: Provided access to the proteomics equipment, conceptual input and coordinated the data processing at the ORNL **ND**: Was involved in the organization and coordination of this study, provided ideas and commented on the manuscript. All authors reviewed and revised the final manuscript before submission.

# Titlepage

**Transcriptomic and proteomic insights into innate immunity and adaptations to a symbiotic lifestyle in the gutless marine worm *Olavius algarvensis***

Juliane Wippler[1*], Manuel Kleiner[1,2*], Christian Lott[1,3], Alexander Gruhl[1], Paul E. Abraham[4], Richard J. Giannone[4], Jacque C. Young[4§], Robert L. Hettich[4], Nicole Dubilier[1]

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

[2]Energy Bioengineering and Geomicrobiology Research Group, University of Calgary, Calgary T2N 1N4, AB, Canada

[3]HYDRA Institute for Marine Sciences, Elba Field Station, Via del Forno 80, 57034 Campo nell' Elba (LI), Italy

[4]Oak Ridge National Laboratory, Chemical Sciences Division, Oak Ridge, Tennessee, 1 Bethel Valley Rd, Oak Ridge, TN 37831, USA

[§]Current address: Saul Ewing LLP, 1500 Market Street, 37th Floor, Philadelphia, PA 19102-2186, USA

*contributed equally, corresponding authors

Corresponding authors:

Juliane Wippler: jwippler@mpi-bremen.de, phone: +49 (421) 2028-907, fax: +49 (421) 2028-580

Manuel Kleiner: manuel.kleiner@ucalgary.ca, phone: +1 (403) 210-6617

## Abstract

**Background:** The gutless marine worm *Olavius algarvensis* has a completely reduced digestive and excretory system, and lives in an obligate nutritional symbiosis with bacterial symbionts. While considerable knowledge has been gained of the symbionts, the host has remained largely unstudied. Here, we generated transcriptomes and proteomes of *O. algarvensis* to better understand how this annelid worm gains nutrition from its symbionts, how it adapted physiologically to a symbiotic lifestyle, and how its innate immune system recognizes and responds to its symbiotic microbiota.

**Results:** Key adaptations to the symbiosis include (i) the expression of gut-specific digestive enzymes despite the absence of a gut, most likely for the digestion of symbionts in the host's epidermal cells; (ii) a modified hemoglobin that may bind hydrogen sulfide produced by two of the worm's symbionts; and (iii) the expression of a very abundant protein for oxygen storage, hemerythrin, that could provide oxygen to the symbionts and the host under anoxic conditions. Additionally, we identified a large repertoire of proteins involved in interactions between the worm's innate immune system and its symbiotic microbiota, such as peptidoglycan recognition proteins, lectins, fibrinogen-related proteins, Toll and scavenger receptors, and antimicrobial proteins.

**Conclusions:** We show how this worm, over the course of evolutionary time, has modified widely-used proteins and changed their expression patterns in adaptation to its symbiotic lifestyle and describe expressed components of the innate immune system in a marine oligochaete. Our results provide further support for the recent realization that animals have evolved within the context of their associations with microbes and that their adaptive responses to symbiotic microbiota have led to biological innovations.

**Keywords:** RNA-Seq, Annelida, Oligochaeta, Phallodrilinae, PGRP, FREP, SRCR, respiratory pigment, carbon monoxide, immunology, chemosynthetic symbiosis

# 1 Background

Most, if not all, animals are associated with a species-specific microbial assemblage that profoundly affects their evolution, ecology, development and health [1, 2, 3]. Animals and their microbiota have evolved molecular mechanisms to recognize and maintain these stable associations, and on the host side, these mechanisms are largely mediated by their immune system [4]. The mechanisms that govern host-symbiont interactions have been studied in a number of model organisms [4, 5], but remain unexplored in many animal phyla.

*Olavius algarvensis* is a gutless oligochaete worm (Annelida; Oligochaeta; Phallodrilinae) that lives in an obligate nutritional symbiosis with at least four bacterial species [6]. These extracellular endosymbionts thrive in a dense bacterial layer between the cuticle and the epidermis of the worm (Figure 1). Over the course of their symbiotic evolution, the gutless oligochaetes have lost their digestive and excretory organs, and rely solely on their bacterial symbionts for nourishment and removal of their waste products [7, 8, 9]. *O. algarvensis* harbors two gammaproteobacterial symbiont species that are chemoautotrophic sulfur oxidizers, and two deltaproteobacterial symbionts that are sulfate-reducing bacteria [6]. Together, these symbionts engage in a syntrophic sulfur cycle that enables autotrophic carbon fixation by the sulfur-oxidizing symbionts and provision of organic carbon to the host [8, 9]. Many worm individuals also harbor a spirochaetal symbiont, whose function has not yet been resolved [10].

Metagenomic and metaproteomic studies of the symbionts have revealed much about their metabolic capabilities, highlighted their immense capacity to use and recycle the host's waste products and led to the discovery of novel, energy-efficient pathways to fix both inorganic and organic carbon into biomass [8, 9]. Research aimed at a better understanding of the host, on the other hand, has been hampered by the fact that the worms are very small (0.1 - 0.2 mm in diameter and 10 - 20 mm in length), cannot be cultivated, and by a lack of sequence data. Recent advances in sequencing technology have made it possible to sequence and assemble

comprehensive *de novo* transcriptomes of uncultured, non-model organisms collected in the environment. These transcriptomes provide a refence database for identifying the proteins organisms express using mass-spectrometry-based proteomic approaches. This methodological advance has opened the door for in depth studies of the molecular repertoire used by *O. algarvensis* and other non-cultivable organisms to establish and maintain a successful symbiosis.

All animals employ mechanisms for selecting and maintaining a specific microbial consortium over the course of their lives, while avoiding overgrowth by their own microbiota or infection by detrimental bacteria from the environment. The innate immune system is crucial in the establishment and maintenance of healthy symbiotic interactions, but has so far not been studied in gutless oligochaetes. These hosts face additional challenges because they obligately rely on their symbionts and therefore must provide conditions under which they can thrive, while also dealing with the physiological challenges caused by their symbiotic lifestyle. For example, *O. algarvensis* must be able to live in anoxic sediment layers for extended periods of time to enable sulfate reduction by its anaerobic sulfate-reducing symbionts [8]. Additionally, it must be able to deal with the hydrogen sulfide that is produced during sulfate reduction. It must also be able to endure the relatively high carbon monoxide concentrations in its environment, which both the sulfate-reducing and sulfur-oxidizing symbionts use as an energy source [9, 11]. Another challenge occurs when *O. algarvensis* inhabits the upper oxygenated sediment layers where it competes for oxygen with its aerobic sulfur-oxidizing symbionts.

Here, we used transcriptomics and proteomics to elucidate how *O. algarvensis* fulfills the physiological requirements outlined above and how it obtains nutrition from its symbionts. We exposed worms collected from the environment to two types of conditions that they naturally encounter to increase transcriptome and proteome coverage. Our identification and analysis of proteins expressed by *O. algarvensis* provide insights into their molecular mechanisms for microbe recognition, interaction and regulation, as well as their physiological adaptations to living in symbiosis with sulfur-oxidizing and sulfate-reducing bacteria.

# 2 Methods

**Sample collection and incubations**

For proteomic analyses, sediment that contained gutless oligochaete worms was collected at 7 meters water depth in the Bay of Sant'Andrea, Elba, Italy (42°48'26''N, 010°08'28''E) in October 2007 and 2008. Worms were carefully washed out of the sediment at the HYDRA field station (Fetovaia, Elba, Italy) by hand (for details see [9]). To increase proteome coverage we treated the worms in the following manner. Worms were either immediately frozen in liquid nitrogen in batches of 150-200 worms (called "fresh" worms in the following) or were kept for 8 days in glass petri dishes filled with a thin layer (2-3 mm) of washed sediment and 0.2 µm-filtered sea water and then frozen in liquid nitrogen (called "starved" worms in the following, because no external electron donor for energy conservation and autotrophic carbon fixation was provided). The sulfur-oxidizing symbionts of *O. algarvensis* store large amounts of sulfur and polyhydroxyalkanoate granules, which give the worms a bright white appearance. Under prolonged exposure to oxygen without access to an electron donor like the sulfide produced anaerobically by the sulfate-reducing symbionts, these storage granules become depleted, the worms turn transparent, and are effectively starved of nutrition. Transparent worms are regularly found in the environment, especially during the reproductive season of the worms. All samples were stored in liquid $N_2$ and later at -80°C until further use.

For transcriptomics, 100-120 worms were collected in April 2012 from the same site as for proteomics. The live worms were kept in washed sediment and transported to the lab in Bremen where they were washed out of the sediment again, washed in petri dishes with filtrated seawater, then flash-frozen in liquid nitrogen and stored at -80°C until they were used to prepare the cDNA library "A". A second collection of worms was used for library "B" to identify genes expressed under prolonged anoxia, a condition that the worms often experience. For library "B", 100-120 live worms were collected in March 2013 from the same site as above,

transported to Bremen in the same way as for library "A", and incubated in anoxic serum bottles for 43 hours. Serum bottles were filled with sediment and sea water from Elba, and were flushed with nitrogen to remove oxygen from the headspace. The sediment and sea water were not sterilized, so that fully anoxic conditions could develop quickly through microbial metabolism. Oxygen concentrations were measured at the end of the incubation with an oxygen microelectrode and were below 0.1 µM. Worms were fixed overnight in RNAlater (Thermo Fisher Scientific, Braunschweig, Germany) at 4°C and stored at -80°C until they were used to prepare the cDNA library “B”.

**Illumina library preparation and sequencing**

Total RNA was isolated using peqGOLD TriFast reagent (PEQLAB, Erlangen, Germany) and treated with DNase. Poly(A)+ RNA was isolated from the total RNA, fragmented with ultrasound (2 pulses of 30 sec at 4°C) and used for cDNA synthesis with random hexamer primers. Illumina TruSeq adaptors were ligated to the ends of the cDNA fragments and amplified according to the manufacturer’s instructions (Illumina Inc., USA). Library DNA fragments of 300-500 bp were eluted from a preparative agarose gel and paired-end sequenced on an Illumina HiSeq2000 sequencer (2x 100 bp). We sequenced ~170 million read pairs from library A, and ~6 million read pairs from library B (Supplementary Table 1). For library B, a much smaller number of reads was sequenced because the purpose of this library was to detect abundant transcripts expressed under anoxic conditions.

***De novo* transcriptome assembly and sequence analysis**

The raw reads were trimmed to remove Illumina adapters, filtered for PhiX174 spike-in DNA and quality trimmed with nesoni clip version 0.109 [12]. The cleaned reads were co-assembled *de novo* using Trinity release 2013-02-25 [13]. Transcripts were quantified with RSEM as implemented in Trinity [14].

*De novo* assembled transcripts were annotated with blast2GO [15]. Transcripts of particular interest were searched against the invertebrate division of EST (Expressed Sequence Tags) and TSA (Transcriptome Shotgun Assembly) sequences of NCBI with tblastx [16] to determine their similarity to genes expressed in other annelids.

Hemoglobin sequences were assigned to families, if possible, based on sequence homology and specific conserved amino acid patterns as described in [17]. The secondary structures of the putative sulfide binding domains in *O. algarvensis* hemoglobin chains were predicted with hydrophobic cluster analysis using the program drawhca [18].

**Host species identification**

Three species of gutless oligochaete co-occur at the sampling site and these species are difficult to distinguish reliably under the dissecting scope. Therefore, we used EMIRGE [19] to estimate the relative abundance of the different species in our samples based on the read coverage of the mitochondrial cytochrome c oxidase I (COI) gene. We determined that the contamination with species other than *Olavius algarvensis* was less than 3.5% in library A and less than 0.1 % in library B.

**2D-LC-MS/MS**

Protein was extracted from frozen worms, and peptides prepared as previously described using a single-tube small processing method [20, 9]. We analyzed three biological replicates for each condition (fresh and starved). All samples were analyzed in technical duplicates via 24 hour nano-2D-LC MS/MS with a split-phase column (RP-SCX-RP) [21, 22] on a hybrid linear ion trap-Orbitrap (Thermo Fischer Scientific), as previously described [9].

**Peptide and protein identifications**

Coding sequences (CDS) were predicted from the transcriptomes using FrameDP [23] and getorf of the EMBOSS package using the standard genetic code [24]. Transcriptome CDS were

combined into a single protein sequence database with the symbiont protein sequence database used by Kleiner et al. 2012 [9]. Redundant CDS were removed from the database using CD-HIT (version 4.5.4, [25]. Experimental peptide fragmentation spectra (MS/MS) generated from Xcalibur v.2.0.7 were compared with theoretical peptide fragmentation spectra obtained from the protein sequence database to which protein sequences of common contaminant proteins (e.g., human keratin and trypsin) were added to a total of 1,318,114 entries. To determine the false-discovery rate (FDR), a decoy database, generated by reversing the sequences of the target database, was appended.

MyriMatch v2.1.111 [26] was configured to derive fully-tryptic peptides with the following parameters: 2 missed cleavages, parent mass tolerance of 10 ppm, and a fragment mass tolerance of 0.5 m/z units. For protein inference, peptide identifications were merged together in IDPicker v.3 [27]. Only protein identifications with at least two identified spectra and a maximum q-value of 0.02 were considered for further analysis. The number of distinct peptides required for identifications was set to 1 to allow for the identification of small antimicrobial proteins and/or small, fragmented protein sequences in the transcriptome assembly. Based on these settings, protein-level FDR was < 3% for all samples.

To deal with sequence redundancy, post-search protein grouping was performed by clustering all protein sequences in the protein sequence database by sequence similarity (≥ 90%) using the UCLUST component of the USEARCH v5.0 software platform [28]. As described previously [29], identified proteins were then consolidated into their defined protein groups. Protein groups were represented by the longest protein sequence (i.e., the seed sequence), which shares ≥ 90% sequence similarity to each member of the protein group. Peptide uniqueness was re-assessed and classified as either unique (i.e., only belonging to one protein group) or non-unique (i.e., shared among multiple protein groups). For shared peptides belonging to multiple protein groups, their spectral counts were recalculated based on the proportion of uniquely identified

peptides between the protein groups sharing the peptide. Following spectra balancing, total spectral counts of a protein group were converted to normalized spectra counts (nSpC) [30], which are derived from normalized spectral abundance factors [31]. Relative protein abundances of host proteins are listed in tables as nSpC values multiplied by 10,000 i.e. the sum of all host protein nSpC values in one sample is 10,000 and the nSpC values are thus given as a fraction of 10,000 ($^{0}/_{000}$).

# 3 Results and Discussion

## 3.1 Transcriptome/proteome measurement metrics

To generate our protein sequence database for host protein identification, we sequenced the transcriptomes of untreated whole worms (library A), and of worms kept under anoxic conditions for 43 hours (library B). We chose these two conditions to obtain a larger range of host transcripts and thus improve protein identification. After trimming and error correction, 159,551,509 (library A) and 5,745,537 (library B) read pairs remained, which were co-assembled into 173,602 contigs (Supplementary Table 1 and 2). Of these contigs, 31913 could be functionally annotated (see Supplementary Figure 1 for annotation summary).

We analyzed proteomes of freshly collected worms, and worms that had been starved for 8 days, that is kept under oxic conditions without an external electron donor for energy conservation and autotrophic carbon fixation (see Methods). The purpose of creating these two conditions was to identify as many proteins as possible, including those expressed in worms that are starved. We identified a total of 4355 proteins, of which 2562 were host proteins and 1793 were symbiont proteins. The annotated host transcriptomes and proteomes were manually screened for sequences relevant for host-symbiont interactions. We identified 316 transcriptome sequences and 60 proteins potentially involved in microbe recognition, microbial growth

regulation, symbiont digestion, immune modulation and physiological interactions (see Table 1 and Figure 1).

## 3.2 Physiological adaptations of the host to the symbiosis

### 3.2.1 Nutrients are transferred from the symbionts to the host via digestion

Previous to this study, it was not clear how gutless oligochaetes gain nutrition from their bacterial symbionts. Two transfer modes, which are not mutually exclusive, have been suggested for symbioses with endosymbionts [32]: (1) "milking" of the symbionts (uptake of small compounds leaked or actively released by the symbionts), and (2) symbiont digestion through endocytosis. Endocytosis can include phagocytosis of symbiont particles or whole cells, as well as uptake of extracellularly digested and dissolved compounds by pinocytosis.

Several results from this study indicate that the main mode of nutrient transfer from the symbionts to *O. algarvensis* is through their digestion. First, we measured significantly less symbiont protein relative to host protein in the proteomes of starved worms compared to freshly collected worms (t-test, $p < 0.01$). In starved worms, symbiont protein accounted for only 18.7% of the total holobiont protein, while freshly collected worms had 29.5% symbiont protein (Table 2 and Supplementary Table 3). We cannot exclude that proteins in the symbionts were also degraded as a result of prolonged starvation. However, this would not explain the discrepancy between host and symbiont protein ratios in fresh compared to starved worms. Therefore, symbiont digestion by the host is the most likely explanation for the reduced amount of symbiont protein in starved worms.

Second, we identified 15 digestive enzymes predicted to occur in lysosomes, indicating their role in endocytosis, and 28 digestive enzymes involved in general secretory pathways, which could be targeted to phagolysosomes or to the extracellular region (Table 3). If secreted extracellularly,

these enzymes would aid in the digestion of symbionts in the extracellular space just below the worm's cuticle, and precede endocytotic digestion by the epidermal cells. The digestive proteins included various proteases for the degradation of polypeptides and oligopeptides, glucosidases with specificity for α1→4, α1→6 and β1→4 glycosidic bonds, and enzymes involved in lipid and peptidoglycan degradation (Table 3).

The third line of evidence that indicates that *O. algarvensis* digests its symbionts is that it expressed three different types of intestinal digestive enzymes, despite the fact that it does not have a mouth or gut. (i) The first type were digestive proteases (Table 3), namely pancreatic carboxypeptidase A, chymotrypsins A and B, cathepsins B, F and L, and pancreatic elastase. These enzymes are most often found in the intestinal tract of animals with a digestive system (Supplementary Table 4). Most of the *O. algarvensis* digestive proteases were highly similar to enzymes expressed in the midgut of the oligochaete *Eisenia andrei* (Supplementary Table 5). (ii) *O. algarvensis* also expressed a number of digestive glucosidases: two alpha amylases, with best BLAST hits to salivary gland and pancreatic amylases, an intestinal sucrase-isomaltase and two enzymes similar to pancreatic acid trehalase (Supplementary Table 6). (iii) *O. algarvensis* expressed five peptidoglycan recognition proteins (PGRPs) with predicted amidase activity (Figure 2) and a lysozyme, all proteins that degrade peptidoglycans. Although PGRPs and lysozyme are known for their role in immune defense [33], they can also aid in the digestion of food bacteria [34, 35]. The five *O. algarvensis* PGRP sequences were highly similar to PGRPs expressed by the annelid *Eisenia andrei* in its midgut (Supplementary Table 5).

Taken together, these results strongly indicate that *O. algarvensis* obtains nutrition from its symbionts by digesting them using a wide range of digestive enzymes, many of which are known to be expressed in the digestive tissues of animals. Given that the symbiotic bacteria are only found in the body wall of their host, it is highly likely that, in adaptation to the symbiosis, the expression of these “intestinal” enzymes has been redirected from the gut to the epidermis. This

assumption is supported by ultrastructural analyses that show the lysis of symbionts in the epidermal cells of the worm [36]. Additional support for the digestion of symbionts instead of “milking” stems from the observation that some of the *O. algarvensis* symbionts abundantly expressed high-affinity uptake transporters for organic substrates [9]. If 'milking' were the main manner in which the hosts gained their nutrition, they would compete with their symbionts for the uptake of small organic compounds.

### 3.2.2 Giant hemoglobins are likely involved in sulfide tolerance and transport

*O. algarvensis* abundantly expressed giant extracellular hemoglobins, which are respiratory pigments produced exclusively by annelids [37]. They are large multiprotein complexes (3.8 MDa in earthworms [38]), each consisting of more than a hundred copies of heme-containing globin chains and non-heme linker chains [37]. We found 12 globin chains and 6 linker chains from giant extracellular hemoglobins in our proteomes and transcriptomes (Supplementary Table 7). A signal peptide was predicted for all complete coding sequences, lending further support that these hemoglobins are indeed extracellular. Of the twelve *O. algarvensis* hemoglobin chain sequences, five could be unequivocally assigned to their respective families (3x family A, 2x family B).

We found that one of the three chains assigned to family A contained a free cysteine residue (Figure 3). Free cysteine residues do not participate in the formation of disulfide bonds in proteins, and therefore may unintentionally react with other blood components and disturb blood homeostasis [39, 40]. Extracellular hemoglobins are therefore under strong selective pressure to avoid the incorporation of free cysteines. The exception are annelids that experience high concentrations of sulfide in their habitats (Figure 3, [17]). In these worms, free cysteine residues in the A2 and B2 hemoglobin chains may allow them to reversibly bind environmental hydrogen sulfide and oxygen simultaneously [41]. It has been argued that this could mitigate the toxic effects of hydrogen sulfide for these worms. In hydrothermal vent tube worms, which also

have free cysteine residues in their hemoglobin, it has been assumed that these also allow them to bind and transport sulfide to their sulfur-oxidizing endosymbionts [42]. In these worms sulfide-binding to hemoglobin could also be mediated by zinc ions rather than free cysteine [43, 44]; however zinc does not appear to play a role in sulfide-binding in other annelids [45].

In *O. algarvensis*, the free cysteine residue is located in the conserved position that allows sulfide binding, and hydrophobic cluster analysis showed that the molecular environment of this free cysteine is highly similar to the sulfide-binding domain of A2 chains in other annelids (Supplementary Figure 2). It is therefore plausible that the *O. algarvensis* hemoglobin can also bind sulfide.

*O. algarvensis* lives in oligotrophic sediments with very low environmental sulfide concentrations [6, 9]. However, its sulfate-reducing symbionts are a considerable internal source of sulfide under anoxic conditions [6]. With its sulfide-binding hemoglobin, the host could store this internally produced sulfide for use by the SOX symbionts once they return to oxic conditions. Furthermore, the sulfide-binding hemoglobin might keep sulfide levels low in sensitive tissues of *O. algarvensis* such as the central nervous system.

### 3.2.3 Hemerythrin may enable respiration in the absence of $O_2$ and in the presence of CO

In addition to hemoglobin, the host expressed two hemerythrins, which are also respiratory proteins, but without heme groups. One of these hemerythrins was by far the most abundant protein in both fresh and starved worms, and accounted for 11 - 15% of total host protein (Supplementary Table 8). In contrast, the second most abundant protein, a histone, accounted only for less than 3%. Both hemerythrins were more highly expressed than any of the hemoglobin chains; expression levels of the most abundant hemerythrin were almost 32 times higher than the most abundant globin chain in the proteome (Supplementary Table 8). Such

abundant expression of hemerythrin is unknown from gut-bearing oligochaetes and other annelids (Supplementary Table 9).

Hemerythrin is an oxygen-carrying protein in sipunculids, priapulids and brachiopods, and also in a few polychaete annelids [46, 47]. In addition to oxygen transport, annelids might use hemerythrins for heavy metal resistance and antibacterial defense, or as an egg yolk protein [48, 49, 50]. In the only study that found hemerythrin expression in an oligochaete, it was assumed to be involved in heavy metal detoxification [49]. Since the environment of the *O. algarvensis* sampled for this study is considered pristine and oligotrophic, and not contaminated with high levels of heavy metals or pathogenic bacteria, and the worms in our experiments were not exposed to such conditions, it is unlikely that the high expression levels of hemerythrin are related to heavy metal resistance or antibacterial defense. We can also exclude its role in egg yolk protein, because the worms for proteomics were sampled in the fall, a time of the year when *O. algarvensis* does not reproduce (Kleiner, Lott, Wippler, unpublished observation). Therefore, it seems most likely that the hemerythrin in *O. algarvensis* is used to bind oxygen. This raises the question why *O. algarvensis* has two abundant oxygen binding proteins - hemoglobin and hemerythrin.

The fact that hemerythrin expression is unusual in oligochaetes suggests that there is a considerable selective advantage for its expression in *O. algarvensis*. One intriguing property of hemerythrin is that it is insensitive to carbon monoxide (CO) [51]. In contrast, heme proteins such as hemoglobin and myoglobin have much higher affinities for CO than for oxygen [52, 53]. This makes CO highly toxic to organisms that rely on heme proteins for oxygen transport. Considerable *in situ* CO concentrations of up to 51 nM were regularly measured in the *O. algarvensis* environment [11], and CO serves as an energy source for its sufur-oxidizing and sulfate-reducing symbionts [9]. Thus, the selective advantage of using hemerythrin for oxygen binding could be that it mitigates the adverse effects of carbon monoxide for the host.

The question remains why hemoglobin is also expressed in *O. algarvensis*, in parallel to hemerythrin. We speculate that hemerythrin and hemoglobin fulfill different functions in these worms. We propose that hemerythrin is used for oxygen storage to bridge the frequent and extended periods of anoxia that *O. algarvensis* is exposed to in the reduced sediment layers it mainly inhabits. Hemerythrin is well suited for oxygen storage because its oxygen binding capacity is stable under varying concentrations of $O_2$, $CO_2$ and protons [54, 55], and has been shown to play a key role in oxygen storage for bridging hypoxic episodes in sipunculids [56]. In contrast, hemoglobin, due to cooperative binding of oxygen and the Bohr effect, is well suited for gas exchange with the environment, which occurs in the upper oxic layer of the sediment where CO concentrations are much lower [11].

Interestingly, hemerythrin was also co-expressed with hemoglobin in the sulfur-oxidizing symbiont-bearing trophosome tissue of the deep-sea hydrothermal vent tube worm *Ridgeia piscesae*, a polychaete annelid that is not closely related to *O. algarvensis* [57]. The function of hemerythrin in *Riftia* is at present unknown. It is intriguing that the two animals currently known to abundantly express both hemoglobin and hemerythrin, *O. algarvensis and R. piscesae*, live in symbiosis with sulfur-oxidizing bacteria.

## 3.3 Interactions between the host innate immune system and its microbiome

We analyzed the proteins of the host innate immune system in our transcriptomes and proteomes, because these receptors, regulators and effectors are essential for sensing and responding to microbes [58], and are thus crucial for establishing and maintaining bacterial symbiosis [4]. The immune system must be able to distinguish beneficial symbionts from harmful intruders, and must respond appropriately, avoiding chronic inflammation in the presence of symbionts, while allowing rapid elimination of non-symbiotic bacteria.

### 3.3.1 Multitude of pattern recognition molecules for differential responses to microbes

Pattern recognition receptors (PRRs) are proteins that recognize microbe-associated molecular patterns (MAMPs) by binding to surface molecules specific to microbes like peptidoglycan or lipopolysaccharide [59]. PRRs are essential for sensing the presence of different microbial species and initiating an appropriate response, either via activation of immune signaling pathways and the synthesis of antimicrobial compounds, or by dampening or silencing the immune response in the case of bacterial symbionts [4]. We identified many different types of classical pattern recognition receptors, as well as proteins potentially involved in pattern recognition via conserved domains (Table 1).

**PGRPs.** Six peptidoglycan recognition proteins (OalgPGRP1-OalgPGRP6) were expressed in the *O. algarvensis* transcriptomes, and one of these was detected in the proteomes (OalgPGRP2, Table 1). PGRPs were first described as an important component of the innate immune defense [60], but are now known to play a major role in many animal-bacteria symbioses, mediating symbiont tolerance [61, 62], controlling symbiont populations [63], and regulating symbiosis establishment and maintenance [62, 64]. Elevated expression of PGRPs was also observed in the symbiont-bearing tissues of hydrothermal vent tube worms and mussels; however their precise function within these symbioses remains unknown [4, 65].

Specific PGRP function can not be determined from sequence information alone and depends on the molecular context of the environment in which they are expressed. However, some assumptions can be made and are discussed in the following. OalgPGRP1, OalgPGRP3 and OalgPGRP5 contained N-terminal transmembrane domains (indicating that they are membrane integral), as well as novel cytoplasmic domains (Figure 2). As is typical for PGRPs, the poorly conserved cytoplasmic domains had no similarity to known sequences [33]. PGRPs that integrate into the cell membrane and carry intracellular domains often induce an antimicrobial response by activating immune signaling pathways like Toll and IMD (immune deficiency) [66, 67].

However, some PGRP receptors bind peptidoglycans, but do not pass on an intracellular signal, thus effectively down-regulating the immune response and mediating tolerance towards resident bacteria [68].

OalgPGRP2 and OalgPGRP4 consisted only of the conserved PGRP domain itself with a signal peptide, indicating that they are secreted (Figure 2). Similar to the transmembrane PGRPs, secreted PGRPs can induce an antimicrobial response by indirectly activating immune signaling [69] or acting as bacterial growth inhibitors or antimicrobials themselves [70, 71]. However, if they possess amidase activity, they also can dampen the immune response, by cleaving peptidoglycan into non-immunogenic fragments [72, 35].

OalgPGRP1, OalgPGRP2, OalgPGRP4 and OalgPGRP5 contained the conserved residues needed to cleave peptidoglycan (Figure 4 [35, 73]). This suggests that they contribute to symbiont tolerance by scavenging immunogenic peptidoglycan fragments, which are released as a by-product of bacterial growth. The sequence of OalgPGRP3 was incomplete, but contained four out of the five residues needed to cleave peptidoglycan (Figure 4). These enzymatically active PGRPs may also play a role in symbiont population control and host nutrition by participating in the digestion of symbionts [74].

The affinities of PGRPs for different types of peptidoglycan stem peptides are determined by specific residues in the PGRP binding groove [75]. OalgPGRP1, OalgPGRP2, OalgPGRP4 and OalgPGRP5 possessed the residues that favor recognition of DAP-type peptidoglycan typical for gram negative bacteria [76], indicating that they could be used for the recognition of the worm's symbionts (which are all gram-negative) (Figure 4). The specificity of OalgPGRP3 could not be assigned because it had an insertion of two amino acids in the binding-groove region, and the OalgPGRP6 fragment did not contain the binding-relevant region.

**Lectins.** We detected six different classes of lectins in the transcriptome and proteome (Table 1, Table 4). They included C-type lectins, R-type lectins, fucolectin, SUEL/rhamnose-binding lectins, galectins, a beta-1,3-glucan binding protein and fibrinogen-like proteins. Lectins are proteins

with widely differing molecular structures and physiological functions. They are unified by their ability to strongly, yet reversibly, bind specific carbohydrate residues on the surfaces of cells and proteins, without exhibiting enzymatic activity [77].

Lectins are often associated with immune functions because of their molecular pattern recognition properties. For instance, they aid in microbe recognition and elimination through agglutination or direct antibacterial activity [78, 79], but, similar to PGRPs, are often also involved in modulating interactions between hosts and their beneficial symbionts. Lectins were, for example, shown to play major roles in symbiont acquisition and maintenance in sponges [80], corals [81, 82], clams [83], mice [84], and stilbonematid nematodes [85]. The sulfur-oxidizing symbionts of stilbonematine nematodes are very closely related to the primary symbionts of gutless oligochaetes [7, 86]. However, the stilbonematine lectins have no notable sequence similarity to the *O. algarvensis* lectins, as expected given the independent evolutionary histories of these two animal groups [86].

The domain architectures of *Olavius* lectins and their potential functions in host-symbiont interaction are summarized in Table 4. C-type lectins were particularly diverse, and 33 different forms were found in the transcriptome. Some of these C-type lectins have significant sequence similarity to lectins implicated in host-microbe interactions (Supplementary Table 10), for example to CD209 antigen-like proteins, macrophage mannose receptors, and C-type lectin receptor B – all MAMP receptors and phagocytosis enhancers of bacteria in vertebrates [87, 88, 89], and to immunolectin A, a microbe-inducible C-type lectin in *Manduca sexta* (tobacco hornworm) that is also involved in phagocytosis [90] .

Another highly diverse group of lectins found in *O. algarvensis* were fibrinogen-related proteins (FREPs), which are almost exclusively involved in host-microbe interactions in invertebrates [91]. They were represented by 27 different unigenes in the transcriptome (Table 1, Table 4). For most of these, several isoforms with varying amino acid sequences were predicted, indicating

that they may form an even more diverse array of proteins, possibly allowing very high specificity in the recognition of microbes.

**Scavenger receptor cysteine rich proteins.** In the transcriptomes we found a large group of sequences containing single or tandem scavenger receptor cysteine rich (SRCR) domains, often in association with other conserved domains, such as C-type lectin, trypsin, epidermal growth factor, low density lipoprotein (LDL) receptor, and immunoglobulin domains (Supplementary Figure 3). One of these proteins, which contained an additional universal stress protein A and four LDL receptor class B domains, was also identified in the proteome (Table 1).

The SRCR domain is an ancient and highly conserved module often found in proteins of the innate immune system that are involved in the recognition of microbial patterns and phagocytosis of bacteria in vertebrates [92]. In invertebrates, SRCR proteins have been implicated in host-symbiont interaction [93] and MAMP recognition [94].

Many SRCR sequences we identified had significant similarity to the MARCO scavenger receptor, DMBT1, CD163/M130, sea urchin scavenger receptors, and lamprey Pema-SRCR protein (Supplementary Table 11); all of these proteins are known or have been implicated to be involved in immune functions [92, 95]. Similar to the *Olavius* FREPs, the SRCR sequences identified in the transcriptome were represented by a considerable number of unigenes (FREP: 27, SRCR: 25), but many more different isoforms were predicted by the assembly. We therefore expect a high variability in the final proteins, possibly supporting highly specific recognition of microbes in *Olavius*, as has been observed in other invertebrates [96].

**Toll-like receptors.** We identified two Toll-like receptors (TLRs) consisting of the typical intracellular Toll/interleukin-1 receptor (TIR) homology domain and extracellular leucine- and cysteine-rich domains [97]. One of them was also detected in the proteome. Furthermore, we identified two sequences with only a TIR domain, one sequence with a TIR and transmembrane domain, and eight sequences containing leucine-rich repeats with high sequence similarity to TLRs from other animals and the variable lymphocyte receptors (VLRs) of agnate fish

(Supplementary Table 12). VLRs are immune receptors that experience somatic recombination and convey a form of adaptive immunity in jawless vertebrates [98].

Toll-like receptors (TLRs) are microbial pattern recognition receptors and intracellular signaling transducers that play a vital role in sensing and responding to microbiota in many animals [99]. They also play a role in many beneficial host-microbe symbioses [100, 101]. TLRs have long been thought to be absent from annelids [102, 103]. However, their presence and importance in host-microbe interactions has recently been recognized in polychaetes, leeches and earthworms [104, 105], where some were shown to be involved in the innate immune response against pathogens [106, 107] or were constitutively expressed in the gut [108].

We identified all the major components of the Toll signaling pathway in *O. algarvensis*, indicating that Toll signaling is active (Supplementary Table 13). We identified SARM (sterile alpha and TIR motif containing protein), an inhibitor of Toll signaling [109], that could aid in down-regulating the immune response against symbionts. Tollip, another inhibitor of Toll signaling [110], was also detected in the proteome, suggesting that these two inhibitors of Toll signaling may protect *O. algarvensis* against constant inflammation in response to its symbionts.

### 3.3.2 Interactions between symbionts and host may be regulated by different immune effectors and modulators

We detected several different types of antimicrobial proteins in the host transcriptome and proteome (Table 1), some of which were very abundant (Supplementary Table 8). The antimicrobials expressed in both transcriptome and proteome were lumbricin, an antimicrobial protein first discovered in earthworms [111], BPI (bactericidal permeability increasing protein), perforin/membrane attack complex-like proteins, insect defensin-like reeler proteins and cysteine-rich secretory proteins (Table 1). Antimicrobials combat infection by pathogenic microbes [112], but are also important in beneficial host-microbe interactions [84, 113], where they are used to modulate and control symbiont populations [114, 115]. In *O. algarvensis* they

might be used to prevent symbionts and pathogens from invading non-symbiotic tissues, or to regulate symbiont growth.

## Conclusions

This study provides insights into the physiological and molecular mechanisms that allow *Olavius algarvensis* to live in a stable beneficial association with its microbial consortium. Our results indicate that these animals have undergone a number of evolutionary changes in adaptation to their symbiotic lifestyle, apart from a complete reduction of the excretory and digestive organs. Examples of such adaptations are host proteins involved in symbiont digestion and nutrient uptake, with likely relocalization of the expression sites of some of these enzymes, and unconventional proteins for gas exchange and storage.

Since a mouth and anus are absent in gutless oligochaetes, foreign microbes can only invade these hosts if they have the ability to penetrate the egg integument, or the cuticle in a juvenile or adult worm. As a result, the complexity of the *O. algarvensis* microbiome is comparatively low and essentially consists of its symbiotic consortium. We found that *O. algarvensis* expresses a highly diverse array of pattern recognition receptors that enable it to recognize and respond to its microbiota. The high number of PAMP recognition proteins expressed in the transcriptome and proteome that clearly originated from different genes demonstrate the need for these hosts to diffentially sense and respond to both their symbiotic microbiota as well as environmental bacteria, although direct contact with the latter may be limited.

This is also the first comprehensive transcriptomic and proteomic analysis of the innate immune system of a marine oligochaete. It shows how genes common to a wide array of invertebrates have evolved to enable the intricate communication and interactions that occur between animals and their symbiotic microbiota. The analyses described here lay the foundation for

future experimental studies of immune processes and physiological responses that are essential in the functioning of this symbiosis.

# Declarations

### Abbreviations

AMP antimicrobial protein/peptide

cDNA complementary DNA

CDS coding DNA sequence

CO carbon monoxide

FREP fibrinogen related protein

PGRP peptidoglycan recognition protein

SRCR scavenger receptor cysteine-rich

TLR Toll-like receptor

### Competing interests

The authors have declared no competing interests.

### Ethics approval and consent to participate

The animals used in this study underlie no national or international guidelines or regulations.

### Consent for publication

Not applicable

### Data availability

The datasets supporting the conclusions of this article are available from the following repositories and its additional files: Assembled transcript sequences are available from the

European Nucleotide Archive (ENA) under the accession numbers HACZ01000001-HACZ01173602 (TSA project HACZ01000000 data, http://www.ebi.ac.uk/ena/data/view/HACZ010000000). Raw reads were deposited under the study ID PRJEB10952 (http://www.ebi.ac.uk/ena/data/view/PRJEB10952). The complete protein sequence database is available from the MassIVE data repository under accession MSV000079512 and available for download via FTP: ftp://MSV000079512@massive.ucsd.edu. All proteomics data sets used in this study were deposited at the MassIVE data repository under accession numbers: MSV000079512 [MassIVE] & PXD003626 [ProteomeXchange] and available for download via FTP: ftp://MSV000079512@massive.ucsd.edu .

**Funding**

The study was funded by the Max Planck Society and by the Gordon and Betty Moore Foundation through Grant GBMF3811 to ND. MK was supported by a PhD scholarship of the Studienstiftung des Deutschen Volkes and a NSERC Banting Postdoctoral Fellowship.

**Authors' contributions**

**JW** and **MK** contributed equally to this manuscript. **JW** conceived and wrote the manuscript, analyzed and interpreted the proteomic data and provided ideas, did in-depth bioinformatic analyses of proteins of interest, prepared all figures and tables, collected all worms for transcriptomics experiments, conceived and performed all transcriptomics experiments, assembled and annotated the transcriptomes, predicted CDS from transcriptome data and did all transcriptome statistics and analyses. **MK** conceived the study and the manuscript, edited the manuscript and provided ideas, conceived proteomics experiments, collected worms for proteomics experiments and performed two of the starvation experiments, compiled the protein

reference database, did statistical analyses and processed and analyzed the proteomics data. **CL** provided ideas and collected worms for proteomics and transcriptomics experiments and performed one of the starvation experiments. **AG** provided all microscopic images used in Figure 1, and commented on the manuscript. **PEA** and **RJG** processed proteomics data and commented on the manuscript. **JCY** provided ideas and ran 2D-LC-MS/MS experiments. **RLH** provided access to the proteomics equipment, provided conceptual input and coordinated the data processing at the ORNL. **ND** was involved in the organization and coordination of this study, provided ideas and commented on the manuscript. All authors reviewed and revised the final manuscript before submission.

**Acknowledgements**

We thank the team of the HYDRA Institute on Elba for their extensive support with sample collection and on site experiments, and Silke Wetzel for excellent technical assistance. Sequencing was performed by the Max Planck Society Genome Center at the Max Planck Institute for Plant Breeding in Cologne, Germany.

# References

[1] Hooper LV, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. Science. 2012;336:1268–1273.

[2] McFall-Ngai M, Hadfield MG, Bosch TC, Carey HV, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF, Hentschel U. Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci U S A. 2013;110:3229–3236.

[3] Gilbert SF, Bosch TC, Ledón-Rettig C. Eco-Evo-Devo: Developmental symbiosis and developmental plasticity as evolutionary agents. Nat Rev Genet. 2015;16:611–622.

[4] Nyholm SV, Song P, Dang J, Bunce C, Girguis PR. Expression and putative function of innate immunity genes under *in situ* conditions in the symbiotic hydrothermal vent tubeworm *Ridgeia piscesae*. PloS One. 2012;7:e38267.

[5] Chu H, Mazmanian SK. Innate immune recognition of the microbiota promotes host-microbial symbiosis. Nat Immunol. 2013;14:668–675.

[6] Dubilier N, Mülders C, Ferdelman T, de Beer D, Pernthaler A, Klein M, Wagner M, Erséus C, Thiermann F, Krieger J, Giere O, Amann R. Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. Nature. 2001;411:298–302.

[7] Dubilier N, Blazejak A, Rühland C. Symbioses between bacteria and gutless marine oligochaetes. In: Overmann J, editor. Molecular Basis of Symbiosis. Berlin Heidelberg: Springer; 2005. p. 251–275.

[8] Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO, Boffelli D, Anderson IJ, Barry KW, Shapiro HJ, Szeto E, Kyrpides NC, Mussmann M, Amann R, Bergin C, Ruehland C, Rubin EM, Dubilier N. Symbiosis insights through metagenomic analysis of a microbial consortium. Nature. 2006;443:950–955.

[9] Kleiner M, Wentrup C, Lott C, Teeling H, Wetzel S, Young J, Chang YJ, Shah M, VerBerkmoes NC, Zarzycki J, Fuchs G, Markert S, Hempel K, Voigt B, Becher D, Liebeke M, Lalk M, Albrecht D, Hecker M, Schweder T, Dubilier, N. Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. Proc Nat Acad Sci U S A. 2012;109:E1173–E1182.

[10] Ruehland C, Blazejak A, Lott C, Loy A, Erséus C, Dubilier N. Multiple bacterial symbionts in two species of co-occurring gutless oligochaete worms from Mediterranean sea grass sediments. Environ Microbiol. 2008;10:3404–3416.

[11] Kleiner M, Wentrup C, Holler T, Lavik G, Harder J, Lott C, Littmann S, Kuypers MM, Dubilier N. Use of carbon monoxide and hydrogen by a bacteria–animal symbiosis from seagrass sediments. Environ Microbiol. 2015;17:5023–5035.

[12] Harrison P, Seemann T. From high-throughput sequencing read alignments to confident, biologically relevant conclusions with Nesoni. 2009. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.509.7046&rep=rep1&type=pdf. Accessed 25 Apr 2016.

[13] Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol. 2011;29:644.

[14] Li W, Li S, Zhong J, Zhu Z, Liu J, Wang W. A novel antimicrobial peptide from skin secretions of the earthworm, *Pheretima guillelmi* (Michaelsen). Peptides. 2011;32:1146–1150.

[15] Conesa A, Götz S, Garcá-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21:3674–3676.

[16] Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–3402.

[17] Bailly X, Jollivet D, Vanin S, Deutsch J, Zal F, Lallier F, Toulmond A. Evolution of the sulfide-binding function within the globin multigenic family of the deep-sea hydrothermal vent tubeworm *Riftia pachyptila*. Mol Biol Evol. 2002;19:1421–1433.

[18] Callebaut I, Labesse G, Durand P, Poupon A, Canard L, Chomilier J, Henrissat B, Mornon J. Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. Cell Mol Life Sci. 1997;53:621–645.

[19] Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. Genome Biol. 2011;12:R44.

[20] Thompson MR, Chourey K, Froelich JM, Erickson BK, VerBerkmoes NC, Hettich RL. Experimental approach for deep proteome measurements from small-scale microbial biomass samples. Anal Chem. 2008;80:9517–9525.

[21] Washburn MP, Wolters D, Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001;19:242–247.

[22] McDonald WH, Ohi R, Miyamoto DT, Mitchison TJ, Yates JR. Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. Int J Mass Spectrom. 2002;219:245–251.

[23] Gouzy J, Carrere S, Schiex T. FrameDP: sensitive peptide detection on noisy matured sequences. Bioinformatics. 2009;25:670–671.

[24] Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. Trends Genet. 2000;16:276–277.

[25] Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–3152.

[26] Tabb DL, Fernando CG, Chambers MC. MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. J Proteome Res. 2007;6:654–661.

[27] Ma ZQ, Dasari S, Chambers MC, Litton MD, Sobecki SM, Zimmerman LJ, Halvey PJ, Schilling B, Drake PM, Gibson BW, Tabb DL. IDPicker 2.0: Improved protein assembly with high discrimination peptide identification filtering. J Proteome Res. 2009;8:3872–3881.

[28] Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–2461.

[29] Abraham P, Adams R, Giannone RJ, Kalluri U, Ranjan P, Erickson B, Shah M, Tuskan GA, Hettich RL. Defining the boundaries and characterizing the landscape of functional genome expression in vascular tissues of *Populus* using shotgun proteomics. J Proteome Res. 2011;11:449–460.

[30] Giannone RJ, Huber H, Karpinets T, Heimerl T, Küper U, Rachel R, Keller M, Hettich RL, Podar M. Proteomic characterization of cellular and molecular processes that enable the *Nanoarchaeum equitans-Ignicoccus hospitalis* relationship. PLoS One. 2011;6:e22942.

[31] Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, Washburn MP. Statistical Analysis of Membrane Proteome Expression Changes in *Saccharomyces cerevisiae*. J Proteome Res. 2006;5:2339–2347.

[32] Cavanaugh CM, McKiness ZP, Newton IL, Stewart FJ. Marine chemosynthetic symbioses. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. The Prokaryotes. Berlin Heidelberg: Springer; 2013. p. 579–607.

[33] Dziarski R, Gupta D. The peptidoglycan recognition proteins (PGRPs). Genome Biol. 2006;7:232.

[34] Van Herreweghe JM, Michiels CW. Invertebrate lysozymes: diversity and distribution, molecular mechanism and in vivo function. J Biosci. 2012;37:327–348.

[35] Mellroth P, Karlsson J, Steiner H. A scavenger function for a *Drosophila* Peptidoglycan recognition protein. J Biol Chem. 2003;278:7059–7064.

[36] Giere O, Erseus C, Stuhlmacher F. A new species of *Olavius* (Tubificidae) from the Algarve coast in Portugal, the first East Atlantic gutless oligochaete with symbiotic bacteria. Zool Anz. 1998;237:209–214.

[37] Royer WE, Sharma H, Strand K, Knapp JE, Bhyravbhatla B. *Lumbricus* erythrocruorin at 3.5 Å resolution: architecture of a megadalton respiratory complex. Structure. 2006;14:1167–1177.

[38] Fushitani K, Matsuura M, Riggs A. The amino acid sequences of chains A, B, and C that form the trimer subunit of the extracellular hemoglobin from *Lumbricus terrestris*. J Biol Chem. 1988;263:6502–6517.

[39] Concha NO, Herzberg O, Rasmussen BA, Bush K. Crystal structures of the cadmium-and mercury-substituted metallo-β-lactamase from *Bacteroides fragilis*. Protein Sci. 1997;6:2671–2676.

[40] Leder A, Wiener E, Lee MJ, Wickramasinghe SN, Leder P. A normal beta-globin allele as a modifier gene ameliorating the severity of alpha-thalassemia in mice. Proc Nat Acad Sci U S A. 1999;96:6291–6295.

[41] Zal F, Leize E, Lallier FH, Toulmond A, Van Dorsselaer A, Childress JJ. S-Sulfohemoglobin and disulfide exchange: the mechanisms of sulfide binding by *Riftia pachyptila* hemoglobins. Proc Nat Acad Sci U S A. 1998;95:8997–9002.

[42] Childress J, Fisher C, Favuzzi J, Kochevar R, Sanders N, Alayse AM. Sulfide-driven autotrophic balance in the bacterial symbiont-containing hydrothermal vent tubeworm, *Riftia pachyptila* Jones. Biol Bull. 1991;180:135–153.

[43] Flores JF, Fisher CR, Carney SL, Green BN, Freytag JK, Schaeffer SW, Royer WE. Sulfide binding is mediated by zinc ions discovered in the crystal structure of a hydrothermal vent tubeworm hemoglobin. Proc Nat Acad Sci U S A. 2005;102:2713–2718.

[44] Flores JF, Hourdez SM. The zinc-mediated sulfide-binding mechanism of hydrothermal vent tubeworm 400-kDa hemoglobin. Cah Biologie Mar. 2006;47:371.

[45] Chabasse C, Bailly X, Rousselot M, Zal F. The multigenic family of the extracellular hemoglobin from the annelid polychaete *Arenicola marina*. Comp Biochem Physiol B. 2006;144:319–325.

[46] Kurtz Jr D. Molecular structure/function relationships of hemerythrins. In: Mangum CP, editor. Blood and Tissue Oxygen Carriers. Berlin Heidelberg: Springer; 1992. p. 151–171.

[47] Manwell C, Baker CA. *Magelona* haemerythrin: tissue specificity, molecular weights and oxygen equilibria. Comp Biochem Physiol B. 1988;89:453–463.

[48] Baert JL, Britel M, Sautière P, Malecha J. Ovohemerythrin, a major 14-kDa yolk protein distinct from vitellogenin in leech. Eur J Biochem. 1992;209:563–569.

[49] Nejmeddine A, Wouters-Tyrou D, Baert JL, Sautière P. Primary structure of a myohemerythrin-like cadmium-binding protein, isolated from a terrestrial annelid oligochaete. C R Acad Sci. 1997;320:459–468.

[50] Deloffre L, Salzet B, Vieau D, Andries JC, Salzet M. Antibacterial properties of hemerythrin of the sand worm *Nereis diversicolor*. Neuroendocrinol Lett. 2003;24:39–45.

[51] Florkin M. Recherches sur les hémérythrines. In: Archives Internationales de Physiologie. Liège: Vaillant-Carmanne; 1933. p. 247-329.

[52] Haldane J, Smith JL. The absorption of oxygen by the lungs. J Physiol. 1897;22:231.

[53] Rossi-Fanelli A, Antonini E. Studies on the oxygen and carbon monoxide equilibria of human myoglobin. Arch Biochem Biophys. 1958;77:478–492.

[54] Kubo M. The oxygen equilibrium of hemerythrin and Bohr effect. Bull Chem Soc Jpn. 1953;26:189–192.

[55] Ordway GA, Garry DJ. Myoglobin: an essential hemoprotein in striated muscle. J Exp Biol. 2004;207:3441–3446.

[56] Mangum CP. Physiological function of the hemerythrins. In: Mangum CP, editor. Blood and Tissue Oxygen Carriers. Berlin Heidelberg: Springer; 1992. p. 173–192.

[57] Sanchez S, Hourdez S, Lallier FH. Identification of proteins involved in the functioning of *Riftia pachyptila* symbiosis by Subtractive Suppression Hybridization. BMC Genomics. 2007;8:1.

[58] Medzhitov R. Recognition of microorganisms and activation of the immune response. Nature. 2007;449:819–826.

[59] Mackey D, McFall AJ. MAMPs and MIMPs: proposed classifications for inducers of innate immunity. Mol Microbiol. 2006;61:1365–1371.

[60] Royet J, Gupta D, Dziarski R. Peptidoglycan recognition proteins: modulators of the microbiome and inflammation. Nat Rev Immunol. 2011;11:837–851.

[61] Wang J, Wu Y, Yang G, Aksoy S. Interactions between mutualist *Wigglesworthia* and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypanosome transmission. Proc Nat Acad Sci U S A. 2009;106:12133–12138.

[62] Troll JV, Bent EH, Pacquette N, Wier AM, Goldman WE, Silverman N, McFall-Ngai MJ. Taming the symbiont for coexistence: a host PGRP neutralizes a bacterial symbiont toxin. Environ Microbiol. 2010;12:2190–2203.

[63] Anselme C, Vallier A, Balmand S, Fauvarque MO, Heddi A. Host PGRP gene expression and bacterial release in endosymbiosis of the weevil *Sitophilus zeamais*. Appl Environ Microbiol. 2006;72:6766–6772.

[64] Saha S, Jing X, Park SY, Wang S, Li X, Gupta D, Dziarski R. Peptidoglycan recognition proteins protect mice from experimental colitis by promoting normal gut flora and preventing induction of interferon-γ. Cell Host Microbe. 2010;8:147–162.

[65] Bettencourt R, Rodrigues M, Barros I, Cerqueira T, Freitas C, Costa V, Pinheiro M, Egas C, Santos RS. Site-related differences in gene expression and bacterial densities in the mussel *Bathymodiolus azoricus* from the Menez Gwen and Lucky Strike deep-sea hydrothermal vent sites. Fish Shellfish Immunol. 2014;39:343–353.

[66] Leulier F, Parquet C, Pili-Floury S, Ryu JH, Caroff M, Lee WJ, Mengin-Lecreulx D, Lemaitre B. The *Drosophila* immune system detects bacteria through specific peptidoglycan recognition. Nature Immunol. 2003;4:478–484.

[67] Kaneko T, Goldman WE, Mellroth P, Steiner H, Fukase K, Kusumoto S, Harley W, Fox A, Golenbock D, Silverman N. Monomeric and polymeric gram-negative peptidoglycan but not purified LPS stimulate the *Drosophila* IMD pathway. Immunity. 2004;20:637–649.

[68] Maillet F, Bischoff V, Vignal C, Hoffmann J, Royet J. The *Drosophila* peptidoglycan recognition protein PGRP-LF blocks PGRP-LC and IMD/JNK pathway activation. Cell Host Microbe. 2008;3:293–303.

[69] Michel T, Reichhart JM, Hoffmann JA, Royet J. *Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. Nature. 2001;414:756–759.

[70] Lu X, Wang M, Qi J, Wang H, Li X, Gupta D, Dziarski R. Peptidoglycan recognition proteins are a new class of human bactericidal proteins. J Biol Chem. 2006;281:5895–5907.

[71] Tydell CC, Yuan J, Tran P, Selsted ME. Bovine peptidoglycan recognition protein-S: antimicrobial activity, localization, secretion, and binding properties. J Immunol. 2006;176:1154–1162.

[72] Zaidman-Rémy A, Hervé M, Poidevin M, Pili-Floury S, Kim MS, Blanot D, Oh BH, Ueda R, Mengin-Lecreulx D, Lemaitre B. The *Drosophila* amidase PGRP-LB modulates the immune response to bacterial infection. Immunity. 2006;24:463–473.

[73] Wang ZM, Li X, Cocklin RR, Wang M, Wang M, Fukase K, Inamura S, Kusumoto S, Gupta D, Dziarski R. Human peptidoglycan recognition protein-L is an N-acetylmuramoyl-L-alanine amidase. J Biol Chem. 2003;278:49044–49052.

[74] Garver LS, Wu J, Wu LP. The peptidoglycan recognition protein PGRP-SC1a is essential for Toll signaling and phagocytosis of *Staphylococcus aureus* in *Drosophila*. Proc Nat Acad Sci U S A. 2006;103:660–665.

[75] Swaminathan CP, Brown PH, Roychowdhury A, Wang Q, Guan R, Silverman N, Goldman WE, Boons GJ, Mariuzza RA. Dual strategies for peptidoglycan discrimination by peptidoglycan recognition proteins (PGRPs). Proc Nat Acad Sci U S A. 2006;103:684–689.

[76] Schleifer KH, Kandler O. Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriol Rev. 1972;36:407.

[77] Varki A, Cummings R, Esko J, Freeze H, Stanley P, Bertozzi CR, Hart GW, Etzler ME. Essentials of Glycobiology. 2nd ed. New York: Cold Spring Harber Laboratory Press; 2009.

[78] Fan C, Zhang S, Li L, Chao Y. Fibrinogen-related protein from amphioxus *Branchiostoma belcheri* is a multivalent pattern recognition receptor with a bacteriolytic activity. Mol Immunol. 2008;45:3338–3346.

[79] Mukherjee S, Zheng H, Derebe MG, Callenberg KM, Partch CL, Rollins D, Propheter DC, Rizo J, Grabe M, Jiang QX, Hooper LV. Antibacterial membrane attack by a pore-forming intestinal C-type lectin. Nature. 2014;505:103–107.

[80] Müller W, Zahn R, Kurelec B, Lucu C, Müller I, Uhlenbruck G. Lectin, a possible basis for symbiosis between bacteria and sponges. J Bacteriol. 1981;145:548–558.

[81] Wood-Charlson EM, Hollingsworth LL, Krupp DA, Weis VM. Lectin/glycan interactions play a role in recognition in a coral/dinoflagellate symbiosis. Cell Microbiol. 2006;8:1985–1993.

[82] Kvennefors ECE, Leggat W, Hoegh-Guldberg O, Degnan BM, Barnes AC. An ancient and variable mannose-binding lectin from the coral *Acropora millepora* binds both pathogens and symbionts. Dev Comp Immunol. 2008;32:1582–1592.

[83] Gourdine JP, Smith-Ravin EJ. Analysis of a cDNA-derived sequence of a novel mannose-binding lectin, codakine, from the tropical clam *Codakia orbicularis*. Fish Shellfish Immunol. 2007;22:498–509.

[84] Cash HL, Whitham CV, Behrendt CL, Hooper LV. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. Science. 2006;313:1126–1130.

[85] Bulgheresi S, Gruber-Vodicka HR, Heindl NR, Dirks U, Kostadinova M, Breiteneder H, Ott JA. Sequence variability of the pattern recognition receptor Mermaid mediates specificity of marine nematode symbioses. ISME J. 2011;5:986–998.

[86] Zimmermann J, Wentrup C, Sadowski M, Blazejak A, Gruber-Vodicka H, Kleiner M, Ott J, Cronholm B, De Wit P, Erséus C, Dubilier N. Closely coupled evolutionary history of ecto-and endosymbionts from two distantly-related animal phyla. Mol Ecol. 2016; doi: 10.1111/mec.13554.

[87] Park CG, Takahara K, Umemoto E, Yashima Y, Matsubara K, Matsuda Y, Clausen BE, Inaba K, Steinman RM. Five mouse homologues of the human dendritic cell C-type lectin, DC-SIGN. Int Immunol. 2001;13:1283–1290.

[88] Ezekowitz R, Sastry K, Bailly P, Warner A. Molecular characterization of the human macrophage mannose receptor: demonstration of multiple carbohydrate recognition-like domains and phagocytosis of yeasts in Cos-1 cells. J Exp Med. 1990;172:1785–1794.

[89] Soanes KH, Figuereido K, Richards RC, Mattatall NR, Ewart KV. Sequence and expression of C-type lectin receptors in Atlantic salmon (*Salmo salar*). Immunogenetics. 2004;56:572–584.

[90] Yu XQ, Gan H, Kanost MR. Immulectin, an inducible C-type lectin from an insect, *Manduca sexta*, stimulates activation of plasma prophenol oxidase. Insect Biochem Mol Biol. 1999;29:585–597.

[91] Hanington PC, Zhang SM. The primary role of fibrinogen-related proteins in invertebrates is defense, not coagulation. J Innate Immun. 2010;3:17–27.

[92] Sarrias MR, Gronlund J, Padilla O, Madsen J, Holmskov U, Lozano F. The Scavenger Receptor Cysteine-Rich (SRCR) domain: an ancient and highly conserved protein module of the innate immune system. Crit Rev Immunol. 2004;24:1–37.

[93] Steindler L, Schuster S, Ilan M, Avni A, Cerrano C, Beer S. Differential gene expression in a marine sponge in relation to its symbiotic state. Mar Biotechnol. 2007;9:543–549.

[94] Liu F, Li J, Fu J, Shen Y, Xu X. Two novel homologs of simple C-type lectin in grass carp (*Ctenopharyngodon idellus*): potential role in immune response to bacteria. Fish Shellfish Immunol. 2011;31:765–773.

[95] Pancer Z. Dynamic expression of multiple scavenger receptor cysteine-rich genes in coelomocytes of the purple sea urchin. Proc Nat Acad Sci. 2000;97:13156–13161.

[96] Buckley KM, Rast JP. Diversity of animal immune receptors and the origins of recognition complexity in the deuterostomes. Dev Comp Immunol. 2015;49:179–189.

[97] Zheng L, Zhang L, Lin H, McIntosh M, Malacrida A. Toll-like receptors in invertebrate innate immunity. Invert Surviv J. 2005;2:105–113.

[98] Pancer Z, Amemiya CT, Ehrhardt GR, Ceitlin J, Gartland GL, Cooper MD. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature. 2004;430:174–180.

[99] Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. Nat Immunol. 2010;11:373–384.

[100] Kubinak JL, Round JL. Toll-like receptors promote mutually beneficial commensal-host interactions. PLoS Pathogens. 2012;8:e1002785.

[101] Venkatesh M, Mukherjee S, Wang H, Li H, Sun K, Benechet AP, Qiu Z, Maher L, Redinbo MR, Phillips RS, Fleet JC. Symbiotic bacterial metabolites regulate gastrointestinal barrier function via the xenobiotic sensor PXR and Toll-like receptor 4. Immunity. 2014;41:296–310.

[102] Cooper E, Kvell K, Engelmann P, Nemeth P. Still waiting for the Toll? Immunol Lett. 2006;104:18–28.

[103] Francis J, Wreesman S, Yong S, Reigstad K, Krutzik S, Cooper EL. Analysis of the earthworm coelomocyte cell surface for the presence of Toll-like immune receptors. Eur J Soil Biol. 2007;43:S92–S96.

[104] Davidson CR, Best NM, Francis JW, Cooper EL, Wood TC. Toll-like receptor genes (TLRs) from *Capitella capitata* and *Helobdella robusta* (Annelida). Dev Comp Immunol. 2008;32:608–612.

[105] Halanych KM, Kocot KM. Repurposed transcriptomic data facilitate discovery of innate immunity toll-like receptor (TLR) genes across lophotrochozoa. Biol Bull. 2014;227:201–209.

[106] Cuvillier-Hot V, Boidin-Wichlacz C, Slomianny C, Salzet M, Tasiemski A. Characterization and immune function of two intracellular sensors, HmTLR1 and HmNLR, in the injured CNS of an invertebrate. Dev Comp Immunol. 2011;35:214–226.

[107] Fjøsne TF, Stenseth EB, Myromslien F, Rudi K. Gene expression of TLR homologues identified by genome-wide screening of the earthworm *Dendrobaena veneta*. Innate Immun. 2014;21:161–166.

[108] Škanta F, Roubalová R, Dvořák J, Procházková P, Bilej M. Molecular cloning and expression of TLR in the *Eisenia andrei* earthworm. Dev Comp Immunol. 2013;41:694–702.

[109] Zhang Q, Zmasek CM, Cai X, Godzik A. TIR domain-containing adaptor SARM is a late addition to the ongoing microbe-host dialog. Dev Comp Immunol. 2011;35:461–468.

[110] Zhang G, Ghosh S. Negative regulation of Toll-like receptor-mediated signaling by Tollip. J Biol Chem. 2002;277:7059–7065.

[111] Cho JH, Park CB, Yoon YG, Kim SC. Lumbricin I, a novel proline-rich antimicrobial peptide from the earthworm: purification, cDNA cloning and molecular characterization. Biochim Biophys Acta. 1998;1408:67–76.

[112] Vizioli J, Salzet M. Antimicrobial peptides from animals: focus on invertebrates. Trends Pharmacol Sci. 2002;23:494–496.

[113] Hooper LV. Do symbiotic bacteria subvert host immunity? Nat Rev Microbiol. 2009;7:367–374.

[114] Vaishnava S, Behrendt CL, Ismail AS, Eckmann L, Hooper LV. Paneth cells directly sense gut commensals and maintain homeostasis at the intestinal host-microbial interface. Proc Natl Acad Sci U S A. 2008;105:20858–20863.

[115] Login FH, Balmand S, Vallier A, Vincent-Monégat C, Vigneron A, Weiss-Gayet M, Rochat D, Heddi A. Antimicrobial peptides keep insect endosymbionts under control. Science. 2011;334:362–365.

[116] Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305:567–580.

[117] Sonnhammer EL, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol. 1998;6:175–182.

[118] Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8:785–786.

[119] Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol. 2004;338:1027–1036.

[120] Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol. 2000;300:1005–1016.

[121] Baú D, Martin AJM, Mooney C, Vullo A, Walsh I, Pollastri G. Distill: a suite of web servers for the prediction of one-, two- and three-dimensional structural features of proteins. BMC Bioinform. 2006;7:402.

[122] Goldberg T, Hecht M, Hamp T, Karl T, Yachdav G, Ahmed N, Altermann U, Angerer P, Ansorge S, Balasz K, Bernhofer M, Betz A, Cizmadija L, Do KT, Gerke J, Greil R, Joerdens V, Hastreiter M, Hembach K, Herzog M, Kalemanov M, Kluge M, Meier A, Nasir H, Neumaier U, Prade V, Reeb J, Sorokoumov A, Troshani I, Vorberg S, Waldraff S, Zierer J, Nielsen H, Rost B. LocTree3 prediction of localization. Nucleic Acids Res. 2014;42:W350–W355.

[123] Pierleoni A, Martelli PL, Fariselli P, Casadio R. BaCelLo: a balanced subcellular localization predictor. Bioinformatics. 2006;22:e408–e416.

[124] Lin WZ, Fang JA, Xiao X, Chou KC. iLoc-Animal: a multi-label learning classifier for predicting subcellular localization of animal proteins. Mol Biosyst. 2013;9:634–644.

[125] Honda S, Kashiwagi M, Miyamoto K, Takei Y, Hirose S. Multiplicity, structures, and endocrine and exocrine natures of eel fucose-binding lectins. J Biol Chem. 2000;275:33151–33157.

[126] Bulgheresi S, Schabussova I, Chen T, Mullin NP, Maizels RM, Ott JA. A new C-type lectin similar to the human immunoreceptor DC-SIGN mediates symbiont acquisition by a marine nematode. Appl Environ Microbiol. 2006;72:2950–2956.

[127] Katzin BJ, Collins EJ, Robertus JD. Structure of ricin A-chain at 2.5 A. Proteins. 1991;10:251–259.

[128] Nakano M, Tabata S, Sugihara K, Kouzuma Y, Kimura M, Yamasaki N. Primary structure of hemolytic lectin CEL-III from marine invertebrate *Cucumaria echinata* and its cDNA: structural similarity to the B-chain from plant lectin, ricin. Biochim Biophys Acta. 1999;1435:167–176.

[129] Hisamatsu K, Nagao T, Unno H, Goda S, Hatakeyama T. Identification of the amino acid residues involved in the hemolytic activity of the *Cucumaria echinata* lectin CEL-III. Biochim Biophys Acta. 2013;1830:4211–4217.

[130] Kim DH, Patnaik BB, Seo GW, Kang SM, Lee YS, Lee BL, Han YS. Identification and expression analysis of a novel R-type lectin from the coleopteran beetle, *Tenebrio molitor*. J Invertebr Pathol. 2013;114:226–229.

[131] Yang RY, Rabinovich GA, Liu FT. Galectins: structure, function and therapeutic potential. Expert Rev Mol Med. 2008;10:e17.

[132] Vasta GR. Galectins as pattern recognition receptors: structure, function, and evolution. Adv Exp Med Biol. 2012;946:21–36.

[133] Bao Y, Shen H, Zhou H, Dong Y, Lin Z. A tandem-repeat galectin from blood clam *Tegillarca granosa* and its induced mRNA expression response against bacterial challenge. Genes Genom. 2013;35:733–740.

[134] Shi XZ, Wang L, Xu S, Zhang XW, Zhao XF, Vasta GR, Wang JX. A galectin from the kuruma shrimp (*Marsupenaeus japonicus*) functions as an opsonin and promotes bacterial clearance from hemolymph. PLoS One. 2014;9:e91794.

[135] Tateno H, Ogawa T, Muramoto K, Kamiya H, Saneyoshi M. Rhamnose-binding lectins from steelhead trout (*Oncorhynchus mykiss*) eggs recognize bacterial lipopolysaccharides and lipoteichoic acid. Biosci Biotechnol Biochem. 2002;66:604–612.

[136] Watanabe Y, Tateno H, Nakamura-Tsuruta S, Kominami J, Hirabayashi J, Nakamura O, Watanabe T, Kamiya H, Naganuma T, Ogawa T, Naudé RJ, Muramoto K. The function of

rhamnose-binding lectin in innate immunity by restricted binding to Gb3. Dev Comp Immunol. 2009;33:187–197.

[137] Mourão PAS. A carbohydrate-based mechanism of species recognition in sea urchin fertilization. Braz J Med Biol Res. 2007;40:5–17.

[138] Ozeki Y, Yokota Y, Kato KH, Titani K, Matsui T. Developmental expression of D-galactoside-binding lectin in sea urchin (*Anthocidaris crassispina*) eggs. Exp Cell Res. 1995;216:318–324.

[139] Yamada S, Hotta K, Yamamoto TS, Ueno N, Satoh N, Takahashi H. Interaction of notochord-derived fibrinogen-like protein with Notch regulates the patterning of the central nervous system of *Ciona intestinalis* embryos. Dev Biol. 2009;328:1–12.

[140] Harada Y, Takagaki Y, Sunagawa M, Saito T, Yamada L, Taniguchi H, Shoguchi E, Sawada H. Mechanism of self-sterility in a hermaphroditic chordate. Science. 2008;320:548–550.

# Tables

## Table 1: Overview of proposed host-symbiont interaction proteins

For more extensive details see Supplementary Table 8.

| Functional category | Protein family | Transcripts [a)] | Proteins [b)] |
|---|---|---|---|
| *Pattern recognition proteins* | | | |
| | peptidoglycan recognition proteins (PGRPs) | 6 (16) | 1 |
| | C-type lectins | 33 (119) | 5 |
| | R-type lectins | 6 (18) | 4 |
| | SUEL/rhamnose-binding lectins | 7 (22) | 1 |
| | galectins | 3 (3) | 1 |
| | fucolectin | 1 (23) | 0 |
| | fibrinogen-related proteins (FREPs) | 27 (161) | 1 |
| | toll-like/variable lymphocyte receptor-like (TLR/VLR) | 13 (52) | 1 |
| | scavenger receptor cysteine-rich (SRCR) domain proteins | 25 (164) | 1 |
| | beta-1,3-glucan binding protein | 1 (1) | 1 |
| | novel immunoglobulin I-set proteins | 16 (17) | 1 |
| | novel immunoglobulin V-set proteins | 9 (22) | 1 |
| *Antimicrobial proteins* | | | |
| | lumbricin | 1 (1) | 1 |
| | invertebrate-type lysozyme | 1 (1) | 0 |
| | bactericidal permeability increasing protein BPI | 1 (3) | 1 |
| | insect defensin/reeler-like proteins | 4 (8) | 1 |
| | cysteine-rich secretory proteins (CRSPs) | 6 (28) | 2 |
| | membrane attack complex/perforin | 2 (23) | 0 |
| *Other immune effectors* | | | |
| | ROS modulator 1 | 2 (3) | 1 |
| | alpha-2-macroglobulin | 10 (24) | 1 |
| | kazal-type serpin | 2 (8) | 0 |
| | kunitz-type serpin | 1 (2) | 0 |
| | leukocyte elastase inhibitors | 5 (25) | 1 |
| | phosphatidylethanolamine-binding protein PEBP | 3 (5) | 1 |
| *Immune response regulators* | | | |
| | Toll/interleukin-1 receptor (TIR) domain proteins | 5 (9) | 0 |
| | NF-kappa-B inhibitor Cactus | 2 (6) | 0 |
| | dorsal protein | 2 (3) | 1 |
| | evolutionarily conserved signaling intermediate in Toll (ECSIT) | 1 (1) | 0 |
| | Pelle protein | 1 (1) | 0 |
| | Relish protein | 1 (6) | 0 |
| | mitogen-activated protein kinase kinase kinase 7 (TAK1) | 1 (1) | 0 |
| | I-kappa-B-kinase alpha (IKK α) | 1 (1) | 0 |
| | I-kappa-B-kinase beta (IKK β) | 1 (5) | 0 |
| | interleukin-1 receptor-associated kinase 1 (IRAK1) | 1 (2) | 0 |
| | mitogen-activated protein kinase kinase kinase 4 (MEKK4) | 1 (1) | 0 |
| | sterile alpha and TIR motif-containing protein (SARM) | 1 (3) | 0 |
| | Toll-interacting protein Tollip | 1 (2) | 1 |

| | | | |
|---|---|---|---|
| | (LPS-induced) tumor necrosis factor (TNF) | 3 (7) | 0 |
| | Tumor necrosis factor (TNF) | 4 (6) | 0 |
| | tumor necrosis factor alpha-induced protein 3 (TNFAIP3) | 1 (3) | 0 |
| | tumor necrosis factor receptor associated proteins (TRAF) | 9 (16) | 0 |
| | IFN regulatory factor | 8 (10) | 0 |
| | IFN regulatory factor-binding protein | 1 (1) | 0 |
| | IFN-induced GTPase | 7 (23) | 1 |
| | macrophage migration inhibitory factor (MIF) | 3 (24) | 1 |
| | ILN enhancer binding factor 2 | 2 (4) | 1 |
| | ILN-16 | 1 (1) | 0 |
| *Digestive enzymes* | | | |
| | carboxypeptidases | 11 (22) | 0 |
| | cathepsins total | 15 (28) | 4 |
| | cathepsin B | 3 (12) | 1 |
| | cathepsin C | 2 (4) | 0 |
| | cathepsin F | 3 (3) | 1 |
| | cathepsin L | 5 (6) | 2 |
| | cathepsin O | 1 (2) | 0 |
| | cathepsin Z | 1 (1) | 0 |
| | chymotrypsins | 3 (17) | 1 |
| | pancreatic elastase | 1 (1) | 1 |
| | alpha amylase | 2 (3) | 1 |
| | lysosomal alpha glucosidase | 1 (3) | 1 |
| | acid trehalase | 2 (4) | 1 |
| | sucrase-isomaltase | 1 (1) | 0 |
| | lysosomal acid lipase | 1 (1) | 0 |
| *Respiration* | | | |
| | hemerythrin | 2 (2) | 2 |
| | giant extracellular hemoglobin, globin chains | 12 (16) | 8 |
| | giant extracellular hemoglobin, linker chains | 6 (18) | 5 |
| sum | | 316 (1032) | 60 |

[a)] Number of transcripts defined as trinity components, which approximately correspond to genes; see [13]; in parentheses: number of contigs (isoforms or fragments)
[b)] Number of unique proteins

## Table 2: Difference in symbiont protein content in fresh worms compared to starved worms

Significant differences between fresh and starved samples were determined with the Student's t-Test; nSpC, normalized spectral counts. For more extensive details see Supplementary Table 3.

| | Symbionts fresh worms | Symbionts starved worms |
|---|---|---|
| average nSpC | 3050.30 | 1869.77 |
| standard deviation | 421.88 | 122.47 |
| # replicates | 3 | 3 |
| p-value t-test | 0.00963 | |

## Table 3: Digestive enzymes expressed in *Olavius algarvensis*

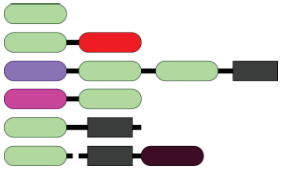
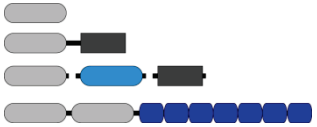
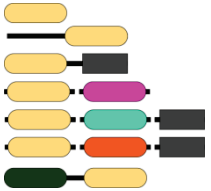
| Transcript ID [a)] | Protein ID | Annotation | Consensus localization evidence [b)] | Substrate or function |
|---|---|---|---|---|
| ***Protein digestion*** | | | | |
| comp310626_c3 | n.d. | pancreatic carboxypeptidase A1 | secreted | release of C-terminal amino acids |
| comp330196_c1 | n.d. | pancreatic carboxypeptidase A2 | secreted | release of C-terminal amino acids |
| comp209868_c0 | n.d. | pancreatic carboxypeptidase A2 | secreted | release of C-terminal amino acids |
| comp329532_c3 | n.d. | pancreatic carboxypeptidase A2 | secreted or membrane | release of C-terminal amino acids |
| comp328734_c12 | n.d. | carboxypeptidase | secreted | peptides |
| comp328734_c1 | n.d. | carboxypeptidase | secreted | peptides |
| comp328734_c4 | n.d. | carboxypeptidase | secreted | peptides |
| comp326419_c0 | n.d. | uncharacterized carboxypeptidase | secreted | peptides |
| comp330196_c2 | n.d. | uncharacterized carboxypeptidase | secreted | peptides |
| comp319717_c3 | n.d. | lysosomal Pro-X carboxypeptidase | lysosomal | proline - amino acid bonds |
| comp320275_c1 | n.d. | chymotrypsin A | secreted | Tyr-/Trp-/Phe-/Leu-\|-Xaa bonds |
| comp306409_c1 | n.d. | chymotrypsin B | secreted | Tyr-/Trp-/Phe-/Leu-\|-Xaa bonds |
| comp334148_c2 | BF11_334148_c2_seq1_11<br>BF11_334148_c2_seq2_10 | chymotrypsin-like protease ctrl-1 | secreted | proteins |
| comp331491_c1 | BF11_331491_c1_seq1_7 | pancreatic elastase | secreted | proteins |
| comp334775_c2 | n.d. | cathepsin B | secreted | Arg – Arg/–Xaa bonds |
| comp335560_c0 | BF11_335560_c0_seq5_5<br>BF11_335560_c0_seq9_5 | cathepsin B | lysosomal | Arg – Arg/–Xaa bonds |
| comp306522_c2 | n.d. | cathepsin F | lysosomal | peptides, cleaves Phe/Leu |
| comp306522_c3 | n.d. | cathepsin F | lysosomal | peptides, cleaves Phe/Leu |
| comp308536_c0 | BF11_308536_c0_seq1_12 | cathepsin F | lysosomal | peptides, cleaves Phe/Leu |
| comp283346_c1 | n.d. | cathepsin L | lysosomal | proteins |
| comp306922_c1 | BF11_306922_c1_seq1_34 | cathepsin L | lysosomal | proteins |
| comp315575_c0 | n.d. | cathepsin L | lysosomal | proteins |
| comp315575_c2 | FD_315575_c2_seq1:64:270:1:+ | cathepsin L | lysosomal | proteins |
| comp328653_c0 | n.d. | cathepsin L | lysosomal | proteins |
| comp329800_c6 | n.d. | cathepsin O | lysosomal | peptides (endopeptidase) |
| comp314408_c1 | n.d. | cathepsin Z | lysosomal | C-terminal amino acids (not Pro) |

| | | | | |
|---|---|---|---|---|
| comp308700_c0 | n.d. | cathepsin C (dipeptidyl peptidase I) | lysosomal | release of an N-terminal dipeptide |
| comp329456_c1 | n.d. | cathepsin C (dipeptidyl peptidase I) | lysosomal | release of an N-terminal dipeptide |
| comp328746_c4 | n.d. | alpha amylase | secreted | α1->4 glycosidic bonds |
| comp324906_c1 | BF11_324906_c1_seq1_9<br>BF11_324906_c1_seq2_12 | alpha amylase | secreted or membrane | α1->4 glycosidic bonds |
| comp335205_c1 | BF11_335205_c1_seq1_20 | lysosomal alpha glucosidase | secreted or membrane | α1->4 glycosidic bonds |
| comp320084_c0 | n.d. | lysosomal beta-mannosidase | lysosomal | cleaves terminal β-D-mannose |
| comp334411_c3 | n.d. | sucrase-isomaltase | secreted or membrane | α1->6 glycosidic bonds |
| comp329957_c8 | BF11_329957_c8_seq2_17 | acid trehalase | secreted or membrane | trehalose -> glucose |
| comp335402_c7 | n.d. | acid trehalase | secreted or membrane | trehalose -> glucose |
| ***Lipid degradation*** | | | | |
| comp22535_c0 | n.d. | lysosomal acid lipase | secreted | hydrolyzes steryl esters |
| comp249291_c0 | n.d. | lysozyme | secreted | peptidoglycan (glycosidic bonds) |
| comp250229_c0 | BF11_250229_c0_seq1_15<br>BF11_250229_c0_seq2_17 | peptidoglycan recognition protein | secreted | peptidoglycan (peptide bonds) |
| comp335695_c10 | n.d. | peptidoglycan recognition protein | secreted or membrane | peptidoglycan (peptide bonds) |
| comp330541_c4 | n.d. | peptidoglycan recognition protein | secreted | peptidoglycan (peptide bonds) |
| comp314994_c0 | n.d. | peptidoglycan recognition protein | secreted or membrane | peptidoglycan (peptide bonds) |
| comp332570_c2 | n.d. | peptidoglycan recognition protein | secreted or membrane | peptidoglycan (peptide bonds) |

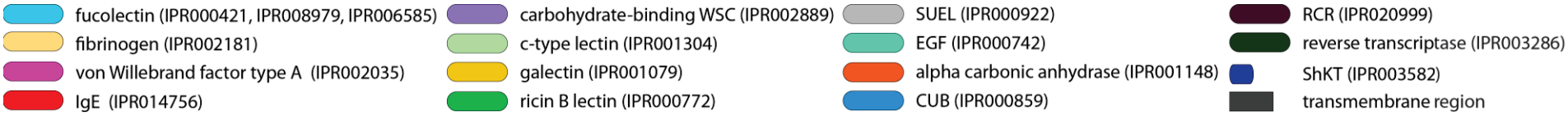
[a)] Defined as trinity components; see [13]

[b)] Probable subcellular localization of proteins based on the results of TMHMM [116, 117], SignalP-4 [118], Phobius [119], TargetP [120], DISTILL [121], LocTree3 [122], BaCelLo [123], and iLoc-Animal [124]. *Secreted* – enters secretory pathway and is either excreted to the extracellular space or confined to non-cytoplasmic insides of intracellular compartments; *membrane* – predicted to be membrane integral; *lysosomal* – predicted to be targeted towards the lysosome. See supplementary table 14 for details on localization evidence.

## Table 4: Lectins expressed in *Olavius algarvensis*.

| Lectin group | Transcripts [a] | Proteins [b] | Domain architectures | Potential functions |
|---|---|---|---|---|
| *fucolectins* | 1 | 0 |  | glycan recognition and host defense [125] |
| *C-type lectins* | 33 | 5 |  | glycan recognition [84], symbiont recognition and acquisition [126] |
| *R-type lectins* | 6 | 4 |  | glycan recognition [127], antimicrobial [128, 129], host defense [130] |
| *galectins* | 3 | 1 |  | [131, 132], host defense [133, 134] |
| *SUEL rhamnose-binding lectins* | 7 | 1 |  | MAMP recognition [135, 136], egg fertilization [137], development [138] |
| *fibrinogen-like proteins* | 27 | 1 |  | [91], MAMP recognition and antimicrobial activity [78], development [139], egg fertilization [140] |

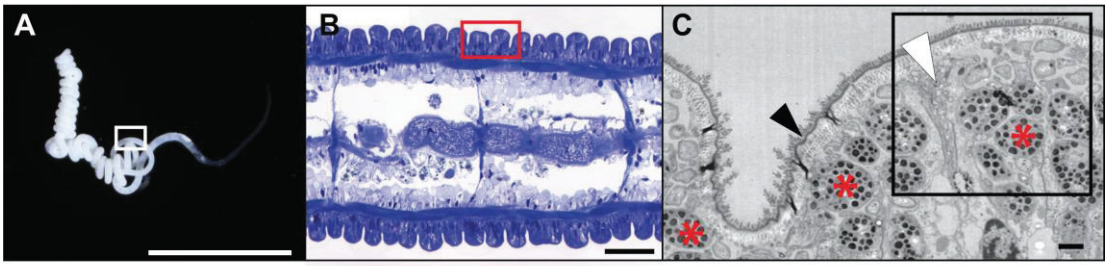
[a] Number of transcripts, defined as trinity components; see [13]
[b] Number of identified proteins in the proteomes

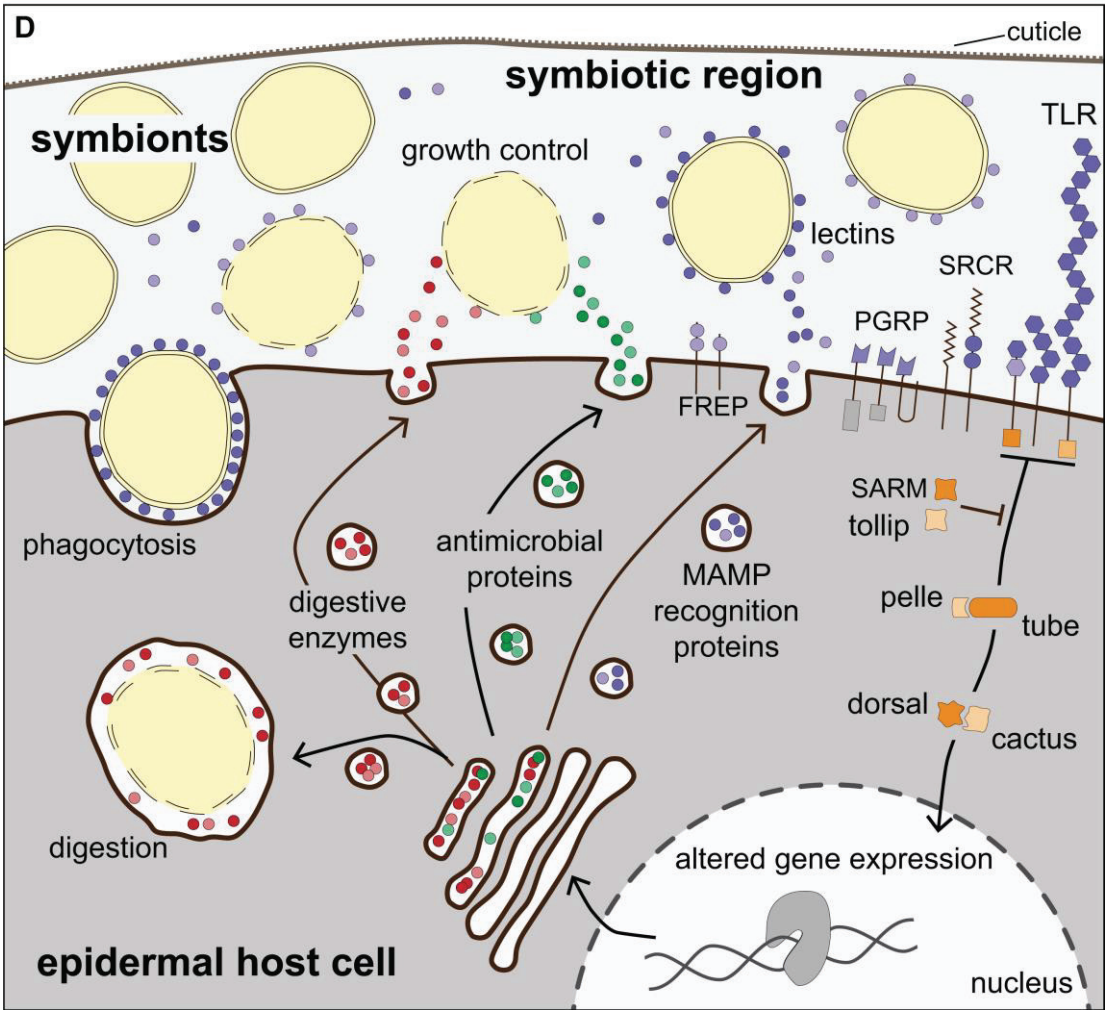


# Figures

## Figure 1: Schematic overview of hypothetical molecular host-symbiont interactions

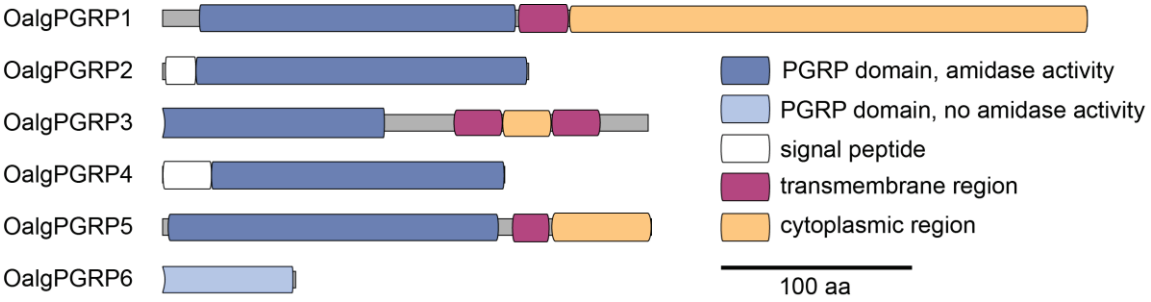
A) Light micrograph of an *Olavius algarvensis* worm. White box frames a region corresponding to the tissue section shown in B. Scale bar 5 mm. B) Light micrograph of a longitudinal section through *O. algarvensis*, tissue stained with toluidine blue. The red box frames a region corresponding to the TEM section shown in C. Scale bar 50 µm. C) Transmission electron micrograph of the symbiotic region, longitudinal section. Black box frames a region corresponding to the schematic representation shown in D. Red asterisks, symbiont cells; black arrow, cuticle; white arrow, epidermal cell extensions. Scale bar 5 µm. Images A, B, and C do not show the same worm specimen. D) Schematic overview of the main groups of expressed pattern recognition molecules, components of the Toll immune signaling pathway and proposed interactions between the host and its symbionts. Ig, immunoglobulin domain proteins; PGRP, peptidoglycan recognition proteins; SRCR scavenger receptor-like cysteine rich proteins; TLR, Toll-like receptors; FREP, fibrinogen-related proteins; AMPs, antimicrobial proteins.



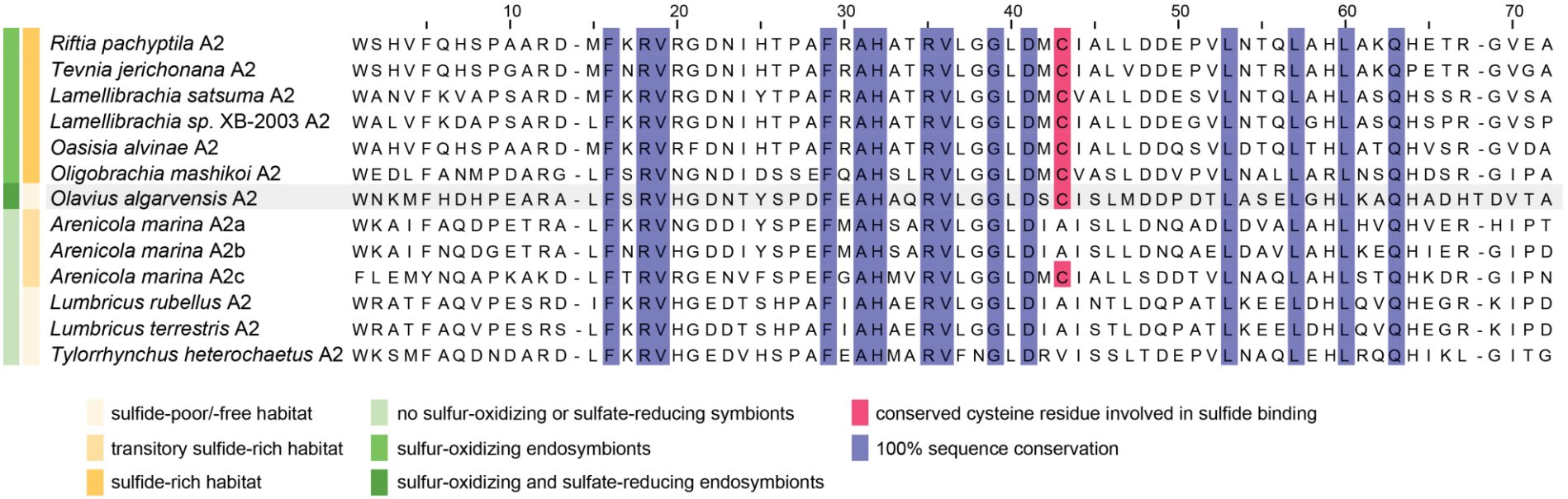
## Figure 2: Domain structures of peptidoglycan recognition proteins

Structure of conserved functional domains in *Olavius algarvensis* peptidoglycan recognition proteins; OalgPGRP1: comp330541_c4; OalgPGRP2: comp250229_c0; OalgPGRP3: comp335695_c10; OalgPGRP4: comp314994_c0; OalgPGRP5: comp332570_c2; OalgPGRP6: comp1100768_c0.



## Figure 3: Protein alignment of hemoglobin A2 chains

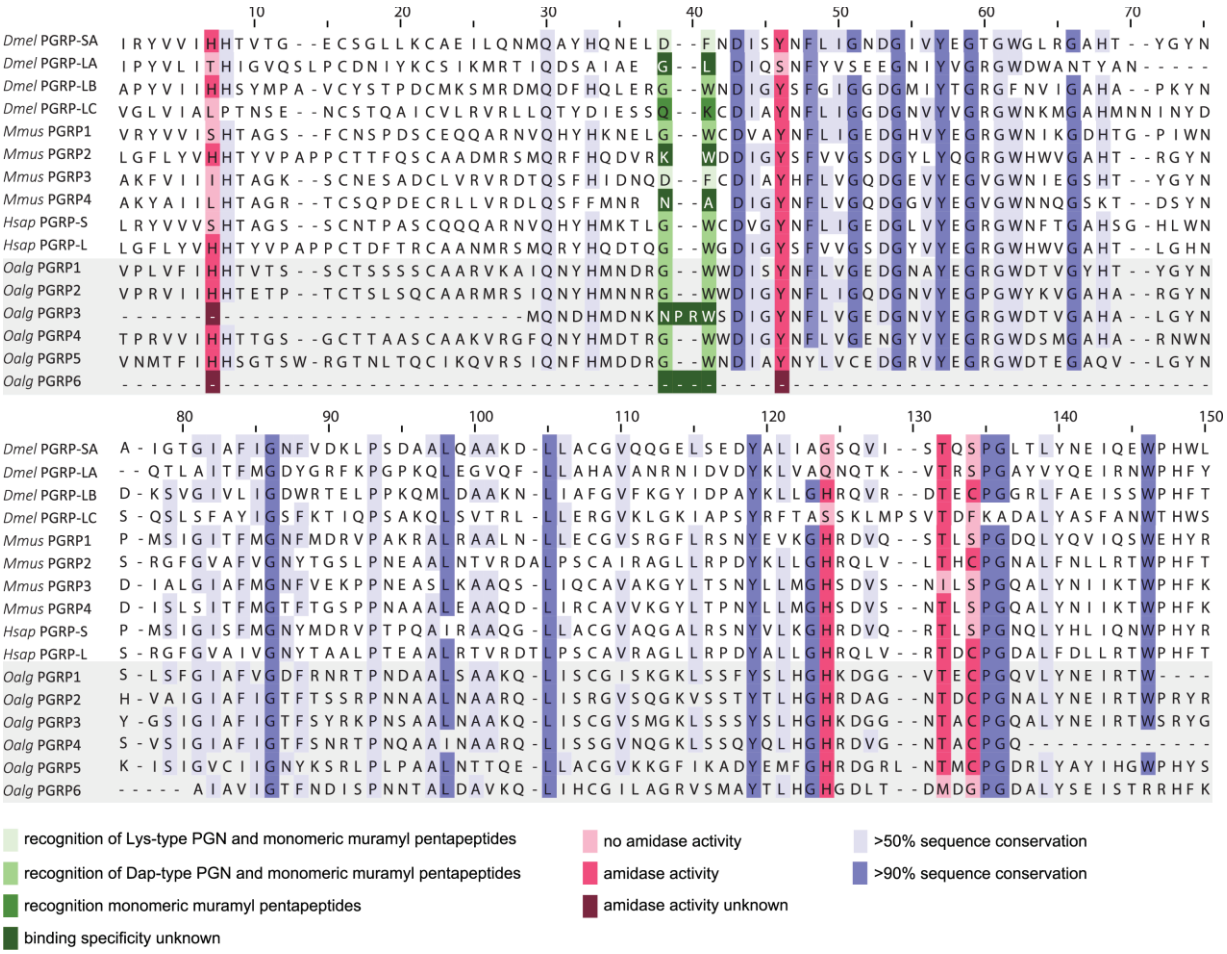
Protein alignment of hemoglobin A2 chains from marine and terrestrial annelids: *Riftia pachyptila* (GenBank accession number: CAD29155), *Tevnia jerichonana* (GenBank accession number: AAP04530), *Lamellibrachia satsuma* (GenBank accession number: BAN58231), *Lamellibrachia sp.* XB-2003 (GenBank accession number: AAP04528), *Oasisia alvinae* (GenBank accession number: AAP04531), *Oligobrachia mashikoi* (GenBank accession number: Q7M413), *Arenicola marina* (GenBank accession numbers: A2a, CAI56308; A2b, CAJ32740; A2c, CAJ32741), *Lumbricus rubellus* (GenBank accession number: BF422675.2), *Lumbricus terrestris* (GenBank accession number: P02218), *Tylorrhynchus heterochaetus* (GenBank accession number: P09966) and *Olavius algarvensis* (comp287449_c0_seq1).



## Figure 4: Protein alignment of peptidoglycan recognition proteins

Protein alignment of PGRP domain sequences from different model organisms and *Olavius algarvensis*; Dmel *Drosophila melanogaster* (GenBank accession numbers: PGRP-SA, Q9VYX7; PGRP-LA, Q95T64; PGRP-LB, Q8INK6; PGRP-LC, Q9GNK5), Mmus *Mus musculus* (GenBank accession numbers: PGRP1, O88593; PGRP2, Q8VCS0; PGRP3, A1A547; PGRP4, Q0VB07), Hsap *Homo sapiens* (GenBank accession numbers: PGRP-S, O75594; PGRP-L, Q96PD5), Oalg *Olavius algarvensis* (OalgPGRP1, comp330541_c4; OalgPGRP2, comp250229_c0; OalgPGRP3, comp335695_c10; OalgPGRP4, comp314994_c0; OalgPGRP5, comp332570_c2; OalgPGRP6, comp1100768_c0). Conserved active-site residues that confer amidase activity are shown in red; mutation of at least one active-site residue (pink) removes amidase activity.

-



# Supplementary Tables

## Supplementary Table 1: Summary of transcriptome sequencing

Library A was prepared from whole worms freshly collected from the environment. Library B was prepared from whole worms after anoxic incubation (see *Experimental Procedures* for details).

| Sequencing statistic | library A | library B |
|---|---|---|
| number of read pairs | 167,402,116 | 6,041,419 |
| read length (bp) | 100 | 100 |
| total bases | 33,815,227,432 | 1,196,200,962 |
| number of read pairs after quality processing | 159,551,509 | 5,745,537 |
| total bases filtered/trimmed reads | 26,801,302,242 | 987,624,581 |

## Supplementary Table 2: Summary of transcriptome assembly and protein database

| **Transcriptome assembly** | |
|---|---|
| number of contigs | 173,602 |
| number of contigs >= 500 bp | 60,369 |
| number of contigs >= 1000 bp | 23,719 |
| N50 | 1236 |
| total size of assembly (bp) | 100,372,073 |
| number of contigs with blast hits (blastx vs ncbi nr, e-value cut-off 1e-6) | 40,860 |
| number of contigs with interpro hits (interproscan*) | 137,596 |
| **CDS prediction** | |
| predicted non-redundant CDS (getorf) after 99% identity clustering | 1,306,981 |
| predicted CDS (FrameDP) after 99% identity clustering | 54,909 |
| Final non-redundant host sequences | 1,359,455 |

* Databases searched: ProDom [1], Prints [2], PIR [3], Pfam [4], Smart [5], TIGRFAM [6], PROSITE [7], HAMAP [8], SuperFamily [8], SignalP [9], TMHMM [10], Panther [11], Gene3D [12], Phobius [13], Coils [14]

## Supplementary Table 3: Total protein of symbionts from fresh compared to starved whole worms

Listed are nSpC (normalized spectral counts) of proteins assigned to each *O. algarvensis* symbiont. Proteins were assigned as described in [15] to each symbiont species based on the symbiont genomes as published in [16]. Each sample corresponds to one biological replicate sample of *O. algarvensis* worms as described in the *Experimental Procedures* section of this paper. Fresh worm samples were prepared from freshly collected, untreated material, while starved worm samples were prepared from worms that had been incubated for 8 days under oxic conditions without external energy sources that would allow the symbionts to fix carbon and grow. Significant differences between fresh and starved samples were determined with a Student's t-Test (significant values marked "*"). Rep, technical replicate; n.d., not determined.

| protein abundance whole worms fresh | sample 1 nSpC | | sample 2 nSpC | | sample 3 nSpC | | |
|---|---|---|---|---|---|---|---|
| | rep. 1 | rep. 2 | rep. 1 | rep. 2 | rep. 1 | rep. 2 | average nSpC |
| total symbiont | 2682.47 | 2609.84 | 3016.81 | n.d. | 3487.93 | n.d. | 2949.26 |
| total *OalgG1* | 775.99 | 784.99 | 714.70 | n.d. | 780.23 | n.d. | 763.98 |
| total *OalgG3* | 107.02 | 109.42 | 110.08 | n.d. | 120.44 | n.d. | 111.74 |
| total *OalgD1* | 36.20 | 33.61 | 27.08 | n.d. | 39.52 | n.d. | 34.10 |
| total *OalgD4* | 12.19 | 18.24 | 42.50 | n.d. | 57.43 | n.d. | 32.59 |
| total unclassified symbionts | 1750.65 | 1662.39 | 2122.45 | n.d. | 2490.31 | n.d. | 2006.45 |

| protein abundance whole worms starved | sample 1 nSpC | | sample 2 nSpC | | sample 3 nSpC | | |
|---|---|---|---|---|---|---|---|
| | rep 1 | rep 2 | rep 1 | rep 2 | rep 1 | rep 2 | average nSpC |
| total symbiont | 1737.92 | 1724.54 | 1859.96 | 1968.93 | 1790.71 | 2136.52 | 1869.76 |
| total *OalgG1* | 385.41 | 391.17 | 446.44 | 474.63 | 437.98 | 492.11 | 437.96 |
| total *OalgG3* | 60.82 | 73.21 | 84.22 | 101.28 | 80.19 | 78.99 | 79.79 |
| total *OalgD1* | 16.28 | 39.66 | 36.06 | 23.84 | 43.64 | 65.27 | 37.46 |
| total *OalgD4* | 28.06 | 47.39 | 45.46 | 53.41 | 34.83 | 35.26 | 40.74 |
| total unclassified symbionts | 1247.35 | 1172.44 | 1247.78 | 1315.77 | 1193.79 | 1464.89 | 1273.67 |

| protein abundance whole worms | average nSpC fresh worms | | | average nSpC starved worms | | | |
|---|---|---|---|---|---|---|---|
| | sample 1 | sample 2 | sample 3 | sample 1 | sample 2 | sample 3 | p-value t-test |
| total symbiont | 2646.16 | 3016.81 | 3487.93 | 1731.23 | 1914.45 | 1963.62 | 0.00963* |
| total *OalgG1* | 780.49 | 714.70 | 780.23 | 388.29 | 460.54 | 465.05 | 0.00064* |
| total *OalgG3* | 108.22 | 110.08 | 120.44 | 67.02 | 92.75 | 79.59 | 0.01655* |
| total *OalgD1* | 34.91 | 27.08 | 39.52 | 27.97 | 29.95 | 54.46 | 0.71551 |
| total *OalgD4* | 15.22 | 42.50 | 57.43 | 37.73 | 49.44 | 35.05 | 0.86642 |
| total unclassified symbionts | 1706.52 | 2122.45 | 2490.31 | 1209.90 | 1281.78 | 1329.34 | 0.02205* |

## Supplementary Table 4: Expression patterns of typical intestinal proteases in other animals

List of digestive enzymes found in *Olavius algarvensis* and their typical expression location in other animals

| enzyme name | species | common name | tissue expression | reference |
|---|---|---|---|---|
| pancreatic carboxypeptidase A1 | *Bos taurus* | cattle | pancreas | [17] |
| | *Homo sapiens* | human | pancreas | [18] |
| | *Rattus norvegicus* | rat | pancreas | [19] |
| pancreatic carboxypeptidase A2 | *Homo sapiens* | human | pancreas | [20] |
| | *Rattus norvegicus* | rat | pancreas | [21] |
| carboxypeptidase A | *Aedes aegypti* | mosquito | gut | [22] |
| | *Helicoverpa armigera* | corn earworm | midgut | [23] |
| | *Lutzomyia longipalpis* | sand fly | midgut | [24] |
| | *Rhodnius prolixus* | kissing bug | gut | [25] |
| | *Tenebrio molitor* | yellow mealworm | gut | [26] |
| | *Trichoplusia ni* | cabbage looper | midgut | [27] |
| zinc carboxypeptidase | *Eisenia andrei* | earthworm | midgut | [28] |
| chymotrypsin A | *Bos taurus* | cattle | pancreas | [29] |
| | *Gadus morhua* | atlantic cod | gut | [30] |
| chymotrypsin B | *Canis lupus* | dog | pancreas | [31] |
| | *Gadus morhua* | atlantic cod | pyloric caeca | [32] |
| | *Homo sapiens* | human | pancreas | [33] |
| | *Rattus norvegicus* | rat | pancreas | [34] |
| chymotrypsin | *Agrotis ipsilon* | black cutworm | midgut | [35] |
| | *Haliotis rufescens* | red abalone | digestive gland | [36] |
| | *Helicoverpa armigera* | corn earworm | gut | [37] |
| | *Helicoverpa zea* | corn earworm | midgut | [35] |
| | *Heliothis virescens* | tobacco budworm | midgut | [38] |
| | *Mayetiola destructor* | hessian fly | midgut | [39] |
| | *Sepia officinalis* | cuttlefish | hepatopancreas | [40] |
| | *Tenebrio molitor* | yellow mealworm | midgut | [41] |
| chymotrypsin-like protease ctrl-1 | *Homo sapiens* | human | pancreas | [42] |

| | | | | |
|---|---|---|---|---|
| pancreatic elastase | *Homo sapiens* | human | pancreas | [43] |
| | *Manduca sexta* | tobacco hawkmoth | midgut | uniprot accession: Q25510 |
| | *Rattus norvegicus* | rat | pancreas | [44] |
| | *Sus scrofa* | pig | pancreas | [45] |
| cathepsin B | *Eisenia andrei* | earthworm | midgut | [28] |
| | *Haliotis discus hannai* | Pacific abalone | digestive gland | [46] |
| | *Litopenaeus vannamei* | white shrimp | midgut | [47] |
| | *Meretrix meretrix* | Asiatic hard clam | digestive gland | [48] |
| | *Necator americanus* | human hookworm | gut | [49] |
| | *Opisthorchis viverrini* | human liver fluke | gut | [50] |
| | *Pandalus borealis* | northern shrimp | hepatopancreas | [51] |
| | *Tityus serrulatus* | scorpion | midgut | [52] |
| cathepsin F | *Clonorchis sinensis* | Chinese liver fluke | intestine | [53] |
| | *Homo sapiens* | human | ubiquitous | [54] |
| | *Mus musculus* | mouse | ubiquitous | [55] |
| | *Paralichthys olivaceus* | olive flounder | ubiquitous | [56] |
| | *Tityus serrulatus* | scorpion | midgut | [52] |
| cathepsin L | *Cristaria plicata* | cockscomb pearl mussel | hepatopancreas | [57] |
| | *Drosophila melanogaster* | fruit fly | midgut | [58] |
| | *Eriocheir sinensis* | Chinese mitten crab | hepatopancreas | [59] |
| | *Homo sapiens* | human | ubiquitous | [60] |
| | *Metapenaeus ensis* | greasyback shrimp | hepatopancreas | [61] |
| | *Mus musculus* | mouse | ubiquitous | [62] |
| | *Pinctada fucata* | pearl oyster | digestive gland | [63] |
| | *Rattus norvegicus* | rat | ubiquitous | [64] |
| | *Tenebrio molitor* | yellow mealworm | midgut | [65] |
| | *Tityus serrulatus* | scorpion | midgut | [52] |

## Supplementary Table 5: High similarity of host digestive proteins to earthworm midgut enzymes

*Olavius algarvensis* digestive enzymes were blasted (tblastn) against the only available gut-specific EST library of an oligochaete, the earthworm *Eisenia andrei*; n.d., not detected; n.a., not available, aa, amino acid.

| transcript ID [a)] | protein ID | annotation | FPKM [b)] | Proteome [c)] | Accession [d)] | e-value | aa similarity |
|---|---|---|---|---|---|---|---|
| ***Protein digestion*** | | | | | | | |
| comp330196_c1 | n.d. | carboxypeptidase a2 | 8.95 | n.d. | BP524493.1 | 1.16E-60 | 68% |
| comp310626_c3 | n.d. | carboxypeptidase a1 | 4.70 | n.d. | BP524493.1 | 3.21E-20 | 62% |
| comp209868_c0 | n.d. | carboxypeptidase a2 | 0.53 | n.d. | no hit | n.d. | n.d. |
| comp329532_c3 | n.d. | carboxypeptidase a2 | 7.24 | n.d. | BP524493.1 | 2.34E-49 | 66% |
| comp328734_c12 | n.d. | carboxypeptidase | 1.98 | n.d. | no hit | n.d. | n.d. |
| comp328734_c1 | n.d. | carboxypeptidase | 0.00 | n.d. | BP524487.1 | 7.13E-10 | 59% |
| comp326419_c0 | n.d. | carboxypeptidase | 14.89 | n.d. | BP524487.1 | 1.64E-28 | 54% |
| comp330196_c2 | n.d. | carboxypeptidase | 8.49 | n.d. | BP524771.1 | 7.1E-9 | 65% |
| comp319717_c3 | n.d. | lysosomal Pro-X carboxypeptidase | 0.31 | n.d. | no hit | n.d. | n.d. |
| comp320275_c1 | n.d. | chymotrypsin A | 10.47 | n.d. | no hit | n.d. | n.d. |
| comp306409_c1 | n.d. | chymotrypsin B | 1.60 | n.d. | BP524415.1 | 9.96E-11 | 67% |
| comp334148_c2 | BF11_334148_c2_seq1_11<br>BF11_334148_c2_seq2_10 | chymotrypsin-like protease ctrl-1 | 29.94 | 4.69 | BP524722.1 | 1.28E-14 | 65% |
| comp331491_c1 | BF11_331491_c1_seq1_7 | pancreatic elastase | 14.03 | 3.92 | BP524414.1 | 1.57E-10 | 55% |
| comp334775_c2 | n.d. | cathepsin B | 1.96 | n.d. | no hit | n.d. | n.d. |
| comp335560_c0 | BF11_335560_c0_seq5_5<br>BF11_335560_c0_seq9_5 | cathepsin B | 355.52 | 1.03 | BP524429.1 | 9.23E-11 | 70% |
| comp306522_c2 | n.d. | cathepsin F | 0.82 | n.d. | no hit | n.d. | n.d. |
| comp306522_c3 | n.d. | cathepsin F | 5.83 | n.d. | no hit | n.d. | n.d. |
| comp308536_c0 | BF11_308536_c0_seq1_12 | cathepsin F | 126.33 | 0.03 | no hit | n.d. | n.d. |
| comp283346_c1 | n.d. | cathepsin L | 0.19 | n.d. | no hit | n.d. | n.d. |
| comp306922_c1 | BF11_306922_c1_seq1_34 | cathepsin L | 498.41 | 1.87 | BP524408.1 | 3.86E-53 | 83% |
| comp315575_c0 | n.d. | cathepsin L | 20.01 | n.d. | no hit | n.d. | n.d. |
| comp315575_c2 | FD_315575_c2_seq1:64:270:1:+ | cathepsin L | 39.31 | n.d. | no hit | n.d. | n.d. |
| comp328653_c0 | n.d. | cathepsin L | 10.39 | n.d. | BP524408.1 | 6.31E-43 | 78% |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| comp329800_c6 | n.d. | cathepsin O | 4.58 | n.d. | BP524408.1 | 9.79E-7 | 51% |
| comp314408_c1 | n.d. | cathepsin Z | 0.79 | n.d. | no hit | n.d. | n.d. |
| ***Carbohydrate digestion*** | | | | | | | |
| comp328746_c4 | n.d. | alpha amylase | 2.71 | n.d. | no hit | n.d. | n.d. |
| comp324906_c1 | BF11_324906_c1_seq1_9<br>BF11_324906_c1_seq2_12 | alpha amylase | 25.14 | 0.00 | no hit | n.d. | n.d. |
| comp335205_c1 | BF11_335205_c1_seq1_20 | lysosomal alpha glucosidase | 30.77 | 2.91 | no hit | n.d. | n.d. |
| comp320084_c0 | n.d. | lysosomal beta-mannosidase | 4.26 | n.d. | no hit | n.d. | n.d. |
| comp334411_c3 | n.d. | sucrase-isomaltase | 0.00 | n.d. | no hit | n.d. | n.d. |
| comp329957_c8 | BF11_329957_c8_seq2_17 | acid trehalase | 20.77 | 0.03 | no hit | n.d. | n.d. |
| comp335402_c7 | n.d. | acid trehalase | 3.32 | n.d. | no hit | n.d. | n.d. |
| ***Lipid degradation*** | | | | | | | |
| comp22535_c0 | n.d. | lysosomal acid lipase | 1.34 | n.d. | no hit | n.d. | n.d. |
| ***Peptidoglycan degradation*** | | | | | | | |
| comp249291_c0 | n.d. | lysozyme | 460.97 | n.d. | BP524379.1 | 3.00E-28 | 78% |
| comp250229_c0 | BF11_250229_c0_seq1_15<br>BF11_250229_c0_seq2_17 | peptidoglycan recognition protein | 180.88 | n.d. | BP524394.1 | 8.43E-35 | 66% |
| comp335695_c10 | n.d. | peptidoglycan recognition protein | 123.41 | n.d. | BP524394.1 | 3.12E-25 | 69% |
| comp330541_c4 | n.d. | peptidoglycan recognition protein | 43.36 | n.d. | BP524394.1 | 4.86E-35 | 70% |
| comp314994_c0 | n.d. | peptidoglycan recognition protein | 7.11 | n.d. | BP524394.1 | 3.09E-32 | 66% |
| comp332570_c2 | n.d. | peptidoglycan recognition protein | 21.15 | n.d. | BP524394.1 | 1.35E-31 | 68% |

[a] Defined as trinity components, see [66]
[b] Average transcript abundance in transcriptomes
[c] Average protein abundance in proteomes in $^{0}/_{000}$
[d] GenBank accession numbers (NCBI EST)

## Supplementary Table 6: Similarity of host digestive glucosidases to intestinal glucosidases of other animals

| transcript ID | conserved domains | annotation | database | Blast Hit Description (HSP) | organism | NCBI accession | e-value | aa sim. |
|---|---|---|---|---|---|---|---|---|
| comp329957_c8_seq1 | glycoside hydrolase family 65 (PF03632) trehalose and maltose hydrolase (COG1554) | acid trehalase | ncbi nr | maltose phosphorylase | *Culex quinquefasciatus* | XP_001850772.1 | 1,41E-68 | 48% |
| | | | ncbi nr | acid trehalase-like protein 1 | *Homo sapiens* | NP_079368.3 | 1,41E-68 | 49% |
| | | | swissprot | acid trehalase-like protein 1 | *Homo sapiens* | Q32M88 | 1,00E-77 | 29% |
| comp329957_c8_seq2 | glycoside hydrolase family 65 (PF03632) trehalose and maltose hydrolase (COG1554) | acid trehalase | ncbi nr | maltose phosphorylase | *Culex quinquefasciatus* | XP_001850772.1 | 1,56E-68 | 48% |
| | | | ncbi nr | acid trehalase-like protein 1 | *Homo sapiens* | NP_079368.3 | 1,56E-68 | 49% |
| | | | swissprot | acid trehalase-like protein 1 | *Homo sapiens* | Q32M88 | 1,00E-77 | 29% |
| comp335402_c7_seq1 | glycoside hydrolase family 65 (PF03632) | acid trehalase | ncbi nr | acid trehalase-like protein 1 | *Danio rerio* | NP_001071193.1 | 1,93E-61 | 66% |
| | | | ncbi nr | acid trehalase-like protein 1 | *Mus musculus* | NP_663362.2 | 1,30E-57 | 63% |
| | | | swissprot | acid trehalase-like protein 1 | *Danio rerio* | A0JMP0 | 3,00E-75 | 48% |
| comp334411_c3_seq1 | alpha-glucosidase (PTHR22762) | sucrase-isomaltase | ncbi nr | sucrase-isomaltase | *Mus musculus* | ACH86012.1 | 4,12E-07 | 61% |
| | | | swissprot | glucoamylase | *Homo sapiens* | Q2M2H8 | 5,00E-08 | 41% |
| | | | swissprot | sucrase-isomaltase, intestinal | *Suncus murinus* | O62653 | 2,00E-07 | 40% |
| comp328746_c4_seq1 | alpha amylase catalytic domain family (cl07893) | pancreatic alpha-amylase | ncbi nr | amylase, alpha 2A, pancreatic | *Danio rerio* | AAH62867.1 | 1,26E-19 | 70% |
| | | | swissprot | salivary alpha-amylase | *Homo sapiens* | P04745 | 7,00E-22 | 56% |
| comp324906_c1_seq1 | alpha amylase, catalytic domain (PFAM00128) | salivary alpha-amylase | ncbi nr | alpha-amylase | *Aedes aegypti* | XP_001656785.1 | 4,27E-80 | 50% |
| | | | ncbi nr | salivary alpha-glucosidase | *Culex tarsalis* | ACJ64288.1 | 9,22E-75 | 53% |
| | | | swissprot | probable maltase | *Aedes aegypti* | P13080 | 6,00E-85 | 34% |
| comp324906_c1_seq2 | alpha amylase, catalytic domain (PFAM00128) | salivary alpha-amylase | ncbi nr | alpha-amylase | *Aedes aegypti* | XP_001656785.1 | 1,35E-89 | 51% |
| | | | ncbi nr | salivary alpha-glucosidase | *Culex tarsalis* | ACJ64288.1 | 4,67E-82 | 54% |
| | | | swissprot | probable maltase | *Aedes aegypti* | P13080 | 8,00E-89 | 35% |

## Supplementary Table 7: *O. algarvensis* giant extracellular hemoglobin sequences

Table lists details on giant hemoglobin (globin and linker) chains identified in *Olavius algarvensis,* n.d. not detected.

| transcript component ID [a] | protein ID | annotation | signalP [b] | conserved domains | FPKM [c] | Proteome [d] |
|---|---|---|---|---|---|---|
| comp254846_c0 | BF11_254846_c0_seq1_18 | hemoglobin chain | no | globin IPR000971 | 39.05 | 1.31 |
| comp287449_c0 | BF11_287449_c0_seq1_4 | hemoglobin chain | yes | none | 911.55 | 32.18 |
| comp300058_c0 | n.d. | hemoglobin chain | yes | globin IPR000971 | 6.40 | n.d. |
| comp300058_c1 | n.d. | hemoglobin chain | no | globin IPR000971 | 6.65 | n.d. |
| comp307848_c0 | BF11_307848_c0_seq1_17 | hemoglobin chain | yes | none | 980.87 | 31.07 |
| comp309954_c0 | n.d. | hemoglobin chain | no | globin IPR000971 | 22.33 | n.d. |
| comp321227_c0 | BF11_321227_c0_seq1_9 | linker chain | yes | LDL receptor class A repeat IPR002172 | 726.5 | 2.46 |
| comp321285_c0 | BF11_321285_c0_seq1_16, BF11_321285_c0_seq2_7 | hemoglobin and linker chains | yes | globin IPR000971 | 5679.8 | 41.16 |
| comp324803_c0 | BF11_324803_c0_seq1_26 | linker chain | no | none | 400.82 | 4.35 |
| comp326155_c0 | n.d. | linker chain | no | LDL receptor class A repeat IPR002172 | 3.51 | n.d. |
| comp327543_c0 | BF11_327543_c0_seq1_6 | hemoglobin chain | no | none | 2.84 | 5.59 |
| comp327543_c2 | BF11_327543_c2_seq1_2 | hemoglobin chain | yes | globin IPR000971 | 3.75 | 0.26 |
| comp327754_c0 | BF11_327754_c0_seq1_26 | hemoglobin chain | yes | globin IPR000971 | 2289.10 | 36.27 |
| comp328129_c0 | BF11_328129_c0_seq1_1 | hemoglobin chain | yes | globin IPR000971 | 38.25 | 1.52 |
| comp331953_c0 | BF11_331953_c0_seq1_4 | linker chain | no | LDL receptor class A repeat IPR002172 | 64.48 | 0.00 |
| comp332275_c3 | BF11_332275_c3_seq1_14 | hemoglobin chain | no | globin IPR000971 | 4.37 | 2.23 |
| comp335909_c2 | BF11_335909_c2_seq2_7 | linker chain | no | LDL receptor class A repeat IPR002172 | 5.86 | 0.00 |

[a] Defined as trinity components; see [66]

[b] Presence/absence of signal peptide as determined with SignalP [9]

[c] Average transcript abundance in transcriptomes

[d] Average protein abundance in proteomes in $^{0}/_{000}$

## Supplementary Table 8: Proteins potentially involved in symbiont interaction in *O. algarvensis*, all transcripts and proteins

Table is too large; refer to supplementary file “Supplementary_Table8.xlsx”

## Supplementary Table 9: Hemerythrin expression in annelids

Out of 67 annelid species for which EST data is available in the public databases, hemerythrin was only expressed in the seven species listed here.

| species | class/subclass | library name | #hemerythrin ESTs/total ESTs | accession |
|---|---|---|---|---|
| *Lumbricus rubellus* | Oligochaeta | LIBEST_007207 Earthworm Lambda Zap Express Library (whole worm adult) | 2/1925 | EL517735.1, EL517681.1 |
| | | LIBEST_014450 Earthworm Cadmium Exposure Library (whole worm adult) | 2/2230 | DR077506.1, DR077413.1 |
| | | LIBEST_015953 Earthworm Head Enriched library (adult) | 5/2569 | DR077018.1, DR076532.1, DR076454.1, DR076360.1, DR075863.1 |
| | | LIBEST_017401 Earthworm Copper Exposure Library (whole worm adult) | 2/1518 | DR009706.1, DR009119.1 |
| | | LIBEST_016202 Juvenile Earthworm Library (whole worm juvenile) | 1/2895 | CV072407.1 |
| | | LIBEST_014289 Lumbricus rubellus Late Cocoon Library 1 (whole worm late cocoon) | 14/2728 | CO046674.1, CO046632.1, CF839199.1, CF810104.1, CF809646.1, CF809632.1, CF426790.2, CF416883.2, CF416571.2, CF416434.2, CF416392.2, CF415962.2, CF415927.2, CF416875.1 |
| *Perionyx excavatus* | Oligochaeta | LIBEST_022800 Perionyx excavatus regenerating tissue cDNA library (head regeneration, adult worms) | 8/1195 | BP998686.1, BP998685.1, BP998684.1, BP998683.1, BP998682.1, BP998681.1, BP998680.1, BP998679.1 |
| *Amynthas koreanus* | Oligochaeta | LIBEST_026527 Amynthas koreanus whole body cDNA library | 3/2150 | FS507482.1, FS507353.1, FS507056.1 |
| *Eisenia fetida* | Oligochaeta | LIBEST_026326 Earthworm SSH cDNA library (*E. coli* challenged adults, whole worms) | 2/394 | HO001559.1, HO001494.1 |
| *Haementeria depressa* | Hirudinea | LIBEST_015769 Leech Haementeria depressa library HDA (salivary gland of adult worms) | 18/891 | CN807659.1, CN807658.1, CN807657.1, CN807656.1, CN807655.1, CN807654.1, CN807653.1, CN807652.1, CN807651.1, CN807650.1, CN807649.1, CN807648.1, CN807647.1, CN807646.1, CN807645.1, CN807644.1, CN807643.1, CN807642.1 |
| Macrobdella decora | Hirudinea | LIBEST_028114 North American medicinal leech Macrobdella decora salivary gland library | 4/1604 | JZ187845.1, JZ186990.1, JZ186989.1, JZ186988.1 |
| *Riftia pachyptila* | Polychaeta | symbiont-bearing trophosome (TR-BW) | 5/59 | [67] |

## Supplementary Table 10: Host c-type lectins with similarity to known immune lectins

| transcript ID | NCBI Accession | Blast Hit Description (HSP) | organism | e-value | aa similarity |
|---|---|---|---|---|---|
| comp289680_c0_seq1 | NP_001178934.1 | CD209 antigen-like protein E | *Rattus norvegicus* | 7.45E-011 | 48% |
| | AAI07190.1 | CD209e antigen | *Mus musculus* | 9.73E-011 | 46% |
| | NP_507557.1 | C-type Lectin family member (clec-42) | *Caenorhabditis elegans* | 1.27E-010 | 46% |
| comp292689_c0_seq1 | XP_221790.4 | CD209 molecule-like | *Rattus norvegicus* | 4.95E-008 | 62% |
| comp329897_c1_seq1 | NP_001186302.1 | CD209 molecule | *Danio rerio* | 2.96E-008 | 53% |
| | NP_002429.1 | macrophage mannose receptor 1 | *Homo sapiens* | 6.60E-008 | 45% |
| comp335421_c2_seq1 | CAZ65474.1 | C-type Lectin | *Caenorhabditis elegans* | 1.58E-008 | 36% |
| | NP_001179599.1 | C-type mannose receptor 2 | *Bos taurus* | 6.02E-008 | 42% |
| comp335421_c2_seq2 | NP_001179599.1 | C-type mannose receptor 2 | *Bos taurus* | 6.16E-008 | 42% |
| | NP_006030.2 | C-type mannose receptor 2 | *Homo sapiens* | 1.37E-007 | 41% |
| comp335421_c2_seq6 | NP_001179599.1 | C-type mannose receptor 2 | *Bos taurus* | 4.52E-008 | 42% |
| | NP_006030.2 | C-type mannose receptor 2 | *Homo sapiens* | 1.01E-007 | 41% |
| comp335598_c3_seq2 | ADD13530.1 | IML1 immunolectin 1 | *Manduca sexta* | 9.14E-010 | 50% |
| | AAC33576.1 | immunolectin-A precursor | *Manduca sexta* | 9.14E-010 | 50% |
| comp336008_c0_seq13 | ABW34402.1 | immunity adhesion receptor L-SIGN | *Nomascus leucogenys* | 7.97E-014 | 52% |
| | AAR04559.1 | L-SIGN variant | *Homo sapiens* | 1.15E-012 | 50% |
| comp336008_c0_seq15 | NP_001178934.1 | CD209 antigen-like protein E | *Rattus norvegicus* | 2.50E-012 | 52% |
| | ABW34402.1 | immunity adhesion receptor L-SIGN | *Nomascus leucogenys* | 2.11E-011 | 51% |
| | AAR04559.1 | L-SIGN variant | *Homo sapiens* | 4.71E-011 | 50% |
| comp336008_c0_seq6 | NP_001178934.1 | CD209 antigen-like protein E | *Rattus norvegicus* | 1.18E-012 | 50% |
| | ABW34402.1 | immunity adhesion receptor L-SIGN | *Nomascus leucogenys* | 1.70E-011 | 51% |
| | AAR04559.1 | L-SIGN variant | *Homo sapiens* | 3.79E-011 | 50% |

## Supplementary Table 11: Host SRCR proteins with similarity to SRCR proteins involved in immune processes

| NCBI accession | transcript | Blast Hit Description (HSP) | organism | accession | e-value | aa similarity |
|---|---|---|---|---|---|---|
| comp306101_c0_seq1 | OalgSRCR2 | macrophage receptor marco | *Mus musculus* | 2OY3 | 7.08E-15 | 68% |
| comp309069_c0_seq1 | OalgSRCR3 | DMBT1-like protein | *Strongylocentrotus purpuratus* | XP_001181916.1 | 4.99E-29 | 63% |
| comp319455_c1_seq1 | OalgSRCR4 | M160/CD163 | *Homo sapiens* | Q9NR16.2 | 1.32E-13 | 72% |
| comp322367_c1_seq1 | OalgSRCR6 | DMBT1-like protein | *Strongylocentrotus purpuratus* | XP_001183040.1 | 1.31E-21 | 68% |
| comp324195_c1_seq2 | OalgSRCR7 | DMBT1-like protein | *Strongylocentrotus purpuratus* | XP_001181916.1 | 7.75E-08 | 66% |
| comp327181_c0_seq1 | OalgSRCR9 | scavenger receptor cysteine-rich type 1 protein M160 | *Callithrix jacchus* | XP_002752336.1 | 3.64E-11 | 77% |
| comp503880_c0_seq1 | OalgSRCR10 | scavenger receptor cysteine-rich protein type 12 | *Strongylocentrotus purpuratus* | XP_001202087.1 | 2.89E-08 | 62% |
| comp740883_c0_seq1 | OalgSRCR11 | SRCR domain, membrane form 2 | *Strongylocentrotus purpuratus* | XP_796328.1 | 1.08E-07 | 73% |
| comp329845_c9_seq3 | OalgSRCR12 | SRCR domain, membrane form 2 | *Geodia cydonium* | CAA75175.1 | 2.66E-22 | 54% |
| comp330058_c4_seq1 | OalgSRCR13 | macrophage receptor marco | *Mus musculus* | 2OY3 | 4.69E-27 | 69% |
| comp331265_c0_seq5 | OalgSRCR15 | scavenger receptor cysteine-rich protein | *Strongylocentrotus purpuratus* | XP_795482.2 | 2.49E-12 | 62% |
| comp332789_c1_seq1 | OalgSRCR17 | scavenger receptor cysteine-rich protein | *Strongylocentrotus purpuratus* | NP_999650.1 | 2.54E-43 | 62% |
| comp334332_c0_seq6 | OalgSRCR18 | Pema-SRCR protein precursor | *Petromyzon marinus* | AAA90990.1 | 1.88E-25 | 67% |
| comp334882_c0_seq1 | OalgSRCR19 | DMBT1-like protein | *Anolis carolinensis* | XP_003228360.1 | 5.22E-19 | 65% |
| comp334995_c0_seq2 | OalgSRCR20 | Pema-SRCR protein precursor | *Petromyzon marinus* | AAA90990.1 | 1.93E-12 | 73% |
| comp334998_c0_seq1 | OalgSRCR21 | M130/CD163 | *Mus musculus* | Q2VLH6.1 | 1.11E-12 | 60% |
| comp334998_c2_seq1 | OalgSRCR22 | scavenger receptor cysteine-rich protein | *Strongylocentrotus purpuratus* | XP_794349.2 | 1.42E-07 | 80% |
| comp333407_c3_seq19 | OalgSRCR25 | SRCR domain, membrane form 2 | *Strongylocentrotus purpuratus* | XP_796328.1 | 2.63E-17 | 63% |

## Supplementary Table 12: Host Toll-like receptor sequences

| transcript ID | Sequence Description | ncbi accession | common name | organism | BlastHit Description (HSP) | e-value | aa sim. |
|---|---|---|---|---|---|---|---|
| comp179841_c0_seq1 | variable lymphocyte receptor | BAF43120.1 | arctic lamprey | *Lethenteron japonicum* | variable lymphocyte receptor | 4.36E-12 | 65% |
| comp244649_c0_seq1 | variable lymphocyte receptor | ABO21284.1 | sea lamprey | *Petromyzon marinus* | variable lymphocyte receptor | 1.84E-07 | 64% |
| comp263075_c0_seq1 | variable lymphocyte receptor | ABO21194.1 | sea lamprey | *Petromyzon marinus* | variable lymphocyte receptor | 7.08E-07 | 62% |
| comp288229_c0_seq1 | toll-like receptor | ABK88278.1 | horseshoe crab | *Carcinoscorpius rotundicauda* | toll-like receptor | 5.70E-12 | 65% |
| comp308544_c0_seq1 | toll-like receptor | XP_551799.2 | mosquito | *Anopheles gambiae* | toll-like receptor | 1.36E-10 | 63% |
| comp321901_c0_seq1 | variable lymphocyte receptor | BAF43195.1 | arctic lamprey | *Lethenteron japonicum* | variable lymphocyte receptor | 2.31E-13 | 72% |
| comp321901_c1_seq1 | variable lymphocyte receptor | ABO85960.1 | sea lamprey | *Petromyzon marinus* | variable lymphocyte receptor | 5.25E-10 | 62% |
| comp324324_c1_seq1 | variable lymphocyte receptor | ABA40047.1 | sea lamprey | *Petromyzon marinus* | variable lymphocyte receptor | 5.43E-07 | 49% |
| comp325011_c2_seq1 | variable lymphocyte receptor | ABO21194.1 | sea lamprey | *Petromyzon marinus* | variable lymphocyte receptor | 3.42E-09 | 64% |
| comp329847_c2_seq10 | toll-like receptor | ABK88278.1 | horseshoe crab | *Carcinoscorpius rotundicauda* | toll-like receptor | 5.36E-15 | 42% |
| comp329847_c2_seq12 | leucine rich repeat containing protein 15 | NP_001128529.2 | human | *Homo sapiens* | leucine-rich repeat-cont. protein 15 | 1.43E-45 | 51% |
| comp329847_c2_seq14 | leucine rich repeat containing protein 15 | NP_659551.1 | rat | *Rattus norvegicus* | leucine-rich repeat-cont. protein 15 | 1.83E-45 | 50% |
| comp329847_c2_seq15 | toll-like receptor | ABK88278.1 | horseshoe crab | *Carcinoscorpius rotundicauda* | toll-like receptor | 9.77E-14 | 42% |
| comp329847_c2_seq16 | variable lymphocyte receptor | BAF43202.1 | arctic lamprey | *Lethenteron japonicum* | variable lymphocyte receptor | 1.93E-07 | 45% |
| comp329847_c2_seq17 | variable lymphocyte receptor | BAI66869.1 | inshore hagfish | *Eptatretus burgeri* | variable lymphocyte receptor A | 1.74E-41 | 63% |
| comp329847_c2_seq1 | toll-like receptor | XP_551799.2 | mosquito | *Anopheles gambiae* | toll-like receptor | 1.97E-09 | 69% |
| comp329847_c2_seq3 | toll-like receptor | ABK88278.1 | horseshoe crab | *Carcinoscorpius rotundicauda* | toll-like receptor | 2.18E-23 | 45% |
| comp329847_c2_seq4 | leucine rich repeat containing protein 15 | NP_659551.1 | rat | *Rattus norvegicus* | leucine-rich repeat-cont. protein 15 | 2.05E-45 | 50% |
| comp329847_c2_seq5 | toll-like receptor | ABK88278.1 | horseshoe crab | *Carcinoscorpius rotundicauda* | toll-like receptor | 2.66E-24 | 45% |
| comp329847_c2_seq7 | leucine rich repeat containing protein 15 | NP_659551.1 | rat | *Rattus norvegicus* | leucine-rich repeat-cont. protein 15 | 1.95E-45 | 50% |
| comp329847_c2_seq8 | toll-like receptor | ABK88278.1 | horseshoe crab | *Carcinoscorpius rotundicauda* | toll-like receptor | 3.87E-19 | 44% |
| comp329847_c2_seq9 | leucine rich repeat containing protein 15 | NP_659551.1 | rat | *Rattus norvegicus* | leucine-rich repeat-cont. protein 15 | 2.09E-45 | 50% |
| comp330118_c1_seq11 | variable lymphocyte receptor | XP_001846467.1 | mosquito | *Culex quinquefasciatus* | leucine-rich repeat-cont. protein 15 | 4.19E-07 | 63% |
| comp330118_c1_seq12 | toll protein | ABK58729.1 | Pacific white shrimp | *Litopenaeus vannamei* | toll protein | 2.19E-08 | 42% |
| **transcript ID** | **Sequence Description** | **ncbi accession** | **common name** | **organism** | **BlastHit Description (HSP)** | **e-value** | **aa** |

| | | | | | | | sim. |
|---|---|---|---|---|---|---|---|
| comp330118_c1_seq13 | toll protein | ABO38434.1 | black tiger shrimp | *Penaeus monodon* | toll receptor | 4.22E-08 | 44% |
| comp330118_c1_seq14 | toll protein | ABO38434.1 | black tiger shrimp | *Penaeus monodon* | toll receptor | 1.22E-07 | 46% |
| comp330118_c1_seq2 | toll protein | ABK58729.1 | Pacific white shrimp | *Litopenaeus vannamei* | toll protein | 2.21E-08 | 42% |
| comp330118_c1_seq5 | variable lymphocyte receptor | ABO15189.1 | sea lamprey | *Petromyzon marinus* | variable lymphocyte receptor | 8.48E-08 | 61% |
| comp330118_c1_seq7 | variable lymphocyte receptor | ABO15189.1 | sea lamprey | *Petromyzon marinus* | variable lymphocyte receptor | 6.34E-08 | 58% |
| comp330178_c1_seq1 | toll-like receptor | AAO53555.1 | goldfish | *Carassius auratus* | toll-like receptor | 7.09E-07 | 57% |
| comp335045_c1_seq6 | toll-like receptor | NP_001133860.1 | salmon | *Salmo salar* | toll-like receptor 13 | 9.38E-07 | 54% |



## Supplementary Table 13 (opposite page): Toll immune signaling pathway in *Olavius*

Further components of the toll pathway in *Olavius algarvensis* were identified by tblastx/blastx searches using previously described sequences from annelids (*Capitella teleta*, *Helobdella robusta*, given are trace archive accession numbers) [68] and model organisms (*Drosophila melanogaster*, *Homo sapiens*) as queries against the *Olavius algarvensis* transcriptome assembly; n.a., protein not identified in *C. teleta*/*H. robusta*; ---------- no query; aa, amino acid.

| | blast query sequence identifiers (annelids: trace archive sequence IDs) | | | respective best blast hits in *Olavius* transcriptome (% aa seq similarity/% seq identity) | | |
|---|---|---|---|---|---|---|
| **Protein** | ***Capitella teleta*** | ***Helobdella robusta*** | **uniprot accession** | ***Capitella teleta*** | ***Helobdella robusta*** | **uniprot** |
| ***Drosophila melanogaster*** | | | | | | |
| Cactus | 1109683694 | 1112642406 | ---------- | comp331734_c2 (60/47) | comp329028_c3 (67/44) | ---------- |
| Dorsal | 1027947919 | 1122215071 | ---------- | comp334686_c3 (82/69) | comp334686_c3 (73/53) | ---------- |
| ECSIT | 1085544040 | 1115381906 | ---------- | comp335923_c4 (73/50) | comp335923_c4 (81/66) | ---------- |
| Pelle | 1112699057 | n.a. | ---------- | comp289883_c1 (79/66) | ---------- | ---------- |
| Relish | n.a. | n.a. | Q94527 | ---------- | ---------- | comp330391_c0 (50/34) |
| Spaetzle | n.a. | n.a. | P48607 | ---------- | ---------- | no hit |
| TAK1 | n.a. | n.a. | Q9V3Q6 | ---------- | ---------- | comp329383_c9 (75/57) |
| TRAF1 | 1028363621 | 1112642913 | ---------- | comp334076_c3 (63/46) | comp333584_c10 (74/47) | ---------- |
| Tube | n.a. | n.a. | P22812 | ---------- | ---------- | no hit |
| **Homo sapiens (MyD88-dependent toll signaling)** | | | | | | |
| ECSIT | 1085544040 | 1115381906 | ---------- | comp335923_c4 (72/54) | comp335923_c4 (76/57) | ---------- |
| IKK β | n.a. | n.a. | O14920 | ---------- | ---------- | comp329208_c1 (50/33) |
| IκB ε | 1074119071 | n.a. | ---------- | comp331734_c2 (70/51) | ---------- | ---------- |
| IKK α | n.a. | n.a. | O15111 | ---------- | ---------- | comp307227_c0 (58/34) |
| IRAK1 | n.a. | n.a. | P51617 | ---------- | ---------- | comp334746_c11 (64/47) |
| MEKK | 1026033689 | n.a. | Q13233 | ---------- | ---------- | comp308092_c0 (54/32) |
| MyD88 | 1320293219 | 1343967404 | ---------- | no hit | no hit | ---------- |
| NEMO | n.a. | n.a. | Q9Y6K9 | ---------- | ---------- | no hit |
| NF-κB p105 | n.a. | n.a. | P19838 | ---------- | ---------- | comp330391_c0 (68/53) |
| NF-κB p65 | 1028298985 | n.a. | Q04206 | comp329272_c3 (81/67) | ---------- | comp329272_c3 (66/50) |
| TRAF5 | 1085417814 | 1140751855 | ---------- | comp333584_c10 (77/57) | comp333584_c10 (88/58) | ---------- |
| **Homo sapiens (MyD88-independent toll signaling)** | | | | | | |
| TICAM1 | n.a. | n.a. | Q8IUC6 | ---------- | ---------- | no hit |
| TICAM2 | n.a. | n.a. | Q86XR7 | ---------- | ---------- | no hit |
| TIRAP | n.a. | n.a. | P58753 | ---------- | ---------- | no hit |

### Supplementary Table 14: Subcellular localization evidence of host digestive proteins

Table is too large; refer to refer to supplementary file "Supplementary_Table14.xlsx"

## Supplementary Figure Legends

### Supplementary Figure 1: Transcriptome annotation statistics

Blast2GO annotation statistics of assembled transcripts; with interproscan: sequences with at least one hit to any of the following databases: ProDom [1], Prints [2], PIR [3], Pfam [4], Smart [5], TIGRFAM [6], PROSITE [7], HAMAP [69], SuperFamily [8], SignalP [9], TMHMM [10], Panther [11], Gene3D [12], Phobius [13], Coils [14]; with blast hits: sequences with at least one significant blastx hit to ncbi nr database (e-value cut-off: 1e-6), but no GO mapping or annotation; with mapping: sequences with GO mapping, but no annotation; with annotation: sequences with blast hits and GO mapping, automatically annotated according to blast2go annotation rules [70].

### Supplementary Figure 2: Hydrophobicity cluster analysis plots of annelid hemoglobin chains

Two-dimensional hydrophobicity cluster analysis (HCA) plots of selected annelid and *Olavius algarvensis* hemoglobin chains, generated with drawhca [71]. *Arenicola marina* (GenBank accession numbers: A2c, CAJ32741; B1, CAJ32742; B2, CAI56309), *Ridgeia piscesae* (GenBank accession numbers: A1, ABD72632; A2, ABD72633; B1a, ABD72634; B2, AAP04527), *Riftia pachyptila* (GenBank accession numbers: A1, ABW24412; A2, CAD29155; B1a, CAD29156; B2, CAD29159), *Lamellibrachia satsuma* (GenBank accession numbers: A1, BAN58230; A2, BAN58231; B1, BAN58232; B2, BAN58233), *Lamellibrachia sp*. XB-2003 (GenBank accession numbers: A1, AAP40327; A2, AAP04528; B1, AAP40328; B2, AAP04529), *Oasisia alvinae* (GenBank accession numbers: A2, AAP04531; B2, AAP40329), *Tevnia jerichonana* (GenBank accession number: A2, AAP04530), *Oligobrachia mashikoi* (GenBank accession numbers: A1, Q7M419; A2, Q7M413; B1, Q5KSB7; B2, Q7M418), *Sabella spallanzanii* (GenBank accession numbers: A2, CAC37412; B2a, CAC37410), *Tylorrhynchus heterochaetus* (GenBank accession numbers: A1, P02219; A2, P09966; B2a, P13578), *Lumbricus terrestris* (GenBank accession numbers: A1, P08924; A2, P02218; B1, P11069; B2, P13579), *Lumbricus rubellus* (GenBank accession numbers: A1a, DR009556; A2, BF422675; B1, CAA09958; B2, BF422540), *Olavius algarvensis*, sequences obtained in this study, accession numbers in figure.

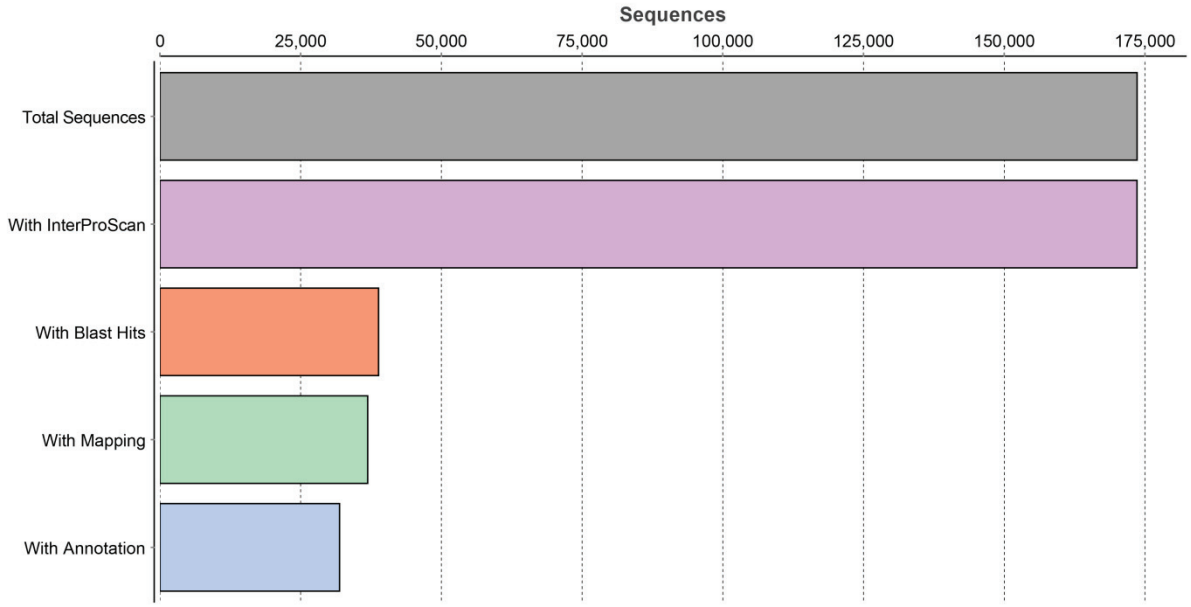
### Supplementary Figure 3: Domain structures of proteins with scavenger domains

Structure of conserved functional domains in *Olavius algarvensis* scavenger receptor cysteine-rich (SRCR) domain containing proteins; Question marks show uncertain sequential arrangement of domains, due to fragmented transcript assembly.

# Supplementary Figures

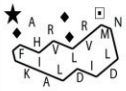
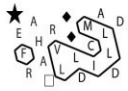
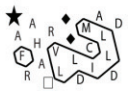
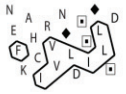
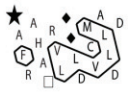
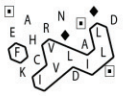
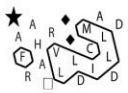
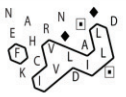
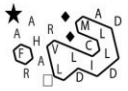
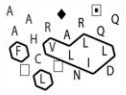
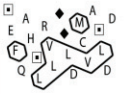
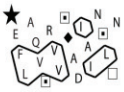
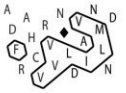
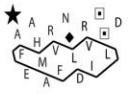
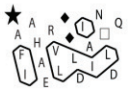
## Supplementary Figure 1: Transcriptome annotation statistics

Blast2GO annotation statistics of assembled transcripts; with interproscan: sequences with at least one hit to any of the following databases: ProDom [1], Prints [2], PIR [3], Pfam [4], Smart [5], TIGRFAM [6], PROSITE [7], HAMAP [69], SuperFamily [8], SignalP [9], TMHMM [10], Panther [11], Gene3D [12], Phobius [13], Coils [14]; with blast hits: sequences with at least one significant blastx hit to ncbi nr database (e-value cut-off: 1e-6), but no GO mapping or annotation; with mapping: sequences with GO mapping, but no annotation; with annotation: sequences with blast hits and GO mapping, automatically annotated according to blast2go annotation rules [70].



## Supplementary Figure 2: Hydrophobicity cluster analysis plots of annelid hemoglobin chains

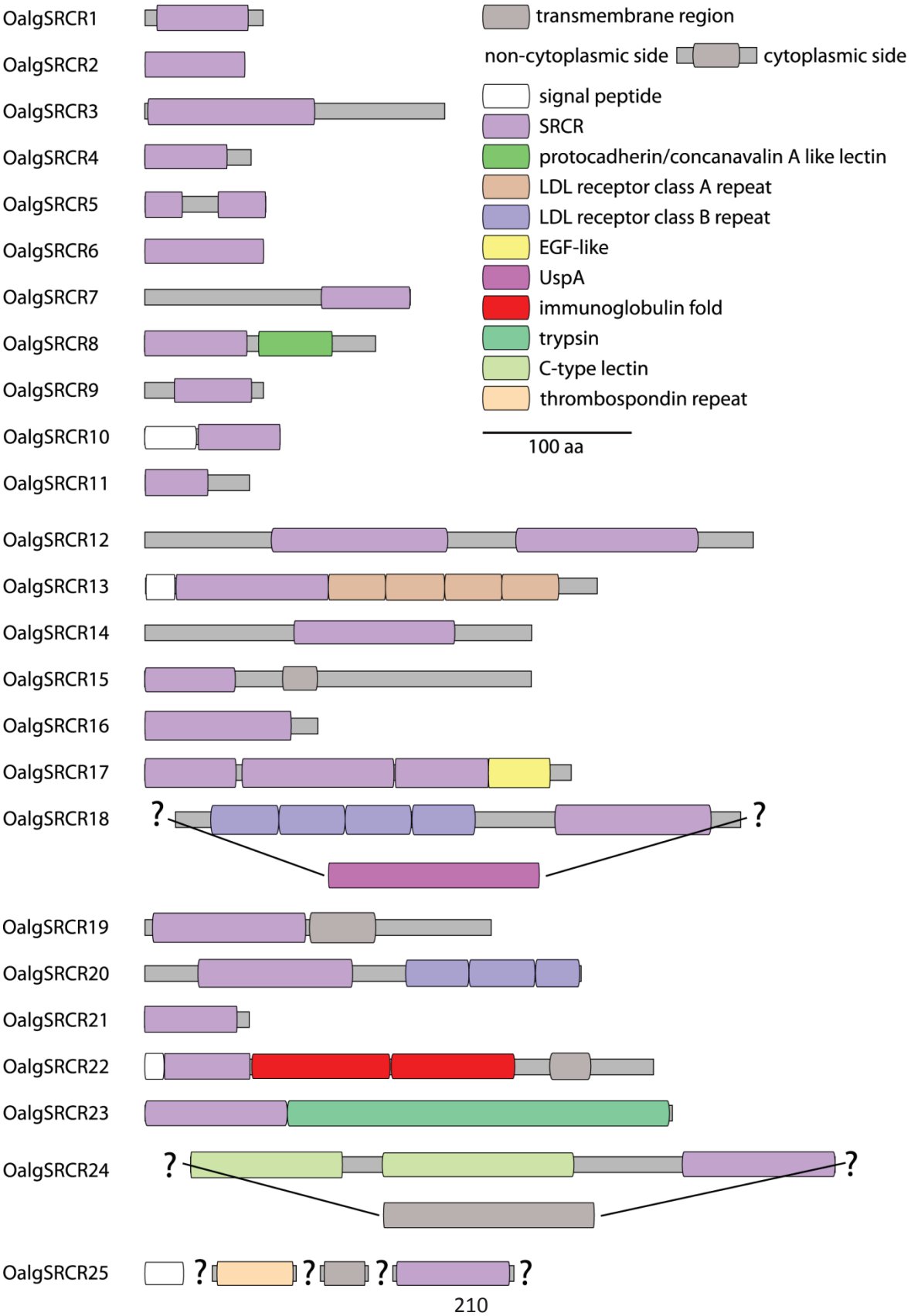
Two-dimensional hydrophobicity cluster analysis (HCA) plots of selected annelid and *Olavius algarvensis* hemoglobin chains, generated with drawhca [71]. *Arenicola marina* (GenBank accession numbers: A2c, CAJ32741; B1, CAJ32742; B2, CAI56309), *Ridgeia piscesae* (GenBank accession numbers: A1, ABD72632; A2, ABD72633; B1a, ABD72634; B2, AAP04527), *Riftia pachyptila* (GenBank accession numbers: A1, ABW24412; A2, CAD29155; B1a, CAD29156; B2, CAD29159), *Lamellibrachia satsuma* (GenBank accession numbers: A1, BAN58230; A2, BAN58231; B1, BAN58232; B2, BAN58233), *Lamellibrachia sp*. XB-2003 (GenBank accession numbers: A1, AAP40327; A2, AAP04528; B1, AAP40328; B2, AAP04529), *Oasisia alvinae* (GenBank accession numbers: A2, AAP04531; B2, AAP40329), *Tevnia jerichonana* (GenBank accession number: A2, AAP04530), *Oligobrachia mashikoi* (GenBank accession numbers: A1, Q7M419; A2, Q7M413; B1, Q5KSB7; B2, Q7M418), *Sabella spallanzanii* (GenBank accession numbers: A2, CAC37412; B2a, CAC37410), *Tylorrhynchus heterochaetus* (GenBank accession numbers: A1, P02219; A2, P09966; B2a, P13578), *Lumbricus terrestris* (GenBank accession numbers: A1, P08924; A2, P02218; B1, P11069; B2, P13579), *Lumbricus rubellus* (GenBank accession numbers: A1a, DR009556; A2, BF422675; B1, CAA09958; B2, BF422540), *Olavius algarvensis*, sequences obtained in this study, accession numbers in figure.

| | chain A1 | chain A2 | chain B1 | chain B2 |
|---|---|---|---|---|
| *Arenicola marina* | not available |  A2c |  |  |
| *Ridgeia piscesae* |  |  |  B1a |  |
| *Riftia pachyptila* |  |  |  B1a |  |
| *Lamellibrachia satsuma* |  |  |  |  |
| *Lamellibrachia sp. XB-2003* |  |  |  |  |
| *Oasisia alvinae* | not available |  | not available |  |
| *Tevnia jerichonana* | not available |  | not available | not available |
| *Oligobrachia mashikoi* |  |  |  |  |
| *Sabella spallanzanii* | not available |  | not available |  B2a |
| *Tylorrhynchus heterochaetus* |  |  | not available |  B2a |
| *Lumbricus terrestris* |  |  |  |  |
| *Lumbricus rubellus* |  A1a |  |  |  |
| *Olavius algarvensis* |  A<br>comp328129_c0_seq1<br> A<br>comp309954_c0_seq1 |  A2<br>comp287449_c0_seq1<br> ?<br>comp321285_c0_seq1 |  B<br>comp327543_c0_seq1<br> ?<br>comp300058_c1_seq1 |  B?<br>comp327754_c0_seq1<br> B<br>comp307848_c0_seq1 |

★ proline ♦ glycine □ threonine ⊡ serine

## Supplementary Figure 3: Domain structures of proteins with scavenger domains

Structure of conserved functional domains in *Olavius algarvensis* scavenger receptor cysteine-rich (SRCR) domain containing proteins; Question marks show uncertain sequential arrangement of domains, due to fragmented transcript assembly.



## Supplementary References

[1] Corpet F, Gouzy J, Kahn D. The ProDom database of protein domain families. Nucleic Acids Res. 1998;26:323–326.

[2] Attwood TK, Beck ME, Bleasby AJ, Parry-Smith DJ. PRINTS–a database of protein motif fingerprints. Nucleic Acids Res. 1994;22:3590–3596.

[3] Wu CH, Huang H, Arminski L, Castro-Alvear J, Chen Y, Hu ZZ, Ledley RS, Lewis KC, Mewes HW, Orcutt BC, Suzek BE, Tsugita A, Vinayaka CR, Yeh LSL, Zhang J, Barker WC. The Protein Information Resource: an integrated public resource of functional annotation of proteins. Nucleic Acids Res. 2002;30:35–37.

[4] Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer ELL, Bateman A. The Pfam protein families database. Nucleic Acids Res. 2008;36:D281–D288.

[5] Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P. SMART 4.0: towards genomic data integration. Nucleic Acids Res. 2004;32:D142–D144.

[6] Haft DH, Selengut JD, White O. The TIGRFAMs database of protein families. Nucleic Acids Res. 2003;31:371–373.

[7] Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M, Sigrist CJA. The PROSITE database. Nucleic Acids Res. 2006;34:D227–D230.

[8] Wilson D, Madera M, Vogel C, Chothia C, Gough J. The SUPERFAMILY database in 2007: families and functions. Nucleic Acids Res. 2007;35:D308–D313.

[9] Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8:785–786.

[10] Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305:567–580.

[11] Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, Rabkin S, Guo N, Muruganujan A, Doremieux O, Campbell MJ, Kitano H, Thomas PD. The PANTHER database of protein families, subfamilies, functions and pathways. Nucleic Acids Res. 2005;33:D284–D288.

[12] Pearl F, Todd A, Sillitoe I, Dibley M, Redfern O, Lewis T, Bennett C, Marsden R, Grant A, Lee D, Akpor A, Maibaum M, Harrison A, Dallman T, Reeves G, Diboun I, Addou S, Lise S, Johnston C, Sillero A, Thornton J, Orengo C. The CATH Domain Structure Database and related resources Gene3D and DHS provide comprehensive domain family information for genome analysis. Nucleic Acids Res. 2005;33:D247–D251.

[13] Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol. 2004;338:1027–1036.

[14] Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. Science. 1991;252:1162–1164.

[15] Kleiner M, Wentrup C, Lott C, Teeling H, Wetzel S, Young J, Chang YJ, Shah M, VerBerkmoes NC, Zarzycki J, Fuchs G, Markert S, Hempel K, Voigt B, Becher D, Liebeke M, Lalk M, Albrecht D, Hecker M, Schweder T, Dubilier, N. Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. Proc Nat Acad Sci U S A. 2012;109:E1173–E1182.

[16] Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO, Boffelli D, Anderson IJ, Barry KW, Shapiro HJ, Szeto E, Kyrpides NC, Mussmann M, Amann R, Bergin C, Ruehland C, Rubin EM, Dubilier, N. Symbiosis insights through metagenomic analysis of a microbial consortium. Nature. 2006;443:950–955.

[17] Pětra PH, Hermodson MA, Walsh KA, Neurath H. Characterization of bovine carboxypeptidase A (Allan). Biochemistry. 1971;10:4023–4025.

[18] Catasús L, Villegas V, Pascual R, Avilés FX, Wicker-Planquart C, Puigserver A. cDNA cloning and sequence analysis of human pancreatic procarboxypeptidase A1. Biochem J. 1992;287:299–303.

[19] Quinto C, Quiroga M, Swain WF, Nikovits W Jr, Standring DN, Pictet RL, Valenzuela P, Rutter WJ. Rat preprocarboxypeptidase A: cDNA sequence and preliminary characterization of the gene. Proc Natl Acad Sci U S A. 1982;79:31–35.

[20] Catasús L, Vendrell J, Avilés FX, Carreira S, Puigserver A, Billeter M. The sequence and conformation of human pancreatic procarboxypeptidase A2. cDNA cloning, sequence analysis, and three-dimensional model. J Biol Chem. 1995;270:6651–6657.

[21] Gardell SJ, Craik CS, Clauser E, Goldsmith EJ, Stewart CB, Graf M, Rutter WJ. A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and evolutionary implications on substrate specificity in the carboxypeptidase gene family. J Biol Chem. 1988;263:17828–17836.

[22] Edwards MJ, Moskalyk LA, Donelly-Doman M, Vlaskova M, Noriega FG, Walker VK, Jacobs-Lorena M. Characterization of a carboxypeptidase A gene from the mosquito, *Aedes aegypti*. Insect Mol Biol. 2000;9:33–38.

[23] Bown DP, Wilkinson HS, Gatehouse JA. Midgut carboxypeptidase from *Helicoverpa armigera* (Lepidoptera: Noctuidae) larvae: enzyme characterisation, cDNA cloning and expression. Insect Biochem Mol Biol. 1998;28:739–749.

[24] Jochim RC, Teixeira CR, Laughinghouse A, Mu J, Oliveira F, Gomes RB, Elnaiem DE, Valenzuela JG. The midgut transcriptome of *Lutzomyia longipalpis*: comparative analysis of cDNA libraries from sugar-fed, blood-fed, post-digested and *Leishmania infantum chagasi*-infected sand flies. BMC Genomics. 2008;9:15.

[25] Ribeiro JMC, Genta FA, Sorgine MHF, Logullo R, Mesquita RD, Paiva-Silva GO, Majerowicz D, Medeiros M, Koerich L, Terra WR, Ferreira C, Pimentel AC, Bisch PM, Leite DC, Diniz MMP, da S G V Junior JL, Da Silva ML, Araujo RN, Gandara ACP, Brosson S, Salmon D, Bousbata S, González-Caballero N, Silber AM, Alves-Bezerra M, Gondim KC, Silva-Neto MAC, Atella GC, Araujo H, Dias FA, Polycarpo C,

Vionette-Amaral RJ, Fampa P, Melo ACA, Tanaka AS, Balczun C, Oliveira JHM, Gonçalves RLS, Lazoski C, Rivera-Pomar R, Diambra L, Schaub GA, Garcia ES, Azambuja P, Braz GRC, Oliveira PL. An insight into the transcriptome of the digestive tract of the bloodsucking bug, *Rhodnius prolixus*. PLoS Negl Trop Dis. 2014;8:e2594.

[26] Prabhakar S, Chen MS, Elpidina EN, Vinokurov KS, Smith CM, Marshall J, Oppert B. Sequence analysis and molecular characterization of larval midgut cDNA transcripts encoding peptidases from the yellow mealworm, *Tenebrio molitor* L. Insect Mol Biol. 2007;16:455–468.

[27] Wang P, Li G, Kain W. Characterization and cDNA cloning of midgut carboxypeptidases from *Trichoplusia ni*. Insect Biochem Mol Biol. 2004;34:831–843.

[28] Lee MS, Cho SJ, Tak ES, Lee JA, Cho HJ, Park BJ, Shin C, Kim DK, Park SC. Transcriptome analysis in the midgut of the earthworm (*Eisenia andrei*) using expressed sequence tags. Biochem Biophys Res Commun. 2005;328:1196–1204.

[29] Hartley BS. Amino-acid sequence of the bovine chymotrypsinogen-A. Nature. 1964;201:1284–1287.

[30] Asgeirsson B, Bjarnason JB. Structural and kinetic properties of chymotrypsin from Atlantic cod (*Gadus morhua*). Comparison with bovine chymotrypsin. Comp Biochem Physiol B. 1991;99:327–335.

[31] Pinsky SD, LaForge KS, Luc V, Scheele G. Identification of cDNA clones encoding secretory isoenzyme forms: sequence determination of canine pancreatic prechymotrypsinogen 2 mRNA. Proc Natl Acad Sci U S A. 1983;80:7486–7490.

[32] Spilliaert R, Gudmundsdóttir A. Molecular cloning of the Atlantic cod chymotrypsinogen B. Microb Comp Genomics. 2000;5:41–50.

[33] Tomita N, Izumoto Y, Horii A, Doi S, Yokouchi H, Ogawa M, Mori T, Matsubara K. Molecular cloning and nucleotide sequence of human pancreatic prechymotrypsinogen cDNA. Biochem Biophys Res Commun. 1989;158:569–575.

[34] Bell GI, Quinto C, Quiroga M, Valenzuela P, Craik CS, Rutter WJ. Isolation and sequence of a rat chymotrypsin B gene. J Biol Chem. 1984;259:14265–14270.

[35] Mazumdar-Leighton S, Broadway RM. Identification of six chymotrypsin cDNAs from larval midguts of *Helicoverpa zea* and *Agrotis ipsilon* feeding on the soybean (Kunitz) trypsin inhibitor. Insect Biochem Mol Biol. 2001;31:633–644.

[36] Degnan BM, Groppe JC, Morse DE. Chymotrypsin mRNA expression in digestive gland amoebocytes: cell specification occurs prior to metamorphosis and gut morphogenesis in the gastropod, *Haliotis rufescens*. Rouxs Arch Dev Biol. 1995;205:97–101.

[37] Chougule NP, Giri AP, Sainani MN, Gupta VS. Gene expression patterns of *Helicoverpa armigera* gut proteases. Insect Biochem Mol Biol. 2005;35:355–367.

[38] Zhu YC, Guo Z, Abel C. Cloning eleven midgut trypsin cDNAs and evaluating the interaction of proteinase inhibitors with Cry1Ac against the tobacco budworm, *Heliothis virescens* (F.) (Lepidoptera: Noctuidae). J Invertebr Pathol. 2012;111:111–120.

[39] Zhu YC, Liu X, Maddur AA, Oppert B, Chen MS. Cloning and characterization of chymotrypsin- and trypsin-like cDNAs from the gut of the Hessian fly [*Mayetiola destructor* (Say)]. Insect Biochem Mol Biol. 2005;35:23–32.

[40] Balti R, Bougherra F, Bougatef A, Hayet BK, Nedjar-Arroume N, Dhulster P, Guillochon D, Nasri M. Chymotrypsin from the hepatopancreas of cuttlefish (*Sepia officinalis*) with high activity in the hydrolysis of long chain peptide substrates: Purification and biochemical characterisation. Food Chem. 2012;130:475–484.

[41] Elpidina EN, Tsybina TA, Dunaevsky YE, Belozersky MA, Zhuzhikov DP, Oppert B. A chymotrypsin-like proteinase from the midgut of *Tenebrio molitor* larvae. Biochimie. 2005;87:771–779.

[42] Reseland JE, Larsen F, Solheim J, Eriksen JA, Hanssen LE, Prydz H. A novel human chymotrypsin-like digestive enzyme. J Biol Chem. 1997;272:8099–8104.

[43] Largman C, Brodrick JW, Geokas MC. Purification and characterization of two human pancreatic elastases. Biochemistry. 1976;15:2491–2500.

[44] Largman C. Isolation and characterization of rat pancreatic elastase. Biochemistry. 1983;22:3763–3770.

[45] Gertler A, Birk Y. Isolation and characterization of porcine proelastase. Eur J Biochem. 1970;12:170–176.

[46] Qiu R, Liu X, Hu YH, Sun BG. Expression characterization and activity analysis of a cathepsin B from Pacific abalone *Haliotis discus hannai*. Fish Shellfish Immunol. 2013;34:1376–1382.

[47] Stephens A, Rojo L, Araujo-Bernal S, Garcia-Carreño F, Muhlia-Almazan A. Cathepsin B from the white shrimp *Litopenaeus vannamei*: cDNA sequence analysis, tissues-specific expression and biological activity. Comp Biochem Physiol B Biochem Mol Biol. 2012;161:32–40.

[48] Wang X, Liu B, Wang G, Tang B, Xiang J. Molecular cloning and functional analysis of cathepsin B in nutrient metabolism during larval development in *Meretrix meretrix*. Aquaculture. 2008;282:41–46.

[49] Ranjit N, Zhan B, Stenzel DJ, Mulvenna J, Fujiwara R, Hotez PJ, Loukas A. A family of cathepsin B cysteine proteases expressed in the gut of the human hookworm, *Necator americanus*. Mol Biochem Parasitol. 2008;160:90–99.

[50] Pinlaor P, Kaewpitoon N, Laha T, Sripa B, Kaewkes S, Morales ME, Mann VH, Parriott SK, Suttiprapa S, Robinson MW, To J, Dalton JP, Loukas A, Brindley PJ. Cathepsin F cysteine protease of the human liver fluke, *Opisthorchis viverrini*. PLoS Negl Trop Dis. 2009;3:e398.

[51] Aoki H, Ahsan MN, Watabe S. Molecular cloning and characterization of cathepsin B from the hepatopancreas of northern shrimp *Pandalus borealis*. Comp Biochem Physiol B Biochem Mol Biol. 2003;134:681–694.

[52] Fuzita FJ, Pinkse MWH, Patane JSL, Juliano MA, Verhaert PDEM, Lopes AR. Biochemical, transcriptomic and proteomic analyses of digestion in the scorpion *Tityus serrulatus*: insights into function and evolution of digestion in an ancient arthropod. PLoS One. 2015;10:e0123841.

[53] Kang JM, Bahk YY, Cho PY, Hong SJ, Kim TS, Sohn WM, Na BK. A family of cathepsin F cysteine proteases of *Clonorchis sinensis* is the major secreted proteins that are expressed in the intestine of the parasite. Mol Biochem Parasitol. 2010;170:7–16.

[54] Wang B, Shi GP, Yao PM, Li Z, Chapman HA, Brömme D. Human cathepsin F. Molecular cloning, functional expression, tissue localization, and enzymatic characterization. J Biol Chem. 1998;273:32000–32008.

[55] Deussing J, Tisljar K, Papazoglou A, Peters C. Mouse cathepsin F: cDNA cloning, genomic organization and chromosomal assignment of the gene. Gene. 2000;251:165–173.

[56] Ahn SJ, Kim NY, Seo JS, Je JE, Sung JH, Lee SH, Kim MS, Kim JK, Chung JK, Lee HH. Molecular cloning, mRNA expression and enzymatic characterization of cathepsin F from olive flounder (*Paralichthys olivaceus*). Comp Biochem Physiol B Biochem Mol Biol. 2009;154:211–220.

[57] Hu X, Hu X, Hu B, Wen C, Xie Y, Wu D, Tao Z, Li A, Gao Q. Molecular cloning and characterization of cathepsin L from freshwater mussel, *Cristaria plicata*. Fish Shellfish Immunol. 2014;40:446–454.

[58] Matsumoto I, Watanabe H, Abe K, Arai S, Emori Y. A putative digestive cysteine proteinase from *Drosophila melanogaster* is predominantly expressed in the embryonic and larval midgut. Eur J Biochem. 1995;227:582–587.

[59] Li WW, Jin XK, He L, Jiang H, Gong YN, Xie YN, Wang Q. Molecular cloning, characterization, expression and activity analysis of cathepsin L in Chinese mitten crab, *Eriocheir sinensis*. Fish Shellfish Immunol. 2010;29:1010–1018.

[60] Gal S, Gottesman MM. Isolation and sequence of a cDNA for human pro-(cathepsin L). Biochem J. 1988;253:303–306.

[61] Hu KJ, Leung PC. Food digestion by cathepsin L and digestion-related rapid cell differentiation in shrimp hepatopancreas. Comp Biochem Physiol B Biochem Mol Biol. 2007;146:69–80.

[62] Portnoy DA, Erickson AH, Kochan J, Ravetch JV, Unkeless JC. Cloning and characterization of a mouse cysteine proteinase. J Biol Chem. 1986;261:14697–14703.

[63] Ma J, Zhang D, Jiang J, Cui S, Pu H, Jiang S. Molecular characterization and expression analysis of cathepsin L1 cysteine protease from pearl oyster *Pinctada fucata*. Fish Shellfish Immunol. 2010;29:501–507.

[64] Ishidoh K, Towatari T, Imajoh S, Kawasaki H, Kominami E, Katunuma N, Suzuki K. Molecular cloning and sequencing of cDNA for rat cathepsin L. FEBS Lett. 1987;223:69–73.

[65] Cristofoletti PT, Ribeiro AF, Terra WR. The cathepsin L-like proteinases from the midgut of *Tenebrio molitor* larvae: sequence, properties, immunocytochemical localization and function. Insect Biochem Mol Biol. 2005;35:883–901.

[66] Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature Biotechnol. 2011;29:644.

[67] Sanchez S, Hourdez S, Lallier FH. Identification of proteins involved in the functioning of *Riftia pachyptila* symbiosis by Subtractive Suppression Hybridization. BMC Genomics. 2007;8:1.

[68] Davidson CR, Best NM, Francis JW, Cooper EL, Wood TC. Toll-like receptor genes (TLRs) from *Capitella capitata* and *Helobdella robusta* (Annelida). Dev Comp Immunol. 2008;32:608–612.

[69] Pedruzzi I, Rivoire C, Auchincloss AH, Coudert E, Keller G, de Castro E, Baratin D, Cuche BA, Bougueleret L, Poux S, Redaschi N, Xenarios I, Bridge A. HAMAP in 2015: updates to the protein family classification and annotation system. Nucleic Acids Res. 2015;43:D1064–D1070.

[70] Conesa A, Götz S, Garcá-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. .2005;21:3674–3676

[71] Callebaut I, Labesse G, Durand P, Poupon A, Canard L, Chomilier J, Henrissat B, Mornon J. Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. Cell Mol Life Sci. 1997;53:621–645.

# Chapter 5: General discussion and perspectives

This thesis contributes to current research in the fields of animal symbiosis, evolution, ecology, and immunology in several ways. Using the model gutless oligochaete species *Olavius algarvensis*, I was able to i) provide new insight into the recent evolutionary relationships between host and symbionts on a population level scale, ii) show that, although the symbionts are obligatory for and highly integrated with the host, and with each other, their transmission fidelity and host specificity can range from very high to low, iii) show that the diversity and flexibility of mutualistic symbionts within the same host species is much higher than previously anticipated, iv) provide the first genomic and functional insights into the physiology of the spirochaetal symbiont and its role within the symbiosis, v) provide the first analysis of the host's immune system and the molecular mechanisms that allow the continued existence of the symbiosis, and vi) discovered several ways of how the host has physiologically adapted to its symbiotic lifestyle.

The major findings of this thesis and their specific discussions are presented as individual manuscripts in chapters 2 (evolution, population genetics, transmission and diversity), 3 (spirochaete symbiont genomics and function), and 4 (host physiological adaptations and immunology). In the following, I will discuss questions that were left open for further exploration and research and will provide some preliminary data and suggestions of how to tackle these questions in the future.

## 5.1 Do *O. algarvensis* symbionts promote host evolution and diversification?

One of the most interesting current questions in evolution and symbiosis research is to what extent mutualistic microbial symbionts are able to contribute to host diversification, and

ultimately, host speciation. In chapter 2, I was able to show that, likely through strict maternal vertical transmission, *OalgG1-A* and *OalgG1-B* were sent on divergent evolutionary paths. Not only did they show genome sequence divergence on the SNP level, but also a divergence in the gene content of their genomes.

While the host showed a clear separation into different haplotypes on the mitochondrial genome level, it is at this point unclear if gene flow between these haplotypes is reduced or prevented, or not. Since the worms cannot be cultivated or brought to mating and egg laying in a controlled manner, judging whether or not reproductive barriers exist, and how permeable they are for occasional hydbrids is impossible without employing culture-independent sequencing methods.

Diversification within the nuclear genome of the host that reflects mitochondrial haplotype diversification could be strong evidence that the haplotypes are not interbreeding, whatever the actual isolation mechanisms may be. With the sequence data made available during this thesis, tracing host divergence on the nuclear gene level was not possible due to i) the lack of a host reference genome, which complicates transcriptomics based approaches even further (e.g. messy *de novo* assemblies, with many fragments, and alternatively spliced transcripts that cannot be resolved), ii) the lack of reference transcriptomes from single worms. Sequence data from single specimens would i) ensure that all sequences are really derived from the species *O. algarvensis,* and don't contain contaminating sequences from co-occuring species, and ii) would allow discerning individual sequence variation from population sequence variation. Towards the very end of my PhD I obtained twelve metagenomes and metatranscriptomes from single *O. algarvensis* worms, six of each haplotype from Sant' Andrea (A and B, the metagenomes are already used in the research of chapter 2). These new transcriptomes would be ideal material to construct a high quality database of host marker sequences (in this case: from CDS (coding sequences) and transcribed microsatellites) for phylogenetic SNP analyses, which would allow the investigation of hundreds of loci at the same time and that could uncover

possible divergent sequence evolution in the nuclear genome of the host [353]. However, care must be taken to avoid various pitfalls and biases in selecting these sequences [354, 355, 356, 357]. Because sequences were obtained from individual specimens, genetic distances between individuals could be calculated and clustered without *a priory* assumption about population structure, similar to the SNP analysis of mitochondrial and symbiont genomes presented in this thesis. Unfortunatley this type of analysis is mostly restricted to coding sequences, which might be under positive or purifying selection pressures and might not evolve neutrally. Additional fixed and frozen *O. algarvensis* specimens are available for sequencing of further specimens and ideally, the analysis would also include *O. algarvensis* specimens from near-by locations in order to avoid false assumptions about population boundaries. However, since the occurrence of *O. algarvensis* is very patchy, and only two confirmed collection sites are known, this might prove to be difficult to achieve in reality. However, the definition of suitable markers from transcriptomic data would already allow the design of PCR-based SNP genotyping assays and the assessment of local population structure and divergence on the nuclear genome level using hundreds of individual worms.

**5.2 Why do several bacterial clades seem prone to form symbioses with gutless oligochaetes?**

Gutless oligochaetes are a very species-rich group of animals, and each species of gutless oligochaete possesses their own specific set of symbionts that is not shared with any other species. However, in most cases, the symbionts do have very close relatives within the same clade that also form symbiotic relationships with other gutless oligochaete species. Examples of such clades include “Gamma1” (i.e. *Candidatus* Thiosymbion), “Gamma3”, “Delta1”, “Alpha1”, etc (see chapter 1, Figure 11, p. 51 and Figure 12, p. 53).

Host species that share symbionts from a certain clade are often not directly related, meaning that these symbionts were most likely not passed down from a common host ancestor, but

acquired anew, either from the environment, or from a co-occuring, but unrelated host. While the Gamma1/*Candidatus* Thiosymbion clade exclusively contains symbionts of either nematodes (Stilbonematinae/Astomonematinae) or gutless oligochaetes, and is relatively far diverged from its last common ancestor with non-host associated species, most other gutless oligochaete symbiont clades appear to be much less derived, and often contain free-living species among symbiont species. This also points to independent origins of gutless oligochaete symbioses with different members of the same bacterial clade. This also means that, from the vast diversity of free-living bacteria that could potentially form mutualistic symbioses with the gutless oligochaetes, only a few select clades actually do, and they did it repeatedly over the course of evolutionary time.

Great examples of this phenomenom are the model species *Olavius algarvensis* and the co-occuring species *Olavius ilvae* from Elba: these two host species share symbionts of the Gamma1, Gamma3, Delta1 and Delta3 clades; of course, each with its own host-exclusive version of each symbiont (see chapter 1, Figure 11, p. 51 and Figure 12, p. 53). The two hosts do not share a direct common ancestor, and neither do their Gamma1, Gamma3, Delta1, and Delta3 symbionts. As suggested by their 16S rRNA gene phylogenies, these symbionts were picked up by the two hosts independently, either from an unrelated host in the case of the Gamma1 symbiont, or from the environment in the case of Gamma3 and Delta1 (chapter 1, Figure 12, p. 53).

This raises the question of why there seem to be certain clades of bacteria that are more prone to form beneficial associations with gutless oligochaetes than others. Since the types of symbionts associated with a particular host species do not necessarily reflect host phylogeny, it seems to be due to factors other than the host genotype. This is expecially interesting since most proteobacterial mutualisms are claimed to have evolved from parasitic ancesters [358], which is clearly not the case in the gutless oligochaete symbionts, which are all derived from free-living environmental bacteria.

One reason for this could simply be that the clades of free-living bacteria, which also contain gutless oligochaete symbionts, are generally very widespread, dominant marine sediment bacteria. In other words, they are more prone to form symbiotic associations, because they are readily available in the habitat oft he worms. This hypothesis is supported by 16S clone libraries that were constructed using DNA extracted from Elba sediment, collected from the same habitat where the worms occur. These clone libraries yielded many sequences from typical gutless oligochaete clades (this chapter, Figure 1). While none of the recovered sequences were identical to symbiont phylotypes, many were very closely related and part of the same bacterial clade (with the exception of the Gamma1 symbiont, for which the existence of a free-living counterpart remains to conclusively demonstrated, [359, 130]).

However, there are other reasonably abundant clades of free-living sulfate-reducers in the sediment as well that could theoretically take the place of the established symbionts (see Figure 1). The same 16S rRNA clone libraries even contained another group of Gammaproteobacteria that forms symbiotic associations with the clam *Thyasira*. Yet, the secondary SOX symbionts of both *O. algarvensis* and *O. ilvae* are derived from the BD7-8 marine group instead.

Comparing the genetic makeup of gutless oligochaete symbionts and their closely-related free-living counterparts could help to better understand what makes certain bacterial clades more likely to form associations with these animals. The fact that gutless oligochaetes form associations with such a large diversity of bacterial clades, provides the unique opportunity to study the differences (and similarities) of symbionts and their close free-living relatives in several bacterial groups that are not directly related, but associated with the same host group. If general patterns or rules for sucessful symbiosis establishment do exist, they should become apparent by studying this host group.

Fairly new technical advances, like single-cell sorting, next generation metagenomic sequencing and improved assembly and binning algorithms make such comparisons feasible. Single cells can be obtained from sediment samples, amplified by MDA (multi displacement amplification) and sequenced to produce draft genomes of closely related, free-living relatives of gutless oligochaete symbionts. This approach has already been successfully employed to obtain draft genomes of marine Sva0081-clade bacteria from the North Sea and East Australia (Marc Mußmann, personal communication). The Sva0081-clade includes all gutless oligochaete Delta1 symbionts and the 16S sequences of these organisms are 96-99% identical to the 16S sequences of the gutless oligochaete Delta1 symbionts.

I propose that a similar approach should be employed to obtain draft genomes of other bacterial clades that contain gutless oligochaete symbionts. During my PhD I was able to obtain sediment samples for single cell sorting from various sites that harbor gutless oligochaetes, including Elba, Bermuda, Egypt and Hawaii, that could be used for this approach. Metagenomic sequencing data of a multitude of gutless oligochaete species from all over the world has recently been obtained through Illumina sequencing. From these, draft genomes of each symbiont could be assembled and binned with similar methods as employed in this thesis. Taken together, these samples and data not only could be used to search for genetic patterns that distinguish symbionts from their free-living relatives, or to find clues to help explain why these clades are prone to form symbioses with gutless oligochaetes, but also give new functional insights into the genomes of important free-living, but so far uncultured and unsequenced bacteria.

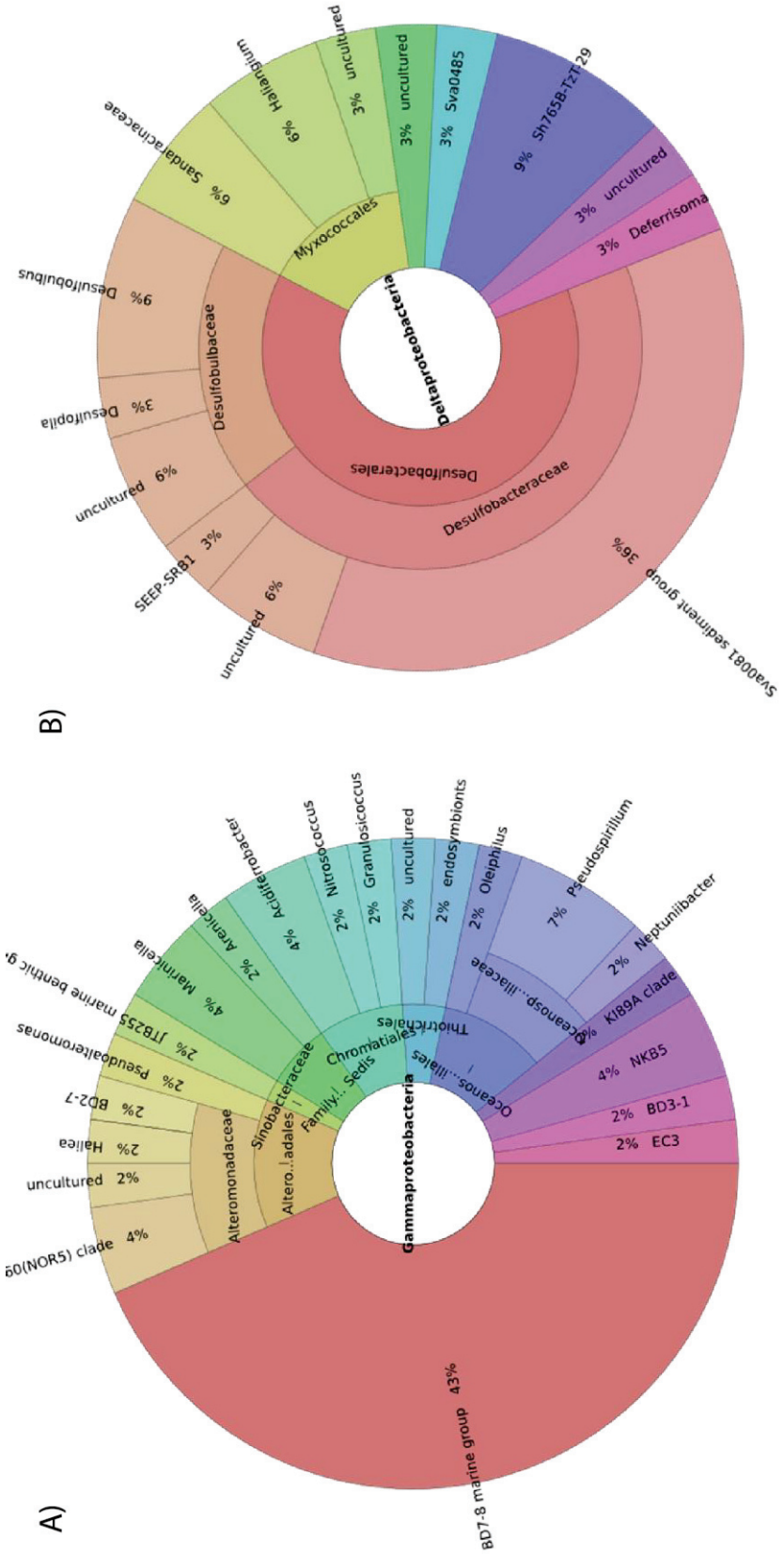



**Figure 2: Proportion of symbiont-related 16S rRNA gene sequences retrieved from clone libraries of DNA extracted from Elba sediment.** **A)** *Gammaproteobacteria*. The most abundant group of Gammaproteobacteria (43% of all Gammaproteobacteria within the clone library), BD7-8, is the clade that contains the Gamma3 symbionts. The group labeled as "endosymbionts" (2%) contains SOX symbionts of the clam *Thyasira*. **B)** *Deltaproteobacteria*. Clades containing gutless oligochaete symbionts are: Sva0081 sediment group (36%), which contains the Delta1 symbionts,

**5.3 Does the spirochaetal symbiont contribute an important function to the symbiosis?**

This thesis provides the first insight into the possible functions of the spirochaete symbiont within the *O. algarvensis* symbiosis, based on its draft genome. However, without information on gene expression, it is difficult to infer which metabolic pathways and other functions are actually active and play an important role in the symbiosis, since the genome itself can only give clues about the functional potential of the organism.

Since this symbiont, like all other gutless oligochaete symbionts, remains to be successfully cultured, one of the simplest ways right now to obtain gene expression information is via proteomics and transcriptomics. Using worms fixed immediately after sampling, both methods would allow investigating which genes are active under *in situ* conditions, while with short-term incubations, one could specifically investigate global gene expression patterns under different controlled laboratory conditions.

Preliminary analysis, using the gene expression data generated in this thesis, seems to support the hypothesis that the spirochaete symbiont is using external sugar sources, such as sucrose, from pore water, since the pathways for sucrose degradation, and fermentation to acetate appear to be more expressed than others (data not shown). Furthermore, several predicted sugar ABC transporters are among the top – ranking genes that have a functional annotation (data not shown). However, this data requires further analysis in order draw any robust conclusions. Preferably, one would set up suitable incubation experiments with whole live worms that are designed with the goal to track particular pathways (e.g. applying sucrose or inositol externally should result in a significant increase in expression levels of the required transporters and pathways).

### 5.4 How do physiological adaptations of the host connect to actual ecology?

The in-depth analysis of *O. algarvensis* proteomes revealed the expression of some unusual and unexpected proteins, including hemoglobin which, based on conserved sequence-patterns, should be able to bind sulfide, and the extremely abundant expression of hemerythrin, a respiratory protein insensitive to carbon monoxide. Both proteins appear to be adaptations to living in symbiosis with sulfide-producing and carbon monoxide oxidizing symbionts, as well as adaptations to experiencing extended periods of anoxia. In order to confirm the binding properties of these proteins, and to relate these findings to the ecology of the worms, further study of the actual proteins beyond their amino acid sequences is necessary. For example, the actual sulfide-binding and oxygen-binding properties of the *O. algarvensis* hemoglobin should be determined experimentally, in order to understand their affinities and binding behavior under varying sulfide or oxygen partial pressures, and in order to determine to what degree this hemoglobin can mitigate negative effects of endogenous sulfide. In the case of hemerythrin, its oxygen binding affinity and binding capacity need to be determined in order to estimate how much oxygen can be stored, and how much aerobic respiration could be supported, while the worms stay in the anoxic layers of the sediment. Expression localization of these proteins, for example using whole mount or thin-section *in situ* mRNA hybridizations should further help to understand the specific role that they play in the physiology and ecology of the host.

### 5.5 How does the immune system interact with each symbiont?

This thesis has established a large list of genes that are part of the host's innate immune system and therefore likely to be involved in molecular host-symbiont interactions in some way. While general putative functions can be inferred from the sequences of these genes, their specific function within the symbiosis remains to be determined for all of them. Some of the proteins

identified in chapter 4 are bound to play crucial roles in establishing and maintaining the symbiosis, and should be further investigated and characterized. However, the host expressed so many immune related genes, that it is impossible to characterize them all within a reasonable timeframe, especially with neither the host, nor the symbionts so far cultivated in the laboratory.

Nonetheless, several experiments are possible that would advance research in this area. For example, different life-stages of the host (mature worms, freshly layn eggs, different egg stages and hatched juveniles) could be screened for genes of the immune system that are only active or particularly active during certain phases of the host life cycle. The differential expression of such genes would indicate that they are especially important during a particular life stage of the host, and would give clues to which proteins are e.g. involved in the intitial recognition of the symbionts and the establishment of the symbiosis in the developing egg.

Another approach would be to localize proteins of particular interest to certain regions within the host (again, a possibility would be using *in situ* mRNA hybridizations). Genes that are likely involved in the interaction with symbionts should be mostly expressed in the epidermal tissue, since it is the only tissue in direct contact with the symbionts.

Futher experiments, like incubations under different conditions, e.g. anoxic conditions vs. oxic conditions, with subsequent transcriptome sequencing or proteomic identification could give indications towards which genes might be involved in regulating the activity of individual symbionts or symbiont groups (e.g. aerobes vs. anaerobes) under different physiological conditions. However, for any analysis that goes beyond descriptive, comparative and correlative research, specific experimental manipulations (e.g. controlled infection with other symbiont strains, genetically manipulated strains, or pathogenic strains) and genetic and physiologcical homogeneity between individual worms should be established first.

Of course, “host-microbe interaction” implies communication and response not only from the host side, but also from the symbionts. Futher analysis of the available symbiont genomes and the genes they express should therefore also incorporate the investigation of genes that might be important factors in establishing and maintaining the symbiosis, as well.

## Bibliography

[1] Sapp, J. (1994) Evolution by association: a history of symbiosis. Oxford University Press, USA

[2] Frank, A. B. (1877) Über die biologischen Verhältnisse des Thallus einiger Krustenflechten. *Beiträge zur Biologie der Pflanzen* 2, 123-200

[3] Bary, A. (1879) Die Erscheinung der Symbiose. Karl J. Trübner Verlag, Straßburg

[4] Martin, B. D. and Schwab, E. (2013) Current usage of symbiosis and associated terminology. *International Journal of Biology* 5(1), 32

[5] Thompson, J. N. (1988) Variation in interspecific interactions. *Annual Review of Ecology and Systematics* 19, 65–87

[6] Bronstein, J. L. (1994) Conditional outcomes in mutualistic interactions. *Trends in Ecology and Evolution* 9(6), 214–217

[7] Daskin, J. H. and Alford, R. A. (2012) Context-dependent symbioses and their potential roles in wildlife diseases. *Proceedings of the Royal Society B* 279, 1457–1465

[8] Garcia, J. R. and Gerardo, N. M. (2014) The symbiont side of symbiosis: do microbes really benefit? *Frontiers in Microbiology* 5, 510

[9] Lowe, C. D., Minter, E. J., Cameron, D. D., and Brockhurst, M. A. (2015) Shining a light on exploitative host control in a photosynthetic endosymbiosis. *Current Biology* 26(2), 207-211

[10] Werren, J. H., Baldo, L., and Clark, M. E. (2008) *Wolbachia*: master manipulators of invertebrate biology. *Nature Reviews Microbiology* 6(10), 741–751

[11] Taylor, M. J., Bandi, C., and Hoerauf, A. (2005) *Wolbachia*. Bacterial endosymbionts of filarial nematodes. *Advances in Parasitology* 60, 245–284

[12] Hosokawa, T., Koga, R., Kikuchi, Y., Meng, X.-Y., and Fukatsu, T. (2010) *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proceedings of the National Academy of Sciences U S A* 107(2), 769–774

[13] Neuhauser, C. and Fargione, J. E. (2004) A mutualism–parasitism continuum model and its application to plant–mycorrhizae interactions. *Ecological Modelling* 177(3), 337–352

[14] Rosenberg, E., Koren, O., Reshef, L., Efrony, R., and Zilber-Rosenberg, I. (2007) The role of microorganisms in coral health, disease and evolution. *Nature Reviwes Microbiology* 5(5), 355–362

[15] Kushmaro, A., Rosenberg, E., Fine, M., and Loya, Y. (1997) Bleaching of the coral *Oculina patagonica* by *Vibrio* AK-1. *Marine Ecology Progress Series* 147(1), 159–165

[16] Rosenberg, E. and Falkovitz, L. (2004) The *Vibrio shiloi*/*Oculina patagonica* model system of coral bleaching. *Annual Review of Microbiology* 58, 143–159

[17] Banin, E., Khare, S. K., Naider, F., and Rosenberg, E. (2001) Proline-rich peptide from the coral pathogen *Vibrio shiloi* that inhibits photosynthesis of zooxanthellae. *Applied and Environmental Microbiology* 67(4), 1536–1541

[18] Hauser, P. M. (2014) Genomic insights into the fungal pathogens of the genus *Pneumocystis*: obligate biotrophs of humans and other mammals. *PLoS Pathogens* 10(11), e1004425

[19] Klose, J., Polz, M. F., Wagner, M., Schimak, M. P., Gollner, S., and Bright, M. (2015) Endosymbionts escape dead hydrothermal vent tubeworms to enrich the free-living population. *Proceedings of the National Academy of Sciences U S A* 112(36), 11300–11305

[20] McFall-Ngai, M. J. (2002) Unseen forces: the influence of bacteria on animal development. *Developmental Biology* 242(1), 1–14

[21] McFall-Ngai, M., Hadfield, M. G., Bosch, T. C., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S. F., et al. (2013) Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences U S A* 110(9), 3229–3236

[22] Gilbert, S. F., Bosch, T. C. G., and Ledón-Rettig, C. (2015) Eco-Evo-Devo: developmental symbiosis and developmental plasticity as evolutionary agents. *Nature Reviews Genetics* 16(10), 611–622

[23] Bordenstein, S. R. and Theis, K. R. (2015) Host biology in light of the microbiome: Ten principles of holobionts and hologenomes. *PLoS Biology* 13(8), e1002226

[24] Sapp, J. (2016) The symbiotic self. *Evolutionary Biology*, 1–8

[25] Overmann, J. G. (2006) Molecular basis of symbiosis, volume 41. Springer Science & Business Media

[26] McFall-Ngai, M. (2008) Are biologists in 'future shock'? Symbiosis integrates biology across domains. *Nature Reviews Microbiology* 6(10), 789–792

[27] Moya, A., Peretó, J., Gil, R., and Latorre, A. (2008) Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nature Reviews Genetics* 9(3), 218–229

[28] Law, R. (1985) Evolution in a mutualistic environment. In: *The biology of Mutualism: Ecology and Evolution*. Oxford University Press, pp. 145–170

[29] Mizrahi, I. (2013) Rumen symbioses. In: *The Prokaryotes*. Springer Berlin Heidelberg, pp. 533–544

[30] Brune, A. (2014) Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews Microbiology* 12(3), 168–180

[31] Hansen, A. K. and Moran, N. A. (2014) The impact of microbial symbionts on host plant utilization by herbivorous insects. *Molecular Ecology* 23(6), 1473–1496

[32] Vincent, J. (1974) Root-nodule symbiosis with *Rhizobium*. In: *Biology of Nitrogen Fixation.* North-Holland Publishing Co., Amsterdam, pp. 265–341

[33] Foster, R. A., Kuypers, M. M., Vagner, T., Paerl, R. W., Musat, N., and Zehr, J. P. (2011) Nitrogen fixation and transfer in open ocean diatom–cyanobacterial symbioses. *The ISME Journal* 5(9), 1484–1493

[34] Roth, M. S. (2014) The engine of the reef: photobiology of the coral–algal symbiosis. *Frontiers in Microbiology* 5, 1-22

[35] Baumann, P. (2005) Biology bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annual Review of Microbiology* 59, 155–189

[36] LeBlanc, J. G., Milani, C., de Giori, G. S., Sesma, F., van Sinderen, D., and Ventura, M. (2013) Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Current Opinion in Biotechnology* 24(2), 160–168

[37] Mushegian, A. A. and Ebert, D. (2016) Rethinking “mutualism”? in diverse host-symbiont communities. *BioEssays* 38(1), 100–108

[38] Polz, M. F., Ott, J. A., Bright, M., and Cavanaugh, C. M. (2000) When bacteria hitch a ride. *ASM News* 66(9), 531–539

[39] Ott, J., Bright, M., and Bulgheresi, S. (2004) Marine microbial thiotrophic ectosymbioses. *Oceanography and Marine Biology – An Annual Review* 42, 95–118

[40] Bright, M., Espada-Hinojosa, S., Lagkouvardos, I., and Volland, J.-M. (2014) The giant ciliate *Zoothamnium niveum* and its thiotrophic epibiont *Candidatus Thiobios zoothamnicoli*: a model system to study interspecies cooperation. *Frontiers in Microbiology* 5(145)

[41] Trench, R. (1993) Microalgal-invertebrate symbioses - a review. *Endocytobiosis and Cell Research* 9(2-3), 135–175

[42] Denison, R. F. and Kiers, E. T. (2011) Life histories of symbiotic rhizobia and mycorrhizal fungi. *Current Biology* 21(18), R775–R785

[43] Lee, K. H. and Ruby, E. G. (1994) Effect of the squid host on the abundance and distribution of symbiotic *Vibrio fischeri* in nature. *Applied and Environmental Microbiology* 60(5), 1565–1571

[44] Skelton, J., Creed, R. P., and Brown, B. L. (2015) A symbiont’s dispersal strategy: condition-dependent dispersal underlies predictable variation in direct transmission among hosts. Proceedings of the Royal Society B 282, 20152081

[45] Giere, O., Conway, N., Gastrock, G., and Schmidt, C. (1991) "Regulation" of gutless annelid ecology by endosymbiotic bacteria. *Marine Ecology Progress Series* 68(3), 287–299

[46] Ott, J., Novak, R., Schiemer, F., Hentschel, U., Nebelsick, M., and Polz, M. (1991) Tackling the sulfide gradient: a novel strategy involving marine nematodes and chemoautotrophic ectosymbionts. *Marine Ecology* 12(3), 261–279

[47] Zientz, E., Dandekar, T., and Gross, R. (2004) Metabolic interdependence of obligate intracellular bacteria and their insect hosts. *Microbiology and Molecular Biology Reviews* 68(4), 745–770

[48] Yellowlees, D., Rees, T. A. V., and Leggat, W. (2008) Metabolic interactions between algal symbionts and invertebrate hosts. *Plant, Cell & Environment* 31(5), 679–694

[49] Ahmadjian, V. (1993) The lichen symbiosis. John Wiley & Sons New York

[50] Rumpho, M. E., Summer, E. J., and Manhart, J. R. (2000) Solar-powered sea slugs. mollusc/algal chloroplast symbiosis. *Plant Physiology* 123(1), 29–38

[51] Blackall, L. L., Wilson, B., and Oppen, M. J. (2015) Coral - the world’s most diverse symbiotic ecosystem. *Molecular Ecology* 24(21), 5330–5347

[52] Dubilier, N., Bergin, C., and Lott, C. (2008) Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nature Reviews Microbiology* 6(10), 725–740

[53] Cavanaugh, C. M., McKiness, Z. P., Newton, I. L., and Stewart, F. J. (2013) Marine chemosynthetic symbioses. In: *The Prokaryotes*, Springer New York, pp. 579–607

[54] Udvardi, M. and Poole, P. S. (2013) Transport and metabolism in legume-rhizobia symbioses. *Annual Review of Plant Biology* 64, 781–805

[55] Wilkinson, C. and Fay, P. (1979) Nitrogen fixation in coral reef sponges with symbiotic cyanobacteria. *Nature* 279, 527-529

[56] Horak, R. E. and Montoya, J. P. (2014) Growth, nitrogen fixation, respiration, and a nitrogen budget for cultures of a cosmopolitan diazotrophic endosymbiont (*Teredinibacter turnerae*) of shipworms. *Journal of the Marine Biological Association of the United Kingdom* 94(01), 177–185

[57] French, J., Turner, G., and Bradbury, J. (1976) Nitrogen fixation by bacteria from the hindgut of termites. *Microbiology* 95(2), 202–206

[58] Akman, L., Yamashita, A., Watanabe, H., Oshima, K., Shiba, T., Hattori, M., and Aksoy, S. (2002) Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nature Genetics* 32(3), 402–407

[59] Woyke, T., Teeling, H., Ivanova, N. N., Huntemann, M., Richter, M., Gloeckner, F. O., Boffelli, D., Anderson, I. J., Barry, K. W., Shapiro, H. J., Szeto, E., Kyrpides, N. C., Mussmann, M., Amann, R., Bergin, C., Ruehland, C., Rubin, E. M., and Dubilier, N. (2006) Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* 443(7114), 950–955

[60] Kleiner, M., Wentrup, C., Lott, C., Teeling, H., Wetzel, S., Young, J., Chang, Y.-J., Shah, M., VerBerkmoes, N. C., Zarzycki, J., Fuchs, G., Markert, S., Hempel, K., Voigt, B., Becher, D., Liebeke, M., Lalk, M., Albrecht, D., Hecker, M., Schweder, T., and Dubilier, N. (2012) Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proceedings of the National Academy of Sciences U S A* 109(19), E1173–E1182

[61] Tanaka, Y., Miyajima, T., Koike, I., Hayashibara, T., and Ogawa, H. (2006) Translocation and conservation of organic nitrogen within the coral-zooxanthella symbiotic system of acropora pulchra, as demonstrated by dual isotope-labeling techniques. *Journal of Experimental Marine Biology and Ecology* 336(1), 110–119

[62] Powell, M. and Somero, G. (1986) Adaptations to sulfide by hydrothermal vent animals: sites and mechanisms of detoxification and metabolism. *The Biological Bulletin* 171(1), 274–290

[63] Kikuchi, Y., Hayatsu, M., Hosokawa, T., Nagayama, A., Tago, K., and Fukatsu, T. (2012) Symbiont-mediated insecticide resistance. *Proceedings of the National Academy of Sciences U S A* 109(22), 8618–8622

[64] Dunbar, H. E., Wilson, A. C., Ferguson, N. R., and Moran, N. A. (2007) Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS Biology* 5(5), e96

[65] Kamada, N., Seo, S.-U., Chen, G. Y., and Núñez, G. (2013) Role of the gut microbiota in immunity and inflammatory disease. *Nature Reviews Immunology* 13(5), 321–335

[66] Currie, C. R., Scott, J. A., Summerbell, R. C., and Malloch, D. (1999) Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. *Nature* 398(6729), 701–704

[67] Jones, B. and Nishiguchi, M. (2004) Counterillumination in the Hawaiian bobtail squid, *Euprymna scolopes* Berry (Mollusca: Cephalopoda). *Marine Biology* 144(6), 1151–1155

[68] Oliver, K. M., Degnan, P. H., Burke, G. R., and Moran, N. A. (2010) Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annual Review of Entomology* 55, 247–266

[69] Haygood, M. G. (1993) Light organ symbioses in fishes. *Critical Reviews in Microbiology* 19(4), 191–216

[70] Forst, S., Dowds, B., Boemare, N., and Stackebrandt, E. (1997) *Xenorhabdus* and *Photorhabdus spp.*: bugs that kill bugs. *Annual Reviews in Microbiology* 51(1), 47–72

[71] Margulis, L. (1970) 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 New Haven

[72] Kutschera, U. and Niklas, K. J. (2005) Endosymbiosis, cell evolution, and speciation. *Theory in Biosciences* 124(1), 1–24

[73] De Duve, C. (2007) The origin of eukaryotes: a reappraisal. *Nature Reviews Genetics* 8(5), 395–403

[74] Martin, W. (2005) Archaebacteria (Archaea) and the origin of the eukaryotic nucleus. *Current Opinion in Microbiology* 8(6), 630–637

[75] Howe, C., Barbrook, A., Nisbet, R., Lockhart, P., and Larkum, A. (2008) The origin of plastids. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 363(1504), 2675–2685

[76] McFadden, G. I. (2001) Primary and secondary endosymbiosis and the origin of plastids. *Journal of Phycology* 37(6), 951–959

[77] Cavalier-Smith, T. (2002) Chloroplast evolution: secondary symbiogenesis and multiple losses. *Current Biology* 12(2), R62–R64

[78] Bolnick, D. I. and Fitzpatrick, B. M. (2007) Sympatric speciation: models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics* 38, 459–487

[79] Rundell, R. J. and Price, T. D. (2009) Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology and Evolution* 24(7), 394–399

[80] Presgraves, D. C. (2010) The molecular evolutionary basis of species formation. *Nature Reviews Genetics* 11(3), 175–180

[81] Crespi, B. and Nosil, P. (2013) Conflictual speciation: species formation via genomic conflict. *Trends in Ecology and Evolution* 28(1), 48–57

[82] Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, Å., et al. (2014) Genomics and the origin of species. *Nature Reviews Genetics* 15(3), 176–192

[83] Knoll, A. H. (2015) Life on a young planet: the first three billion years of evolution on earth. Princeton University Press, New Jersey

[84] Servedio, M. R., Van Doorn, G. S., Kopp, M., Frame, A. M., and Nosil, P. (2011) Magic traits in speciation: “magic” but not rare? *Trends in Ecology and Evolution* 26(8), 389–397

[85] Mayr, E., Mayr, E., Mayr, E., and Mayr, E. (1963) Animal species and evolution, volume 797. Belknap Press of Harvard University Press Cambridge, Massachusetts

[86] Doebeli, M. and Dieckmann, U. (2003) Speciation along environmental gradients. *Nature* 421(6920), 259–264

[87] Monahan, W. B., Pereira, R. J., and Wake, D. B. (2012) Ring distributions leading to species formation: a global topographic analysis of geographic barriers associated with ring species. *BMC Biology* 10(1), 1

[88] Mayr, E. (1954) Change of genetic environment and evolution. In: *Evolution as a Process*. Allen & Unwin, London, pp. 157-180

[89] Edwards, C. J., Suchard, M. A., Lemey, P., Welch, J. J., Barnes, I., Fulton, T. L., Barnett, R., O’Connell, T. C., Coxon, P., Monaghan, N., et al. (2011) Ancient hybridization and an Irish origin for the modern polar bear matriline. *Current Biology* 21(15), 1251–1258

[90] Mackie, R. I. (2002) Mutualistic fermentative digestion in the gastrointestinal tract: diversity and evolution. *Integrative and Comparative Biology* 42(2), 319–326

[91] Hawthorne, D. J. and Via, S. (2001) Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature* 412(6850), 904–907

[92] Favret, C. and Voegtlin, D. J. (2004) Speciation by host-switching in *Pinyon cinara* (Insecta: Hemiptera: Aphididae). *Molecular Phylogenetics and Evolution* 32(1), 139–151

[93] Tsuchida, T., Koga, R., and Fukatsu, T. (2004) Host plant specialization governed by facultative symbiont. *Science* 303(5666), 1989

[94] Lozier, J. D., Roderick, G. K., and Mills, N. J. (2007) Genetic evidence from mitochondrial, nuclear, and endosymbiont markers for the evolution of host plant associated species in the aphid genus *Hyalopterus* (Hemiptera: Aphididae). *Evolution* 61(6), 1353–1367

[95] Janson, E. M., Stireman, J. O., Singer, M. S., and Abbot, P. (2008) Phytophagous insect-microbe mutualisms and adaptive evolutionary diversification. *Evolution* 62(5), 997–1012

[96] Nardon, P. and Grenier, A.-M. (1991) Serial endosymbiosis theory and weevil evolution: The role of symbiosis. In: *Symbiosis as a source of evolutionary innovation: Speciation and morphogenesis*. The MIT Press, Massachusetts, pp. 153-169

[97] O’Connor, T. K., Humphrey, P. T., Lapoint, R. T., Whiteman, N. K., and O’Grady, P. M. (2014) Microbial interactions and the ecology and evolution of Hawaiian Drosophilidae. *Frontiers in Microbiology* 5, 616

[98] Takiya, D. M., Tran, P. L., Dietrich, C. H., and Moran, N. A. (2006) Co-cladogenesis spanning three phyla: leafhoppers (Insecta: Hemiptera: Cicadellidae) and their dual bacterial symbionts. *Molecular Ecology* 15(13), 4175–4191

[99] Hosokawa, T., Kikuchi, Y., Nikoh, N., Shimada, M., and Fukatsu, T. (2006) Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. *PLoS Biology* 4(10), e337

[100] Chen, X., Li, S., and Aksoy, S. (1999) Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. *Journal of Molecular Evolution* 48(1), 49–58

[101] Russell, J. A., Moreau, C. S., Goldman-Huertas, B., Fujiwara, M., Lohman, D. J., and Pierce, N. E. (2009) Bacterial gut symbionts are tightly linked with the evolution of herbivory in ants. *Proceedings of the National Academy of Sciences U S A* 106(50), 21236–21241

[102] Bordenstein, S. R. (2003) Symbiosis and the origin of species. *Insect Symbiosis* 1, 347

[103] Brucker, R. M. and Bordenstein, S. R. (2012) Speciation by symbiosis. *Trends in Ecology and Evolution* 27(8), 443–451

[104] Sharon, G., Segal, D., Ringo, J. M., Hefetz, A., Zilber-Rosenberg, I., and Rosenberg, E. (2010) Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences U S A* 107(46), 20051–20056

[105] Koukou, K., Pavlikaki, H., Kilias, G., Werren, J. H., Bourtzis, K., and Alahiotis, S. N. (2006) Influence of antibiotic treatment and *Wolbachia* curing on sexual isolation among *Drosophila melanogaster* cage populations. *Evolution* 60(1), 87–96

[106] Miller, W. J., Ehrman, L., and Schneider, D. (2010) Infectious speciation revisited: impact of symbiont-depletion on female fitness and mating behavior of *Drosophila paulistorum*. *PLoS Pathogens* 6(12), e1001214

[107] Leal, W. S. (1998) Chemical ecology of phytophagous scarab beetles. *Annual Review of Entomology* 43, 39–61

[108] Reusch, T. B., HaÈberli, M. A., Aeschlimann, P. B., and Milinski, M. (2001) Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. *Nature* 414(6861), 300–302

[109] Penn, D. J. (2002) The scent of genetic compatibility: sexual selection and the major histocompatibility complex. *Ethology* 108(1), 1–21

[110] Forsberg, L. A., Dannewitz, J., Petersson, E., and Grahn, M. (2007) Influence of genetic dissimilarity in the reproductive success and mate choice of brown trout–females fishing for optimal MHC dissimilarity. *Journal of Evolutionary Biology* 20(5), 1859–1869

[111] Schwensow, N., Eberle, M., and Sommer, S. (2008) Compatibility counts: MHC-associated mate choice in a wild promiscuous primate. *Proceedings of the Royal Society of London B: Biological Sciences* 275(1634), 555–564

[112] Eizaguirre, C., Lenz, T. L., Traulsen, A., and Milinski, M. (2009) Speciation accelerated and stabilized by pleiotropic major histocompatibility complex immunogenes. *Ecology Letters* 12(1), 5–12

[113] Werren, J. H., Windsor, D., and Guo, L. (1995) Distribution of *Wolbachia* among neotropical arthropods. *Proceedings of the Royal Society of London B: Biological Sciences* 262(1364), 197–204

[114] Bouchon, D., Rigaud, T., and Juchault, P. (1998) Evidence for widespread *Wolbachia* infection in isopod crustaceans: molecular identification and host feminization. *Proceedings of the Royal Society of London B: Biological Sciences* 265(1401), 1081–1090

[115] Jeyaprakash, A. and Hoy, M. (2000) Long PCR improves *Wolbachia* DNA amplification: wsp sequences found in 76% of sixty-three arthropod species. *Insect Molecular Biology* 9(4), 393–405

[116] Zchori-Fein, E., Perlman, S. J., Kelly, S. E., Katzir, N., and Hunter, M. S. (2004) Characterization of a *'Bacteroidetes'* symbiont in *Encarsia* wasps (Hymenoptera: Aphelinidae): proposal of *'Candidatus* Cardinium Hertigii'. *International Journal of Systematic and Evolutionary Microbiology* 54(Pt 3), 961–968

[117] Gotoh, T., Noda, H., and Ito, S. (2007) *Cardinium* symbionts cause cytoplasmic incompatibility in spider mites. *Heredity* 98(1), 13–20

[118] Adachi-Hagimori, T., Miura, K., and Stouthamer, R. (2008) A new cytogenetic mechanism for bacterial endosymbiont-induced parthenogenesis in hymenoptera. *Proceedings of the Royal Society B* 275(1652), 2667–2673

[119] Gordon, I. J., Ireri, P., and Smith, D. A. (2014) Hologenomic speciation: synergy between a male-killing bacterium and sex-linkage creates a “magic trait” in a butterfly hybrid zone. *Biological Journal of the Linnean Society* 111(1), 92–109

[120] Chafee, M. E., Zecher, C. N., Gourley, M. L., Schmidt, V. T., Chen, J. H., Bordenstein, S. R., Clark, M. E., and Bordenstein, S. R. (2011) Decoupling of host-symbiont-phage coadaptations following transfer between insect species. *Genetics* 187(1), 203–215

[121] Brucker, R. M. and Bordenstein, S. R. (2013) The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus *Nasonia*. *Science* 341(6146), 667–669

[122] Fritz, R. S., Moulia, C., and Newcombe, G. (1999) Resistance of hybrid plants and animals to herbivores, pathogens, and parasites. *Annual Review of Ecology and Systematics* , 565–591

[123] Ispolatov, I. and Doebeli, M. (2009) Speciation due to hybrid necrosis in plant–pathogen models. *Evolution* 63(12), 3076–3084

[124] Traw, M. B. and Bergelson, J. (2010) Plant immune system incompatibility and the distribution of enemies in natural hybrid zones. *Current Opinion in Plant Biology* 13(4), 466–471

[125] Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R., and Gordon, J. I. (2008) Worlds within worlds: Evolution of the vertebrate gut microbiota. *Nature Reviews Microbiology* 6(10), 776–788

[126] Atherton, J. C. and Blaser, M. J. (2009) Coadaptation of *Helicobacter pylori* and humans: ancient history, modern implications. *The Journal of Clinical Investigation* 119(9), 2475–2487

[127] Oh, P. L., Benson, A. K., Peterson, D. A., Patil, P. B., Moriyama, E. N., Roos, S., and Walter, J. (2010) Diversification of the gut symbiont *Lactobacillus reuteri* as a result of host-driven evolution. *The ISME Journal* 4(3), 377–387

[128] Sachs, J. L., Skophammer, R. G., and Regus, J. U. (2011) Evolutionary transitions in bacterial symbiosis. *Proceedings of the National Academy of Sciences U S A* 108(Supplement 2), 10800–10807

[129] Hongoh, Y. (2010) Diversity and genomes of uncultured microbial symbionts in the termite gut. *Bioscience, Biotechnology, and Biochemistry* 74(6), 1145–1151

[130] Zimmermann, J., Wentrup, C., Sadowski, M., Blazejak, A., Gruber-Vodicka, H., Kleiner, M., Ott, J., Cronholm, B., De Wit, P., Erséus, C., and Dubilier, N. (2016) Closely coupled evolutionary history of ecto- and endosymbionts from two distantly-related animal phyla. *Molecular Ecology,* doi: 10.1111/mec.13554

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

[132] Toft, C. and Andersson, S. G. E. (2010) Evolutionary microbial genomics: insights into bacterial host adaptation. *Nature Reviews Genetics* 11(7), 465–475

[133] McCutcheon, J. P. and Moran, N. A. (2012) Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology* 10(1), 13–26

[134] Dale, C., Wang, B., Moran, N., and Ochman, H. (2003) Loss of DNA recombinational repair enzymes in the initial stages of genome degeneration. *Molecular Biology and Evolution* 20(8), 1188–1194

[135] Parkhill, J., Sebaihia, M., Preston, A., Murphy, L. D., Thomson, N., Harris, D. E., Holden, M. T., Churcher, C. M., Bentley, S. D., Mungall, K. L., et al. (2003) Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nature Genetics* 35(1), 32–40

[136] Garnier, T., Eiglmeier, K., Camus, J.-C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsempe, C., et al. (2003) The complete genome sequence of *Mycobacterium bovis*. *Proceedings of the National Academy of Sciences U S A* 100(13), 7877–7882

[137] Seshadri, R., Paulsen, I. T., Eisen, J. A., Read, T. D., Nelson, K. E., Nelson, W. C., Ward, N. L., Tettelin, H., Davidsen, T. M., Beanan, M. J., et al. (2003) Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proceedings of the National Academy of Sciences U S A* 100(9), 5455–5460

[138] Moran, N. A. and Plague, G. R. (2004) Genomic changes following host restriction in bacteria. *Current Opinion in Genetics and Development* 14(6), 627–633

[139] Moran, N. A. (1996) Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences U S A* 93(7), 2873–2878

[140] Rio, R. V. M., Lefevre, C., Heddi, A., and Aksoy, S. (2003) Comparative genomics of insect-symbiotic bacteria: influence of host environment on microbial genome composition. *Applied and Environmental Microbiology* 69(11), 6825–6832

[141] Kresse, A. U., Dinesh, S. D., Larbig, K., and Römling, U. (2003) Impact of large chromosomal inversions on the adaptation and evolution of *Pseudomonas aeruginosa* chronically colonizing cystic fibrosis lungs. *Molecular Microbiology* 47(1), 145–158

[142] Hinchliffe, S. J., Isherwood, K. E., Stabler, R. A., Prentice, M. B., Rakin, A., Nichols, R. A., Oyston, P. C., Hinds, J., Titball, R. W., and Wren, B. W. (2003) Application of DNA microarrays to study the evolutionary genomics of *Yersinia pestis* and *Yersinia pseudotuberculosis*. *Genome Research* 13(9), 2018–2029

[143] Moran, N. A., McLaughlin, H. J., and Sorek, R. (2009) The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science* 323(5912), 379–382

[144] Wu, M., Sun, L. V., Vamathevan, J., Riegler, M., Deboy, R., Brownlie, J. C., McGraw, E. A., Martin, W., Esser, C., Ahmadinejad, N., et al. (2004) Phylogenomics of the reproductive parasite *Wolbachia pipientis* w Mel: a streamlined genome overrun by mobile genetic elements. *PLoS Biology* 2(3), e69

[145] Bordenstein, S. R. and Reznikoff, W. S. (2005) Mobile DNA in obligate intracellular bacteria. *Nature Reviews Microbiology* 3(9), 688–699

[146] Cordaux, R., Pichon, S., Ling, A., Pérez, P., Delaunay, C., Vavre, F., Bouchon, D., and Grève, P. (2008) Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont. *Molecular Biology and Evolution* 25(9), 1889–1896

[147] Bright, M. and Bulgheresi, S. (2010) A complex journey: transmission of microbial symbionts. *Nature Reviews Microbiology* 8(3), 218–230

[148] Nyholm, S. V. and Graf, J. (2012) Knowing your friends: invertebrate innate immunity fosters beneficial bacterial symbioses. *Nature Reviews Microbiology* 10(12), 815–827

[149] Ruby, E. G. (2008) Symbiotic conversations are revealed under genetic interrogation. *Nature Reviews Microbiology* 6(10), 752–762

[150] Dowds, C. M., Blumberg, R. S., and Zeissig, S. (2015) Control of intestinal homeostasis through crosstalk between natural killer T cells and the intestinal microbiota. *Clinical Immunology* 159(2), 128–133

[151] Hooper, L. V., Littman, D. R., and Macpherson, A. J. (2012) Interactions between the microbiota and the immune system. *Science* 336(6086), 1268–1273

[152] Chu, H. and Mazmanian, S. K. (2013) Innate immune recognition of the microbiota promotes host-microbial symbiosis. *Nature Immunology* 14(7), 668–675

[153] Hentschel, U., Steinert, M., and Hacker, J. (2000) Common molecular mechanisms of symbiosis and pathogenesis. *Trends in Microbiology* 8(5), 226–231

[154] Joyce, S. A., Watson, R. J., and Clarke, D. J. (2006) The regulation of pathogenicity and mutualism in *Photorhabdus*. *Current Opinion in Microbiology* 9(2), 127–132

[155] McFall-Ngai, M., Heath-Heckman, E. A. C., Gillette, A. A., Peyer, S. M., and Harvie, E. A. (2012) The secret languages of coevolved symbioses: insights from the *Euprymna scolopes-Vibrio fischeri* symbiosis. *Seminars in Immunology* 24(1), 3–8

[156] Casadevall, A. and Pirofski, L.-a. (2014) Microbiology: Ditch the term pathogen. *Nature* 516(7530), 165–166

[157] Söderhäll, K. (2011) Invertebrate immunity. Springer Science and Business Media

[158] Paul, W. E. (2013) *Fundamental Immunology 7th Edition*. Wolters Kluwer

[159] Medzhitov, R. and Janeway, C., Jr (2000) Innate immune recognition: mechanisms and pathways. *Immunological Reviews* 173, 89–97

[160] Medzhitov, R. (2007) Recognition of microorganisms and activation of the immune response. *Nature* 449(7164), 819–826

[161] Schulenburg, H., Boehnisch, C., and Michiels, N. K. (2007) How do invertebrates generate a highly specific innate immune response? *Molecular Immunology* 44(13), 3338–3344

[162] Kurtz, J. and Armitage, S. A. O. (2006) Alternative adaptive immunity in invertebrates. *Trends in Immunology* 27(11), 493–496

[163] Huang, S., Yuan, S., Guo, L., Yu, Y., Li, J., Wu, T., Liu, T., Yang, M., Wu, K., Liu, H., Ge, J., Yu, Y., Huang, H., Dong, M., Yu, C., Chen, S., and Xu, A. (2008) Genomic analysis of the immune gene repertoire of *Amphioxus* reveals extraordinary innate complexity and diversity. *Genome Research* 18(7), 1112–1126

[164] Ghosh, J., Lun, C. M., Majeske, A. J., Sacchi, S., Schrankel, C. S., and Smith, L. C. (2011) Invertebrate immune diversity. *Developmental and Comparative Immunology* 35(9), 959–974

[165] Cerenius, L. and Söderhäll, K. (2013) Variable immune molecules in invertebrates. *Journal of Experimental Biology* 216(Pt 23), 4313–4319

[166] Fraune, S. and Bosch, T. C. (2007) Long-term maintenance of species-specific bacterial microbiota in the basal metazoan *Hydra*. *Proceedings of the National Academy of Sciences U S A* 104(32), 13146–13151

[167] Nyholm, S. V., Stewart, J. J., Ruby, E. G., and McFall-Ngai, M. J. (2009) Recognition between symbiotic *Vibrio fischeri* and the haemocytes of *Euprymna scolopes*. *Environmental Microbiology* 11(2), 483–493

[168] Bulgheresi, S., Gruber-Vodicka, H. R., Heindl, N. R., Dirks, U., Kostadinova, M., Breiteneder, H., and Ott, J. A. (2011) Sequence variability of the pattern recognition receptor Mermaid mediates specificity of marine nematode symbioses. The *ISME Journal* 5(6), 986–998

[169] Wang, D., Yang, S., Tang, F., and Zhu, H. (2012) Symbiosis specificity in the legume–rhizobial mutualism. *Cellular Microbiology* 14(3), 334–342

[170] Nyholm, S. V. and McFall-Ngai, M. J. (2004) The winnowing: establishing the squid-*Vibrio* symbiosis. *Nature Reviews Microbiology* 2(8), 632–642

[171] Franzenburg, S., Walter, J., Künzel, S., Wang, J., Baines, J. F., Bosch, T. C. G., and Fraune, S. (2013) Distinct antimicrobial peptide expression determines host species-specific bacterial associations. *Proceedings of the national Academy of Sciences U S A* 110(39), E3730–E3738

[172] Kurtz, J. and Franz, K. (2003) Innate defence: evidence for memory in invertebrate immunity. *Nature* 425(6953), 37–38

[173] Kurtz, J. (2005) Specific memory within innate immune systems. *Trends in Immunology* 26(4), 186–192

[174] Rowley, A. F. and Powell, A. (2007) Invertebrate immune systems – specific, quasi-specific, or nonspecific? *The Journal of Immunology* 179(11), 7209–7214

[175] Hoffmann, J. A., Kafatos, F. C., Janeway, C. A., and Ezekowitz, R. A. (1999) Phylogenetic perspectives in innate immunity. *Science* 284(5418), 1313–1318

[176] Loker, E. S., Adema, C. M., Zhang, S.-M., and Kepler, T. B. (2004) Invertebrate immune systems – not homogeneous, not simple, not well understood. *Immunological Reviews* 198, 10–24

[177] Dishaw, L. J. and Litman, G. W. (2013) Changing views of the evolution of immunity. *Frontiers in Immunology* 4(122)

[178] Smith, V. J. (2001) Immunology of invertebrates: Cellular. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester

[179] Smith, L. C., Azumi, K., and Nonaka, M. (1999) Complement systems in invertebrates. the ancient alternative and lectin pathways. *Immunopharmacology* 42(1-3), 107–120

[180] Nappi, A. J. and Ottaviani, E. (2000) Cytotoxicity and cytotoxic molecules in invertebrates. *Bioessays* 22(5), 469–480

[181] Kulkarni, A. B., Huh, C.-G., Becker, D., Geiser, A., Lyght, M., Flanders, K. C., Roberts, A. B., Sporn, M. B., Ward, J. M., and Karlsson, S. (1993) Transforming growth factor beta 1 null mutation in mice causes excessive inflammatory response and early death. *Proceedings of the National Academy of Sciences U S A* 90(2), 770–774

[182] Lai, Y., Di Nardo, A., Nakatsuji, T., Leichtle, A., Yang, Y., Cogen, A. L., Wu, Z.-R., Hooper, L. V., Schmidt, R. R., von Aulock, S., Radek, K. A., Huang, C.-M., Ryan, A. F., and Gallo, R. L. (2009) Commensal bacteria regulate Toll-like receptor 3-dependent inflammation after skin injury. *Nature Medicine* 15(12), 1377–1382

[183] Wells, J. M., Loonen, L. M., and Karczewski, J. M. (2010) The role of innate signaling in the homeostasis of tolerance and immunity in the intestine. *International Journal of Medical Microbiology* 300(1), 41–48

[184] Royet, J., Gupta, D., and Dziarski, R. (2011) Peptidoglycan recognition proteins: modulators of the microbiome and inflammation. *Nature Reviews Immunology* 11(12), 837–851

[185] Medzhitov, R. and Janeway, C. A. (2002) Decoding the patterns of self and nonself by the innate immune system. *Science* 296(5566), 298–300

[186] Koropatnick, T. A., Engle, J. T., Apicella, M. A., Stabb, E. V., Goldman, W. E., and McFall-Ngai, M. J. (2004) Microbial factor-mediated development in a host-bacterial mutualism. *Science* 306(5699), 1186–1188

[187] Fraune, S., Augustin, R., Anton-Erxleben, F., Wittlieb, J., Gelhaus, C., Klimovich, V. B., Samoilovich, M. P., and Bosch, T. C. G. (2010) In an early branching metazoan, bacterial colonization of the embryo is controlled by maternal antimicrobial peptides. *Proceedings of the National Academy of Sciences U S A* 107(42), 18067–18072

[188] Franzenburg, S., Fraune, S., Künzel, S., Baines, J. F., Domazet-Loso, T., and Bosch, T. C. G. (2012) Myd88-deficient *Hydra* reveal an ancient function of TLR signaling in sensing bacterial colonizers. *Proceedings of the National Academy of Sciences U S A* 109(47), 19374–19379

[189] Login, F. H., Balmand, S., Vallier, A., Vincent-Monégat, C., Vigneron, A., Weiss-Gayet, M., Rochat, D., and Heddi, A. (2011) Antimicrobial peptides keep insect endosymbionts under control. *Science* 334(6054), 362–365

[190] Botos, I., Segal, D. M., and Davies, D. R. (2011) The structural biology of Toll-like receptors. *Structure* 19(4), 447–459

[191] Kawai, T. and Akira, S. (2011) Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity* 34(5), 637–650

[192] Kawai, T. and Akira, S. (2010) The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nature Immunology* 11(5), 373–384

[193] Leulier, F. and Lemaitre, B. (2008) Toll-like receptors - taking an evolutionary approach. *Nature Reviews Genetics* 9(3), 165–178

[194] Kubinak, J. L. and Round, J. L. (2012) Toll-like receptors promote mutually beneficial commensal-host interactions. *PLoS Pathogens* 8(7), e1002785

[195] Venkatesh, M., Mukherjee, S., Wang, H., Li, H., Sun, K., Benechet, A. P., Qiu, Z., Maher, L., Redinbo, M. R., Phillips, R. S., Fleet, J. C., Kortagere, S., Mukherjee, P., Fasano, A., Le Ven, J., Nicholson, J. K., Dumas, M. E., Khanna, K. M., and Mani, S. (2014) Symbiotic bacterial metabolites regulate gastrointestinal barrier function via the xenobiotic sensor PXR and Toll-like receptor 4. *Immunity* 41(2), 296–310

[196] Wang, Y., Telesford, K. M., Ochoa-Repáraz, J., Haque-Begum, S., Christy, M., Kasper, E. J., Wang, L., Wu, Y., Robson, S. C., Kasper, D. L., and Kasper, L. H. (2014) An intestinal commensal symbiosis factor controls neuroinflammation via TLR2-mediated CD39 signalling. *Nature Communications* 5, 4432

[197] Valanne, S., Wang, J.-H., and Rämet, M. (2011) The *Drosophila* Toll signaling pathway. *Journal of Immunology* 186(2), 649–656

[198] Myllymäki, H., Valanne, S., and Rämet, M. (2014) The *Drosophila* IMD signaling pathway. *Journal of Immunology* 192(8), 3455–3462

[199] Kaneko, T. and Silverman, N. (2005) Bacterial recognition and signalling by the Drosophila IMD pathway. *Cellular Microbiology* 7(4), 461–469

[200] Dong, C., Davis, R. J., and Flavell, R. A. (2002) MAP kinases in the immune response. *Annual Review of Immunology* 20(1), 55–72

[201] Zarubin, T. and Jiahuai, H. (2005) Activation and signaling of the p38 MAP kinase pathway. *Cell Research* 15(1), 11–18

[202] Anselme, C., Vallier, A., Balmand, S., Fauvarque, M.-O., and Heddi, A. (2006) Host PGRP gene expression and bacterial release in endosymbiosis of the weevil *Sitophilus zeamais*. *Applied and Environmental Microbiology* 72(10), 6766–6772

[203] Anselme, C., Pérez-Brocal, V., Vallier, A., Vincent-Monegat, C., Charif, D., Latorre, A., Moya, A., and Heddi, A. (2008) Identification of the weevil immune genes and their expression in the bacteriome tissue. *BMC Biology* 6(1), 1

[204] Nehme, N. T., Liégeois, S., Kele, B., Giammarinaro, P., Pradel, E., Hoffmann, J. A., Ewbank, J. J., and Ferrandon, D. (2007) A model of bacterial intestinal infections in *Drosophila melanogaster*. *PLoS Pathogens* 3(11), e173

[205] Ha, E.-M., Lee, K.-A., Seo, Y. Y., Kim, S.-H., Lim, J.-H., Oh, B.-H., Kim, J., and Lee, W.-J. (2009) Coordination of multiple dual oxidase–regulatory pathways in responses to commensal and infectious microbes in *Drosophila* gut. *Nature Immunology* 10(9), 949–957

[206] Kleino, A., Myllymäki, H., Kallio, J., Vanha-aho, L.-M., Oksanen, K., Ulvila, J., Hultmark, D., Valanne, S., and Rämet, M. (2008) Pirk is a negative regulator of the *Drosophila* IMD pathway. *The Journal of Immunology* 180(8), 5413–5422

[207] Zaidman-Rémy, A., Hervé, M., Poidevin, M., Pili-Floury, S., Kim, M.-S., Blanot, D., Oh, B.-H., Ueda, R., Mengin-Lecreulx, D., and Lemaitre, B. (2006) The *Drosophila* amidase PGRP-LB modulates the immune response to bacterial infection. *Immunity* 24(4), 463–473

[208] Ryu, J.-H., Nam, K.-B., Oh, C.-T., Nam, H.-J., Kim, S.-H., Yoon, J.-H., Seong, J.-K., Yoo, M.-A., Jang, I.-H., Brey, P. T., et al. (2004) The homeobox gene caudal regulates constitutive local expression of antimicrobial peptide genes in *Drosophila* epithelia. *Molecular and Cellular Biology* 24(1), 172–185

[209] Ruby, E. G. and Lee, K. H. (1998) The *Vibrio fischeri-Euprymna scolopes* light organ association: Current ecological paradigms. *Applied and Environmental Microbiology* 64(3), 805–812

[210] Lee, K. H. and Ruby, E. G. (1992) Detection of the light organ symbiont, *Vibrio fischeri*, in Hawaiian seawater by using lux gene probes. *Applied and Environmental Microbiology* 58(3), 942–947

[211] Visick, K. L. and Ruby, E. G. (2006) *Vibrio fischeri* and its host: it takes two to tango. *Current Opinion in Microbiology* 9(6), 632–638

[212] McFall-Ngai, M., Nyholm, S. V., and Castillo, M. G. (2010) The role of the immune system in the initiation and persistence of the *Euprymna scolopes–Vibrio fischeri* symbiosis. *Seminars in Immunology* 22(1), 48–53

[213] Troll, J. V., Bent, E. H., Pacquette, N., Wier, A. M., Goldman, W. E., Silverman, N., and McFall-Ngai, M. J. (2010) Taming the symbiont for coexistence: a host PGRP neutralizes a bacterial symbiont toxin. *Environmental Microbiology* 12(8), 2190–2203

[214] Reife, R. A., Coats, S. R., Al-Qutub, M., Dixon, D. M., Braham, P. A., Billharz, R. J., Howald, W. N., and Darveau, R. P. (2006) *Porphyromonas gingivalis* lipopolysaccharide lipid A heterogeneity: differential activities of tetra-and penta-acylated lipid A structures on E-selectin expression and TLR4 recognition. *Cellular Microbiology* 8(5), 857–868

[215] Herath, T. D., Wang, Y., Seneviratne, C. J., Lu, Q., Darveau, R. P., Wang, C.-Y., and Jin, L. (2011) *Porphyromonas gingivalis* lipopolysaccharide lipid A heterogeneity differentially modulates the expression of IL-6 and IL-8 in human gingival fibroblasts. *Journal of Clinical Periodontology* 38(8), 694–701

[216] Kawahara, K., Tsukano, H., Watanabe, H., Lindner, B., and Matsuura, M. (2002) Modification of the structure and activity of lipid A in *Yersinia pestis* lipopolysaccharide by growth temperature. *Infection and Immunity* 70(8), 4092–4098

[217] Mason, K. L., Stepien, T. A., Blum, J. E., Holt, J. F., Labbe, N. H., Rush, J. S., Raffa, K. F., and Handelsman, J. (2011) From commensal to pathogen: translocation of *Enterococcus faecalis* from the midgut to the hemocoel of *Manduca sexta*. *MBio* 2(3), e00065–11

[218] Lotze, M. T., Zeh, H. J., Rubartelli, A., Sparvero, L. J., Amoscato, A. A., Washburn, N. R., DeVera, M. E., Liang, X., Tör, M., and Billiar, T. (2007) The grateful dead: damage-associated molecular pattern molecules and reduction/oxidation regulate immunity. *Immunological Reviews* 220(1), 60–81

[219] Kono, H. and Rock, K. L. (2008) How dying cells alert the immune system to danger. *Nature Reviews Immunology* 8(4), 279–289

[220] Bosch, T. C. G., Augustin, R., Anton-Erxleben, F., Fraune, S., Hemmrich, G., Zill, H., Rosenstiel, P., Jacobs, G., Schreiber, S., Leippe, M., Stanisak, M., Grötzinger, J., Jung, S., Podschun, R., Bartels, J., Harder, J., and Schröder, J.-M. (2009) Uncovering the evolutionary history of innate immunity: the simple metazoan *Hydra* uses epithelial cells for host defence. *Developmental and Comparative Immunology* 33(4), 559–569

[221] Markwardt, F. (1994) The development of hirudin as an antithrombotic drug. *Thrombosis Research* 74(1), 1–23

[222] Sanchez-Hernandez, J. (2006) Earthworm biomarkers in ecological risk assessment. In: *Reviews of environmental contamination and toxicology*. Springer, pp. 85–126

[223] Cooper, E. L. (2006) Comparative immunology. *Integrative Zoology* 1(1), 32–43

[224] Satchell, J. (2012) *Earthworm ecology: from Darwin to vermiculture*. Springer Science & Business Media

[225] Tambuwala, A., Kaul, D., Bohra, T., Antulay, R. A., et al. (2016) Leech therapy in the management of congested free flaps. *Universal Research Journal of Dentistry* 6(1), 25

[226] Cooper, E. L. and Roch, P. (1993) Immunological profile of annelids: transplantation immunity. In: *Immunology of Annelids. CRC Press, Boca Raton*, pp. 201–244

[227] Tasiemski, A. and Salzet, M. (2010) Leech immunity: from brain to peripheral responses. In: *Invertebrate Immunity*. Springer US, pp. 80–104

[228] Salzet, M., Tasiemski, A., and Cooper, E. (2006) Innate immunity in lophotrochozoans: the annelids. *Current Pharmaceutical Design* 12(24), 3043–3050

[229] Bilej, M., Procházková, P., Šilerová, M., and Josková, R. (2010) Earthworm immunity. In: *Invertebrate Immunity*. Springer US, pp. 66–79

[230] Bilej, M., De Baetselier, P., and Beschin, A. (2000) Antimicrobial defense of the earthworm. *Folia Microbiologica* 45(4), 283–300

[231] Davidson, C. R., Best, N. M., Francis, J. W., Cooper, E. L., and Wood, T. C. (2008) Toll-like receptor genes (TLRs) from *Capitella capitata* and *Helobdella robusta* (Annelida). *Developmental and Comparative Immunology* 32(6), 608–612

[232] Nyholm, S. V., Song, P., Dang, J., Bunce, C., and Girguis, P. R. (2012) Expression and putative function of innate immunity genes under in situ conditions in the symbiotic hydrothermal vent tubeworm *Ridgeia piscesae*. *PLoS One* 7(6), e38267

[233] Cerenius, L. and Söderhäll, K. (2004) The prophenoloxidase-activating system in invertebrates. *Immunological Reviews* 198(1), 116–126

[234] Procházková, P., Šilerová, M., Stijlemans, B., Dieu, M., Halada, P., Josková, R., Beschin, A., Baetselier, P. D., and Bilej, M. (2006) Evidence for proteins involved in prophenoloxidase cascade *Eisenia fetida* earthworms. *Journal of Comparative Physiology B* 176(6), 581–587

[235] Cuvillier-Hot, V., Boidin-Wichlacz, C., Slomianny, C., Salzet, M., and Tasiemski, A. (2011) Characterization and immune function of two intracellular sensors, HmTLR1 and HmNLR, in the injured CNS of an invertebrate. *Developmental and Comparative Immunology* 35(2), 214–226

[236] Škanta, F., Roubalová, R., Dvořák, J., Procházková, P., and Bilej, M. (2013) Molecular cloning and expression of TLR in the *Eisenia andrei* earthworm. *Developmental and Comparative Immunology* 41(4), 694–702

[237] Fjøsne, T. F., Stenseth, E.-B., Myromslien, F., and Rudi, K. (2014) Gene expression of TLR homologues identified by genome-wide screening of the earthworm *Dendrobaena veneta*. *Innate Immunity* 21, 161–166

[238] Wojdani, A., Stein, E. A., Lemmi, C. A., and Cooper, E. L. (1982) Agglutinins and proteins in the earthworm, *Lumbricus terrestris*, before and after injection of erythrocytes, carbohydrates, and other materials. *Developmental and Comparative Immunology* 6(4), 613–624

[239] Bilej, M., De Baetselier, P., Van Dijck, E., Stijlemans, B., Colige, A., and Beschin, A. (2001) Distinct carbohydrate recognition domains of an invertebrate defense molecule recognize gram-negative and gram-positive bacteria. *The Journal of Biological Chemistry* 276(49), 45840–45847

[240] Wang, X., Chang, L., and Sun, Z. (2011) Differential expression of genes in the earthworm *Eisenia fetida* following exposure to *Escherichia coli* O157:H7. *Developmental and Comparative Immunology* 35(5), 525–529

[241] Beschin, A., Bilej, M., Hanssens, F., Raymakers, J., Van Dyck, E., Revets, H., Brys, L., Gomez, J., De Baetselier, P., and Timmermans, M. (1998) Identification and cloning of a glucan- and lipopolysaccharide-binding protein from *Eisenia foetida* earthworm involved in the activation of prophenoloxidase cascade. *The Journal of Biological Chemistry* 273(38), 24948–24954

[242] Deloffre, L., Salzet, B., Vieau, D., Andries, J.-C., and Salzet, M. (2003) Antibacterial properties of hemerythrin of the sand worm *Nereis diversicolor*. *Neuroendocrinology Letters* 24(1-2), 39–45

[243] Tasiemski, A. (2008) Antimicrobial peptides in annelids. *Invertebrate Survival Journal* 5, 75–82

[244] Minic, Z. and Hervé, G. (2004) Biochemical and enzymological aspects of the symbiosis between the deep-sea tubeworm *Riftia pachyptila* and its bacterial endosymbiont. *European Journal of Biochemistry* 271(15), 3093–3102

[245] Stewart, F. J. and Cavanaugh, C. M. (2006) Symbiosis of thioautotrophic bacteria with *Riftia pachyptila*. *Progress in Molecular and Subcellular Biology* 41, 197–225

[246] Robidart, J. C., Bench, S. R., Feldman, R. A., Novoradovsky, A., Podell, S. B., Gaasterland, T., Allen, E. E., and Felbeck, H. (2008) Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed through metagenomics. *Environmental Microbiology* 10(3), 727–737

[247] Myokai, F., Takashiba, S., Lebo, R., and Amar, S. (1999) A novel lipopolysaccharide-induced transcription factor regulating tumor necrosis factor α gene expression: molecular

cloning, sequencing, characterization, and chromosomal assignment. *Proceedings of the National Academy of Sciences U S A* 96(8), 4518–4523

[248] Marra, M. N., Wilde, C. G., Griffith, J. E., Snable, J. L., and Scott, R. W. (1990) Bactericidal/permeability-increasing protein has endotoxin-neutralizing activity. *Journal of Immunology* 144(2), 662–666

[249] Elsbach, P. and Weiss, J. (1998) Role of the bactericidal/permeability-increasing protein in host defence. *Current Opinion in Immunology* 10(1), 45–49

[250] Wu, L. P. and Anderson, K. V. (1998) Regulated nuclear import of Rel proteins in the *Drosophila* immune response. *Nature* 392(6671), 93–97

[251] Li, C., Chen, Y.-X., Zhang, S., Lü, L., Chen, Y.-H., Chai, J., Weng, S., Chen, Y.-G., He, J., and Xu, X. (2012) Identification, characterization, and function analysis of the cactus gene from *Litopenaeus vannamei*. *PloS One* 7(11), e49711

[252] Lund, M. B., Kjeldsen, K. U., and Schramm, A. (2014) The earthworm-*Verminephrobacter* symbiosis: an emerging experimental system to study extracellular symbiosis. *Frontiers in Microbiology* 5, 128

[253] Pinel Peláez, N. (2009) Physiological and Genomic Insights into the Biology of *Verminephrobacter eiseniae*, a Bacterial Symbiont of the Earthworm *Eisenia fetida*. Ph.D. thesis, University of Washington, Seattle, USA

[254] Lund, M. B., Holmstrup, M., Lomstein, B. A., Damgaard, C., and Schramm, A. (2010) Beneficial effect of *Verminephrobacter* nephridial symbionts on the fitness of the earthworm *Aporrectodea tuberculata*. *Applied and Environmental Microbiology* 76(14), 4738–4743

[255] Dulla, G. F. J., Go, R. A., Stahl, D. A., and Davidson, S. K. (2012) *Verminephrobacter eiseniae* type IV pili and flagella are required to colonize earthworm nephridia. *The ISME Journal* 6(6), 1166–1175

[256] Nelson, M. C. and Graf, J. (2012) Bacterial symbioses of the medicinal leech *Hirudo verbana*. *Gut Microbes* 3(4), 322–331

[257] Graf, J., Kikuchi, Y., and Rio, R. V. M. (2006) Leeches and their microbiota: naturally simple symbiosis models. *Trends in Microbiology* 14(8), 365–371

[258] Tasiemski, A., Massol, F., Cuvillier-Hot, V., Boidin-Wichlacz, C., Roger, E., Rodet, F., Fournier, I., Thomas, F., and Salzet, M. (2015) Reciprocal immune benefit based on complementary production of antibiotics by the leech *Hirudo verbana* and its gut symbiont *Aeromonas veronii*. *Scientific Reports* 5, 17498

[259] Graf, J. (2002) The effect of symbionts on the physiology of *Hirudo medicinalis*, the medicinal leech. *Invertebrate Reproduction and Development* 41(1-3), 269–275

[260] Janda, J. M. and Abbott, S. L. (2010) The genus *Aeromonas*: taxonomy, pathogenicity, and infection. *Clinical Microbiology Reviews* 23(1), 35–73

[261] Silver, A. C., Kikuchi, Y., Fadl, A. A., Sha, J., Chopra, A. K., and Graf, J. (2007) Interaction between innate immune cells and a bacterial type III secretion system in mutualistic and pathogenic associations. *Proceedings of the National Academy of Sciences U S A* 104(22), 9481–9486

[262] Silver, A. C., Rabinowitz, N. M., Küffer, S., and Graf, J. (2007) Identification of *Aeromonas veronii* genes required for colonization of the medicinal leech, *Hirudo verbana*. *Journal of Bacteriology* 189(19), 6763–6772

[263] Maltz, M. and Graf, J. (2011) The type II secretion system is essential for erythrocyte lysis and gut colonization by the leech digestive tract symbiont *Aeromonas veronii*. *Applied and Environmental Microbiology* 77(2), 597–603

[264] Kikuchi, Y. and Graf, J. (2007) Spatial and temporal population dynamics of a naturally occurring two-species microbial community inside the digestive tract of the medicinal leech. *Applied and Environmental Microbiology* 73(6), 1984–1991

[265] Macagno, E. R., Gaasterland, T., Edsall, L., Bafna, V., Soares, M. B., Scheetz, T., Casavant, T., Da Silva, C., Wincker, P., Tasiemski, A., and Salzet, M. (2010) Construction of a medicinal leech transcriptome database and its application to the identification of leech homologs of neural and innate immune genes. *BMC Genomics* 11, 407

[266] Bomar, L. and Graf, J. (2012) Investigation into the physiologies of *Aeromonas veronii* in vitro and inside the digestive tract of the medicinal leech using RNA-Seq. *Biological Bulletin* 223(1), 155–166

[267] Kirk, J. T. (1994) Light and photosynthesis in aquatic ecosystems. Cambridge University Press, UK

[268] Eiler, A. (2006) Evidence for the ubiquity of mixotrophic bacteria in the upper ocean: implications and consequences. *Applied and Environmental Microbiology* 72(12), 7431–7437

[269] Béja, O., Spudich, E. N., Spudich, J. L., Leclerc, M., and DeLong, E. F. (2001) Proteorhodopsin phototrophy in the ocean. *Nature* 411(6839), 786–789

[270] Siegl, A., Kamke, J., Hochmuth, T., Piel, J., Richter, M., Liang, C., Dandekar, T., and Hentschel, U. (2011) Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. The *ISME Journal* 5(1), 61–70

[271] Sucharita, K., Shiva Kumar, E., Sasikala, C., Panda, B. B., Takaichi, S., and Ramana, C. V. (2010) *Marichromatium fluminis sp. nov*., a slightly alkaliphilic, phototrophic gammaproteobacterium isolated from river sediment. *International Journal of Systematic and Evolutionary Microbiology* 60(Pt 5), 1103–1107

[272] Overmann, J. (2008) Green Nonsulfur Bacteria. John Wiley & Sons, Ltd

[273] Giovannoni, S. J., Bibbs, L., Cho, J.-C., Stapels, M. D., Desiderio, R., Vergin, K. L., Rappé, M. S., Laney, S., Wilhelm, L. J., Tripp, H. J., Mathur, E. J., and Barofsky, D. F. (2005) Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* 438(7064), 82–85

[274] Fuchs, B. M., Spring, S., Teeling, H., Quast, C., Wulf, J., Schattenhofer, M., Yan, S., Ferriera, S., Johnson, J., Glöckner, F. O., and Amann, R. (2007) Characterization of a marine gammaproteobacterium capable of aerobic anoxygenic photosynthesis. *Proceedings of the National Academy of Sciences U S A* 104(8), 2891–2896

[275] Kelly, D. P., Wood, A. P., and Stackebrandt, E. (2015) Thiobacillus. John Wiley & Sons, Ltd

[276] Gommers, P. and Kuenen, J. (1988) *Thiobacillus* strain Q, a chemolithoheterotrophic sulphur bacterium. *Archives of Microbiology* 150(2), 117–125

[277] Vacelt, J., Fiala-Medioni, A., Fisher, C., and Boury-Esnault, N. (1997) Symbiosis between methane-oxidizing bacteria and a deep-sea carnivorous cladorhizid sponge. *Oceanographic Literature Review* 6(44), 618

[278] Childress, J. J., Fisher, C. R., Brooks, J. M., Kennicutt, M., Bidigare, R., and Anderson, A. E. (1986) A methanotrophic marine molluscan (Bivalvia, Mytilidae) symbiosis: mussels fueled by gas. *Science* 233(4770), 1306–1308

[279] Ott, J. A., Bright, M., and Schiemer, F. (1998) The ecology of a novel symbiosis between a marine peritrich ciliate and chemoautotrophic bacteria. *Marine Ecology* 19(3), 229–243

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

[281] Giere, O., Windoffer, R., and Southward, E. C. (1995) The bacterial endosymbiosis of the gutless nematode, *Astomonema southwardorum*: ultrastructural aspects. *Journal of the Marine Biological Association of the United Kingdom* 75(01), 153–164

[282] Felbeck, H. (1983) Sulfide oxidation and carbon fixation by the gutless clam *Solemya reidi*: an animal-bacteria symbiosis. *Journal of Comparative Physiology* 152(1), 3–11

[283] Duperron, S., Fiala-Médioni, A., Caprais, J.-C., Olu, K., and Sibuet, M. (2007) Evidence for chemoautotrophic symbiosis in a Mediterranean cold seep clam (Bivalvia: Lucinidae): comparative sequence analysis of bacterial 16S rRNA, APS reductase and RuBisCO genes. *FEMS Microbiology Ecology* 59(1), 64–70

[284] Dufour, S. C. (2005) Gill anatomy and the evolution of symbiosis in the bivalve family Thyasiridae. *The Biological Bulletin* 208(3), 200–212

[285] Newton, I. L. G., Woyke, T., Auchtung, T. A., Dilly, G. F., Dutton, R. J., Fisher, M. C., Fontanez, K. M., Lau, E., Stewart, F. J., Richardson, P. M., Barry, K. W., Saunders, E., Detter, J. C., Wu, D., Eisen, J. A., and Cavanaugh, C. M. (2007) The *Calyptogena magnifica* chemoautotrophic symbiont genome. *Science* 315(5814), 998–1000

[286] Pruski, A. M., Rousse, N., Fiala-Médioni, A., and Boulègue, J. (2002) Sulphur signature in the hydrothermal vent mussel *Bathymodiolus azoricus* from the Mid-Atlantic Ridge. *Journal of the Marine Biological Association of the UK* 82(03), 463–468

[287] Stein, J. L., Cary, S. C., Hessler, R. R., Vetter, R. D., Felbeck, H., Ohta, S., and Childress, J. J. (1988) Chemoautotrophic symbiosis in a hydrothermal vent gastropod. *The Biological Bulletin* 174(3), 373–378

[288] Endow, K. and Ohta, S. (1989) The symbiotic relationship between bacteria and a mesogastropod snail, *Alviniconcha hessleri*, collected from hydrothermal vents of the Mariana Back-Arc Basin. *Bulletin of Japanese Society of Microbial Ecology* 3(2), 73–82

[289] Burgh, M. and Singla, C. (1984) Bacterial colonization and endocytosis on the gill of a new limpet species from a hydrothermal vent. *Marine Biology* 84(1), 1–6

[290] McLean, J. H. (1988) New archaeogastropod limpets from hydrothermal vents; superfamily Lepetodrilacea I. Systematic descriptions. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 319(1192), 1–32

[291] Campbell, B. J., Stein, J. L., and Cary, S. C. (2003) Evidence of chemolithoautotrophy in the bacterial community associated with *Alvinella pompejana*, a hydrothermal vent polychaete. *Applied and Environmental Microbiology* 69(9), 5070–5078

[292] Cavanaugh, C. M., Gardiner, S. L., Jones, M. L., Jannasch, H. W., and Waterbury, J. B. (1981) Prokaryotic cells in the hydrothermal vent tube worm *Riftia pachyptila* Jones: Possible chemoautotrophic symbionts. *Science* 213(4505), 340–342

[293] Felbeck, H. (1981) Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila* Jones (Vestimentifera). *Science* 213(4505), 336–338

[294] Dubilier, N., Mülders, C., Ferdelman, T., de Beer, D., Pernthaler, A., Klein, M., Wagner, M., Erséus, C., Thiermann, F., Krieger, J., Giere, O., and Amann, R. (2001) Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. *Nature* 411(6835), 298–302

[295] Van Dover, C., Fry, B., Grassle, J., Humphris, S., and Rona, P. (1988) Feeding biology of the shrimp *Rimicaris exoculata* at hydrothermal vents on the Mid-Atlantic Ridge. *Marine Biology* 98(2), 209–216

[296] Temara, A., De Ridder, C., Kuenen, J., and Robertson, L. (1993) Sulfide-oxidizing bacteria in the burrowing echinoid, *Echinocardium cordatum* (Echinodermata). *Marine Biology* 115(2), 179–185

[297] Petersen, J. M., Zielinski, F. U., Pape, T., Seifert, R., Moraru, C., Amann, R., Hourdez, S., Girguis, P. R., Wankel, S. D., Barbe, V., Pelletier, E., Fink, D., Borowski, C., Bach, W., and Dubilier, N. (2011) Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* 476(7359), 176–180

[298] Hentschel, U., Cary, S. C., and Felbeck, H. (1993) Nitrate respiration in chemoautotrophic symbionts of the bivalve *Lucinoma aequizonata*. *Marine Ecology Progress Series* 94, 35–41

[299] Hentschel, U. and Felbeck, H. (1993) Nitrate respiration in the hydrothermal vent tubeworm *Riftia pachyptila*. *Nature* 366(6453), 338–340

[300] Lonsdale, P. (1977) Clustering of suspension-feeding macrobenthos near abyssal hydrothermal vents at oceanic spreading centers. *Deep Sea Research* 24(9), 857–863

[301] Van Dover, C. (2000) *The ecology of deep-sea hydrothermal vents*. Princeton University Press

[302] Polz, M. F. and Cavanaugh, C. M. (1995) Dominance of one bacterial phylotype at a Mid-Atlantic Ridge hydrothermal vent site. *Proceedings of the National Academy of Sciences U S A* 92(16), 7232–7236

[303] Haddad, A., Camacho, F., Durand, P., and Cary, S. (1995) Phylogenetic characterization of the epibiotic bacteria associated with the hydrothermal vent polychaete *Alvinella pompejana*. *Applied and Environmental Microbiology* 61(5), 1679–1687

[304] Urakawa, H., Dubilier, N., Fujiwara, Y., Cunningham, D. E., Kojima, S., and Stahl, D. A. (2005) Hydrothermal vent gastropods from the same family (Provannidae) harbour epsilon- and gamma-proteobacterial endosymbionts. *Environmental Microbiology* 7(5), 750–754

[305] Jamieson, B. (1977) Marine meiobenthic oligochaeta from Heron and Wistari Reefs (Great Barrier Reef) of the genera *Clitellio*, *Limnodriloides* and *Phallodrilus* (Tubificidae) and *Grania* (Enchytraeidae). *Zoological Journal of the Linnean Society* 61(4), 329–349

[306] Erséus, C. (1979) *Inanidrilus bulbosus gen. et sp. n*., a marine tubificid (Oligochaeta) from Florida, USA. *Zoologica Scripta* 8(1-4), 209–210

[307] Giere, O. (1981) The gutless marine oligochaete *Phallodrilus leukodermatus*. Structural studies on an aberrant tubificid associated with bacteria. *Marine Ecology Progress Series* 5, 353–357

[308] Felbeck, H., Liebezeit, G., Dawson, R., and Giere, O. (1983) $CO_2$ fixation in tissues of marine oligochaetes (*Phallodrilus leukodermatus* and *P. planus*) containing symbiotic, chemoautotrophic bacteria. *Marine Biology* 75(2-3), 187–191

[309] Giere, O. (1985) The gutless marine tubificid *Phallodrilus planus*, a flattened oligochaete with symbiotic bacteria. Results from morphological and ecological studies. *Zoologica Scripta* 14(4), 279–286

[310] Dubilier, N., Blazejak, A., and Rühland, C. (2006) Symbioses between bacteria and gutless marine oligochaetes. *Progress in Molecular and Subcellular Biology* 41, 251–275

[311] Giere, O. (2013) Meiobenthology: the microscopic fauna in aquatic sediments. Springer Science & Business Media

[312] Erseus, C. (1984) Taxonomy and phylogeny of the gutless Phallodrilinae (Oligochaeta, Tubificidae), with descriptions of one new genus and twenty-two new species. *Zoologica Scripta* 13(4), 239–272

[313] Giere, O., Nieser, C., Windoffer, R., and Erséus, C. (1995) A comparative structural study on bacterial symbioses of Caribbean gutless Tubificidae (Annelida, Oligochaeta). *Acta Zoologica* 76(4), 281–290

[314] Liebezeit, G., Felbeck, H., Dawson, R., and Giere, O. (1983) Transepidermal uptake of dissolved carbohydrates by the gutless marine oligochaete *Phallodrilus leukodermatus* (Annelida). *Oceanis* 9, 205–211

[315] Giere, O., Felbeck, H., Dawson, R., and Liebezeit, G. (1984) The gutless oligochaete *Phallodrilus leukodermatus* Giere, a tubificid of structural, ecological and physiological relevance. In: *Aquatic Oligochaeta*. Springer Netherlands, pp. 83–89

[316] Panchal, B., Bagdadi, A., and Roy, I. (2013) Polyhydroxyalkanoates: The natural polymers produced by bacterial fermentation. In: *Advances in natural polymers*. Springer Berlin Heidelberg, pp. 397–421

[317] Giere, O. and Langheld, C. (1987) Structural organisation, transfer and biological fate of endosymbiotic bacteria in gutless oligochaetes. *Marine Biology* 93(4), 641–650

[318] Giere, O., Wirsen, C., Schmidt, C., and Jannasch, H. (1988) Contrasting effects of sulfide and thiosulfate on symbiotic $CO_2$-assimilation of *Phallodrilus leukodermatus* (Annelida). *Marine Biology* 97(3), 413–419

[319] Kleiner, M., Woyke, T., Ruehland, C., and Dubilier, N. (2011) The *Olavius algarvensis* metagenome revisited: Lessons learned from the analysis of the low-diversity microbial

consortium of a gutless marine worm. In: *Handbook of Molecular Microbial Ecology II: Metagenomics in Different Habitats*, pp. 319–333

[320] Giere, O. and Erséus, C. (2002) Taxonomy and new bacterial symbioses of gutless marine Tubificidae (Annelida, Oligochaeta) from the Island of Elba (Italy). *Organisms Diversity & Evolution* 2(4), 289–297

[321] Schimak, M. P. (2015) Transmission of bacterial symbionts in the gutless oligochaete *Olavius algarvensis*. Ph.D. thesis, University of Bremen, Germany

[322] Erséus, C. (2003) The gutless Tubificidae (Annelida: Oligochaeta) of the Bahamas. *Meiofauna Marina* 12, 59–84

[323] Eréus, C. (1981) Taxonomic studies of Phallodrilinae (Oligochaeta, Tubificidae) from the Great Barrier Reef and the Comoro Islands with descriptions of ten new species and one new genus. *Zoologica Scripta* 10(1), 15–31

[324] Erséus, C. and Baker, H. (1982) New species of the gutless marine genus *Inanidrilus* (Oligochaeta, Tubificidae) from the Gulf of Mexico and Barbados. *Canadian Journal of Zoology* 60(12), 3063–3067

[325] Erséus, C. (1993) The marine Tubificidae (Oligochaeta) of Rottnest Island, Western Australia. In: *Proceedings of the Fifth International Marine Biological Workshop: The Marine Flora and Fauna of Rottnest Island, Western Australia, Western Australia Museum, Perth*, pp. 331–390

[326] Finogenova, N. (1986) Six new species of marine Tubificidae (Oligochaeta) from the continental shelf off Peru. *Zoologica Scripta* 15(1), 45–51

[327] Erséus, C. (1991) Two new deep-water species of the gutless genus *Olavius* (Oligochaeta: Tubificidae) from sides of North America. *Proceedings of the Biological Society of Washington* 104(3), 627–630

[328] Giere, O., Liebezeit, G., and Dawson, R. (1982) Habitat conditions and distribution pattern of the gutless oligochaete *Phallodrilus leukodermatus*. *Marine Ecology Progress Series* 8(3), 291–299

[329] Krieger, J. (2000) Transmission in *Inanidrilus leukodermatus*. Ph.D. thesis, University of Hamburg, Germany

[330] Perner, M. (2003) Biogeochemische und mikrobiologische Charakterisierung mariner Sedimente vor Elba - ein Beitrag zur ökosystemaren Analyse bakteriensymbiontischer Oligochaeten. Master's thesis, University of Bremen, Germany

[331] Grieshaber, M. K., Hardewig, I., Kreutzer, U., and Pörtner, H. O. (1994) Physiological and metabolic responses to hypoxia in invertebrates. *Reviews of Physiological Biochemistry and Pharmacology* 125, 43–147

[332] Levin, L., Gutiérrez, D., Rathburn, A., Neira, C., Sellanes, J., Munoz, P., Gallardo, V., and Salamanca, M. (2002) Benthic processes on the Peru Margin: a transect across the oxygen minimum zone during the 1997–98 El Niño. *Progress in Oceanography* 53(1), 1–27

[333] Bergin, C. (2009) Phylogenetic diversity and metabolic versatility of the bacterial endosymbionts in marine gutless oligochaete worms. Ph.D. thesis, University of Bremen, Germany

[334] Erséus, C. (1992) A generic revision of the Phallodrilinae (Oligochaeta, Tubificidae). *Zoologica Scripta* 21(1), 5–48

[335] Nylander, J. A., Erséus, C., and Källersjö, M. (1999) A test of monophyly of the gutless Phallodrilinae (Oligochaeta, Tubificidae) and the use of a 573-bp region of the mitochondrial cytochrome oxidase I gene in analysis of annelid phylogeny. *Zoologica Scripta* 28(3-4), 305–313

[336] Erséus, C., Prestegaard, T., and Källersjö, M. (2000) Phylogenetic analysis of Tubificidae (Annelida, Clitellata) based on 18S rDNA sequences. *Molecular Phylogenetics and Evolution* 15(3), 381–389

[337] Ruehland, C. (2010) Characterization of bacterial endo- and ectosymbionts of oligochaete worms from marine sediments: Phylogeny and metabolic potential. Ph.D. thesis, University of Bremen, Germany

[338] Blazejak, A., Erséus, C., Amann, R., and Dubilier, N. (2005) Coexistence of bacterial sulfide oxidizers, sulfate reducers, and spirochetes in a gutless worm (Oligochaeta) from the Peru Margin. *Applied and Environmental Microbiology* 71(3), 1553–1561

[339] Ruehland, C., Blazejak, A., Lott, C., Loy, A., Erséus, C., and Dubilier, N. (2008) Multiple bacterial symbionts in two species of co-occurring gutless oligochaete worms from Mediterranean sea grass sediments. *Environmental Microbiology* 10(12), 3404–3416

[340] Bulgheresi, S., Schabussova, I., Chen, T., Mullin, N. P., Maizels, R. M., and Ott, J. A. (2006) A new C-type lectin similar to the human immunoreceptor DC-SIGN mediates symbiont acquisition by a marine nematode. *Applied and Environmental Microbiology* 72(4), 2950–2956

[341] Dubilier, N., Amann, R., Erseus, C., Muyzer, G., Park, S., Giere, O., and Cavanaugh, C. M. (1999) Phylogenetic diversity of bacterial endosymbionts in the gutless marine oligochete *Olavius loisae* (Annelida). *Marine Ecology Progress Series* 178, 271–280

[342] Giere, O. and Milligan, M. (1989) The gutless, bacteria-symbiotic tubificid *Inanidrilus bulbosus*: structural analysis of an oligochaete with aberrant reproductive organs. *Zoologischer Anzeiger* 223(3-4), 174–185

[343] Ortega Noriega, B. (2011) Symbiont transmission in *Olavius algarvensis*. Master's thesis, Universidad Nacional Autónoma de México, Mexico

[344] Blazejak, A. (2005) Phylogenetic and functional characterization of symbiotic bacteria in gutless marine worms (Annelida, Oligochaeta). Ph.D. thesis, University of Bremen, Germany

[345] Giere, O., Erséus, C., and Stuhlmacher, F. (1998) A new species of *Olavius* (Tubificidae) from the Algarve coast in Portugal, the first East Atlantic gutless oligochaete with symbiotic bacteria. *Zoologischer Anzeiger* 237(2), 209–214

[346] Kleiner, M. (2012) Metabolism and evolutionary ecology of chemosynthetic symbionts from marine invertebrates. Ph.D. thesis, University of Bremen, Germany

[347] Larkum, A. W., Orth, R. J., and Duarte, C. M. (2006) Seagrasses: Biology, Ecology and Conservation. Springer Netherlands

[348] Kleiner, M., Wentrup, C., Holler, T., Lavik, G., Harder, J., Lott, C., Littmann, S., Kuypers, M. M. M., and Dubilier, N. (2015) Use of carbon monoxide and hydrogen by a bacteria-animal symbiosis from seagrass sediments. *Environmental Microbiology* 17(12), 5023-5035

[349] Dorman, D. C., Moulin, F. J.-M., McManus, B. E., Mahle, K. C., James, R. A., and Struve, M. F. (2002) Cytochrome oxidase inhibition induced by acute hydrogen sulfide inhalation: correlation with tissue sulfide concentrations in the rat brain, liver, lung, and nasal epithelium. *Toxicological Sciences* 65(1), 18–25

[350] Haldane, J. and Smith, J. L. (1897) The absorption of oxygen by the lungs. *Journal of Physiology* 22(3), 231–258

[351] Rossi-Fanelli, A. and AntoniniN, E. (1958) Studies on the oxygen and carbon monoxide equilibria of human myoglobin. *Archives of Biochemistry and Biophysics* 77(2), 478–492

[352] Barbazuk, W. B. and Schnable, P. S. (2011) SNP discovery by transcriptome pyrosequencing. In: *cDNA Libraries: Methods and Applications.* Springer, pp. 225–246

[353] Kuhner, M. K., Beerli, P., Yamato, J., and Felsenstein, J. (2000) Usefulness of single nucleotide polymorphism data for estimating population parameters. *Genetics* 156(1), 439–447

[354] Brumfield, R. T., Beerli, P., Nickerson, D. A., and Edwards, S. V. (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution* 18(5), 249–256

[355] Akey, J. M., Zhang, K., Xiong, M., and Jin, L. (2003) The effect of single nucleotide polymorphism identification strategies on estimates of linkage disequilibrium. *Molecular Biology and Evolution* 20(2), 232–242

[356] Nothnagel, M., Wolf, A., Herrmann, A., Szafranski, K., Vater, I., Brosch, M., Huse, K., Siebert, R., Platzer, M., Hampe, J., et al. (2011) Statistical inference of allelic imbalance from transcriptome data. *Human Mutation* 32(1), 98–106

[357] Sachs, J. L., Skophammer, R. G., Bansal, N., and Stajich, J. E. (2014) Evolutionary origins and diversification of proteobacterial mutualists. *Proceedings of the Royal Society B* 281(1775), 20132146

[358] Heindl, N. R., Gruber-Vodicka, H. R., Bayer, C., Lücker, S., Ott, J. A., and Bulgheresi, S. (2011) First detection of thiotrophic symbiont phylotypes in the pelagic marine environment. *FEMS Microbiology Ecology* 77(1), 223–227

## Acknowledgements

I would like to thank a number of people without whom this thesis would not have been possible. I am very grateful to:

**Prof. Dr. Nicole Dubilier,** for giving me the opportunity to do this thesis in her lab and for being my PhD advisor. Thank you for your support, for giving me so much freedom and responsibility, and for allowing me to grow both scientifically and personally at my own pace. Thank you also, for letting me participate in so many fun and interesting workshops and field trips, even towards the end of my PhD when I should have been writing ;)

**Prof. Dr. Monika Bright**, for great discussions and fun at the Symbiomics Workshop on Elba, and for kindly reviewing my thesis and serving as a member of the Prüfungsausschuss.

**Prof. Dr. Michael Friedrich**, **Dr. Pierre Offre**, **Oliver Jäckle** and **Dolma Michellod** for graciously serving as members of my Prüfungsausschuss. Dolma, thank you so much for taking my thesis on the cruise and delivering it to Nicole for review!!

**Dr. Manuel Kleiner**, for your continuing support during this PhD thesis, in particular in the last two years. Thank you so much for your excellent mentoring, fruitful discussions, valuable comments and moral support, for always being there and bringing back my optimism. I truly enjoy working with you.

**Dr. Cecilia Wentrup**, **Dr. Anne-Christin Kreutzmann**, **Dr. Alexander Gruhl, Dr. Manuel Liebeke**, **Dr. Harald Gruber-Vodicka**, **Brandon Seah** and **Dr. Maggie Sogin** for absolutely excellent tea kitchen discussions about science, life and the universe and your helpful advices.

**Silke Wetzel**, for exceptional technical and personal support! Thank you for all your support and for your caring friendship! And for delivering this thesis to Nicole!!

**Martina Meyer** and **Miriam Sadowski** for excellent technical assistance

The **IT department**, in particular **Carsten John**, **Oliver Artmann** and **Olaf Gundermann** for excellent technical support

**Miriam Weber**, **Christian Lott** and the **HYDRA-Team** for their invaluable field support “an allen Ecken und Enden”, fruitful discussions, amazing catering and organization during the workshops (and always amazing organization), and of course, for your encouragement and friendship. No matter how many hours we spent sorting worms, you always made our stays fun and exciting.

**Bernd Stickfort** for retrieving any and all obscure publications almost instantly

**Dr. Mario Schimak L. v. M.**, **Adrien Assié**, **Oli Jäckle** (#puglife), **Brandon Seah**, and **Burak Avci** for insanity, hilarity and your close friendship over the years. Adrien, thank you so much for teaching me Illustrator! Let’s all be old and insane, but endearing professors someday! I can already picture it in my head!

All my colleagues in the **Symbiosis Department and beyond** for being a great team to work with!

Very special thanks go to Chriss, my mum and Amandine. You all know how much you mean to me and that without you, I would not have come that far. Thank you for supporting me in anything I do.

# Appendix

**Digital supplement**

**Supplementary Files of Chapter 2:**

- **Supplementary file 1**: Multiple alignment of host COI sequences from Sant' Andrea
  -> Contains all host COI sequences obtained in this study by PCR

- **Supplementary file 2**: Multiple alignment of *OalgG1* 16S sequences from clone library
  -> contains all *OalgG1* 16S sequences from clone library

- **Supplementary file 3**: Multiple alignment of *OalgG3* 16S sequences reconstructed from metagenomes
  -> contains all 16S sequences of *OalgG3* symbiont reconstructed from metagenomes with EMIRGE

- **Supplementary file 4**: Multiple alignment of deltaproteobacterial symbiont 16S sequences reconstructed from metagenomes
  -> contains all 16S sequences of deltaproteobacterial symbionts reconstructed from metagenomes with EMIRGE

- **Supplementary file 5**: Multiple alignment of *OalgS1* 16S sequences reconstructed from metagenomes
  -> contains all 16S sequences of *OalgS1* symbiont reconstructed from metagenomes with EMIRGE

- **Supplementary file 6**: Confirmation of *OalgG1* phylotype-specific genes by read mapping analysis
  -> contains a list of putative phylotype-specific genes and the number of mapped reads from each metagenome for each gene

**Supplementary Files of Chapter 4:**

- **Supplementary_Table8.xls**
- **Supplementary_Table14.xls**

Universität Bremen

**Name:** Juliane Wippler

**Ort, Datum:** 18. April 2016

**Anschrift:** Schwachhauser Heerstr. 245a, 28211 Bremen

**ERKLÄRUNG**

Hiermit erkläre ich, dass ich die Doktorarbeit mit dem Titel:

„Characterization of host-symbiont molecular interactions and evolutionary relationships in the gutless oligochaete *Olavius algarvensis*"

selbstständig verfasst und geschrieben habe und außer den angegebenen Quellen keine weiteren Hilfsmittel verwendet habe.

Ebenfalls erkläre ich hiermit, dass es sich bei den von mir abgegebenen Arbeiten um 3 identische Exemplare handelt.

…….

(Unterschrift)