



universität
wien

DISSERTATION

Titel der Dissertation

**Genomic Insights into Molecular Interactions of two
Bacteroidetes Symbionts with their Eukaryotic Hosts**

Verfasser

Mag. rer. nat. Thomas Penz

angestrebter akademischer Grad

Doktor der Naturwissenschaften (Dr.rer.nat.)

Wien, 2014

Studienkennzahl lt. Studienblatt:	A 091 441
Dissertationsgebiet lt. Studienblatt:	441 Genetik - Mikrobiologie
Betreuerin / Betreuer:	Univ.-Prof. Dr. Matthias Horn

Contents

Chapter I	Introduction	1
	Outline and Contributions	19
Chapter II	Comparative genomics suggests an independent origin of cytoplasmic incompatibility in <i>Cardinium hertigii</i>	30
Chapter III	A bacterial genome in transition - an exceptional enrichment of IS elements but lack of evidence for recent transposition in the symbiont <i>Amoebophilus asiaticus</i>	42
Chapter IV	The genome of the amoeba symbiont " <i>Candidatus</i> <i>Amoebophilus asiaticus</i> " encodes an afp-like prophage possibly used for protein secretion	59
Chapter V	Host adaptation of a symbiont: The biphasic life cycle of <i>Amoebophilus asiaticus</i> and its phage derived protein secretion system	65
Chapter VI	The endosymbiont <i>Amoebophilus asiaticus</i> encodes an S-adenosylmethionine carrier that compensates for its missing methylation cycle	142
Chapter VII	Conclusion	154
Chapter VIII	Summary	160
	Zusammenfassung	162
	Curriculum Vitae	164
	Acknowledgments	167

Chapter I

Introduction

Outline and contributions

Introduction

The Success of Symbioses

Symbioses are intimate associations between two or more species and are present everywhere in nature. According to the endosymbiosis theory, all eukaryotic organisms on earth arose from symbiotic associations between ancient prokaryotes two billion years ago (Margulis, 1970). Today, endless numbers of symbioses can be found between and within the three domains of life (Bacteria, Archaea and Eukarya). The holobiont concept is a very good example of how omnipresent symbioses are. This concept describes the relations of multicellular eukaryotes with its colonies of persistent symbionts and postulates that no eukaryotic organism is a biological individual (Gilbert et al, 2012).

In 1879 Heinrich Anton de Bary introduced the symbiotic concept, which led to a new scientific field. He defined symbiosis as any (long-term) association between organisms. However, symbiosis, in Greek living together, can be used to describe various degrees of relationships between different organisms. The three major categories in which symbiosis can be divided based on the gained costs and benefits for each symbiotic partner are parasitism, commensalism and mutualism (Table 1).

Table 1. The three major categories of symbioses with their effects on their partners

Interaction	Partner A	Partner B
parasitism	receives benefit	harmed
commensalism	receives benefit	not affected
mutualism	receives benefit	receives benefit

Although symbiotic interactions span all realms of life, symbioses between prokaryotes and eukaryotic organisms are the most frequent ones (Moran, 2006). In prokaryote-eukaryote symbioses, the bacterium is usually referred to as symbiont, while the eukaryote is referred to as host. In eukaryote-prokaryote symbioses the local relation between the partners plays an important role and can thus be characterised further. Symbionts that live extracellularly on the surface of their host are referred to as ectosymbionts (e.g. chemolithoautotrophic sulfur-oxidizing bacteria on the surface of marine ciliates (Rinke et al, 2006)), whereas symbionts that live inside a cell are described as endosymbionts (environmental chlamydia in amoebae (Horn, 2008)).

Nature shows many fascinating examples where symbiotic partners have such an intimate relationship that neither of them could live without the other. The *Gammaproteobacterium Buchnera aphidicola* and its aphid host *Acyrtosiphon pisum* (Buchner, 1965) for instance possess a very intimate obligate symbiosis. The aphid *Acyrtosiphon pisum* feeds on plant phloem exclusively. The plant phloem is rich in sugar, but lacks amino acids, which are essential for the aphid (Douglas, 2006). Interestingly, the *Gammaproteobacterium Buchnera* is able to provide essential amino acids and compensates for the poor diet of the aphid. Both, *Buchnera* and the aphid would not be able to live without each other for a long period of time. The consequence of removing the symbiont from the aphid would be infertility and reduced growth of the aphid. The removal of the host is even more dramatic; it results in the immediate death of the symbiont (Buchner, 1965; Douglas, 1996; Houk & Griffiths, 1980).

Taken together, symbiosis is a key principle in nature, which is one of the driving forces behind evolution. The omnipresence of symbiotic associations in nature leads to a new awareness of the importance to study and understand interactions between organisms.

***Acanthamoebae* as Hosts for Bacterial Endosymbionts**

Acanthamoebae are ubiquitous protozoa that can be found in various natural environments (Khan, 2006; Rodriguez-Zaragoza, 1994). Besides being opportunistic human pathogens, causing blinding keratitis or fatal encephalitis (Khan, 2006), amoebae are predators of bacteria (Rodriguez-Zaragoza, 1994) and have thus a great impact on microbial community composition. However, some bacteria have developed mechanisms to resist amoebal phagocytosis and are able to use the intracellular environment of an amoeba as a niche for survival, multiplication and spreading (Horn & Wagner, 2004). Amoebae harbor two different categories of symbionts. Along with facultative intracellular bacteria of amoebae including the human pathogens *Legionella pneumophila*, *Mycobacterium avium*, *Francisella tularensis*, *Chlamydia pneumoniae*, *Listeria monocytogenes* and *Pseudomonas aeruginosa* (Albert-Weissenberger et al, 2007; Greub & Raoult, 2004; Thomas & McDonnell, 2007) also obligate intracellular bacterial symbionts of amoebae have been observed. Obligate intracellular bacteria of amoebae are affiliated with only three bacterial phyla belonging to five bacterial lineages: (i) the *Alphaproteobacteria* (Birtles et al, 2000), (ii) the *Betaproteobacteria* (Horn et al, 2002), (iii) the *Gammaproteobacteria* (Horn et al, 1999), (iv) the *Chlamydiales* (Amann et al, 1997) and the (v) *Bacteroidetes* (Horn et al, 2001).

The obligate *Acanthamoeba* Endosymbiont *Amoebophilus asiaticus*

The obligate intracellular symbiont of *Acanthamoebae*, *Amoebophilus asiaticus* (Figure 1) belongs to the diverse phylum of *Bacteroidetes* (Horn et al, 2001). Within the *Bacteroidetes* different symbiotic lifestyles with various hosts (e.g. protists, insects, mammals) are observed and vary from free-living to endosymbiotic ones.

Phylogenetic analyses place *Amoebophilus* together with the insect endosymbiont *Cardinium hertigii* (Zchori-Fein et al, 2001) and the nematode symbiont *Candidatus Paenicardinium endonii* (Noel & Atibalentja, 2006) within one phylogenetic clade (Gruwell et al, 2007). Other sequences clustering into the “*Amoebophilus/Cardinium* clade” are retrieved from coral samples (Sunagawa et al, 2010). The well-known obligate *Bacteroidetes* insect endosymbionts *Blattabacterium* sp. and *Sulcia* sp. are only distantly related to this clade.

However, among prokaryotic genomes the genome of *Amoebophilus* is unique. The genome contains a large fraction of transposase genes (n=354; 23% of all coding sequences) and genes encoding putative host-cell interaction proteins (n=129; 8% of all coding sequences) (Schmitz-Esser et al, 2010). The latter includes proteins with typical eukaryotic protein-protein interaction motifs such as ankyrin-repeats (ANKs), tetratricopeptide-repeats (TPRs), leucine-rich repeats (LRRs), as well as proteins with F- and U-box domains and ubiquitin-specific proteases that enable the bacterium to interfere with the host’s ubiquitin system (Schmitz-Esser et al, 2010).

Although the genome of *Amoebophilus* is only moderately reduced in size (Schmitz-Esser et al, 2010), it does not encode most amino acid, cofactor and vitamin biosynthesis pathways (Schmitz-Esser et al, 2010). To compensate for its reduced biosynthetic capabilities, *Amoebophilus* harbors an arsenal of transport proteins to gain energy from its host and is thus not a mutualistic symbiont but rather a parasite that exploits its *Acanthamoeba* host (Schmitz-Esser et al, 2010).

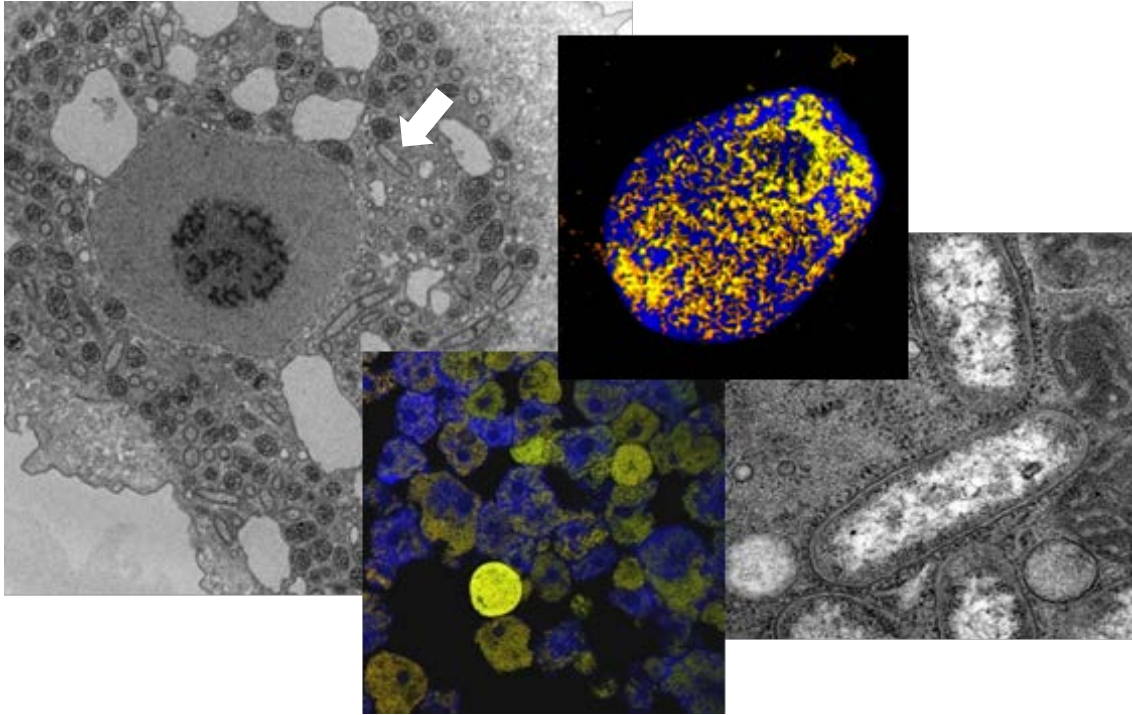


Figure 1: from left to right; electron micrograph of an *Acanthamoeba* infected with *Amoebophilus*, with the white arrow indicating *Amoebophilus* in the vicinity to mitochondria, image modified from (Schmitz-Esser et al, 2010); Fluorescence in situ hybridization (FISH) image of an *Acanthamoeba* culture infected with *Amoebophilus*, *Acanthamoebae* are indicated in blue and *Amoebophilus* in yellow; FISH image of an *Acanthamoeba* infected with *Amoebophilus*, the *Acanthamoeba* is indicated in blue and *Amoebophilus* in yellow; electron micrograph of *Amoebophilus* in the cytoplasm of an *Acanthamoeba* showing that *Amoebophilus* is associated with ribosomes (rough endoplasmic reticulum), image modified from (Schmitz-Esser et al, 2010)

Insects as Hosts for Bacterial Endosymbionts

With six to ten million species, insects are the most successful and abundant animal group on earth (Raven & Yeates, 2007). Most probably, every insect is associated with bacteria. There are conservative estimates that up to 20% of all insects harbor symbiotic bacteria (Buchner, 1965). Most of these bacteria are vertically transmitted, from mother to offspring and, based

on their features and host effects they can be divided into three major categories (Table 2): (i) the obligate symbionts that provide nutrients; (ia) the facultative symbionts that provide protection against stress or natural enemies and (iib) facultative symbionts that are able to manipulate the reproduction of their hosts, also known as reproductive manipulators (Moran et al, 2008). From the perspective of the host, obligate symbionts are indispensable for the survival of the host, while facultative symbionts are not.

Table 2 Features of obligate and facultative insect symbionts, modified from (Moran et al, 2008).

Features of Obligate Insect Symbionts	Features of Facultative Insect Symbionts	
housed within special host organ (bacteriome)	invades various cell and tissue types of hosts	
long evolutionary history of diversification with host lineages	short evolutionary history in current host lineage	
no horizontal transfer	horizontal transfer within and between host species	
supplies nutrients to hosts	provides protection against stress or natural enemies	manipulates reproduction to favor its own host matriline
extreme genome reduction, <<1 MB	moderate genome reduction and gene inactivation, >1MB	
lack of gene uptake, phage, mobile elements or genome rearrangements	dynamic genomes with bacteriophage, mobile elements, rearrangements	

Obligate Insect Endosymbionts

Obligate insect symbionts (also known as primary or P-endosymbionts) are restricted to a specialized organ called bacteriome. These bacteriomes contain specialized cells called bacteriocytes in which the obligate symbionts are housed (Baumann & Baumann, 2005). The major function of these obligate insect symbionts is the provision of nutrients to compensate

the poor diet of their insect hosts. Besides being important nutrient providers, obligate insect symbionts have a long evolutionary history with their hosts (Moran et al, 2008). On a non-individual scale, the association between the well-known vertically transmitted *gammaproteobacterial* insect symbiont *Buchnera aphidicola* and its aphid host has persisted through many generations for at least 160 million years (Burke et al, 2010). An even more extreme example of duration of interaction between a bacterial symbiont and its insect hosts is the symbiosis of the *Bacteroidetes* symbiont *Candidatus Sulcia muelleri*. *Sulcia* is associated with insects since the mid-Permian for about 260 million years (Moran et al, 2005). On an individual scale, the association between these two insect endosymbionts, *Buchnera* and *Sulcia*, and their hosts lasts for a lifetime.

An outstanding feature of bacterial insect endosymbionts is their genomic evolution. The so far smallest described bacterial genome can be found within this group and is only 139 kb in size. This incredibly small genome belongs to a mealy bug endosymbiont called *Candidatus Tremblaya princeps* (McCutcheon & Moran, 2011). Many other members of this group have very small genomes too e.g. *Candidatus Hodgekinia cicadicola* (McCutcheon et al, 2009b), *Candidatus Carsonella ruddii* (Nakabachi et al, 2006), *Candidatus Zinderia insecticola* (McCutcheon & Moran, 2010) and *Candidatus Sulcia muelleri* (Woyke et al, 2010) (Table 3). In comparison, the smallest genome of a free living bacterium, which is the genome of *Mycoplasma genitalium* (Bak et al, 1969), is twice as big as the largest genome of any of the bacterial insect endosymbionts mentioned above (Table 3).

Facultative Insect Endosymbionts

In contrast to obligate insect endosymbionts, facultative insect endosymbionts (also known as secondary or S-endosymbionts) are not restricted to specialized organs. They can be found in

various tissues and cells of their hosts. In comparison to obligate insect endosymbionts, the genomes of facultative insect endosymbionts are only moderately reduced. Based on the increased number of phages and mobile genetic elements in their genomes, the process of gene inactivation and gene loss is still in progress. (McCutcheon & Moran, 2011). Thus, genomes of facultative insect endosymbionts are still in transition. Facultative endosymbionts can be divided into two major groups. The first group are (iia) facultative endosymbionts that provide benefits to their hosts and the second group (iib) are able to manipulate the reproduction of their hosts and are thus known as reproductive manipulators.

An endosymbiont that belongs to the first group is the aphid and whitefly endosymbiont *Hamiltonella defensa* (McCutcheon & Moran, 2011). *Hamiltonella* is able to provide protection against parasitic wasps. Parasitic wasps lay their eggs into living aphids so that subsequently wasp larvae develop within the living aphid. Interestingly, aphids harbouring *Hamiltonella* are able to block the larval development of the endoparasitoid wasps (Oliver et al, 2003).

The second group, reproductive manipulators, are fascinating examples of how bacteria manage to influence the reproduction of their eukaryotic hosts in a very dramatic and severe way. The two best-studied reproductive manipulators of arthropods are *Cardinium hertigii* (Hunter et al, 2003) and *Wolbachia pipientis* (Werren et al, 2008). These reproductive manipulators are able to induce four different phenotypes in their insect hosts: male killing, parthenogenesis, feminization and cytoplasmic incompatibility, which is the most observed phenotype (Figure 2). All four phenotypes increase the relative fitness of infected females and consequently lead to an enormous spread of the symbiont in the host population. Due to the strict maternal transmission of the reproductive manipulator, all four phenotypes favour infected arthropod females.

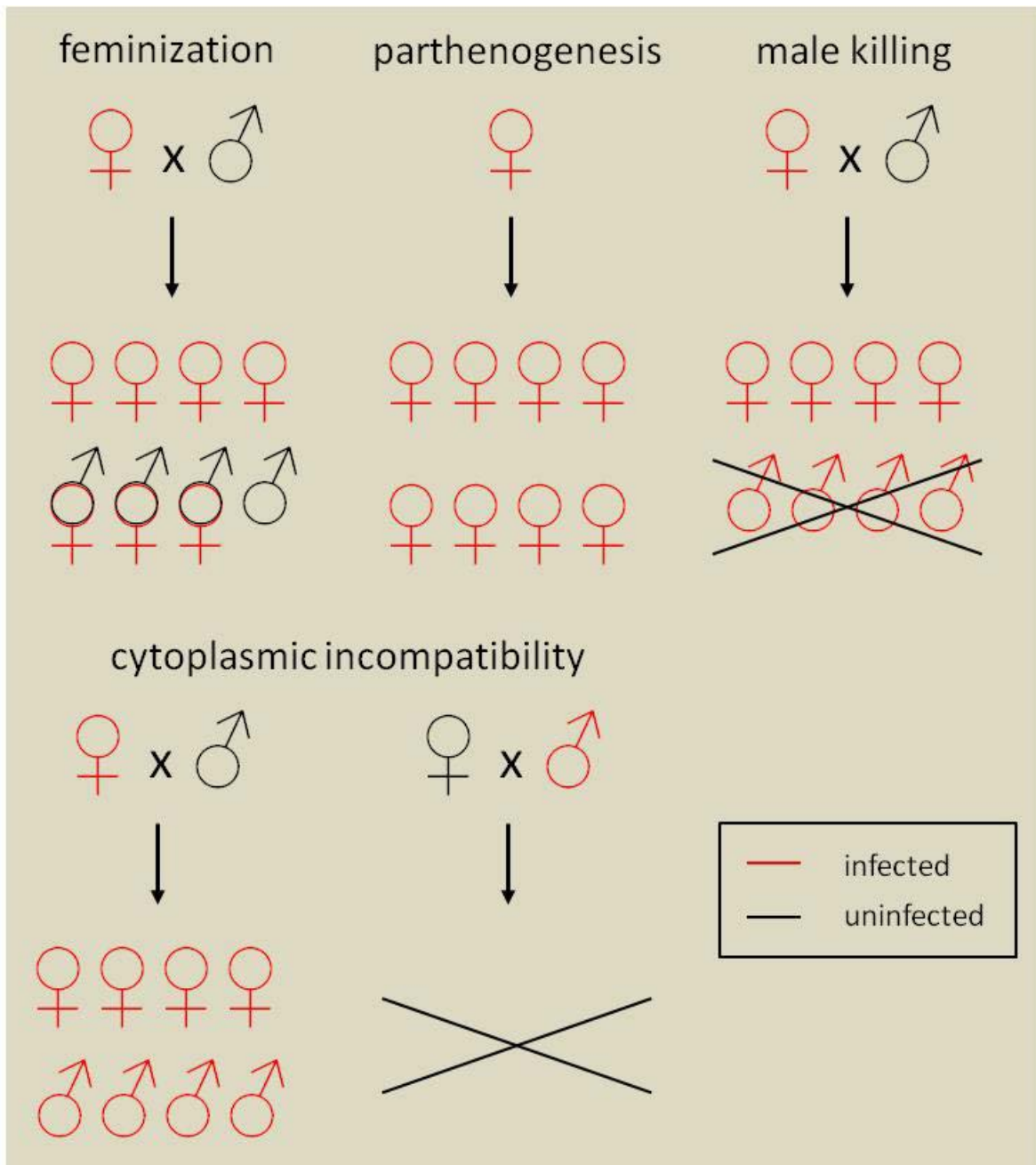


Figure 2: Schematic overview of feminization, parthenogenesis, male killing and cytoplasmic incompatibility. In feminization, genetically male insects develop as females. In parthenogenesis, males are not needed for reproduction at all. Male killing results in the elimination of symbiont-infected males. In cytoplasmic incompatibility, the outcome of crosses between symbiont-infected males and uninfected females is reproductive failure.

The Reproductive Manipulator of Arthropods *Cardinium hertigii*

Phylogenetic analyses place *Cardinium hertigii* like the *Acanthamoeba* endosymbiont *Amoebophilus asiaticus* as sister lineages within the *Bacteroidetes*. *Cardinium* has been found in the arthropod groups Hymenoptera, Hemiptera, Diptera, Protura, Acari and Araneae. It is estimated that 6–7% of all arthropods are infected with these bacteria (Dallai et al, 2011; Duron et al, 2008; Nakamura et al, 2009). Different *Cardinium* strains are able to manipulate their host's reproduction by inducing parthenogenesis and cytoplasmic incompatibility (CI). The CI-inducing *Cardinium* strain *cEper1* is a symbiont of the parasitic wasp *Encarsia pergandiella* (Zchori-Fein et al, 2004). *Cardinium cEper1* is involved in a tripartite symbiosis. The tiny wasp *Encarsia pergandiella* (Figure 3), which has only 1/1000 of the weight of a *Drosophila* spp. (~18 µg), lays its eggs in whiteflies and the wasps' larvae develop and emerge as adults from the whitefly. Interestingly, it was observed that a *Cardinium* strain could have a significantly positive effect on its host, as *Cardinium*-infected white-backed plant hoppers *Sogatella furcifera* reached adulthood earlier than uninfected plant hoppers (Zhang 2012). The reason for the faster development of infected compared to uninfected insects could be the provision of nutrients by the symbiont.



Figure 3: the parasitic wasp *Encarsia pergandiella* laying its eggs in a whitefly nymph and an electron micrograph (Penz et al, 2012) showing *Cardinium*, the endosymbiont of the parasitic wasp.

Evolution and Genomics of Bacterial Endosymbionts

Free-living bacteria have to undergo a number of evolutionary processes to become an endosymbiont. Facultative intracellular bacteria, which are still able to replicate outside their hosts, need to acquire many host-cell interaction factors to survive in an intracellular environment. There are several lines of evidence that simple protozoa, such as amoebae, contribute to the adaption of bacteria to new intracellular environments in higher eukaryotic organisms like insects and mammals (Toft & Andersson, 2010). In amoebae, bacteria learn how to escape from the hosts' immune system and how to modify the host for their needs. Thus, amoebae can be thought of as “training grounds” for intracellular pathogenic bacteria (Molmeret et al, 2005).

In this process, mobile genetic elements such as plasmids, phages or transposons provide very effective means to acquire genes important for host-cell interaction (Toft & Andersson, 2010). In fact, genes important for host adaption tend to be located in genomic islands surrounded by or in the vicinity of these mobile genetic elements (Dobrindt et al, 2004). Many genomic studies show that mobile genetic elements are enriched in the genomes of facultative endosymbionts, which recently became endosymbionts, while mobile genetic elements are almost completely lost in obligate intracellular endosymbionts (Bordenstein & Reznikoff, 2005) that are associated with their hosts for a long time.

As already mentioned, the genomes of endosymbiotic bacteria show some outstanding features (Moran et al, 2008). Genome-based studies revealed that bacterial endosymbionts have the smallest and most-evolved genomes (McCutcheon et al, 2009a; McCutcheon et al, 2009b; McCutcheon & von Dohlen, 2011; Nakabachi et al, 2006). The smallest discovered prokaryotic genome (Table 3) is the genome of the mealy bug endosymbiont *Candidatus Tremblaya princeps* (McCutcheon & von Dohlen, 2011). Most likely, the genome of *Tremblaya* is not even the smallest genome. The theoretically calculated minimal genome size of an intracellular bacterial endosymbiont is approximately 73 kb (McCutcheon & Moran, 2011).

On the evolution to an obligate intracellular endosymbiont, dramatic genomic mutations occur in genes encoding for outer surface structures. Once intracellularly established, endosymbionts tend to lose more and more genes involved in the production of fatty acids, phospholipids, peptidoglycan and the maintenance of cell shape (McCutcheon & Moran, 2011). Genes involved in these processes are under weak selection pressure, because the intracellular symbiont is well protected from the extracellular environment so that outer surface structures are not needed anymore.

Table 3. Bacteria with small genome sizes. obligate intracellular bacteria are shown in orange, the smallest genome of a free-living bacterium in blue and facultative intracellular bacteria in green.

Bacterium	Genome size in bp	GC content in %	Host	Reference
<i>Candidatus Tremblaya princeps</i>	138,927	58.8	<i>Planococcus citri</i> (mealybug)	(McCutcheon & von Dohlen, 2011)
<i>Candidatus Hodgkinia cicadicola</i>	143,795	58.4	<i>Diceroprocta semicineta</i> (cicada)	(McCutcheon et al, 2009b)
<i>Candidatus Carsonella ruddii</i>	159,662	16.5	<i>Pachypsylla venusta</i> (psyllid)	(Nakabachi et al, 2006)
<i>Candidatus Zinderia insecticola</i> CARI	208,564	13.5	<i>Clasoptera arizonana</i> (spittlebug)	(McCutcheon & Moran, 2010)
<i>Candidatus Sulcia muelleri</i> DMIN	243,933	22.5	<i>Draeculacephala minerva</i> (sharpshooter)	(Woyke et al, 2010)
<i>Buchnera aphidicola</i> BCc	416,380	20.1	<i>Cinara cedri</i> (aphid)	(Perez-Brocail et al, 2006)
<i>Mycoplasma genitalium</i> G37	580,076	31,7	human	(Glass et al, 2006)
<i>Cardinium hertigii</i> cEper1	887,130	36.6	<i>Encarsia pergandiella</i> (wasp)	(Penz et al, 2012)
<i>Amoebophilus asiaticus</i> 5a2	1,884,364	35.0	<i>Acanthamoeba</i> spp. (amoeba)	(Schmitz-Esser et al, 2010)

Another characteristic feature of endosymbiontal genomes, with the exceptions of *Candidatus Tremblaya princeps* (McCutcheon & von Dohlen, 2011) and *Candidatus Hodgkinia cicadicola* (McCutcheon et al, 2009b), is the nucleotide base composition bias towards adenine and thymine (A+T) (Table 2). The rapid sequence evolution of these genomes results in accelerated rates of amino acid substitution in all protein-coding genes (Moran et al, 2008; Toft & Andersson, 2010). Evolutionary forces that drive these genomes faster towards this direction are the small population size of endosymbionts and their asexuality (Mira et al, 2001; Moran, 1996). The lack of recombination between endosymbionts of different hosts leads to high levels of genetic drift, inactivation and deletion of genes that are mildly beneficial, but not essential (Moran et al, 2008).

In summary, obligate intracellular symbionts possess the most compact and evolved genomes. These genomes are free of mobile genetic elements and have very reduced biosynthetic capabilities. Mainly genes involved in central cellular processes such as DNA replication, transcription or translation, as well as genes important for the symbiosis remain present (Moran et al, 2008).

Molecular Interactions between Symbiotic Bacteria and their Hosts

Effector Proteins of Symbiotic Bacteria

Symbiotic bacteria interact with their hosts. Intracellular symbionts modify their intracellular host environment via so called effectors. Effectors are macromolecules delivered by the symbiont and are secreted into the environment or into the host cell. On their site of action effector proteins are able to modify the host for the needs of the symbiont. Many of these effectors are bacterial proteins containing typical eukaryotic domains. These eukaryotic domains are important for the interaction with their host cell. Interestingly, there is a significant enrichment of proteins containing eukaryotic domains in the proteome of amoeba-associated bacteria (Schmitz-Esser et al, 2010). The most highly enriched proteins in amoeba-associated bacteria are proteins containing ankyrin- (ANKs), tetratricopeptide- (TPR/Sel1), leucine-rich (LRRs) repeats, as well as F- and U-box proteins. The domains of these proteins are able to mediate protein-protein interactions and are particularly important for the interaction of various intracellular bacterial pathogens with their eukaryotic host cells. Effector proteins containing a protein-protein interaction motif are involved in many cellular processes, such as cytoskeleton integrity, cell cycle control, transcriptional regulation, cell signaling, development and differentiation, apoptosis, cellular scaffolding, vesicular

trafficking, inflammatory response and bacterial invasion (Andrade et al, 2001; Forrer et al, 2003).

Secretion Systems of Symbiotic Bacteria

Many genomes of bacterial symbionts encode an arsenal of effector proteins that are putatively able to modulate the host. But how do these effector proteins get into the host cell? Protein secretion systems are able to modulate these interactions. So far, six secretions systems (type one to type six) have been described in Gram-negative bacteria (Tseng et al, 2009). These bacterial secretion systems span over the two bacterial membranes and are able to translocate effector proteins into the extracellular environment or across the host's plasma membrane into the host's cytosol. Once in the host, these proteins modulate the host-cell interaction. Many of the bacterial secretion systems are encoded in so-called pathogenicity islands. In pathogenic and symbiotic bacteria, pathogenicity islands are usually organized as clusters of genes important for virulence or host-cell interaction. From an evolutionary point of view, some of the bacterial secretion machineries show structural similarities to bacteriophage tails. These bacteriophage-like injection machineries including the type six secretion system (T6SS), R-type pyocins, rhabdosomes and antifeeding prophages are composed of a contractile tail, similar to those of bacteriophages consisting of an outer sheath and an inner tube. Most of these contractile tails are anchored in the bacterial membrane in a structure that is similar to a phage baseplate and possess a VgrG-related protein that is able to puncture holes into membranes of the target organism. For the T6SS it was demonstrated that the contractile tail remains in a ready-to-fire conformation, contracts in the right moment to penetrate the adjacent target cell membrane (Basler et al, 2012) and delivers the bacterial effector proteins into the host cell. In addition to its structural similarities, these contractile

tail-based secretion machineries (T6SS, R-type pyocins, rhabdosomes and antifeeding prophages) share a common evolutionary origin with bacteriophages (Leiman et al, 2009).

Methods and Challenges of Symbiosis Research

Back in the 19th century symbiotic associations could only be described without modern molecular methods. In the mid 20th century symbiotic interactions were described on the basis of a few genes. The development and revolution of modern, molecular high-throughput methods opens many new opportunities to study the molecular interaction between symbiotic partners. Many new high-throughput techniques, such as next generation sequencing platforms, can answer questions regarding these interactions.

One challenge in symbiosis research is the low amount of symbiont material that is gained for molecular experiments. Many host-symbiont systems are not easily accessible, because they are not trivial to cultivate or at least only cultivable with huge efforts.

An example for a system accessible with difficulties is the tripartite symbiosis of the insect endosymbiont *Cardinium cEper1*. The access to the symbiont is not straight forward, because cowpea plants need to be grown first, which serve as food for the whitefly that in turn can be used by the minute parasitoid wasp *Encarsia pergandiella* as nursery ground for their offspring. In contrast to obligate insect endosymbionts that occur usually at high titers in specialized organs called bacteriomes, facultative insect endosymbionts such as *Cardinium* occur only at low titers in various host tissues (Oliver et al, 2010). Depending on the life stage of the host, the genome of the obligate aphid endosymbionts *Buchnera* for instance is present in a ratio of more than 70 per aphid genome (Vogel & Moran, 2011), while the number of genomes of facultative insect endosymbionts are on average 100 times lower than the number of *Buchnera* in the same aphid individual (Wilkinson et al, 2007).

Applied Symbiosis Research

One extremely fascinating example of practical applications in symbiosis research is the controlling of insect pest populations utilizing reproductive manipulators. In field experiments three different groups of insects are needed: (i) reproductive, manipulator-free insect females, (ii) reproductive, manipulator-free insect males and (iii) insect males that are infected with reproductive manipulators, which are able to induce cytoplasmic incompatibility. The presence of these three insect-symbiont combinations is able to drastically reduce or even suppress insect pest populations.

Cage experiments of *Wolbachia*-infected medfly lines containing different ratios of uninfected females to uninfected males to infected males (1:1:0, 1:1:1, 1:1:10, 1:1:20, 1:1:30, 1:1:50) showed that the number of hatched eggs decreased with the increase of infected males in the population. At a ratio of 1:1:50, the number of hatched eggs was almost zero (Zabalou et al, 2004).

The major advantage of this method is that potentially harmful chemicals or pesticides are not needed. Bacteria able to influence the reproduction of its host are already in nature for millions of years. The reproductive manipulators in insect males, which are used to induce the CI phenotype, cannot be transmitted to other insects, because they are strictly maternally transmitted and males are a one-way street. These reproductive manipulators die with the insect male. The environmentally friendly tool of insect pest control was already used to control disease vectors such as the mosquito *Culex pipiens* (Laven, 1967), as well as agricultural pests such as the European cherry fruit fly, *Rhagoletis cerasi* and the almond moth *Cadra (Ephestia) cautella*. (Zabalou et al, 2004).

Not only disease vectors such as Tsetse flies and mosquitoes can be controlled by symbiotic bacteria. There is evidence that the reproductive manipulator *Wolbachia* is able to reduce

transmission of human pathogens, including the dengue virus (DENV) (Walker et al, 2011) and the parasitic protozoa *Plasmodium* (Kambris et al, 2010). In DENV-containing Tsetse flies, *Wolbachia* induces the reactive oxygen species-dependent Toll pathway, which is essential in mediating the expression of antioxidants to counterbalance oxidative stress. The Toll immune pathway is responsible for the activation of antimicrobial peptides such as defensins and cecropins. These antimicrobial peptides are involved in the inhibition of DENV proliferation in *Wolbachia*-infected mosquitoes (Pan et al, 2012).

The interactions of bacterial endosymbionts with their eukaryotic host cells are manifold and have a major impact on biology, ecology and evolution. They open important possibilities and opportunities e.g. as an environment-friendly pest control tool and as a factor that is able to reduce insect-transmitted human diseases.

Outline and Contributions

Chapter I provides a general overview of symbioses and describes different symbioses of bacteria with eukaryotic hosts. It gives insights into the outstanding genomic features of bacterial symbionts, their evolution and the molecular interaction between eukaryotic hosts and bacterial symbionts. It highlights applied examples of symbiosis research and gives an introduction into the challenges of symbiosis research.

Chapter II compares two genomes of cytoplasmic incompatibility-inducing reproductive manipulators: the genome of the *Bacteroidetes Cardinium hertigii* and the *Alphaproteobacterium Wolbachia* sp.. It highlights both, similarities and differences of these two distantly related bacteria and provides a novel comparative context for understanding the

mechanistic basis of cytoplasmic incompatibility. Hence, this study substantially increases our knowledge on reproductive manipulator symbionts.

Penz T, Schmitz-Esser S, Kelly SE, Cass BN, Muller A, Woyke T, Malfatti SA, Hunter MS, Horn M (2012) Comparative genomics suggests an independent origin of cytoplasmic incompatibility in *Cardinium hertigii*. *PLoS Genet* **8**: e1003012

Contributions: I analyzed the next generation sequencing data, conducted control experiments for finishing the draft genome including DNA isolation, PCR primer design and PCR. In addition, I performed the phylogenetic sequence analysis, annotated the genome and wrote the first draft paper.

Chapter III focuses on mobile genetic elements, which are highly enriched in the genome of the *Acanthamoeba* endosymbiont *Amoebophilus asiaticus* (23% of all coding sequences). This chapter describes the mobile genetic elements found in *Amoebophilus* and shows their potential role in genome reduction.

Schmitz-Esser S*, **Penz T***, Spang A, Horn M (2011) A bacterial genome in transition - an exceptional enrichment of IS elements but lack of evidence for recent transposition in the symbiont *Amoebophilus asiaticus*. *BMC Evol Biol* **11**: 270

*both authors contributed equally to this study

***Contributions:** I conducted experiments including cultivation of *Acanthamoeba* hosts and bacterial symbionts, performed DNA and RNA isolations, reverse transcription of RNA, PCR primer design and PCR, Southern hybridizations and analyzed the experimental data.*

***Chapter IV** describes a putatively new secretion system encoded in several genomes of *Bacteroidetes* that are associated with eukaryotic hosts. The description of the putative secretion system is based on genomic sequence analysis.*

Penz T, Horn M, Schmitz-Esser S (2011) The genome of the amoeba symbiont "*Candidatus Amoebophilus asiaticus*" encodes an *afp*-like prophage possibly used for protein secretion. *Virulence* **1**: 541-545

***Contributions:** I performed the sequence analysis including phylogenetic analyses and wrote the paper.*

***Chapter V** provides insights into the expression of a putative novel secretion apparatus during the infection cycle of *Amoebophilus* in its *Acanthamoeba* host. In addition, we provide first molecular insights into the fibril-like structures of the secretion apparatus.*

Penz T, Harreither A, Tsao HF, Aistleitner K, Kostanjsek R, Pilhofer M, Jensen GJ, Schmitz-Esser S, Horn M (2013) Host adaptation of a symbiont: The biphasic life cycle of

Amoebophilus asiaticus and its phage-derived protein secretion system. *Manuscript in preparation.*

Contributions: *I designed the study, as well as PCR and qPCR primers, conducted experimental work, including Acanthamoeba host and symbiont cultivation, symbiont isolation, infection experiments, prophage tail sheath preparations, DNA and RNA isolation, reverse transcription of RNA, fluorescence in situ hybridization, analyzed the data including sequence analysis and phylogenetic analysis and wrote the draft manuscript.*

Chapter VI focuses on the metabolic interactions between the *Bacterioidetes* endosymbiont *Amoebophilus* and its *Acanthamoeba* host. Here we functionally characterized one S-adenosylmethionine (SAM) transport protein of *Amoebophilus*. For the first time, we provide direct evidence for a proton-driven S-adenosylmethionine/S-adenosylhomocysteine antiport mechanism by bacterial SAM-transporters. With this SAM transport protein *Amoebophilus* might compensate for its missing methylation cycle.

Haferkamp I, **Penz T**, Geier M, Ast M, Mushak T, Horn M, Schmitz-Esser S (2013) The endosymbiont *Amoebophilus asiaticus* encodes an S-adenosylmethionine carrier that compensates for its missing methylation cycle. *J Bacteriol.* 2013 Jul;195(14):3183-92

Contributions: *I performed sequence analysis including PCR primer design and conducted experimental work, including DNA and RNA isolation, reverse transcription, expression analysis and cloning of the transport protein.*

Chapter VII provides a concluding discussion about the thesis and highlights its relevance in symbiosis research.

Chapter VIII includes summaries in both, English and German, as well as my curriculum vitae and my acknowledgements.

References

Albert-Weissenberger C, Cazalet C, Buchrieser C (2007) *Legionella pneumophila* - a human pathogen that co-evolved with fresh water protozoa. *Cellular and Molecular Life Sciences* **64**: 432-448

Amann R, Springer N, Schonhuber W, Ludwig W, Schmid EN, Muller KD, Michel R (1997) Obligate intracellular bacterial parasites of acanthamoebae related to *Chlamydia* spp. *Appl Environ Microbiol* **63**: 115-121

Andrade MA, Perez-Iratxeta C, Ponting CP (2001) Protein repeats: structures, functions, and evolution. *J Struct Biol* **134**: 117-131

Bak AL, Black FT, Christiansen C, Freundt EA (1969) Genome size of mycoplasmal DNA. *Nature* **224**: 1209-1210

Basler M, Pilhofer M, Henderson GP, Jensen GJ, Mekalanos JJ (2012) Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* **483**: 182-186

Baumann L, Baumann P (2005) Cospeciation between the primary endosymbionts of mealybugs and their hosts. *Curr Microbiol* **50**: 84-87

Birtles RJ, Rowbotham TJ, Michel R, Pitcher DG, Lascola B, Alexiou-Daniel S, Raoult D (2000) '*Candidatus* *Odyssella thessalonicensis*' gen. nov., sp. nov., an obligate intracellular parasite of *Acanthamoeba* species. *Int J Syst Evol Microbiol* **50 Pt 1**: 63-72

Bordenstein SR, Reznikoff WS (2005) Mobile DNA in obligate intracellular bacteria. *Nat Rev Microbiol* **3**: 688-699

Buchner P (1965) *Endosymbiosis of animals with plant microorganisms*, Rev. Eng. edn. New York,; Interscience Publishers.

Burke GR, McLaughlin HJ, Simon JC, Moran NA (2010) Dynamics of a Recurrent *Buchnera* Mutation That Affects Thermal Tolerance of Pea Aphid Hosts. *Genetics* **186**: 367-U577

Dallai R, Mercati D, Giusti F, Gottardo M, Carapelli A (2011) A *Cardinium*-like symbiont in the proturan *Acerella muscorum* (Hexapoda). *Tissue Cell* **43**: 151-156

Dobrindt U, Hochhut B, Hentschel U, Hacker J (2004) Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* **2**: 414-424

Douglas AE (1996) Reproductive failure and the free amino acid pools in pea aphids (*Acyrtosiphon pisum*) lacking symbiotic bacteria. *J Insect Physiol* **42**: 247-255

Douglas AE (2006) Phloem-sap feeding by animals: problems and solutions. *J Exp Bot* **57**: 747-754

Duron O, Hurst GD, Hornett EA, Josling JA, Engelstadter J (2008) High incidence of the maternally inherited bacterium *Cardinium* in spiders. *Mol Ecol* **17**: 1427-1437

Forrer P, Stumpp MT, Binz HK, Pluckthun A (2003) A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. *FEBS Lett* **539**: 2-6

Gilbert SF, Sapp J, Tauber AI (2012) A symbiotic view of life: we have never been individuals. *Q Rev Biol* **87**: 325-341

Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA, 3rd, Smith HO, Venter JC (2006) Essential genes of a minimal bacterium. *Proc Natl Acad Sci U S A* **103**: 425-430

Greub G, Raoult D (2004) Microorganisms resistant to free-living amoebae. *Clin Microbiol Rev* **17**: 413-433

Gruwell ME, Morse GE, Normark BB (2007) Phylogenetic congruence of armored scale insects (Hemiptera : Diaspididae) and their primary endosymbionts from the phylum *Bacteroidetes*. *Molecular Phylogenetics and Evolution* **44**: 267-280

Horn M (2008) *Chlamydiae* as symbionts in eukaryotes. *Annu Rev Microbiol* **62**: 113-131

Horn M, Fritsche TR, Gautom RK, Schleifer KH, Wagner M (1999) Novel bacterial endosymbionts of *Acanthamoeba* spp. related to the *Paramecium caudatum* symbiont *Caedibacter caryophilus*. *Environ Microbiol* **1**: 357-367

Horn M, Fritsche TR, Linner T, Gautom RK, Harzenetter MD, Wagner M (2002) Obligate bacterial endosymbionts of *Acanthamoeba* spp. related to the *beta-Proteobacteria*: proposal of '*Candidatus Procabacter acanthamoebae*' gen. nov., sp. nov. *Int J Syst Evol Microbiol* **52**: 599-605

Horn M, Harzenetter MD, Linner T, Schmid EN, Muller KD, Michel R, Wagner M (2001) Members of the *Cytophaga-Flavobacterium-Bacteroides* phylum as intracellular bacteria of acanthamoebae: proposal of '*Candidatus Amoebophilus asiaticus*'. *Environ Microbiol* **3**: 440-449

Horn M, Wagner M (2004) Bacterial endosymbionts of free-living amoebae. *J Eukaryot Microbiol* **51**: 509-514

Houk EJ, Griffiths GW (1980) Intracellular Symbiotes of the Homoptera. *Annu Rev Entomol* **25**: 161-187

Hunter MS, Perlman SJ, Kelly SE (2003) A bacterial symbiont in the *Bacteroidetes* induces cytoplasmic incompatibility in the parasitoid wasp *Encarsia pergandiella*. *Proc Biol Sci* **270**: 2185-2190

Kambris Z, Blagborough AM, Pinto SB, Blagrove MS, Godfray HC, Sinden RE, Sinkins SP (2010) Wolbachia stimulates immune gene expression and inhibits plasmodium development in *Anopheles gambiae*. *PLoS Pathog* **6**: e1001143

Khan NA (2006) *Acanthamoeba*: biology and increasing importance in human health. *FEMS Microbiol Rev* **30**: 564-595

Laven H (1967) Eradication of *Culex pipiens fatigans* through cytoplasmic incompatibility. *Nature* **216**: 383-384

Leiman PG, Basler M, Ramagopal UA, Bonanno JB, Sauder JM, Pukatzki S, Burley SK, Almo SC, Mekalanos JJ (2009) Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc Natl Acad Sci U S A* **106**: 4154-4159

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*, New Haven,: Yale University Press.

McCutcheon JP, McDonald BR, Moran NA (2009a) Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc Natl Acad Sci U S A* **106**: 15394-15399

McCutcheon JP, McDonald BR, Moran NA (2009b) Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet* **5**: e1000565

McCutcheon JP, Moran NA (2010) Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biol Evol* **2**: 708-718

McCutcheon JP, Moran NA (2011) Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* **10**: 13-26

McCutcheon JP, von Dohlen CD (2011) An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr Biol* **21**: 1366-1372

Mira A, Ochman H, Moran NA (2001) Deletional bias and the evolution of bacterial genomes. *Trends Genet* **17**: 589-596

Molmeret M, Horn M, Wagner M, Santic M, Abu Kwaik Y (2005) Amoebae as training grounds for intracellular bacterial pathogens. *Appl Environ Microbiol* **71**: 20-28

Moran NA (1996) Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A* **93**: 2873-2878

Moran NA (2006) Symbiosis. *Curr Biol* **16**: R866-871

Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* **42**: 165-190

Moran NA, Tran P, Gerardo NM (2005) Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the bacterial phylum *Bacteroidetes*. *Appl Environ Microbiol* **71**: 8802-8810

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

Nakamura Y, Kawai S, Yukuhiro F, Ito S, Gotoh T, Kisimoto R, Yanase T, Matsumoto Y, Kageyama D, Noda H (2009) Prevalence of *Cardinium* Bacteria in Planthoppers and Spider Mites and Taxonomic Revision of "*Candidatus Cardinium hertigii*" Based on Detection of a New *Cardinium* Group from Biting Midges. *Applied and Environmental Microbiology* **75**: 6757-6763

Noel GR, Atibalentja N (2006) '*Candidatus Paenicardinium endonii*', an endosymbiont of the plant-parasitic nematode *Heterodera glycines* (Nemata: *Tylenchida*), affiliated to the phylum *Bacteroidetes*. *Int J Syst Evol Microbiol* **56**: 1697-1702

Oliver KM, Degnan PH, Burke GR, Moran NA (2010) Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annu Rev Entomol* **55**: 247-266

Oliver KM, Russell JA, Moran NA, Hunter MS (2003) Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proc Natl Acad Sci U S A* **100**: 1803-1807

Pan X, Zhou G, Wu J, Bian G, Lu P, Raikhel AS, Xi Z (2012) Wolbachia induces reactive oxygen species (ROS)-dependent activation of the Toll pathway to control dengue virus in the mosquito *Aedes aegypti*. *Proc Natl Acad Sci U S A* **109**: E23-31

Penz T, Schmitz-Esser S, Kelly SE, Cass BN, Muller A, Woyke T, Malfatti SA, Hunter MS, Horn M (2012) Comparative genomics suggests an independent origin of cytoplasmic incompatibility in *Cardinium hertigii*. *PLoS Genet* **8**: e1003012

Perez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ, Moya A, Latorre A (2006) A small microbial genome: the end of a long symbiotic relationship? *Science* **314**: 312-313

Raven PH, Yeates DK (2007) Australian biodiversity: threats for the present, opportunities for the future. *Aust J Entomol* **46**: 177-187

Rinke C, Schmitz-Esser S, Stoecker K, Nussbaumer AD, Molnar DA, Vanura K, Wagner M, Horn M, Ott JA, Bright M (2006) "Candidatus Thiobios zoothamnicoli," an ectosymbiotic bacterium covering the giant marine ciliate *Zoothamnium niveum*. *Appl Environ Microbiol* **72**: 2014-2021

Rodriguez-Zaragoza S (1994) Ecology of free-living amoebae. *Crit Rev Microbiol* **20**: 225-241

Schmitz-Esser S, Tischler P, Arnold R, Montanaro J, Wagner M, Rattei T, Horn M (2010) The genome of the amoeba symbiont "Candidatus Amoebophilus asiaticus" reveals common mechanisms for host cell interaction among amoeba-associated bacteria. *J Bacteriol* **192**: 1045-1057

Sunagawa S, Woodley CM, Medina M (2010) Threatened corals provide underexplored microbial habitats. *PLoS One* **5**: e9554

Thomas V, McDonnell G (2007) Relationship between mycobacteria and amoebae: ecological and epidemiological concerns. *Letters in Applied Microbiology* **45**: 349-357

Toft C, Andersson SG (2010) Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat Rev Genet* **11**: 465-475

Tseng TT, Tyler BM, Setubal JC (2009) Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *Bmc Microbiol* **9**

Vogel KJ, Moran NA (2011) Sources of variation in dietary requirements in an obligate nutritional symbiosis. *Proc Biol Sci* **278**: 115-121

Walker T, Johnson PH, Moreira LA, Iturbe-Ormaetxe I, Frentiu FD, McMeniman CJ, Leong YS, Dong Y, Axford J, Kriesner P, Lloyd AL, Ritchie SA, O'Neill SL, Hoffmann AA (2011) The *wMel* *Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations. *Nature* **476**: 450-453

Werren JH, Baldo L, Clark ME (2008) *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol* **6**: 741-751

Wilkinson TL, Koga R, Fukatsu T (2007) Role of host nutrition in symbiont regulation: impact of dietary nitrogen on proliferation of obligate and facultative bacterial endosymbionts of the pea aphid *Acyrtosiphon pisum*. *Appl Environ Microbiol* **73**: 1362-1366

Woyke T, Tighe D, Mavromatis K, Clum A, Copeland A, Schackwitz W, Lapidus A, Wu D, McCutcheon JP, McDonald BR, Moran NA, Bristow J, Cheng JF (2010) One bacterial cell, one complete genome. *PLoS One* **5**: e10314

Zabalou S, Riegler M, Theodorakopoulou M, Stauffer C, Savakis C, Bourtzis K (2004) *Wolbachia*-induced cytoplasmic incompatibility as a means for insect pest population control. *Proc Natl Acad Sci U S A* **101**: 15042-15045

Zchori-Fein E, Gottlieb Y, Kelly SE, Brown JK, Wilson JM, Karr TL, Hunter MS (2001) A newly discovered bacterium associated with parthenogenesis and a change in host selection behavior in parasitoid wasps. *Proc Natl Acad Sci U S A* **98**: 12555-12560

Zchori-Fein E, Perlman SJ, Kelly SE, Katzir N, Hunter MS (2004) Characterization of a 'Bacteroidetes' symbiont in *Encarsia* wasps (Hymenoptera: Aphelinidae): proposal of 'Candidatus Cardinium hertigii'. *Int J Syst Evol Microbiol* **54**: 961-968

Chapter II

Comparative genomics suggests an independent origin of cytoplasmic incompatibility in *Cardinium hertigii*

published in PLoS Genet 8(10): e1003012. PMID: 23133394

Comparative Genomics Suggests an Independent Origin of Cytoplasmic Incompatibility in *Cardinium hertigii*

Thomas Penz¹*, Stephan Schmitz-Esser^{1,2}*, Suzanne E. Kelly³, Bodil N. Cass⁴, Anneliese Müller², Tanja Woyke⁵, Stephanie A. Malfatti⁵, Martha S. Hunter^{3*}, Matthias Horn^{1*}

1 Department of Microbial Ecology, University of Vienna, Vienna, Austria, **2** Institute for Milk Hygiene, University of Veterinary Medicine Vienna, Vienna, Austria, **3** Department of Entomology, The University of Arizona, Tucson, Arizona, United States of America, **4** Graduate Interdisciplinary Program in Entomology and Insect Science, The University of Arizona, Tucson, Arizona, United States of America, **5** U.S. Department of Energy Joint Genome Institute, Walnut Creek, California, United States of America

Abstract

Terrestrial arthropods are commonly infected with maternally inherited bacterial symbionts that cause cytoplasmic incompatibility (CI). In CI, the outcome of crosses between symbiont-infected males and uninfected females is reproductive failure, increasing the relative fitness of infected females and leading to spread of the symbiont in the host population. CI symbionts have profound impacts on host genetic structure and ecology and may lead to speciation and the rapid evolution of sex determination systems. *Cardinium hertigii*, a member of the *Bacteroidetes* and symbiont of the parasitic wasp *Encarsia pergandiella*, is the only known bacterium other than the *Alphaproteobacteria Wolbachia* to cause CI. Here we report the genome sequence of *Cardinium hertigii* cEper1. Comparison with the genomes of CI-inducing *Wolbachia pipientis* strains wMel, wRi, and wPip provides a unique opportunity to pinpoint shared proteins mediating host cell interaction, including some candidate proteins for CI that have not previously been investigated. The genome of *Cardinium* lacks all major biosynthetic pathways but harbors a complete biotin biosynthesis pathway, suggesting a potential role for *Cardinium* in host nutrition. *Cardinium* lacks known protein secretion systems but encodes a putative phage-derived secretion system distantly related to the antifeeding prophage of the entomopathogen *Serratia entomophila*. Lastly, while *Cardinium* and *Wolbachia* genomes show only a functional overlap of proteins, they show no evidence of laterally transferred elements that would suggest common ancestry of CI in both lineages. Instead, comparative genomics suggests an independent evolution of CI in *Cardinium* and *Wolbachia* and provides a novel context for understanding the mechanistic basis of CI.

Citation: Penz T, Schmitz-Esser S, Kelly SE, Cass BN, Müller A, et al. (2012) Comparative Genomics Suggests an Independent Origin of Cytoplasmic Incompatibility in *Cardinium hertigii*. PLoS Genet 8(10): e1003012. doi:10.1371/journal.pgen.1003012

Editor: Nancy A. Moran, Yale University, United States of America

Received: June 20, 2012; **Accepted:** August 22, 2012; **Published:** October 25, 2012

Copyright: © 2012 Penz et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by Austrian Science Fund (FWF) grants Y277-B03 and P22703-B17 to MH and SS-E, respectively, and by National Science Foundation grant DEB-1020460 and USDA AFRI 2010-03752 to MSH. The genome sequencing effort was conducted by the U.S. Department of Energy Joint Genome Institute (Community Sequencing Program project no. 776895) and supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Additional support from the European Cooperation in Science and Technology (COST) Action FA0701 Arthropod Symbiosis is acknowledged. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mhunter@ag.arizona.edu (MSH); horn@microbial-ecology.net (MH)

☞ These authors contributed equally to this work.

Introduction

Bacterial symbionts of terrestrial arthropods are common, influential associates, known to affect fundamental aspects of the host life history, ecology, and evolution. These maternally inherited bacteria may, for example, provide essential nutrients supplementing their host's diet, confer protection against natural enemies, increase stress resistance, or influence host plant suitability [1–4]. Others have evolved sophisticated means of manipulating the arthropod host's reproduction in ways that cause the symbiont to spread within the host population [5–6]. Infection with reproductive manipulators may drive rapid evolution of host sex determination [7], affect genetic population structure, including reproductive isolation and speciation [8], as well as influence the evolution of sexual traits [9]. Reproductive manipulator symbionts may also be powerful tools in pest management for suppression or transformation of pest or vector populations [10–11].

The most common symbiont-induced reproductive manipulation, cytoplasmic incompatibility (CI), is also perhaps the most enigmatic. CI occurs, in the simplest case, when a symbiont-infected male host mates with an uninfected female. Affected host embryos die in early development. The symbiont spreads because of the decreased fitness of uninfected relative to infected female hosts [5]. The CI manipulation has been studied most extensively in *Wolbachia pipientis*, a member of the *Alphaproteobacteria* established in as many as 40% of terrestrial arthropod species [12] and in filarial nematodes [13]. The verbal model that best describes CI has been termed “modification/rescue” [14], where a factor that is important for the normal development of the insect embryo is modified in sperm cells and can be rescued only if a related strain is present in the eggs. In the fertilized oocyte of an incompatible mating of *Drosophila* or the parasitic wasp *Nasonia vitripennis*, CI *Wolbachia* leads to asynchrony of the timing of maternal and paternal chromosome condensation and segregation during the first embryonic mitotic division, disrupting

Author Summary

Many arthropods are infected with bacterial symbionts that are maternally transmitted and have a great impact on their hosts' biology, ecology, and evolution. One of the most common phenotypes of facultative symbionts appears to be cytoplasmic incompatibility (CI), a type of reproductive failure in which bacteria in males modify sperm in a way that reduces the reproductive success of uninfected female mates. In spite of considerable interest, the genetic basis for CI is largely unknown. *Cardinium hertigii*, a symbiont of tiny parasitic wasps, is the only bacterial group other than the well-studied *Wolbachia* that is known to cause CI. Analysis of the *Cardinium* genome indicates that CI evolved independently in *Wolbachia* and *Cardinium*. However, a suite of shared proteins was likely involved in mediating host cell interactions, and CI shows functional overlap in both lineages. Our analysis suggests the presence of an unusual phage-derived, putative secretion system and reveals that *Cardinium* encodes biosynthetic pathways that suggest a potential role in host nutrition. Our findings provide a novel comparative context for understanding the mechanistic basis of CI and substantially increase our knowledge on reproductive manipulator symbionts that do not only severely affect population genetic structure of arthropods but may also serve as powerful tools in pest management.

embryonic development [15–16]. However, the molecular basis of CI in this uncultivable microbe remains largely unknown [5].

Genome analysis and expression studies of genes of diverse CI *Wolbachia* strains have revealed a number of genes with a potential role in CI [17–22], but our inability to cultivate these bacteria in a host-free environment, the lack of methods to genetically manipulate *Wolbachia*, and the absence of an independently evolved CI lineage with which to make comparisons has limited the progress in this area. Here we describe the genome of the only CI-inducing symbiont known that is distantly related to *Wolbachia*. *Cardinium hertigii* is a member of the *Bacteroidetes*, and the strain *cEper1* infecting the parasitic wasp *Encarsia pergandiella* causes CI [23]. The tiny parasitic wasp host (~18 µg, 1/1000 of the weight of *Drosophila* spp.) lays eggs in whiteflies, and larval wasps develop at the whiteflies' expense, emerging as adults from the whitefly remains. Related *Cardinium* strains have also been found in the arthropod groups Hymenoptera, Hemiptera, Diptera, Protura, Acari and Araneae, and an estimated 6–7% of all arthropods are infected with these bacteria [24–26]. The most recent analysis also places the nematode symbiont 'Candidatus Paenicardinium endonii' within the *Cardinium* clade, and *Cardinium* as sister group to the *Acanthamoeba* endosymbiont *Amoebophilus asiaticus* [26–28]. The *Cardinium/Amoebophilus* clade is only distantly related to other known insect symbiont lineages within the *Bacteroidetes*.

The genome sequence of *Cardinium hertigii cEper1* reveals a highly reduced genome, both in terms of genome size and metabolic pathways, and a 58 kb cryptic plasmid. *Cardinium* encodes a set of proteins with the potential to interfere with eukaryotic cell cycle regulation. These proteins, some of which also occur in CI-inducing *Wolbachia* strains, are good candidates for effectors mediating CI. Despite its metabolically restricted genome, *Cardinium* encodes a complete biotin biosynthesis pathway, which suggests a potential role of *Cardinium* in host nutrition. Lastly, several lines of evidence suggest that protists have served as hosts for the progenitor of *Cardinium* before its adaptation to insects.

Results/Discussion

A highly reduced genome with features of both facultative symbionts and obligate nutritional symbionts of arthropods

The genome of *Cardinium hertigii cEper1* consists of a single 887 kb chromosome and a 58 kb plasmid (pCher), with 841 protein coding genes (CDS) (Figure 1, Table 1). It is thus not only smaller than the genomes of free-living bacteria but also reduced compared to the genomes of the CI-inducing *Wolbachia* strains *wMel*, *wRi*, and *wPip* (1.27–1.48 Mb; [20–22]). The size of the *Cardinium* genome is actually closer in size to the described genomes of obligate (mutualist) symbionts of diverse insect hosts, which are typically highly reduced and range from 140 kb to 790 kb (Table S1) [29–30]. Other genomic features of *Cardinium* such as a low G+C content (36.6%) and a single (unlinked) set of rRNA genes are also common characteristics of intracellular bacterial symbionts. *Cardinium* differs from obligate symbionts in its abundance of transposable genetic elements (n = 104; 12.4% of all CDSs; Table S2), a feature more typical of facultative symbionts, which generally show a broader host range than obligate symbionts and are not required for host reproduction [29,31]. In addition, while some obligate insect symbionts harbor small plasmids [32], *Cardinium* possesses a large cryptic plasmid. pCher contains 65 CDSs, most of which code for transposases and proteins with unknown function (Figure 1, Table 1). Plasmids of similar size have been reported from several rickettsial symbionts infecting arthropods [33–35].

The representation of functional categories in the *Cardinium* genome based on the assignment of CDSs to NCBI clusters of orthologous genes (COGs, [36]) is similar to that of other endosymbionts with small genomes (Figure S1). For example, the gene set required for DNA repair and recombination is similarly reduced as in other facultative symbionts. While several proteins involved in recombination are not encoded (RecBCD, RecF, RecN, RecR), *Cardinium* has retained RecA, which is missing in most obligate symbionts [32]. The presence of this and other important components suggests that homologous recombination is still possible in *Cardinium*. The biosynthetic capabilities of *Cardinium* are very limited, similar to other intracellular insect symbionts and *Cardinium's* closest sequenced relative, *Amoebophilus* [37]. *Cardinium* is not able to synthesize most cofactors or any amino acids or nucleotides *de novo*. The tricarboxylic acid cycle is missing completely; an F-type ATPase is present but other components of a respiratory chain are lacking. Only the pay-off phase of glycolysis for the generation of ATP and NADH is present (Table S3, Figure 2). To compensate for its reduced metabolic capabilities *Cardinium* encodes 60 transport proteins (Table S4), facilitating the uptake of oligopeptides and amino acids via an oligopeptide transport system Opp A-F (CAHE_0240-0242, 0244 and 0245), ATP and other nucleotides via nucleotide transport proteins (CAHE_0018, 0158, 0160 and 0789), dicarboxylates via a C4-dicarboxylate transporter DcuAB (CAHE_0645 and 0647), and S-adenosylmethionine via an S-adenosylmethionine transporter (CAHE_0109), among others. Clearly, *Cardinium* is highly dependent on its intracellular environment and gains most key metabolites and energy in the form of ATP from its eukaryotic host cell.

Potential role of retained biosynthetic pathways in host nutrition

Virtually the only complete biosynthetic pathways in the *Cardinium* genome are those for lipoate and biotin (Figure 2, Table S3). Lipoate is a highly conserved sulfur-containing cofactor

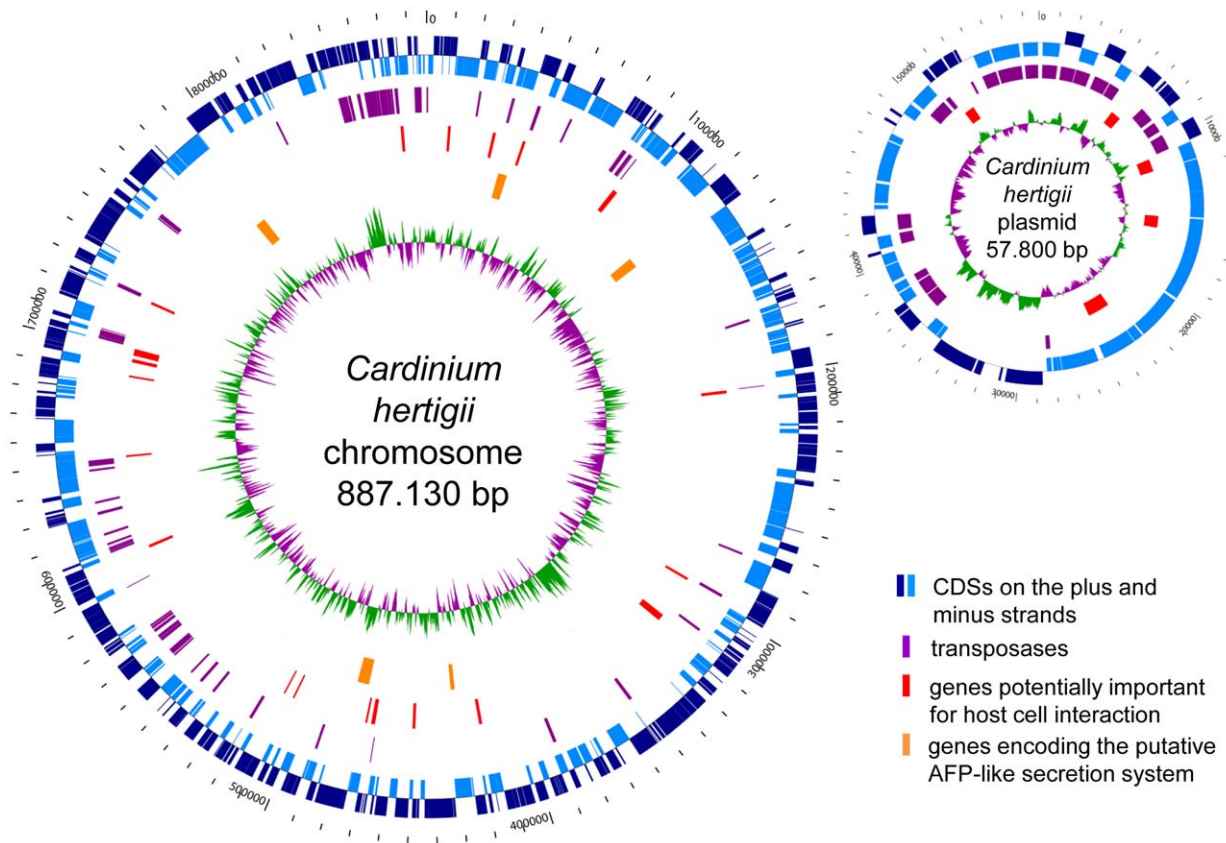


Figure 1. Circular maps of the *Cardinium hertigii* cEper1 chromosome and plasmid pCher. The distribution of protein coding genes (CDSs), mobile genetic transposases, genes potentially important for host cell interaction including ankyrin repeat containing proteins, tetratricopeptide repeat containing proteins and others, and the genes encoding the putative antifeeding prophage-derived secretion system is shown. The innermost green and violet circles represent the GC-skew (purple: below average, green: above average). doi:10.1371/journal.pgen.1003012.g001

involved in oxidative reactions, and also associated with pathogenesis and virulence of microbial pathogens [38]. Biotin is important for carboxylation reactions and cannot be synthesized by many multicellular eukaryotes, including insects. This B-vitamin is thus an indispensable nutritional factor for insect growth and metamorphosis [39]. Vertebrate blood is deficient in B-vitamins and a complete biotin pathway is also present in the genome of a

number of symbionts of blood-feeding hosts including the tsetse fly endosymbiont *Wigglesworthia* and the tick-associated *Ehrlichia*, *Anaplasma*, and *Rickettsia* species [35,40–41]. It was also experimentally shown that the *Wolbachia* strain of the bedbug *Cimex lectularius* supplies various B-vitamins, including biotin, to compensate for the lack of these compounds in their insect host's blood diet [42]. The presence of the biotin pathway in *Cardinium* cEper1

Table 1. General features of the genome of *Cardinium hertigii* cEper1 and its closest sequenced relative *Amoebophilus asiaticus* 5a2.

	<i>Cardinium hertigii</i> cEper1		<i>Amoebophilus asiaticus</i> 5a2
	chromosome	plasmid pCher	chromosome
size (bp)	887,130*	57,800	1,884,364
GC content (%)	36.6	31.5	35.0
CDS	841	65	1557
average CDS length (bp)	911	733	990
coding density (%)	85.5	82.1	81.8
rRNA gene set	1	-	1
tRNA genes	37	-	35
reference	this study	this study	[37]

The genome sequence of *Cardinium* contains a single gap that could not be closed due to repetitive elements (*). doi:10.1371/journal.pgen.1003012.t001

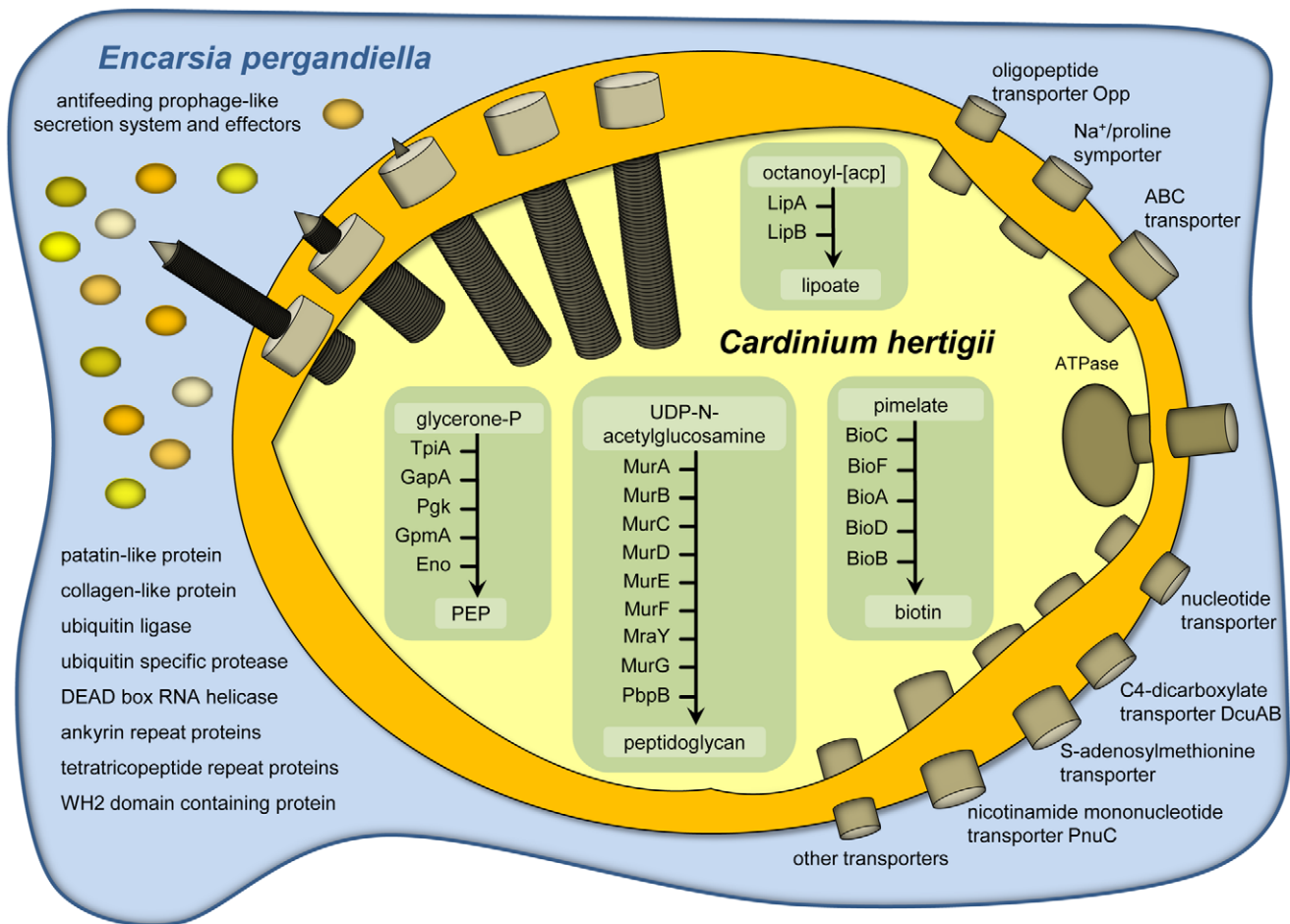


Figure 2. Metabolism, transport capabilities, and host cell interaction of *Cardinium hertigi* cEper1. All predicted complete metabolic pathways and major transport proteins encoded on the genome are indicated. *Cardinium* lacks most biosynthetic pathways and imports nearly all essential metabolites from its host cell by employing a variety of transport proteins. Host cell interaction is mediated by secretion of effector proteins although no evidence for known protein secretion systems was found in the genome. A putative antifeeding prophage-derived secretion system could be used for translocation of proteins directly into the insect host cell by a contraction mechanism similar to type VI secretion systems [103]. doi:10.1371/journal.pgen.1003012.g002

despite of the lack or truncation of almost all other metabolic pathways is puzzling given the hosts' predaceous larval lifestyle, and that antibiotic curing of *Cardinium* does not lead to obvious fitness deficits in its host [23]. This does not rule out a possible benefit of supplemental B-vitamin provision that could partially compensate for what appear to be moderately severe fecundity costs (~15%) to *Cardinium* infection [43]. It appears reasonably common for facultative, reproductive manipulator symbionts to simultaneously confer host fitness benefits [44–45]. On the other hand biotin is also essential for bacteria, and in the absence of alternative sources this pathway might be equally beneficial for *Cardinium* and its host.

A putative phage-derived protein secretion system

While many obligate symbionts of insects lack dedicated protein secretion systems, several facultative symbionts, including *Wolbachia* and *Rickettsia* species, *Hamiltonella defensa* and *Sodalis glossinidius* encode protein secretion systems well known from pathogenic microbes [46–48]. In *Wolbachia*, a type four (IV) secretion system is likely involved in mediating CI or other effects on their insect hosts [48–49]. No known protein secretion system is present in the genome of *Cardinium*, but we identified 16 genes arranged in five different genome regions that show highest similarity to antifeeding

prophage (AFP)-like genes recently identified in *Amoebophilus* (amino acid sequence identity between 24% and 76%; E-value $\leq 1e^{-10}$; Figure 3C; Tables S5, S6) [50]. These AFP-like genes are somewhat similar to the putative defective prophage of the entomopathogen *Serratia entomophila*, which delivers toxins into the hemocytes of its insect host [51]. AFP-like genes are encoded also in other *Bacteroidetes* [52], with the phage tail sheath protein SCFP from the algicidal bacterium *Saprospira* sp. being one of the few characterized components. This protein forms characteristic cytoplasmic fibril structures in *Saprospira* [53]. Interestingly, transmission electron microscopy shows similar subcellular structures in *Cardinium* (Figure 3A, 3B) [54–56], suggesting the presence of an intact protein secretion system encoded by the AFP-like genes. The *Cardinium* AFP gene cluster lacks putative toxins that are the substrates of the *Serratia* and *Photothabdus* AFPs. Instead, the AFP-like genes of *Cardinium* may encode a more general secretion system for proteins that are important for manipulation of the insect host cell, taking over the function of the type IV secretion system found in other reproductive manipulators such as *Wolbachia*. We were able to detect by PCR the three most highly conserved AFP-like genes (CAHE_0458, 0763, 0760) in four other *Cardinium* strains from three different *Encarsia* host species (Figure 3C, Table S7), suggesting that AFP-like genes are conserved among *Cardinium* strains displaying

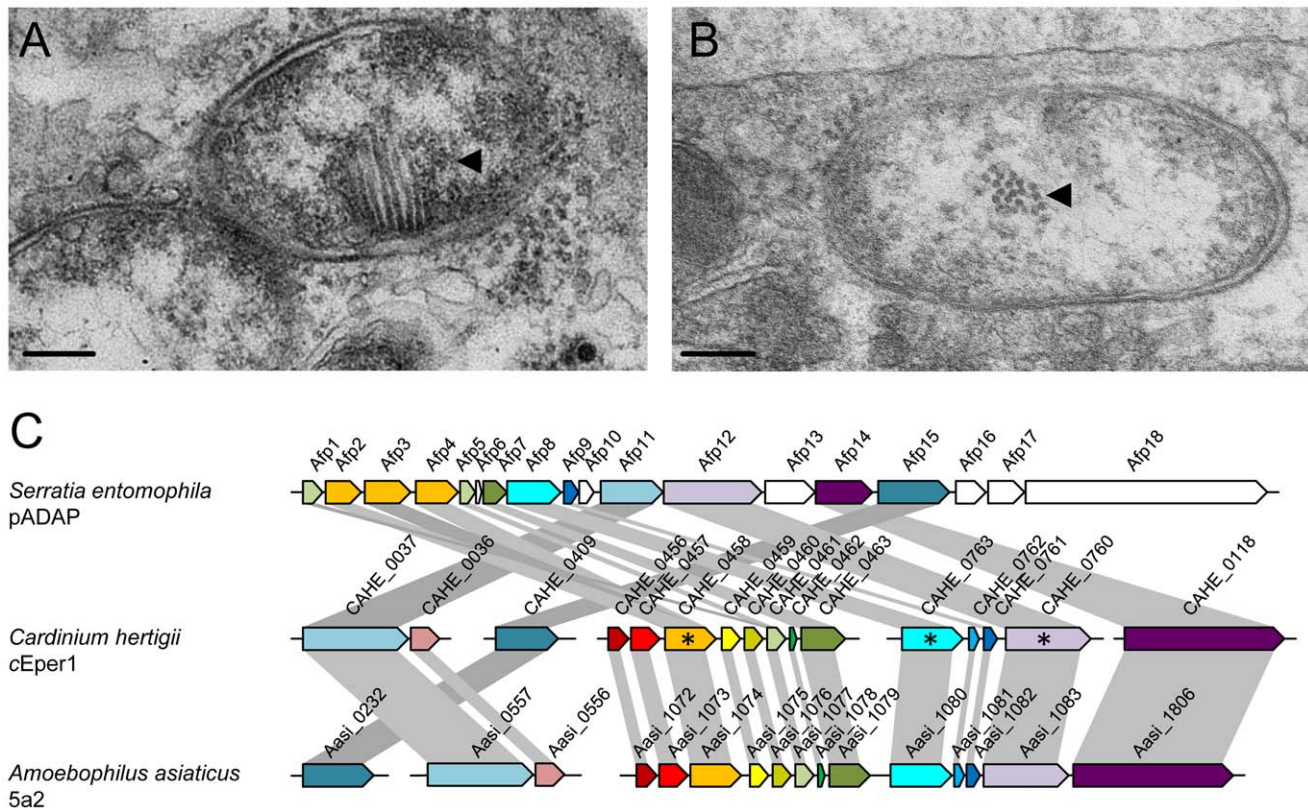


Figure 3. The putative phage derived protein secretion system of *Cardinium hertigi* cEper1. Electron micrographs showing *Cardinium* in *Encarsia pergandiella* ovaries within a nurse cell (A) and a follicle cell (B), respectively. Arrows point to the antifeeding prophage (AFP) like fibril structures in longitudinal view (A) and cross section (B) representing the putative secretion system for translocation of effector proteins into the host cells; bars, 200 nm. (C) A schematic representation of the genomic organization of the AFP-like gene cluster of *Cardinium* compared to those of *Serratia entomophila* and *Amoebophilus asiaticus*. Locus tags and gene names are indicated. Homologous proteins are shown in the same color and connected with grey bars. Genes labeled with an asterisk are conserved among five different *Cardinium* strains tested by PCR (Tables S5, S6, S7). doi:10.1371/journal.pgen.1003012.g003

different phenotypes and likely serve an important function. Our hypothesis of a phage-derived protein secretion system in *Cardinium* parallels the finding that the type six (VI) secretion system shares a common origin with phage tail-associated protein complexes [57–58].

Candidate proteins for CI, host cell interaction, and host cell modulation

Typically, bacterial proteins for host cell interaction contain domains that are known to function in the context of a eukaryotic cell [59], including tetratricopeptide repeats (TPR), ankyrin repeats (ANK), leucine-rich repeats, and F- and U-box domains. Several *Cardinium* proteins contain characteristic TPR and ANK eukaryotic protein-protein interaction motifs (Table S8). In eukaryotic cells TPRs are often associated with multiprotein complexes and play important roles in the functioning of chaperones, transcription and protein transport complexes [60]. Proteins containing TPRs are also involved in the regulation of the eukaryotic cell cycle as components of the anaphase promoting complex (APC), a multi-subunit E3 ubiquitin ligase [61]. Proteins containing TPRs are also present in *Amoebophilus* and in CI-inducing *Wolbachia* strains, as well as in the mutualistic nematode-associated *Wolbachia* strain *wBm* [62].

ANK proteins play important roles in a variety of cellular processes in eukaryotes such as cell cycle regulation, cytoskeleton regulation, developmental and transcriptional regulation [63]. For

example, the ANK protein PLUTONIUM has an important role in the regulation of DNA replication in early *Drosophila* development [64]. ANK proteins are also known from pathogenic intracellular bacteria such as *Legionella pneumophila*, *Anaplasma phagocytophilum*, and *Coxiella burnetii*, which use type IV secretion systems to translocate these bacterial effectors into their eukaryotic host cells [65–66]. Notably, among bacteria, *Amoebophilus* and CI-inducing *Wolbachia* strains encode the largest number of ANK proteins (54 ANK proteins in *Amoebophilus*, 60 in *wPip*, 35 in *wRi*, and 23 in *wMel*), and, while ANK proteins are virtually absent in other sequenced *Bacteroidetes* genomes and the mutualist *Wolbachia* strain *wBm* (five ANK proteins; [62]) *Cardinium* encodes 19 ANK proteins (14 encoded on the chromosome, five on the plasmid pCher). This overrepresentation of ANK proteins in CI-inducing but only distantly related *Cardinium* and *Wolbachia* strains suggests that this class of proteins comprises important mediators of host cell interaction possibly involved in CI. Indeed, it has been frequently suggested earlier that ANK proteins could play a role in *Wolbachia* CI [22,67], although the evidence has been equivocal [19,68].

Cardinium encodes a DEAD box RNA helicase (CAHE_0677). Eukaryotic homologs of this protein promote chromosome segregation in concert with the RNA interference pathway [69]. The DEAD box RNA helicase in *Cardinium* is conserved among five different *Cardinium* strains (Table S7), and shows the greatest similarity to *Amoebophilus* and to intracellular *Alphaproteobacteria*, including *Wolbachia*. In addition, the gene encoding this protein is

located in a predicted operon with a gene (CAHE_0676) coding for a cold shock DNA-binding protein that is also conserved in CI-inducing *Wolbachia* strains.

Ubiquitination is a key regulatory process specific to eukaryotes and absent in bacteria. It is thus interesting that *Cardinium* encodes a protein with a putative RING domain ubiquitin ligase activity (CAHE_p0026; Figure S2) and an ubiquitin specific protease (USP, CAHE_0028; Figure S3). USPs are effector proteins that in bacteria are known in only a few pathogens and symbionts [70–71]. The *Cardinium* USP is conserved among five different strains (Table S7) and belongs to the CA clan of cysteine proteases; the three key domains, the catalytic cysteine box and two histidine boxes, are highly conserved among known and functionally characterized eukaryotic USPs [72]. This high degree of sequence conservation suggests that the *Cardinium* USP functions in the context of a eukaryotic cell and is able to manipulate the host's ubiquitin system. Ubiquitin proteases are involved in stabilizing/destabilizing proteins, signaling, DNA repair, histone structure, and cell-cycle progression [70,73]. Among other proteins, eukaryotic USPs interact with cyclin-dependent kinases (CDKs) and with CDK inhibitor proteins (CKI). CDKs are associated with DNA replication initiation in the S-phase, nuclear envelope breakdown, chromosome condensation, assembly of mitotic spindle and changes in microtubule behavior in the M-phase [74]. In CI induced by *Wolbachia*, delayed nuclear envelope breakdown and histone H3 phosphorylation of mitotic male pronuclei relative to female pronuclei indicates a delayed activity of Cdk1 in the male pronuclei of insect embryos. As a consequence, male pronuclear chromosomes do not segregate properly during mitotic anaphase [5]. Interference of bacterial effectors with CDKs is thus one way in which reproductive incompatibility could be accomplished. If *Cardinium* used a similar mechanism for induction of CI as *Wolbachia*, this could be directly achieved via secretion of the *Cardinium* encoded USP and the counteracting ubiquitin ligase. In *Wolbachia* strains, which appear to lack USPs, this could be performed through other effectors targeting host USPs, for example ANK proteins [48,67].

Although orthologs of some of these proteins were also detected in *Cardinium* strains that cause other phenotypes (Table S7), they are still likely to be good candidates for CI involvement. In addition, *Cardinium* encodes a number of other more general host interaction proteins. One such protein contains a WH2 motif and a proline-rich domain at the N-terminus (CAHE_0010). These two features are commonly found in actin binding proteins, such as the Sca2 protein in *Rickettsia* of the spotted fever group, used for bacterial motility within the eukaryotic host cell [75]. Similar proteins are also present in *Wolbachia*. Other known virulence factors present in *Cardinium* include a patatin-like phospholipase (CAHE_0286) that is most similar to patatin-like proteins encoded in WO prophages in *Wolbachia* [76], and a collagen-like protein containing collagen triple helix repeats (CAHE_0706). Collagen is mainly found in multicellular eukaryotes, but is also present in pathogenic bacteria and viruses [77] and has been associated with adhesion and invasion of eukaryotic cells [78].

Evolution from an ancestor in amoebae

Cardinium shares a number of genome characteristics with its closest sequenced relative, the amoeba symbiont *Amoebophilus*. Sixty-seven percent of all CDSs ($n = 561$) show similarity with *Amoebophilus* proteins (at least 25% sequence identity, at least 80% similarity in size). Further, their metabolic pathways are similarly truncated, encode similar transporters for the import of host-derived metabolites, and contain a notably large fraction of transposases or remnants of IS elements compared to other

bacteria. The similarity of these genome features between *Cardinium* and *Amoebophilus* is striking considering the low degree of 16S rRNA sequence similarity (91%) between these symbionts, indicative of a large evolutionary distance. Consistent with its smaller size (47% relative to *Amoebophilus*) the *Cardinium* genome represents a subset of the *Amoebophilus* genome, with fewer CDSs (841 versus 1557), a greater degree of truncation of metabolic pathways (Figure S1), and the fewer functional transposase genes; 71% (74 out of 104) of the transposase genes are truncated or contain a frame shift compared to 43% in *Amoebophilus* [79]. Transposable elements are key mediators of genome plasticity; they are able to disrupt genes and to induce rearrangements such as inversions, duplications and deletions. They also play important roles in the shaping of symbiont genomes and in genome size reduction [29,80–81]. The irregular genomic GC skew of *Cardinium* (Figure 1) is indicative of past activity of transposable elements. Distortion of the compositional strand bias is well known from other bacteria containing large numbers of transposases, including *Wolbachia* [20–22,82–83]. The presence of a large proportion of transposase genes in the genomes of *Cardinium* and *Amoebophilus* is also consistent with the low degree of synteny in these relatives, indicating extensive reshuffling during the evolution of these bacteria from their last common ancestor (Figure S4).

The reduction in the capabilities of the *Cardinium* genome relative to *Amoebophilus* is also illustrated by cell wall biosynthesis. Both *Cardinium* and *Amoebophilus* are able to generate peptidoglycan, but they lack lipopolysaccharide (LPS) and show truncated phospholipid biosynthesis pathways. While *Amoebophilus* still encodes the complete MreBCD complex, RodA, and IspA considered necessary for a rod-shaped morphology [29], *Cardinium* lacks all of these genes with the exception of *mreB* (CAHE_0369) and indeed has a more coccoid appearance compared to *Amoebophilus*, a pattern also observed in other insect endosymbionts [29]. In general, the *Cardinium* genome represents a subset of the larger genome of the amoeba symbiont *Amoebophilus*. The large amount of inactivated transposase genes in the *Cardinium* genome suggests that it is undergoing further degradation and reduction.

In the *Cardinium* genome, we identified 68 genes (8% of all CDSs) that were possibly involved in past horizontal gene transfer (HGT) events (Table S9). A prominent example are the genes encoding the biotin synthesis pathway. Phylogenetic analysis suggests that *Cardinium* has originally lost all genes involved in biotin synthesis, and acquired the complete gene cluster by horizontal gene transfer, putatively from a donor related to rickettsiae (Figure S5). HGT among intracellular bacteria may occur among bacteria infecting the same hosts [84–86], and thus document ecological niches inhabited during the organism's evolutionary history. We used phylogenetic analysis to determine the putative HGT partners (donors or recipients) and infer additional possible hosts of the bacterial lineage leading to *Cardinium*. As expected, *Cardinium* contains a number of HGT-affected genes shared with partners generally found in arthropod hosts (38% of all HGT affected genes, Figures S6, S7; Table S9). In addition, there are many genes shared with a diverse assemblage of bacteria, and a few eukaryotic genes. Notably, 14% of the HGT-affected genes of *Cardinium* are shared with bacteria known to be associated with amoebae, e.g. *Simkania negevensis* and *Legionella drancourtii*, and 24% are shared with bacteria that have been reported to infect both amoebae and arthropods. The most likely explanation for the presence of genes from amoeba-associated bacteria is that prior to the adaptation to its arthropod host, *Cardinium* (or its ancestor) lived as a symbiont of amoebae or other protists, in which HGT with other amoeba-associated bacteria was facilitated. This notion is consistent with

our observation that the *Cardinium* genome represents a subset of the genome of the sister lineage to *Cardinium*, the amoeba symbiont *Amoebophilus*. It is thus likely that the common ancestor of *Cardinium* and *Amoebophilus* lived as a symbiont of an amoeba or a protist. These unicellular eukaryotes are known to have contributed to the development of key features for survival in eukaryotic host cells by other intracellular bacteria [84,87–88].

Independent origin of CI

At some point during its evolutionary history, *Cardinium* made the transition from amoebae to insect hosts and became a reproductive manipulator able to induce CI to facilitate its spread in host populations. Although *Cardinium* and *Wolbachia* share this phenotype, it is unknown whether the molecular mechanisms leading to CI are identical. If they were, and if the ability to cause CI originated in either one of the two groups and subsequently was acquired by the other through HGT during coinfection of the same host [89–91], one would expect to observe a set of genes in *Cardinium* and CI-inducing *Wolbachia* that likely mediate this phenotype and share a common evolutionary origin. Among the orthologous genes shared by *Cardinium* and *Wolbachia* there is not a single obvious case of a gene encoding a candidate effector involved in CI. Apart from the patatin-like phospholipase, which is considered a more general virulence factor, we identified only one orthologous gene (CAHE_0604) that was exclusive to *Cardinium* and some rickettsiae including the CI-inducing *Wolbachia* strains. This gene encodes a predicted integral membrane protein without any known functional domains and is thus unlikely to mediate CI. This suggests that there is no common evolutionary origin of CI in *Cardinium* and *Wolbachia*, and that the molecular mechanism of CI is either different in these two groups, or convergent.

It is striking, however, that comparison of the genomes of CI-inducing *Wolbachia* strains with the CI lineage of *Cardinium* revealed in both genomes a large number of proteins that contain eukaryotic domains and likely mediate host cell interaction and CI. These include a DEAD box RNA helicase, and many ANK and TPR proteins that are highly unusual in bacterial genomes and good candidates for CI effectors manipulating the eukaryotic cell cycle. Most of these proteins are highly divergent and show no sequence similarity beyond the presence of eukaryotic domains. This indicates an independent origin of genes involved in CI, most likely through independent HGT events and acquisition of host genes. This notion is further supported by the presence of ubiquitin modifying proteins in *Cardinium*, which might be involved in CI, and the absence of these in CI-inducing *Wolbachia* strains. Taken together, CI seems to be based on the exploitation of eukaryotic domains for host cell manipulation, and there is strong evidence for an independent emergence of the molecular mechanisms underlying CI in these two groups. In general, the *Cardinium* genome points to the utility of a comparative context for analysis of reproductive manipulation in symbiotic bacteria that are refractory to direct genetic manipulation, a fertile area for research in the coming years.

Materials and Methods

Nomenclature of *Cardinium* strains

No strain nomenclature has previously been adopted for *Cardinium hertigi*. In an effort to create a convenient and consistent system, strains have been named in this study following the strain nomenclature of *Wolbachia pipientis* [5]. Thus the genome reference strain is “cEper1”, where “c” refers to *Cardinium*, “Eper” refers to the host *Encarsia pergandiella*, and “1” simply denotes the first named strain from this host.

Rearing of *Encarsia pergandiella* wasps harboring *Cardinium*

Cardinium hertigi cEper1 is a symbiont of the minute parasitoid wasp *Encarsia pergandiella* (~18 µg) that attacks whiteflies [23]. Wasps were originally collected from the whitefly *Bemisia tabaci* near Weslaco, Texas in October 2006, and kept in culture on *B. tabaci* on cowpea. Males of *E. pergandiella* develop as hyperparasites and were reared on another whitefly primary parasitoid, *Eretmocerus eremicus*. Prior to purification of *Cardinium* cells, wasps were reared on *B. tabaci* that were not infected with *Rickettsia* spp.

Purification of *Cardinium* cells and DNA isolation

For *Cardinium* purification, wasps were reared on dozens of whitefly-infested plants. Approximately 8,000 adult wasps were collected from emergence jars. The *Cardinium* purification protocol was modified from [92]. Wasps were surface-sterilized with 2.6% sodium hypochlorite and 0.5% SDS for 1 min, washed with sterile water, and homogenized by hand in buffer A (250 mM EDTA, 35 mM Tris-HCl, 250 mM sucrose, 25 mM KCl, 10 mM MgCl₂) using a Dounce tissue grinder (Wheaton). The homogenate was transferred to a 1.5 ml centrifuge tube with an additional 1 ml of buffer A. Cellular debris was pelleted for 5 min at 80 g, 4°C. The supernatant was centrifuged for 5 min at 4000 g, 4°C. The resulting pellet was carefully resuspended in 1 ml of buffer A, then vortexed for 3 sec. Following a 5 min centrifugation at 300 g, the supernatant was loaded onto a 13 mm diameter filter cassette holder (Swinnex filter holder, Millipore) containing a 0.8 to 8 µm pore size glass fiber prefilter (Millipore) and a strong protein binding 5 µm pore-size mixed cellulose ester membrane (Millipore). The supernatant was slowly pushed through the filter with a syringe. The filter cassette holder was washed with buffer A (without EDTA) until 1.5 ml of filtrate was obtained. The filtrate was centrifuged for 5 min at 5000 g, 4°C. Following resuspension of the pellet in buffer A (without EDTA), 10 units of DNase 1 (Roche) were added to the cell suspension and incubated for 30 min at 4°C to remove insect host DNA. The reaction was stopped with 100 µl 0.5 M EDTA. The tube was spun down for 5 min at 4100 g, 4°C, the pellet washed with 1 ml buffer A, then spun down again. The cell pellet was resuspended in 250 µl of TE buffer.

The purified *Cardinium* cells were mixed with 675 µl of DNA extraction buffer (100 mM Tris/HCl, 100 mM EDTA, 100 mM sodium phosphate, 1.5 M NaCl, 1% cetyltrimethylammonium bromide (CTAB) (w/v), 200 µg/ml proteinase K, pH 8.0; [93]), 10 µl of 20 mg/ml proteinase K (Roche) was added and the tube was incubated for 30 min at 37°C. Then 75 µl of 20% SDS was added, the tube was shaken and incubated at 65°C for 1 h, with gentle inversions every 15 to 20 min. Following the incubation, 1 ml of chloroform/isoamylalcohol (24:1 v/v) was mixed in. The aqueous phase was recovered following centrifugation. Nucleic acids were precipitated by adding 0.6 volumes of isopropanol, holding at room temperature for 1 h, then centrifuging at 16,000 g for 20 min, 4°C. The pellet was washed with cold 70% ethanol and spun down for 5 min at max speed, 4°C. Ethanol was removed and the pellet allowed to air dry. The DNA pellet was resuspended in TE buffer with 7 units RNase/ml (RNaseA, Qiagen), and incubated for 20 min at 37°C.

Whole-genome amplification

The extracted *Cardinium* DNA was quantified using PicoGreen (Invitrogen), totaling approximately 2 ng, which was insufficient for library generation and sequencing, thus requiring amplification. To minimize bias, multiple displacement amplification

(MDA) was performed on eight replicate reactions as follows. Approximately 0.1 ng of template DNA was denatured using alkaline solution and amplified using the Repli-g UltraFast Mini Kit (Qiagen) according to the manufacturer's instructions. MDA was performed overnight and the eight resulting MDA products were pooled prior to library generation and sequencing.

Sequencing, assembly, and gap closure

A combination of Illumina and 454 shotgun sequencing was performed on the pooled symbiont MDA DNA product. Two differing 454 standard libraries (one un-normalized, one normalized) were generated and sequenced totaling 300,490,911 bp. In addition, we generated and sequenced two 454 paired end libraries totaling 106,933,881 bp. An Illumina GAii shotgun library was constructed and sequenced (run mode 2×76 bp) generating 1,371,155,520 bp. All general aspects of library construction and sequencing can be found at <http://www.jgi.doe.gov/>. The Illumina GAii sequencing data was assembled with Velvet (<http://genome.cshlp.org/content/18/5/821.short>) with a hash length of 61 and with the following parameters `-exp_cov 130 -cov_cutoff 1 -min_contig_lgth 100`. The consensus sequences were shredded into 1.5 Kbp overlapped fake reads and assembled together with the 454 data. The velvet contig fake reads (17,983 reads, 9.2 Mbp) and the 454 pyrosequencing reads (400.3 Mbp) were assembled using the Newbler assembler version 2.4 (Roche) using the parameters `-acc -g -mi 98 -ml 80 -rip`. The Newbler assembly consisted of 20,306 contigs in 1,154 scaffolds. Illumina reads were additionally used to correct potential base errors and increase consensus quality using the software SeqMan NGen from DNASTAR. One scaffold consisting of 78 contigs was identified as the *Cardinium* chromosome based on BLAST searches against the ribosomal rRNA database Silva (release_102); another scaffold (6 contigs) was identified representing the *Cardinium* plasmid based on BLAST searches against the non-redundant sequence dataset (nr) at GenBank/EMBL/DBJ. The gaps in both scaffolds were closed by manual refinement of the assembly and by PCR and Sanger sequencing in house and by LGC Genomics (Berlin, Germany).

Genome annotation and analysis

The genome was analyzed and automatically annotated using the Microbial Genome Analysis and Annotation Platform MaGe [94]. The automatic annotation was further refined by blastp against Swiss-Prot and UniProt using an E-value of $1e^{-5}$, a minimum amino acid identity of 30%, and minimum alignment overlaps of 40% as threshold values, and by manual annotation of selected genes. The circular view of the genome (Figure 1) was generated using the software GenVision (DNASTAR); the GC skew was calculated using the program CGView [95] with a sliding window size of 887 bp. Transposable genetic elements were identified using blastp. Data for NCBI clusters of orthologous genes (COGs, [36]) analysis were taken from the MaGe [94]. Biochemical pathway reconstruction was performed using KEGG [96] integrated in MaGe [94]. Classification of transport proteins into Transport Classification Database (TCDB) families was done using BLAST (<http://www.tcdb.org/index.php>) [97]. The anti-feeding prophage (AFP)-like cluster was first identified by using blastp with proteins encoded on the AFP-like gene cluster of *Amoebophilus* and then by using either blastp or psi-blast with proteins of the AFP from *Serratia*. Putative host cell interaction proteins were further analyzed using blastp; protein domains were predicted using PFAM [98] and SMART [99]. Multiple amino acid sequence alignments were done using MAFFT [100]. Putative horizontal gene transfer candidate proteins were predicted by blastp of all *Cardinium* proteins against the non-redundant protein

GenBank/EMBL/DBJ sequence database (nr). *Cardinium* proteins with ten best blast hits to proteins from organisms outside the bacterial phylum *Bacteroidetes* were considered to potentially be involved in a past horizontal gene transfer (HGT) events. To further investigate this, the top 50 blast hits were used for amino acid sequence alignments with MUSCLE [101], and phylogenetic trees were reconstructed using the software MEGA5 [102]. Trees were calculated using the neighbor joining algorithm (2000 bootstrap resamplings) and the maximum likelihood algorithm (100 bootstrap resamplings). The nearest neighbor of putatively HGT affected genes of *Cardinium* was identified by the lowest number of internal nodes in the calculated trees. If there were several neighbors with the same number of nodes, the minimum sum of branch lengths was used as criterion. The sequences described in this paper have been deposited at GenBank/EMBL/DBJ under accession numbers HE983995 (chromosome) and HE983996 (plasmid pCher). All contigs from the original *Encarsia* metagenome from which the *Cardinium* genome was reconstructed are also available at GenBank/EMBL/DBJ.

PCR screening for putative host cell interaction genes in different *Cardinium* strains

Approximately 100 wasps from five *Encarsia* spp. cultures harboring different *Cardinium hertigii* strains, including the reference strain *cEper1* were each collected in a 1.5 ml reaction tube, resuspended in 180 μ l buffer ATL (QIAGEN DNeasy blood and tissue kit) and homogenized with a pellet pestle suitable for 1.5 ml microcentrifuge tubes. DNA from homogenized wasps was isolated with QIAGEN DNeasy blood and tissue kit as recommended in the manufacturer's protocol with the exception of the usage of 400 μ g proteinase K (Roche) resuspended in 20 μ l ddH₂O instead of the proteinase K recommended by the manufacturer. A standard PCR cycling program with 35 cycles with primers specific for different *Cardinium* genes was used for the amplification (for primer sequences and annealing temperatures see Table S10). PCR included New England Biolabs Taq DNA Polymerase at a concentration of 0.8 units/20 μ l reaction with ThermoPol Buffer. dNTPs were used at a final concentration of 1 mM. Primers were used at a concentration of 0.4 μ M; BSA was added at 0.6 μ g/ μ l.

Transmission electron microscopy of *Cardinium* cells

Transmission electron microscopy of *Cardinium* cells was performed as described elsewhere [55]. Ovaries of adult *E. pergandiella* wasps were fixed in 4% glutaraldehyde in 0.05 M cacodylate buffer overnight at 4°C. After postfixation in 2% OsO₄ for 2 h, the samples were washed, *en bloc*-stained in 2% uranyl acetate, and dehydrated through an ethanol series (50, 70, 95, and 100%). The samples were then placed in propylene oxide and embedded in Epon. Serial sections were cut with an RMC MT7000 ultra microtome. The grids were stained with saturated uranyl acetate and lead citrate and viewed under a Philips Electronic Instruments CM12 transmission electron microscope.

Supporting Information

Figure S1 Representation of clusters of orthologous gene (COG) categories in selected genomes of obligate and facultative bacterial symbionts.
(PDF)

Figure S2 Conservation of the RING-like domain encoded in CAHE_p0026. Comparison of the domain found in CAHE_p0026 with RING-like domains showing E3 ubiquitin ligase

activity according to [44]. The domains RING-HC and RING-H2 represent the two major subcategories of RING finger domains (depending on whether a Cys or His occupies the fifth coordination site); Mdm2, murine double minute 2 protein; RBQ-1, retinoblastoma binding protein 6 (RBBP6); RBX1, RING-box protein 1; Cnot4, CCR4-NOT transcription complex subunit 4. Only conserved amino acid residues indicative for the RING finger domain are shown. Cys, cysteine; His, histidine; X, any amino acid; subscript number corresponds to number of amino acid.

(PDF)

Figure S3 Multiple sequence alignment of selected ubiquitin-specific proteases (USPs) with CAHE_0028, the USP of *Cardinium hertigi*. An amino acid sequence alignment of the catalytic core domains of selected USPs is shown. The alignment was done with MAFFT [45], shading of conserved amino acid residues was performed with Boxshade available at the Swiss EMBnet server (http://www.ch.embnet.org/software/BOX_form.html). Data for important amino acid residues are taken from [46–47]. Amino acid residues forming the catalytic triad are highlighted in red. Amino acid residues that have been shown to be involved in van der Waals contact with ubiquitin are highlighted in green. Amino acid residues that are involved in direct inter-molecular hydrogen bond interactions using their side chains and main chains are highlighted in blue. Amino acid residues are only highlighted if they were present in all aligned sequences. Regions of high sequence conservation within characterized USPs are underlined: Cys-box (215–229), QDE-box (292–305), His-box (446–468, 477–486, 512–520); numbering according to UBP7_HUMAN residues. The consensus is displayed at the bottom of each alignment block, asterisks indicate identical positions, dots indicate similar positions. Abbreviations and accession numbers: UBP2_HUMAN (human, O75604), UBP7_HUMAN (human, Q93009), UBP14_HUMAN (human, P54578), UBP4_YEAST (*S. cerevisiae*, CAA86791), UBP8_YEAST (*S. cerevisiae*, P50102), UBP15_YEAST (*S. cerevisiae*, P50101), Aasi_0770 (*A. asiaticus*, YP_001957879), Aasi_1805 (*A. asiaticus*, YP_003573189), USP_Cardinium (*C. hertigi*, CAHE_0028).

(PDF)

Figure S4 Synteny between *Cardinium hertigi* and *Amoebophilus asiaticus*. Syntons comprising at least three genes are indicated by green lines if the orientation is conserved or by red lines in case of inversions. In total, 284 *Cardinium* CDSs are arranged in 106 syntons (larger than three genes) with *Amoebophilus*.

(PDF)

Figure S5 Phylogenetic analysis of the biotin biosynthesis cluster of *Cardinium hertigi*. Tree calculations were performed using the maximum likelihood algorithm (1000 bootstrap resamplings) with a concatenated dataset of six biotin synthesis proteins (BioB, BioF, BioH, BioC, BioD and BioH; Table S11) of bacteria from eight different phyla. Genes and their genomic organization are indicated as colored boxes. Breaks in the black bars denote noncontiguous genes. Boxes above and below the black bars indicate genes encoded on the plus and minus strand, respectively. Bootstrap values are indicated at the respective node. Note that the *Cardinium* genes are syntenic with those of the putative rickettsial donors.

(PDF)

Figure S6 HGT-affected genes in *Cardinium hertigi* and its putative donors/recipients. Only HGT candidates with a bootstrap value higher than 75% and a consistent grouping in both neighbor joining and maximum likelihood trees (shown in

Figure S6) were included from the list of HGT candidate genes (Table S9).

(PDF)

Figure S7 Phylogenetic relationships of candidate HGT genes of *Cardinium hertigi*. Phylogenetic trees are based on amino acid sequences and were calculated with MEGA using the neighbor-joining algorithm (NJ) with 2000× bootstrapping and maximum-likelihood algorithm (ML) with 100× bootstrapping. Bootstrap values are indicated at the respective nodes. GenBank accession numbers are indicated.

(PDF)

Table S1 Genome sizes of selected endosymbionts. Obligate (primary) symbionts are shaded in grey; obligate symbionts are indicated with a section sign; members of the *Bacteroidetes* are indicated by an asterisk; plasmids were not taken into account.

(DOCX)

Table S2 Transposases in the genome of *Cardinium hertigi*.

(DOCX)

Table S3 *Cardinium hertigi* proteins involved in biotin biosynthesis, glycolysis, peptidoglycan biosynthesis, and lipoate biosynthesis.

(DOCX)

Table S4 Transport proteins in the genome of *Cardinium hertigi*.

(DOCX)

Table S5 Comparison of the *Amoebophilus asiaticus* AFP-like gene cluster (as query) with the *Cardinium hertigi* AFP-like gene cluster and the *Serratia entomophila* AFP on the pADAP plasmid by blast. Blast results obtained using psi-blast are labeled with an asterisk; I, amino acid identity to best blast hit; E, E-value; n.d., not determined.

(DOCX)

Table S6 Comparison of the *Serratia entomophila* AFP gene cluster (as query) with the AFP-like gene cluster of *Cardinium hertigi* and *Amoebophilus asiaticus* by blastp. I, amino acid identity to best blast hit; E, E-value; n.d., not determined.

(DOCX)

Table S7 Phenotypes of different *Cardinium hertigi* strains, their *Encarsia* wasp hosts, and presence of selected genes detected by PCR. CI (cytoplasmic incompatibility inducing), PI (parthenogenesis inducing).

(DOCX)

Table S8 Proteins of *Cardinium hertigi* likely involved in host cell interaction. I, amino acid identity to best blast hit; E, E-value; n.a., not applicable; n.p., not present.

(DOCX)

Table S9 *Cardinium hertigi* genes putatively involved in past horizontal gene transfer events. Nearest neighbors in phylogenetic trees are indicated (neighbor-joining trees, 2000 bootstrap replications; maximum-likelihood trees, 100 bootstrap replications; Figure S6). Genes encoding transposases, repeat proteins and Na⁺/proline symporters, and genes shared with *Amoebophilus* are not listed. AM, amoeba associated bacteria; AA, *Rickettsia* that are able to multiply in amoebae and arthropods; ART, arthropod associated bacteria; E, eukaryotes; X, other bacteria; n.a., not applicable. Nodes with a bootstrap higher than 75 and the same group are indicated with an asterisk.

(DOCX)

Table S10 Primers used for the detection of putative host cell interaction genes in different *Cardinium hertigi* strains (Table S7).

(DOCX)

Table S11 NCBI accession numbers of proteins from the biotin biosynthesis pathway used for a concatenated data set for the calculation of a phylogenetic tree with the maximum likelihood algorithm. (DOCX)

Acknowledgments

We gratefully acknowledge Thomas Rattei and Thomas Weinmaier for help with bioinformatics analysis; Claudine Médigue, David Vallenet, and

the MicroScope Team for help with the annotation platform MaGe; Cara Gibson for help with *Cardinium* purification and DNA isolation; and Bill Sullivan and Scott O'Neill for discussions.

Author Contributions

Conceived and designed the experiments: MSH SS-E MH. Performed the experiments: TP SS-E SEK BNC AM TW SAM. Analyzed the data: TP SS-E TW MSH MH. Wrote the paper: TP MSH MH.

References

- Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* 42: 165–190.
- Hedges LM, Brownlie JC, O'Neill SL, Johnson KN (2008) *Wolbachia* and virus protection in insects. *Science* 322: 702.
- Oliver KM, Russell JA, Moran NA, Hunter MS (2003) Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proc Natl Acad Sci U S A* 100: 1803–1807.
- Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, et al. (2006) The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* 314: 267.
- Werren JH, Baldo L, Clark ME (2008) *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol* 6: 741–751.
- Serbus LR, Casper-Lindley C, Landmann F, Sullivan W (2008) The genetics and cell biology of *Wolbachia*-host interactions. *Annu Rev Genet* 42: 683–707.
- Cordaux R, Bouchon D, Greve P (2011) The impact of endosymbionts on the evolution of host sex-determination mechanisms. *Trends Genet* 27: 332–341.
- Telschow A, Flor M, Kobayashi Y, Hammerstein P, Werren JH (2007) *Wolbachia*-induced unidirectional cytoplasmic incompatibility and speciation: mainland-island model. *PLoS ONE* 2: e701. doi:10.1371/journal.pone.0000701.
- Jiggins FM, Hurst GDD, Majerus MEN (2000) Sex-ratio-distorting *Wolbachia* causes sex-role reversal in its butterfly host. *Proceedings of the Royal Society B-Biological Sciences* 267: 69–73.
- Zabalou S, Riegler M, Theodorakopoulou M, Stauffer C, Savakis C, et al. (2004) *Wolbachia*-induced cytoplasmic incompatibility as a means for insect pest population control. *Proc Natl Acad Sci U S A* 101: 15042–15045.
- Walker T, Johnson PH, Moreira LA, Iturbe-Ormaetxe I, Frentiu FD, et al. (2011) The *wMel* *Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations. *Nature* 476: 450–453.
- Zug R, Hammerstein P (2012) Still a host of hosts for wolbachia: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PLoS ONE* 7: e38544. doi:10.1371/journal.pone.0038544.
- Taylor MJ, Hoerauf A (1999) *Wolbachia* bacteria of filarial nematodes. *Parasitol Today* 15: 437–442.
- Werren JH (1997) Biology of *Wolbachia*. *Annu Rev Entomol* 42: 587–609.
- Lassy CW, Karr TL (1996) Cytological analysis of fertilization and early embryonic development in incompatible crosses of *Drosophila simulans*. *Mech Dev* 57: 47–58.
- Tram U, Sullivan W (2002) Role of delayed nuclear envelope breakdown and mitosis in *Wolbachia*-induced cytoplasmic incompatibility. *Science* 296: 1124–1126.
- Gavotte L, Henri H, Stouthamer R, Charif D, Charlat S, et al. (2007) A Survey of the bacteriophage WO in the endosymbiotic bacteria *Wolbachia*. *Mol Biol Evol* 24: 427–435.
- Sinkins SP, Walker T, Lynd AR, Steven AR, Makepeace BL, et al. (2005) *Wolbachia* variability and host effects on crossing type in *Culex mosquitoes*. *Nature* 436: 257–260.
- Papafotiou G, Oehler S, Savakis C, Bourtzis K (2011) Regulation of *Wolbachia* ankyrin domain encoding genes in *Drosophila* gonads. *Res Microbiol* 162: 764–772.
- Klasson L, Walker T, Sebahia M, Sanders MJ, Quail MA, et al. (2008) Genome evolution of *Wolbachia* strain *wPip* from the *Culex pipiens* group. *Mol Biol Evol* 25: 1877–1887.
- Klasson L, Westberg J, Sapountzis P, Naslund K, Lutnaes Y, et al. (2009) The mosaic genome structure of the *Wolbachia* *wRi* strain infecting *Drosophila simulans*. *Proc Natl Acad Sci U S A* 106: 5725–5730.
- Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, et al. (2004) Phylogenomics of the reproductive parasite *Wolbachia pipiensis* *wMel*: a streamlined genome overrun by mobile genetic elements. *PLoS Biol* 2: e69. doi:10.1371/journal.pbio.0020069.
- Hunter MS, Perlman SJ, Kelly SE (2003) A bacterial symbiont in the *Bacteroidetes* induces cytoplasmic incompatibility in the parasitoid wasp *Encarsia pergandiella*. *Proc Biol Sci* 270: 2185–2190.
- Duron O, Hurst GD, Hornett EA, Josling JA, Engelstadter J (2008) High incidence of the maternally inherited bacterium *Cardinium* in spiders. *Mol Ecol* 17: 1427–1437.
- Dallai R, Mercati D, Giusti F, Gottardo M, Carapelli A (2011) A *Cardinium*-like symbiont in the proturan *Acerella muscorum* (Hexapoda). *Tissue Cell* 43: 151–156.
- Nakamura Y, Kawai S, Yukuhiro F, Ito S, Gotoh T, et al. (2009) Prevalence of *Cardinium* bacteria in planthoppers and spider mites and taxonomic revision of “*Candidatus Cardinium hertigii*” based on detection of a new *Cardinium* group from biting midges. *Applied and Environmental Microbiology* 75: 6757–6763.
- Noel GR, Atibalentja N (2006) ‘*Candidatus Paenicardinium endonii*’, an endosymbiont of the plant-parasitic nematode *Heterodera glycines* (Nemata: *Tylenchida*), affiliated to the phylum *Bacteroidetes*. *Int J Syst Evol Microbiol* 56: 1697–1702.
- Horn M, Harzenetter MD, Linner T, Schmid EN, Muller KD, et al. (2001) Members of the *Cytophaga-Flavobacterium-Bacteroides* phylum as intracellular bacteria of acanthamoebae: proposal of ‘*Candidatus Amoebophilus asiaticus*’. *Environ Microbiol* 3: 440–449.
- McCutcheon JP, Moran NA (2011) Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 10: 13–26.
- Moya A, Pereto J, Gil R, Latorre A (2008) Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat Rev Genet* 9: 218–229.
- Newton IL, Bordenstein SR (2011) Correlations between bacterial ecology and mobile DNA. *Curr Microbiol* 62: 198–208.
- Wu D, Daugherty SC, Van Aken SE, Pai GH, Watkins KL, et al. (2006) Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol* 4: e188. doi:10.1371/journal.pbio.0040188.
- Baldrige GD, Burkhardt NY, Felsheim RF, Kurtz TJ, Munderloh UG (2008) Plasmids of the pRM/pRF family occur in diverse *Rickettsia* species. *Appl Environ Microbiol* 74: 645–652.
- Ogata H, Renesto P, Audic S, Robert C, Blanc G, et al. (2005) The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. *PLoS Biol* 3: e248. doi:10.1371/journal.pbio.0030248.
- Gillespie JJ, Joardar V, Williams KP, Driscoll T, Hostetler JB, et al. (2011) A *Rickettsia* genome overrun by mobile genetic elements provides insight into the acquisition of genes characteristic of obligate intracellular lifestyle. *J Bacteriol* 194: 376–394.
- Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 28: 33–36.
- Schmitz-Esser S, Tischler P, Arnold R, Montanaro J, Wagner M, et al. (2010) The genome of the amoeba symbiont ‘*Candidatus Amoebophilus asiaticus*’ reveals common mechanisms for host cell interaction among amoeba-associated bacteria. *J Bacteriol* 192: 1045–1057.
- Spalding MD, Prigge ST (2010) Lipoid acid metabolism in microbial pathogens. *Microbiology and Molecular Biology Reviews* 74: 200–228.
- Lipke H, Fraenkel G (1956) Insect nutrition. *Annual Review of Entomology* 1: 17–44.
- Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, et al. (2002) Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nat Genet* 32: 402–407.
- Dunning Hotopp JC, Lin M, Madupu R, Crabtree J, Angiuoli SV, et al. (2006) Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genet* 2: e21. doi:10.1371/journal.pgen.0020021.
- Hosokawa T, Koga R, Kikuchi Y, Meng XY, Fukutsu T (2010) *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proc Natl Acad Sci U S A* 107: 769–774.
- Perlman SJ, Kelly SE, Hunter MS (2008) Population biology of cytoplasmic incompatibility: maintenance and spread of *Cardinium* symbionts in a parasitic wasp. *Genetics* 178: 1003–1011.
- Brownlie JC, Cass BN, Riegler M, Witsenburg JJ, Iturbe-Ormaetxe I, et al. (2009) Evidence for metabolic provisioning by a common invertebrate endosymbiont, *Wolbachia pipiensis*, during periods of nutritional stress. *PLoS Pathog* 5: e1000368. doi:10.1371/journal.ppat.1000368.
- Himpler AG, Adachi-Hagimori T, Bergen JE, Kozuch A, Kelly SE, et al. (2011) Rapid Spread of a Bacterial Symbiont in an Invasive Whitefly Is Driven by Fitness Benefits and Female Bias. *Science* 332: 254–256.
- Preston GM (2007) Metropolitan Microbes: Type III Secretion in Multihost Symbionts. *Cell Host Microbe* 2: 291–294.

47. Coombes BK (2009) Type III secretion systems in symbiotic adaptation of pathogenic and non-pathogenic bacteria. *Trends in Microbiology* 17: 89–94.
48. Dale C, Moran NA (2006) Molecular interactions between bacterial symbionts and their hosts. *Cell* 126: 453–465.
49. Rancès E, Voronin D, Tran-Van V, Mavingui P (2008) Genetic and functional characterization of the type IV secretion system in *Wolbachia*. *J Bacteriol* 190: 5020–5030.
50. Penz T, Horn M, Schmitz-Esser S (2011) The genome of the amoeba symbiont “*Candidatus Amoebophilus asiaticus*” encodes an *afp*-like prophage possibly used for protein secretion. *Virulence* 1: 541–545.
51. Hurst MR, Beard SS, Jackson TA, Jones SM (2007) Isolation and characterization of the *Serratia entomophila* antifeeding prophage. *FEMS Microbiol Lett* 270: 42–48.
52. Persson OP, Pinhassi J, Riemann L, Marklund BI, Rhen M, et al. (2009) High abundance of virulence gene homologues in marine bacteria. *Environ Microbiol* 11: 1348–1357.
53. Furusawa G, Yoshikawa T, Takano Y, Mise K, Furusawa I, et al. (2005) Characterization of cytoplasmic fibril structures found in gliding cells of *Saprosira* sp. *Can J Microbiol* 51: 875–880.
54. Bigliardi E, Sacchi L, Genchi M, Alma A, Pajoro M, et al. (2006) Ultrastructure of a novel *Cardinium* sp. symbiont in *Scaphoideus titanus* (Hemiptera: Cicadellidae). *Tissue Cell* 38: 257–261.
55. Zchori-Fein E, Gottlieb Y, Kelly SE, Brown JK, Wilson JM, et al. (2001) A newly discovered bacterium associated with parthenogenesis and a change in host selection behavior in parasitoid wasps. *Proc Natl Acad Sci U S A* 98: 12555–12560.
56. Zchori-Fein E, Perlman SJ, Kelly SE, Katzir N, Hunter MS (2004) Characterization of a ‘*Bacteroidetes*’ symbiont in *Encarsia* wasps (Hymenoptera: Aphelinidae): proposal of ‘*Candidatus Cardinium hertigii*’. *Int J Syst Evol Microbiol* 54: 961–968.
57. Bonemann G, Pietrosiuk A, Mogk A (2010) Tubules and donuts: a type VI secretion story. *Mol Microbiol* 76: 815–821.
58. Leiman PG, Basler M, Ramagopal UA, Bonanno JB, Sauder JM, et al. (2009) Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc Natl Acad Sci U S A* 106: 4154–4159.
59. Jehl MA, Arnold R, Rattai T (2011) Effective - a database of predicted secreted bacterial proteins. *Nucleic Acids Res* 39: D591–595.
60. Blatch GL, Lassel M (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* 21: 932–939.
61. Schreiber A, Stengel F, Zhang Z, Enchev RI, Kong EH, et al. (2011) Structural basis for the subunit assembly of the anaphase-promoting complex. *Nature* 470: 227–232.
62. Foster J, Ganatra M, Kamal I, Ware J, Makarova K, et al. (2005) The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol* 3: e121. doi:10.1371/journal.pbio.0030121.
63. Li J, Mahajan A, Tsai MD (2006) Ankyrin repeat: a unique motif mediating protein-protein interactions. *Biochemistry* 45: 15168–15178.
64. Axton JM, Shamanski FL, Young LM, Henderson DS, Boyd JB, et al. (1994) The inhibitor of DNA replication encoded by the *Drosophila* gene plutonium is a small, ankyrin repeat protein. *EMBO J* 13: 462–470.
65. Pan X, Luhrmann A, Satoh A, Laskowski-Arce MA, Roy CR (2008) Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* 320: 1651–1654.
66. Voth DE (2011) ThANKs for the repeat: Intracellular pathogens exploit a common eukaryotic domain. *Cell Logist* 1: 128–132.
67. Iturbe-Ormaetxe I, Burke GR, Riegler M, O’Neill SL (2005) Distribution, expression, and motif variability of ankyrin domain genes in *Wolbachia pipientis*. *J Bacteriol* 187: 5136–5145.
68. Duron O, Boureux A, Echaubard P, Berthomieu A, Berticat C, et al. (2007) Variability and expression of ankyrin domain genes in *Wolbachia* variants infecting the mosquito *Culex pipiens*. *J Bacteriol* 189: 4442–4448.
69. Pek JW, Kai T (2011) DEAD-box RNA helicase Belle/DDX3 and the RNA interference pathway promote mitotic chromosome segregation. *Proc Natl Acad Sci U S A* 108: 12007–12012.
70. Rytönen A, Holden DW (2007) Bacterial interference of ubiquitination and deubiquitination. *Cell Host Microbe* 1: 13–22.
71. Wilkes TE, Darby AC, Choi JH, Colbourne JK, Werren JH, et al. (2010) The draft genome sequence of *Arsenophonus nasoniae*, son-killer bacterium of *Nasonia vitripennis*, reveals genes associated with virulence and symbiosis. *Insect Mol Biol* 19 Suppl 1: 59–73.
72. Hu M, Li P, Li M, Li W, Yao T, et al. (2002) Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with Ubiquitin Aldehyde. *Cell* 111: 1041–1054.
73. Nijman SM, Luna-Vargas MP, Velds A, Brummelkamp TR, Dirac AM, et al. (2005) A genomic and functional inventory of deubiquitinating enzymes. *Cell* 123: 773–786.
74. Stüfeler LA, Ji JY, Trautmann S, Trusty C, Schubiger G (1999) Cyclin A and B functions in the early *Drosophila* embryo. *Development* 126: 5505–5513.
75. Haglund CM, Choe JE, Skau CT, Kovar DR, Welch MD (2010) *Rickettsia* Sca2 is a bacterial formin-like mediator of actin-based motility. *Nat Cell Biol* 12: 1057–1063.
76. Kent BN, Funkhouser IJ, Setia S, Bordenstein SR (2011) Evolutionary genomics of a temperate bacteriophage in an obligate intracellular bacteria (*Wolbachia*). *PLoS ONE* 6: e24984. doi:10.1371/journal.pone.0024984.
77. Rasmussen M, Jacobsson M, Björck L (2003) Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J Biol Chem* 278: 32313–32316.
78. Paterson GK, Nieminen L, Jefferies JM, Mitchell TJ (2008) PclA, a pneumococcal collagen-like protein with selected strain distribution, contributes to adherence and invasion of host cells. *FEMS Microbiol Lett* 285: 170–176.
79. Schmitz-Esser S, Penz T, Spang A, Horn M (2011) A bacterial genome in transition - an exceptional enrichment of IS elements but lack of evidence for recent transposition in the symbiont *Amoebophilus asiaticus*. *BMC Evol Biol* 11: 270.
80. Siguié P, Filee J, Chandler M (2006) Insertion sequences in prokaryotic genomes. *Curr Opin Microbiol* 9: 526–531.
81. Touchon M, Rocha EP (2007) Causes of insertion sequences abundance in prokaryotic genomes. *Mol Biol Evol* 24: 969–981.
82. Nakayama K, Yamashita A, Kurokawa K, Morimoto T, Ogawa M, et al. (2008) The Whole-genome sequencing of the obligate intracellular bacterium *Orientia tsutsugamushi* revealed massive gene amplification during reductive genome evolution. *DNA Res* 15: 185–199.
83. Cho NH, Kim HR, Lee JH, Kim SY, Kim J, et al. (2007) The *Orientia tsutsugamushi* genome reveals massive proliferation of conjugative type IV secretion system and host-cell interaction genes. *Proc Natl Acad Sci U S A* 104: 7981–7986.
84. Bordenstein SR, Reznikoff WS (2005) Mobile DNA in obligate intracellular bacteria. *Nat Rev Microbiol* 3: 688–699.
85. Georgiades K, Merhej V, El Karkouri K, Raoult D, Pontarotti P (2011) Gene gain and loss events in *Rickettsia* and *Orientia* species. *Biol Direct* 6: 6.
86. Gimenez G, Bertelli C, Moliner C, Robert C, Raoult D, et al. (2011) Insight into cross-talk between intra-amoebal pathogens. *BMC Genomics* 12.
87. Molmeret M, Horn M, Wagner M, Santic M, Abu Kwaik Y (2005) Amoebae as training grounds for intracellular bacterial pathogens. *Appl Environ Microbiol* 71: 20–28.
88. Toft C, Andersson SG (2010) Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat Rev Genet* 11: 465–475.
89. White JA, Kelly SE, Cockburn SN, Perlman SJ, Hunter MS (2011) Endosymbiont costs and benefits in a parasitoid infected with both *Wolbachia* and *Cardinium*. *Heredity* 106: 585–591.
90. Skaljic M, Zanic K, Ban SG, Kontsedalov S, Ghanim M (2010) Co-infection and localization of secondary symbionts in two whitefly species. *BMC Microbiol* 10: 142.
91. Sirvio A, Pamilo P (2010) Multiple endosymbionts in populations of the ant *Formica cinerea*. *BMC Evol Biol* 10: 335.
92. Braig HR, Zhou WG, Dobson SL, O’Neill SL (1998) Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. *J Bacteriol* 180: 2373–2378.
93. Zhou J, Bruns MA, Tiedje JM (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* 62: 316–322.
94. Vallenet D, Labarre L, Rouy Z, Barbe V, Bocs S, et al. (2006) MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res* 34: 53–65.
95. Stothard P, Wishart DS (2005) Circular genome visualization and exploration using CGView. *Bioinformatics* 21: 537–539.
96. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, et al. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 36: D480–484.
97. Saier MH, Jr., Yen MR, Noto K, Tamang DG, Elkan C (2009) The transporter classification database: recent advances. *Nucleic Acids Res* 37: D274–278.
98. Finn RD, Tate J, Mistry J, Coghill PC, Sammut SJ, et al. (2008) The Pfam protein families database. *Nucleic Acids Res* 36: D281–288.
99. Letunic I, Doerks T, Bork P (2009) SMART 3: recent updates and new developments. *Nucleic Acids Res* 37: D229–232.
100. Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9: 286–298.
101. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792–1797.
102. Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 9: 299–306.
103. Basler M, Pilhofer M, Henderson GP, Jensen GJ, Mekalanos JJ (2012) Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* 483: 182–186.

Chapter III

A bacterial genome in transition - an exceptional enrichment of IS elements but lack of evidence for recent transposition in the symbiont *Amoebophilus asiaticus*

Published in BMC Evol Biol. 2011 Sep 26;11:270. PMID: 21943072

RESEARCH ARTICLE

Open Access

A bacterial genome in transition - an exceptional enrichment of IS elements but lack of evidence for recent transposition in the symbiont *Amoebophilus asiaticus*

Stephan Schmitz-Esser^{1,3*†}, Thomas Penz^{1†}, Anja Spang² and Matthias Horn¹

Abstract

Background: Insertion sequence (IS) elements are important mediators of genome plasticity and are widespread among bacterial and archaeal genomes. The 1.88 Mbp genome of the obligate intracellular amoeba symbiont *Amoebophilus asiaticus* contains an unusually large number of transposase genes ($n = 354$; 23% of all genes).

Results: The transposase genes in the *A. asiaticus* genome can be assigned to 16 different IS elements termed ISCaa1 to ISCaa16, which are represented by 2 to 24 full-length copies, respectively. Despite this high IS element load, the *A. asiaticus* genome displays a GC skew pattern typical for most bacterial genomes, indicating that no major rearrangements have occurred recently. Additionally, the high sequence divergence of some IS elements, the high number of truncated IS element copies ($n = 143$), as well as the absence of direct repeats in most IS elements suggest that the IS elements of *A. asiaticus* are transpositionally inactive. Although we could show transcription of 13 IS elements, we did not find experimental evidence for transpositional activity, corroborating our results from sequence analyses. However, we detected contiguous transcripts between IS elements and their downstream genes at nine loci in the *A. asiaticus* genome, indicating that some IS elements influence the transcription of downstream genes, some of which might be important for host cell interaction.

Conclusions: Taken together, the IS elements in the *A. asiaticus* genome are currently in the process of degradation and largely represent reflections of the evolutionary past of *A. asiaticus* in which its genome was shaped by their activity.

Keywords: insertion sequence element, endosymbiont, *Bacteroidetes*, genome evolution

Background

Mobile genetic elements such as phages, plasmids and transposable elements play a vital role in horizontal gene transfer and genome rearrangement in bacteria and archaea [1]. Among transposable elements, insertion sequence (IS) elements are particularly widespread within bacterial and archaeal genomes, and are considered the most abundant and ubiquitous genes in nature [2-6]. IS elements can have profound effects on chromosome structure and evolution. Due to their ability to

disrupt genes and to induce rearrangements such as inversions, duplications and deletions they are key mediators of genome plasticity [2,3,7-9]. Although IS elements are perceived primarily as genomic parasites, their activity can also be beneficial. As composite transposons IS elements are able to mobilize adjacent genes, thereby mediating the spread of antibiotic resistance genes and genes involved in the catabolism of complex xenobiotics [10,11]. IS elements may also promote adaptation of their host genomes as demonstrated in experimental evolution experiments [12-15]. In addition, IS elements can influence or activate the expression of adjacent genes, e.g. by forming hybrid or fusion

* Correspondence: stephan.schmitz-esser@vetmeduni.ac.at

† Contributed equally

¹Department of Microbial Ecology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria

Full list of author information is available at the end of the article

promoters or by containing outward-directed promoters [16-22].

IS elements are usually less than 2.5 kbp in length and have a relatively simple genetic organization. Most IS elements are flanked both by inverted and direct repeats and generally encode no function other than those involved in mobility, which is mediated by transposases [16]. IS elements have been classified into several families based on the degree of sequence conservation of their transposases and its catalytic site, similar genetic organization such as size, number of open reading frames (ORFs) and potential coding sequences (CDSs), inverted repeats, and genome target sites [2,16]. The majority of IS elements encode transposases containing the so-called DDE-motif consisting of the three amino acids aspartic acid, aspartic acid, and glutamic acid. These residues form the catalytic triad necessary for transposition. They are found in three regions (N2, N3, and C1) of the transposase amino acid sequence separated by spacers of various lengths [2,16].

Although IS elements are found in the majority of sequenced bacterial and archaeal genomes [2-4], their distribution is patchy, and their occurrence within single genomes is usually below 3% [2,6]. IS elements are very rare in the genomes of most ancient host-restricted symbionts or pathogens such as mutualistic insect and clam symbionts or chlamydiae [23,24]. On the other hand, elevated numbers of IS elements has been observed in the genomes of bacteria which adapted only recently to an intracellular or pathogenic lifestyle [8,25-27]. However, this view has been challenged by the recent detection of IS element-rich genomes in ancient symbionts such as *Wolbachia* spp. or *Orientia tsutsugamushi* [9,28-31]. Interestingly, the genomes containing the highest percentages of IS elements are from obligate intracellular bacteria: *Orientia tsutsugamushi* [28,29], the γ 1 symbiont of the marine oligochaete *Olavivus algarvensis* [32], the symbionts of grain weevils [26,33], and the amoeba symbiont *Amoebophilus asiaticus* 5a2 [34].

Amoebophilus asiaticus is a Gram-negative, obligate intracellular symbiont, which has been discovered within an amoeba isolated from alkaline lake sediment [35]. Highly similar *A. asiaticus* strains have been recovered from various sources worldwide [35-38]. *A. asiaticus* shows highest 16S rRNA similarity to '*Candidatus Cardinium hertigii*', an obligate intracellular parasite of arthropods able to manipulate the reproduction of its hosts [39]. Both organisms belong to the phylum *Bacteroidetes* and form a monophyletic lineage in 16S rRNA-based phylogenetic trees [35], consisting only of symbionts and sequences retrieved from coral samples [40]. The *A. asiaticus* genome is only moderately reduced in size compared to many other obligate intracellular

bacteria [41,42] but nevertheless, its biosynthetic capabilities are extremely limited [34]. The *A. asiaticus* genome encodes a hitherto unparalleled high number of proteins with eukaryotic domains such as ankyrin repeats, TPR/SEL1 repeats, leucine-rich repeats and domains from the eukaryotic ubiquitin system, and it contains an unusually large number of transposase genes ($n = 354$) corresponding to 23% of all genes [34].

Here, we report on the in-depth analysis of the IS elements in the *A. asiaticus* genome. We classified them and describe their main characteristics. We demonstrated that other symbionts closely related to *A. asiaticus* contain highly similar IS elements, and we could show that although they are transcribed, they exhibited no transpositional activity on a population level during a time period of almost 1,000 days. Taking into account evidence that no major rearrangements have occurred recently in the *A. asiaticus* genome, this suggests that the IS elements are evolutionary older components of the *A. asiaticus* genome, which likely played an important role during genome reduction and adaptation to an obligate intracellular life style.

Results

Diversity of IS elements in the *A. asiaticus* 5a2 genome

IS elements make up 183 kbp (10%) of the *A. asiaticus* genome. In total, 354 transposase genes (corresponding to 23% of all CDSs) were identified in the detailed and manually curated analysis performed here (Tables 1, 2). Compared to other sequenced prokaryotic genomes, the percentage of IS elements as well as the number of IS elements per megabase genome is among the highest in *A. asiaticus* (Additional file 1, Figures S1, S2). We were able to assign the vast majority of these transposase genes ($n = 329$, 93%; including partial IS element copies) to 16 different IS elements (ISCaa1 to ISCaa16), which belong to eight different IS element families, with IS5 family IS elements being most abundant in the *A. asiaticus* genome (Table 2). Each of the 16 IS elements is present in 2 to 24 full-length copies in the *A. asiaticus* genome, the only exception being ISCaa1, which was identified earlier by the ISFinder website [43] and is only present as a single full-length copy (Table 2). This results in a total copy number of 122 full-length IS elements that are evenly spread across the *A. asiaticus* genome [34]. A high number of IS elements in *A. asiaticus* is truncated ($n = 143$), and in some cases (e.g. ISCaa5, ISCaa6 and ISCaa11) there are more truncated than full-length copies present (Table 2). Truncation sites were generally not conserved, i.e. truncations occurred in different regions, and truncated IS elements show varying lengths (Additional file 1, Figure S3). For most of the full-length IS element copies ($n = 101$, 83%) we could not identify direct repeats (Tables 1, 2). In the

Table 1 IS element statistics for the genome of *A. asiaticus*

No. of protein coding genes	1557
No. of transposase encoding genes	354 (23% of all protein coding genes)
No. of transposase encoding genes assigned to IS elements	329 (93% of all transposase genes)
No. of full-length IS element copies*	122
No. of partial IS element copies*	143
No. of full-length IS element copies with functional transposase gene	106 (87% of all full-length IS element copies)
No. of full-length IS element copies without direct repeats	101 (83% of all full-length IS element copies)

* Note that IS elements can consist of more than one transposase gene

following sections we shortly describe few selected IS elements of *A. asiaticus* in more detail.

ISCaa4

ISCaa4 is the most abundant IS element in *A. asiaticus*. It is present in 24 full-length copies, 21 of these copies should be able to produce an intact, functional transposase. ISCaa4 belongs to the IS1 family and shows a typical IS1 family DDE-motif [2,44]. Similar to other IS1 family members, the ISCaa4 transposase is encoded by two overlapping ORFs, which are probably translated into a 226 amino acid transposase by -1 ribosomal frameshifting (Table 2, Additional file 1, Figure S4). Translational frameshifting is often found in IS elements and represents an important mechanism regulating the expression of the transposases at a translational level [16,45]. Translation starts at the first ORF (*orfA*) and shifts to the -1 reading frame at the so-called slippery site and continues in a second overlapping ORF (*orfB*) resulting in a transframe ORFAB protein. The predicted frameshift site in ISCaa4 (AAAAAAG) is highly shift-prone in bacteria such as *A. asiaticus* that have only a single tRNA^{Lys} (anticodon: UUU) and lack the tRNA^{Lys} with the anticodon UUC [45,46]. In ISCaa4 five nucleotides downstream of the putative slippery site a stem-loop structure is predicted (ΔG -6.3 kcal/mol) (Additional file 1, Figure S4). Such stem-loop structures have been shown to be stimulatory for -1 ribosomal frameshifting [45,46]. Interestingly, ISCaa4 shows highest amino acid sequence identity (46 to 51%) to uncharacterized IS elements from methanogenic archaea of the family *Methanosarcinaceae*; the similarity to other transposases is lower than 40%. In phylogenetic trees, ISCaa4 forms a stable monophyletic group with these archaeal transposases, indicating interdomain horizontal gene transfer between methanogenic archaea and *A. asiaticus* (Figure 1, Additional file 1, Figure S5).

ISCaa3

ISCaa3 is present in ten full-length copies in *A. asiaticus* and belongs to ISL2 group within the IS5 family (based

on the presence of a typical DDE-motif) whose transposases typically consist of a single ORF [2,16]. The transposase of ISCaa3 however, is encoded by two overlapping ORFs most likely translated into a 275 amino acid protein by -1 ribosomal frameshifting. In contrast to other IS elements with canonical slippery sites like ISCaa4 and ISCaa9, no stimulatory stem-loop structure possibly enhancing ribosomal frameshifting is predicted downstream of the slippery site in ISCaa3 (Additional file 1, Figure S6). ISCaa3 shows highest amino acid sequence identity (57 to 66%) to ISCaa2 and IS elements found in the intracellular bacteria *Orientia tsutsugamushi*, *Legionella drancourtii*, *Regiella insecticola*, and *Parachlamydia acanthamoebae*. In phylogenetic analyses ISCaa2, ISCaa3 and related IS elements from intracellular bacteria consistently group together in all treeing methods applied, suggesting horizontal transfer of IS elements between these intracellular bacteria (Figure 1, Additional file 1, Figure S7). Interestingly, a number of cyanobacterial IS elements form a sister group with the ISCaa3-related IS elements.

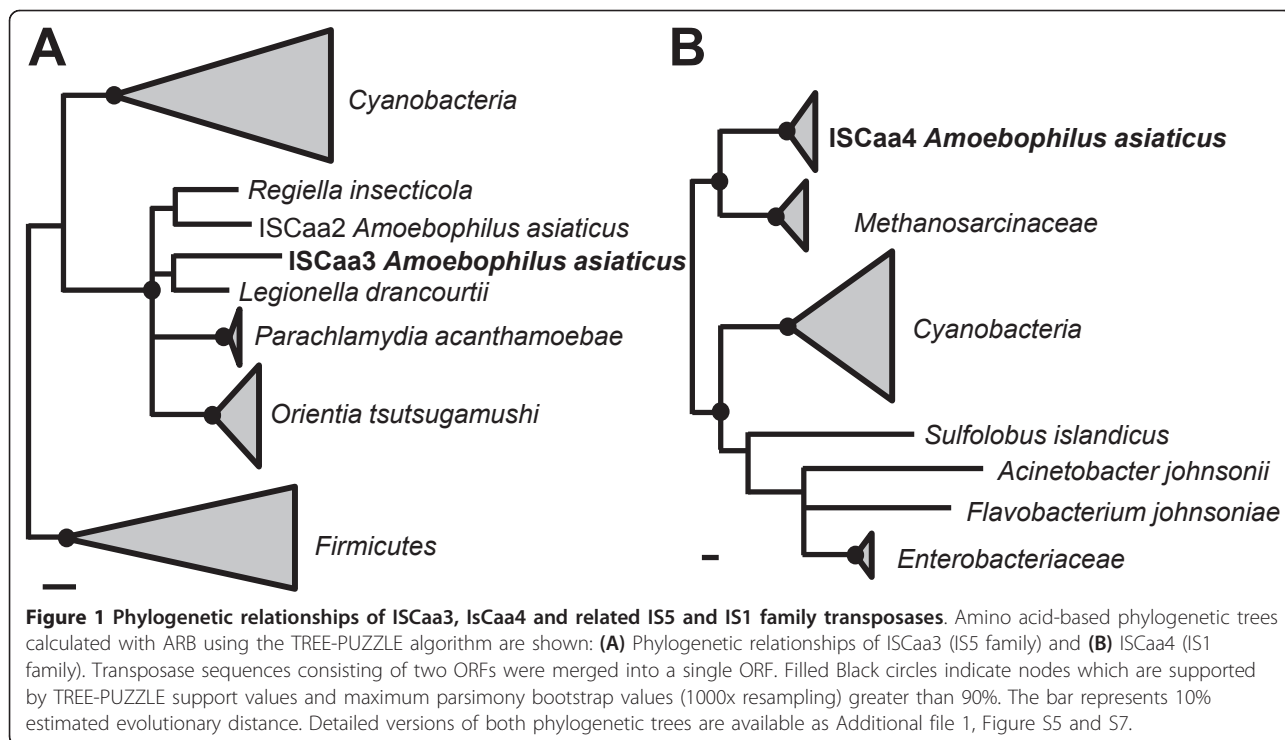
ISCaa9

ISCaa9 is an 881 bp IS element which is present in 15 almost identical copies (the differences occur only in the inverted repeats). ISCaa9 belongs to the IS5 family and shows highest amino acid sequence identity (45%) to ISMac15 from *Methanosarcina acetivorans* C2A, and 40% amino acid identity to ISWpi1, an IS element found in many *Wolbachia* strains [47,48]. The ISCaa9 transposase is encoded by three consecutive and overlapping ORFs which are translated into a 253 amino acid protein (Table 2, Additional file 1, Figure S8). We propose a stop codon read-through to occur at the stop codon (UGA) at nucleotide positions 263 to 265, which is supported by the presence of the stop codon in all 15 ISCaa9 copies in *A. asiaticus*, the absence of a stem-loop structure indicative of a terminator downstream of the stop codon, and the observation that UGAA is a weak stop codon quartet [49,50]. We predict that the

Table 2 IS elements in the *A. asiaticus* 5a2 genome

IS element family	IS	Number of ORFs (predicted translational frameshift)	Length of IS element [bp]	Inverted repeats [bp] ^a	Direct repeats [bp] ^b	G+C content of IS element [%] (range)	Length of transposase [amino acids]	Number of full-length IS element copies (conservation on DNA level)	Number of partial IS element copies	Number of full-length IS element copies with intact transposase genes (conservation on protein level)
ISCaa1^c	IS1	1	759	17/21	0	35.7	232	1	6	1
ISCaa2	IS5, ISL2 group	2 (-1)	916	19/20	0	36.3	275	3 (99-100%)	3	3 (99-100%)
ISCaa3	IS5, ISL2 group	2 (-1)	914	22/23	0	36.5	275	10 (99-100%)	5	9 (99-100%)
ISCaa4	IS1	2 (-1)	732	17/22	8/10	37.4 (36.2 - 38.0)	226	24 (85-100%)	8	21 (96-100%)
ISCaa5	IS982	1	932	18/21	0	38.2	274	10 (99-100%)	24	8 (99-100%)
ISCaa6	IS5, ISL2 group	1	991	15/19	0	36.6 (35.3 - 36.6)	275	18 (86-100%)	41	17 (88-100%)
ISCaa7	IS110	1	1483	0	0	31.9	343	3 (100%)	4	3 (100%)
ISCaa8	IS5, IS1031 group	2 (+1)	893	18/22	0	39.5	264	6 (99-100%)	6	6 (99-100%)
ISCaa9	IS5 (-1 ORFBC)	3	881	18/21	0	38.8	253	15 (100%)	3	15 (100%)
ISCaa10	IS200/IS605/IS200 group	1	527	0	0	38.5	147	7 (99-100%)	2	7 (99-100%)
ISCaa11	IS481	1	1031	10/11	6/3	38.9 (37.4 - 39.7)	314	10 (83-100%)	27	3 (87-100%)
ISCaa12	IS481	1	1210	29/34	6/2	37.6	364	3 (100%)	3	3 (100%)
ISCaa13	IS5, IS427 group	2 (+1)	860	17/21	0	40.7	253	2 (99%)	7	1
ISCaa14	IS110	1	1256	0	0	38.1	326	2 (97%)	0	1
ISCaa15	IS1182	1	1434	18/18	4/2	35.9	457	3 (100%)	3	3 (100%)
ISCaa16	IS6	1	837	15/18	0	37.4 (34.4 - 37.4)	235	5 (82-100%)	1	5 (85-100%)

^a number of base pairs conserved between left and right end repeats/length of the repeat^b length of direct repeat/number of isoforms with direct repeat^c ISCaa1 was identified by the ISFinder website <http://www-isbionet.fr/>



stop codon is recoded into tryptophane (UGG), a common feature of UGAA stop codon quartets [50]. In addition, the majority of ISCaa9-related transposases encodes a tryptophane at the position of the stop codon read-through in ISCaa9 (Additional file 1, Figure S9). We predict a translational -1 frameshifting at a slippery site (AAAAAAG) between *orfB* and *orfC* (Additional file 1, Figure S7). Five nucleotides downstream of this putative slippery site, a stem-loop structure (ΔG -12.6 kcal/mol) is predicted in ISCaa9 (Additional file 1, Figure S8). The ISCaa9 transposase contains a DDE-motif, which is most similar to the IS1031 group within the IS5 family, the transposases of this group however, are usually encoded by a single ORF [2].

ISCaa10

ISCaa10 is with a length of 527 bp a very short IS element that contains a single ORF encoding a 147 amino acid transposase. It belongs to the IS200/IS605 family and IS200 group of IS elements comprising the shortest known transposases [2]. Members of the IS200 group are unusual IS elements because their transposases do not contain the DDE-motif found in most transposases. Instead they belong to the Y1 transposases with a catalytic tyrosine residue and a conserved HuH motif (consisting of a histidine, a hydrophobic amino acid, and another histidine) [51,52]. Interestingly, this motif is present only in two of seven ISCaa10 copies; in the others, the second histidine is replaced by tyrosine, which

might render these copies nonfunctional. Other unusual features of IS200 IS elements, that are also found in ISCaa10, are the absence of both direct and terminal inverted repeats and the presence of secondary structures leading to low transcriptional and transpositional activity [51-53]. For example, IS200 from *Salmonella typhimurium* LT2 forms two stem-loop structures: The first is a transcriptional repressor terminating impinging transcripts, the second acts at the translational level and occludes the ribosome binding site [53]. Similarly, a stem-loop structure is predicted ten nucleotides upstream of the start codon of the ISCaa10 transposase and close to the 3' end of ISCaa10 (ΔG -13 kcal/mol and -20.3 kcal/mol, respectively). ISCaa10 shows highest amino acid sequence identity (76%) to (uncharacterized) IS200 family transposases from *Xenorhabdus nematophila* (GenBank accession no: YP_003712757).

Unclassified IS elements

Twenty-five transposase genes could not be assigned to either of the 16 *A. asiaticus* IS elements under the criteria applied here. Among these unclassified full-length transposases two transposases belong to the IS110 family (Aasi_1379 and Aasi_1284); and to the Tn3 family (Aasi_0096, Aasi_0545); one belongs to the IS3 family (probably consisting of the two consecutive ORFs Aasi_1748 and Aasi_0907); and two belong to the YhgA-like family of putative transposases (Aasi_0894, Aasi_1306; PFAM-family PF04754).

Conservation of IS elements among different *A. asiaticus* strains

In order to analyze whether the IS elements found in the genome of *A. asiaticus* 5a2 are also present in closely related *A. asiaticus* strains, we performed PCR using primers targeting the 13 most abundant IS elements (Additional file 2, Table S1) with genomic DNA from *A. asiaticus* strain EIDS3 [35] as well as from two novel *A. asiaticus* isolates, *A. asiaticus* US1 and *A. asiaticus* WR. These strains show 98.9%, 99.2%, and 98.5% 16S rRNA sequence similarity to *A. asiaticus* 5a2, respectively, corresponding to strain and species level diversity, respectively. Six out of the 13 IS elements analyzed here were detected in all four *A. asiaticus* isolates. Cloning and sequencing of PCR products obtained from *A. asiaticus* EIDS3 revealed nucleic and amino acid sequence identities to consensus sequences of the *A. asiaticus* 5a2 IS elements of 87% to 98% (Table 3). The lack of PCR products for some IS elements indicates either the absence of these IS elements in the investigated *A. asiaticus* strains or a low degree of conservation and hence the absence of or mismatches with the primer target sites.

Transcription but lack of transpositional activity of the *A. asiaticus* IS elements

The large copy number and the high degree of conservation of some IS elements identified in the *A. asiaticus*

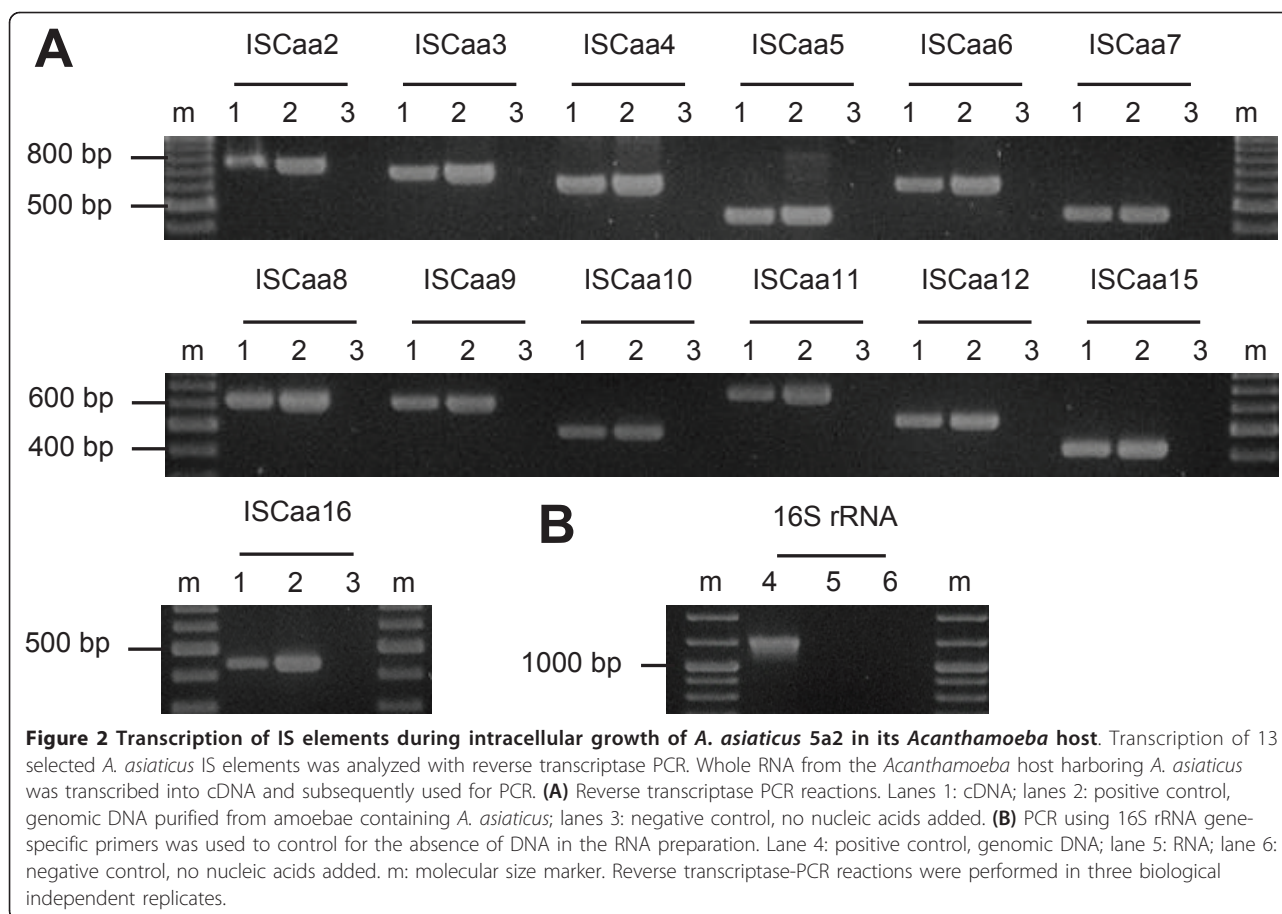
5a2 genome might indicate that they are transpositionally active. To investigate this, we first asked whether the IS elements are transcribed during intracellular replication of *A. asiaticus* in its amoeba host. Using reverse transcriptase (RT)-PCR, we analyzed the transcription of those 13 IS elements that are present in at least three copies in the genome (ISCaa2 to ISCaa12, ISCaa15 and ISCaa16). The detection of transcripts of all 13 IS elements demonstrates that at least one copy each is actively transcribed (Figure 2). Next, we used Southern hybridizations to check for chromosomal rearrangements resulting from transposition events [12,14,54]. We analyzed the same 13 IS elements for which we could show transcription and compared DNA from the same *A. asiaticus* culture isolated in November 2006 and in July 2009, respectively, a period of 984 days. We could not detect differences in the banding pattern indicative for chromosomal rearrangements in Southern hybridizations for any of the IS elements tested (Figure 3).

Contiguous transcription of IS elements and their downstream genes

Some of the *A. asiaticus* IS elements are in close proximity to their downstream genes (with distances less than 50 bp). As previous reports have shown that IS elements can influence the transcription of neighboring

Table 3 Occurrence of IS elements in four different *A. asiaticus* strains based on PCR.

IS element in <i>A. asiaticus</i> 5a2	<i>A. asiaticus</i> EIDS3 (amino acid identity to <i>A. asiaticus</i> 5a2 element)	<i>A. asiaticus</i> WR	<i>A. asiaticus</i> US1
ISCaa2	+ >(95%)	-	-
ISCaa3	+ >(97%)	+	+
ISCaa4	-	+	-
ISCaa5	+ >(94%)	-	+
ISCaa6	+ >(92%)	-	+
ISCaa7	-	-	-
ISCaa8	+ >(91%)	+	+
ISCaa9	+ >(94%)	+	+
ISCaa10	+ >(98%)	+	+
ISCaa11	+ >(90%)	+	+
ISCaa12	+ >(98%)	+	+
ISCaa15	+ >(87%)	+	-
ISCaa16	-	-	-



genes [17-21], we investigated whether contiguous transcripts between *A. asiaticus* IS elements and downstream genes occur. We analyzed ten selected loci where IS elements and their downstream genes are encoded on the same strand and have the same orientation (Figure 4). Using RT-PCR we could show contiguous transcripts of the investigated IS elements with their downstream genes at 9 out of 10 analyzed loci (Figure 5). We performed two control experiments in order to exclude that the observed transcripts from RT-PCR derive from unspecific background noise transcriptional read-through. One control targeted an unlikely contiguous transcript between two genes located on different strands and oriented in opposite directions (Aasi_1200/1201, Figure 4). We could not detect transcripts in this control reaction (Figure 5), indicating that the observed transcripts from the nine loci of IS elements and their downstream genes are above unspecific read-through transcription. This is further supported by a second, semi-quantitative control experiment in which we compared RT-PCR products (using the same conditions) from contiguous transcripts between IS elements and their downstream genes with the products from RT-PCR reactions targeting only the downstream genes

(Additional file 1, Figure S10). In all cases the obtained bands were of similar intensity, providing further evidence that the observed contiguous transcripts are above unspecific transcriptional read-through.

Discussion

Mobile genetic elements such as IS elements move within and between genomes. Owing to its intracellular lifestyle in free-living amoebae *A. asiaticus* is, however, largely shielded from other bacteria. Although horizontal gene transfer seems unlikely to occur under these circumstances, previous studies proposed that amoebae may serve as hot spots for horizontal gene transfer among intracellular bacteria [34,55], and according to the 'intracellular arena' hypothesis genetic material may move in and out of communities of obligate intracellular bacteria that co-infect the same intracellular host environment [23]. We identified four IS elements in *A. asiaticus* that were likely involved in horizontal gene transfer although the direction of the transfer cannot be inferred (ISCaa2, ISCaa3, ISCaa4, ISCaa12; Figure 1, Additional file 1, Figure S5, S7). Three of these IS elements group with IS elements from several other intracellular bacteria related to rickettsiae, legionellae and

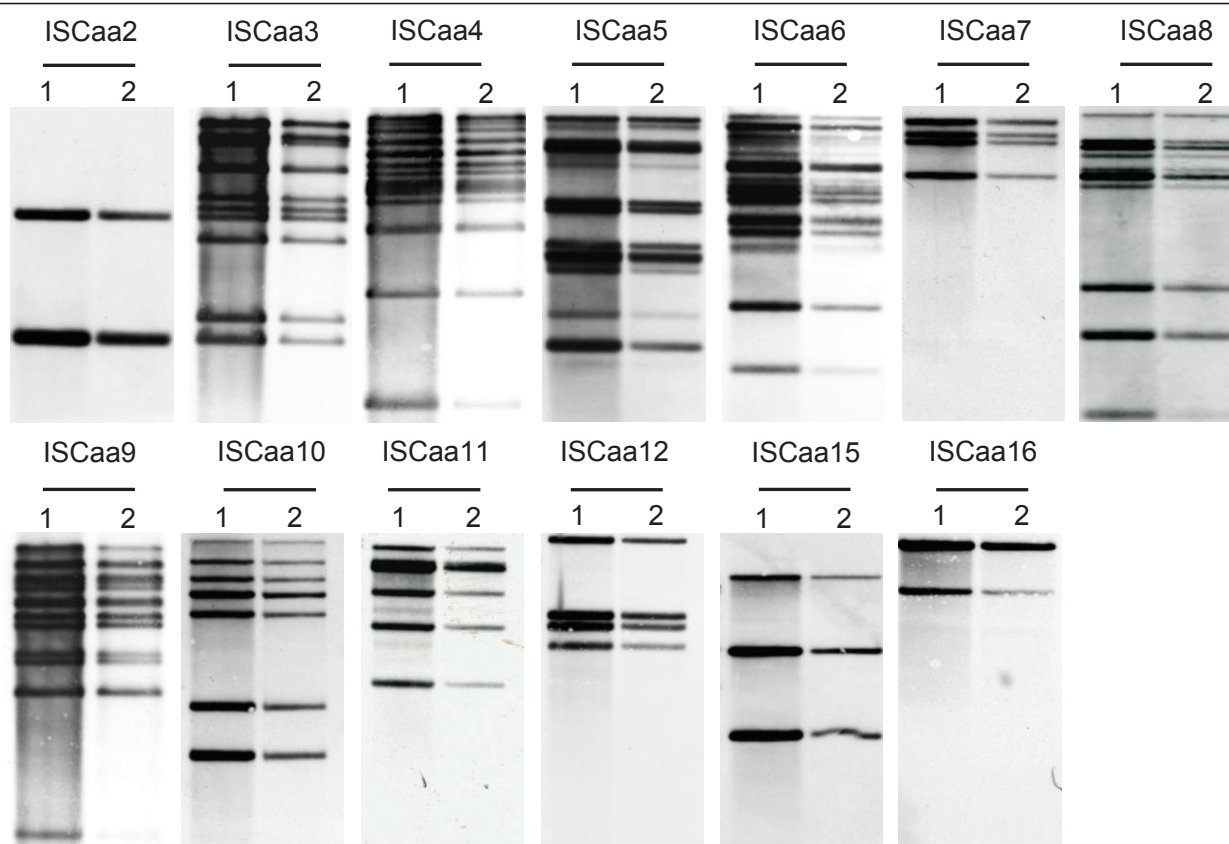
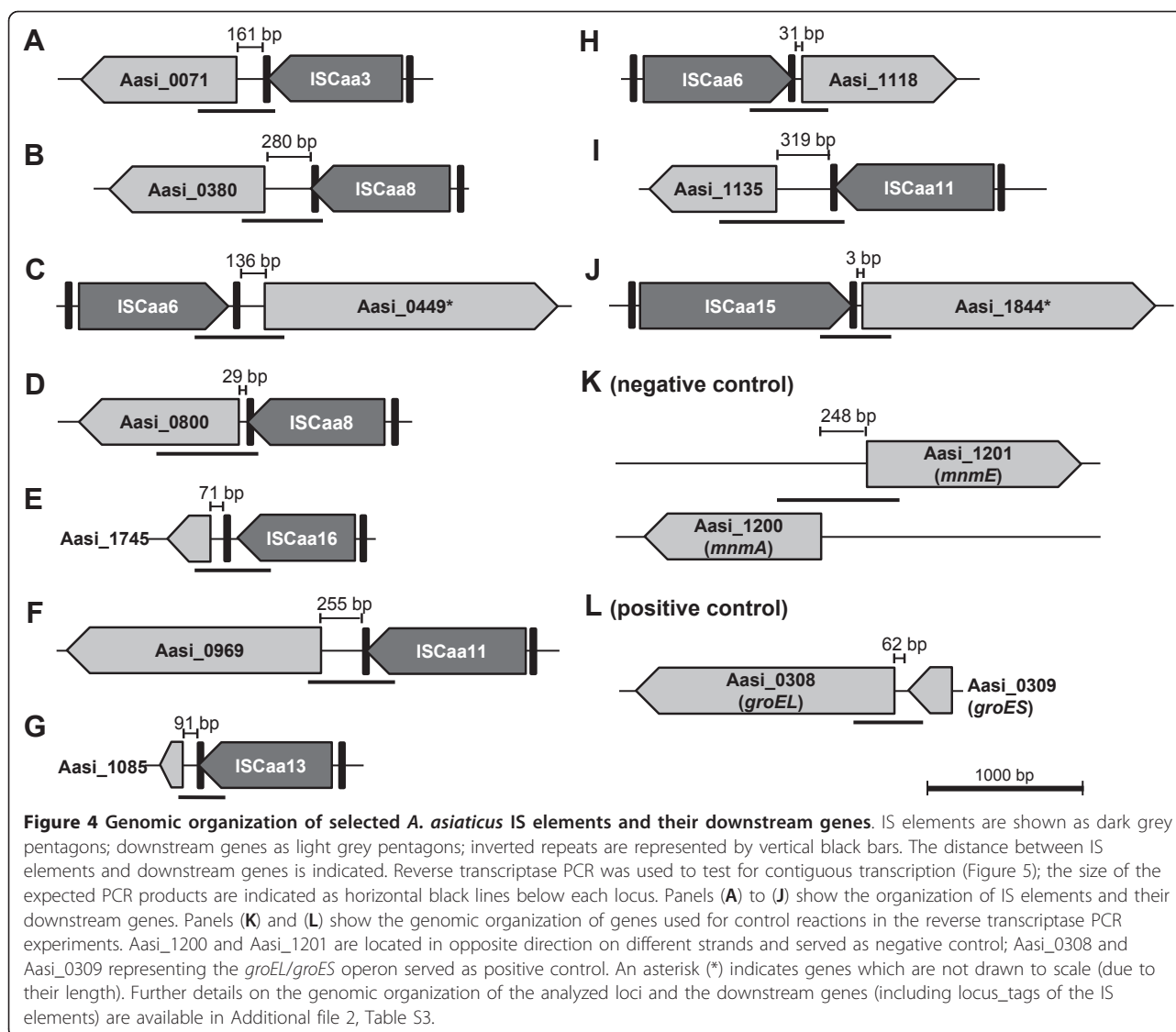


Figure 3 Analysis of transpositional activity of the most abundant IS elements of *A. asiaticus* 5a2. Transposition of IS elements was analyzed with Southern hybridizations using IS element-specific probes and DNA purified from the same *A. asiaticus* 5a2 culture in November 2006 and July 2009, respectively. DNA was digested with Eco32I (except for ISCa2, where HindIII was used). Each visible band corresponds to at least one IS element copy on the respective DNA fragment, as the restriction endonucleases do not cut within the IS elements. IS elements are indicated above each hybridization; lanes 1: DNA isolated November 2006; lanes 2: DNA isolated July 2009. The absence of changes in the banding patterns between both time points indicates that no (major) chromosomal rearrangements due to IS element transposition has occurred.

chlamydiae, and consistent with previous findings it is conceivable that amoebae or other protozoa served as a common habitat for these microbes. One IS element of *A. asiaticus* is most closely related to IS elements found in free-living methanogenic archaea (*Methanosarcinaceae*). Anoxic aquatic sediments, where free-living amoebae and methanogenic archaea can be found, might represent a possible shared habitat facilitating horizontal gene transfer [56-58]. Horizontal gene transfer of IS elements between distantly related organisms is rather rare [59]. Hence the discovery of related IS elements in three different bacterial phyla (*Bacteroidetes*, *Proteobacteria*, *Chlamydiae*) and the *Archaea* might be surprising. However, a recent study based on the analysis of 800 bacterial and archaeal genomes showed that although the majority of horizontal gene transfer events occur between closely related organisms there is a considerable number of large-distance horizontal gene transfer events [60]. Our observations expand our view on the

extent of horizontal gene transfer of IS elements among distantly related microbes, and they provide a glimpse into past interactions of *A. asiaticus* with other microbes during its evolutionary history.

Several lines of evidence point to an ancient origin of many IS elements in *A. asiaticus*. First, ISCa4, ISCa6, ISCa11, and ISCa16, which together make up 46% of all full-length *A. asiaticus* IS elements, show a remarkably low degree of sequence conservation among their different copies (Table 2). This is in contrast to high sequence similarities expected if IS elements have entered a genome and spread only recently [4,61]. Second, the high number ($n = 143$) of truncated IS element copies suggests that these IS elements have been present in the *A. asiaticus* genome for extended time periods during which they disintegrated slowly. Third, the GC-content of the *A. asiaticus* IS elements (37.3% on average, range: 31.9 to 40.7%) is similar to the overall GC content of the *A. asiaticus* genome (35.0%). This

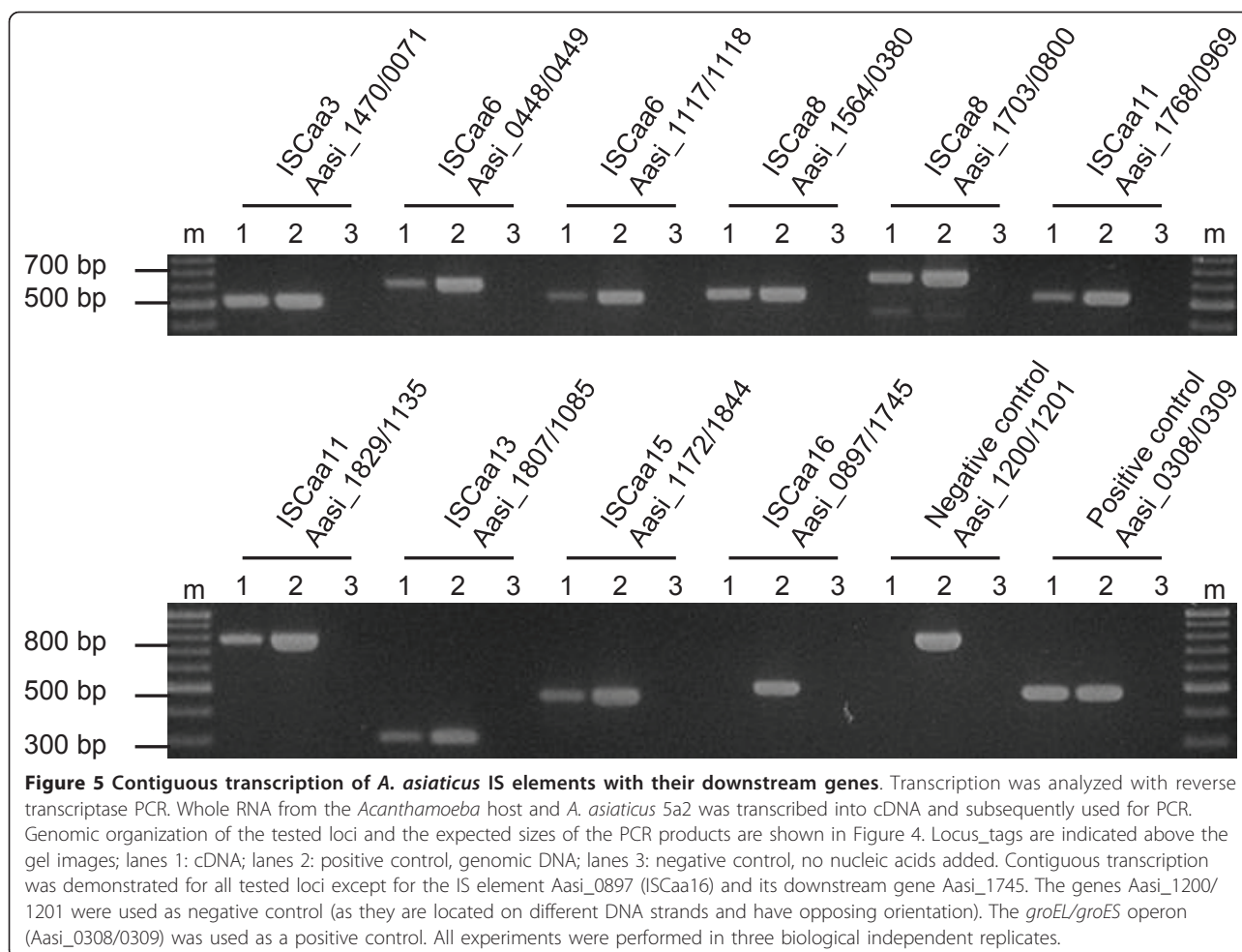


suggests that considerable time has elapsed to allow the base composition of the IS elements to adapt towards the general base composition of the *A. asiaticus* genome [62]. Finally, at least six IS elements are conserved among four different *A. asiaticus* strains, some of which show a relatively high divergence (Table 3). Taking into account that our PCR based screening likely underestimates the actual number of shared IS elements (due to mismatches at the primer binding sites in more diverged homologs), this indicates that many - if not most - *A. asiaticus* IS elements were already present in the last common ancestor of the *A. asiaticus* strains investigated here. Taken together, there is compelling evidence that the IS elements have been residing in the *A. asiaticus* genome for considerable evolutionary time periods.

We noted previously that the *A. asiaticus* genome shows a GC skew pattern typical for most bacterial

genomes with two major shifts at the origin and terminus of replication and only few local deviations, which are indicative of recent genome rearrangements [34]. This is remarkable because with the exception of *Lactobacillus helveticus* DPC 4571 and *Shigella sonnei* Ss046 (whose genomes contain significantly lower percentages of IS elements than *A. asiaticus*; Additional file 1, Figures S1, S2)[27,63], all other bacteria with high numbers of IS elements do not show such a regular genomic GC skew pattern (Additional file 1, Figure S11). Thus, despite of the high number of IS elements, the *A. asiaticus* genome has not been reshuffled extensively recently, which indicates that most IS elements are transpositionally inactive and also that recombination events between highly similar IS element copies have not occurred.

A mechanism by which apparently inactive, non-functional IS elements can be maintained in bacterial



genomes is gene conversion, which was described recently for the genome of *Wolbachia* wBm, a mutualistic symbiont of the nematode *Brugia malayi* whose genome contains a number of highly similar IS element copies rendered non-functional by multiple stop codons and frame shifts [64,65]. In contrast to *Wolbachia* wBm, *A. asiaticus* still encodes intact copies of each IS element, and many IS elements show relatively high sequence divergence (Table 2). In addition, the transposase genes of different non-functional IS element copies show variable pseudogenization states. This largely rules out gene conversion as the main mechanism for maintenance of IS elements in *A. asiaticus*.

Rather unexpectedly, we detected transcription of 13 *A. asiaticus* 5a2 IS elements during intracellular growth in amoebae (Figure 2). Generally, IS elements are among the lowest expressed genes due to their potentially detrimental effects on the host genome [16,61,66-70]. For several *A. asiaticus* IS elements (ISCaa2, ISCaa9, ISCaa10, ISCaa14, ISCaa15; data not shown) stable hairpin structures within the first 50 bp

of the IS elements are predicted, which might interfere with expression both at the transcriptional and the translational level, thus controlling the activity of these IS elements. In addition, evidence for programmed translational frameshifting, another regulatory mechanism, can be found in six *A. asiaticus* IS elements (Table 2). Translational frameshifting acts at the level of translation elongation between two consecutive (and partially overlapping) open reading frames where the ribosome slides one basepair up- or downstream at a so-called slippery site [16,45,66]. For several IS elements of *A. asiaticus*, the occurrence of frameshifting is supported by the presence of a canonical slippery site, of stimulatory secondary structures downstream of the slippery site and, most importantly, the merged amino acid sequences of the IS elements transposase ORFs show more significant Blast hits than the single ORFs alone (data not shown). In summary, transcription of several IS elements occurs in *A. asiaticus*, but there is evidence that many IS elements are tightly regulated both at the transcriptional and the translational level.

Southern hybridizations demonstrated the absence of major transposition events and genome rearrangements for *A. asiaticus* during a time period of 984 days (Figure 3). With an estimated generation time of *Acanthamoeba* sp. 5a2 infected with *A. asiaticus* of 19 h (data not shown), this time period corresponds to approximately 1200 generations of the *Acanthamoeba* host. Although the generation time of *A. asiaticus* is unknown, it must be shorter than that of its amoeba host (due to the high number of symbionts per amoeba cell [35,37]). The analyzed time period thus corresponds to considerably more than 1200 *A. asiaticus* generations. For *E. coli* and *Lactococcus lactis*, the first IS element-mediated genomic changes (insertions, deletions, duplications) occurred already after 400 to 500 generations [12-15]. This indicates that the time period monitored in our study should be sufficient to detect IS element-mediated genomic rearrangements. However, in contrast to our experiment, in these studies bacterial cultures were exposed to environmental stress conditions with respect to nutrient availability, temperature, or oxygen, facilitating adaptive changes. Although no genome rearrangements were observed for *A. asiaticus*, we cannot exclude the possibility that transposition events occurred in individual *A. asiaticus* cells which subsequently became not fixed at the population level and would thus be undetectable by our experimental approach. However, Southern blot is a highly sensitive method [71], and we have estimated that we should be able to monitor changes in Southern blot patterns in subpopulations consisting of only a few to a few hundred of amoeba host cells (Additional file 2, Table S2). Taking into account typical densities of *Acanthamoeba* sp. 5a2 infected with *A. asiaticus* during in vitro cultivation of 10^5 up to 10^7 cells/ml, the sensitivity of our assay should thus be sufficient to detect variations even in very small subpopulations. The IS elements in *A. asiaticus* are therefore most likely transpositional inactive. Their abundance is explained by transpositional activity in the evolutionary past of *A. asiaticus*, and while still being transcriptionally active, most IS elements are transpositionally inactive in extant *A. asiaticus*. In addition to a tight transcriptional and (post-) translational control there are several other conceivable explanations for this observation. For example, *A. asiaticus* might lack host factors required for transposition activity of IS elements although most of those are specific for certain IS elements; they act at different steps and cellular processes and their exact role in transposition is still largely unclear [16,66,72]. Alternatively, the small, reduced genome of *A. asiaticus*, which is highly adapted to the intracellular life style and optimized for host cell interactions, might not allow for major rearrangements as most transposition events would be deleterious rendering the cell nonviable.

One reason why some IS elements were retained in the *A. asiaticus* genome despite of the apparent lack of transpositional activity might be their influence on the transcription of downstream genes. Indeed, we could show contiguous transcripts of IS elements with their downstream genes at 9 out of 10 tested loci (Figure 4, Figure 5). In some cases, the distance between the IS element and the start codon of the downstream gene is too short to include known *Bacteroidetes* Shine-Dalgarno sequences, which are located at -33 and -7 bp relative to the transcription initiation site [73,74]. Expression of the respective downstream genes might thus depend on promoter sequences located within the upstream IS element (e.g. in the inverted repeats), a feature often found in IS elements [16,66], or on the endogenous promoter of the IS element. In other cases the distance between the analyzed IS elements and their downstream genes was larger (up to 300 bp). Similar polycistronic mRNAs starting from IS elements including downstream genes have been described recently for two IS elements in *Francisella tularensis* [17] and in *Mycobacterium tuberculosis* IS6110 [21]. It is striking that many of the genes whose transcription is affected by the presence of IS elements in *A. asiaticus* likely play an important role (Additional file 2, Table S3). For example, Aasi_1844 is an uncharacterized membrane protein conserved among most *Bacteroidetes* and *Chlorobi*; Aasi_1118 contains six TPR/SEL1 repeats, eukaryotic domains that can be involved in host cell interaction [75], and Aasi_0380 is a ferritin homolog involved in iron storage. Furthermore, a genomic organization of IS elements and downstream genes similar to the loci analyzed in this study was found in 44 other regions on the *A. asiaticus* genome (data not shown), suggesting that contiguous transcripts between IS elements and downstream genes are even more widespread and represent a more general feature of *A. asiaticus*.

Genome reduction is an important process during the adaptation of bacteria to an obligate intracellular life style, and IS elements are considered to be important in this process [2,8]. The genome of *A. asiaticus* is only moderately reduced compared to other obligate intracellular bacteria [41,42]. Its genome size is with 1.9 Mbp notably larger than that of other related symbionts in the *Bacteroidetes* (0.2 to 1.1 Mbp), but smaller than those of free-living relatives (2.2 to 9.1 Mbp, Additional file 1, Figure S12). The genome of *A. asiaticus* thus represents a transitional stage in genome reduction. We argue that the IS elements in the *A. asiaticus* genome are evolutionary remnants. They have been present in the *A. asiaticus* genome for extended time periods and reflect the organism's evolutionary history. The IS elements proliferated and were important during the adaptation of *A. asiaticus* to the intracellular life style, but

they became increasingly redundant. The *A. asiaticus* genome thus represents a snapshot of a bacterial genome which was shaped by the activity of IS elements but whose IS elements are largely inactive and in the process of further degradation at the present stage.

Conclusion

Analysis and characterization of the *A. asiaticus* IS elements provides evidence for an extremely IS element-rich genome, which seems to be evolutionary surprisingly stable - a feature not found in other IS element-rich genomes. The presence of contiguous transcripts between IS elements and their downstream genes indicates that these IS elements influence the transcription of their downstream genes, most of which likely play an important role for *A. asiaticus*. Proliferation of IS elements in the evolutionary past of *A. asiaticus* might thus have been an important process during the adaptation of *A. asiaticus* to an intracellular life style in which its genome was shaped by their activity.

Methods

Sequence analyses

The genome sequence of *A. asiaticus* 5a2 has recently been determined and analyzed [34] and is available at GenBank under accession no. CP001102. For identification of IS elements we first compiled a list of candidate transposase genes by keyword, PFAM and InterPro domain search available in the genome annotation software Pedant [76]. We then manually inspected this list in order to verify the evidence for each gene to encode a putative transposase. In addition, further transposase genes were identified by manually analyzing each predicted gene in the *A. asiaticus* genome (e.g by using Blast against the NCBI nr dataset (provided by the annotation software Pedant), Blast against the ISfinder database <http://www-is.biotoul.fr/>) In order to classify the transposases into groups of homologs we performed Blast (BlastP, BlastN) searches against the *A. asiaticus* genome. In order to identify full-length IS elements, the gene sequences of the transposases and surrounding genomic regions were aligned and the full-length IS elements were then manually identified based on these alignments. Partial IS element copies were identified by BlastN and BlastP searches and alignment of full-length IS element copies against the *A. asiaticus* genome. Inverted repeats were identified with the EMBOSS software palindrome and einverted [77]. Nucleic acid sequences of IS elements and amino acid sequences of transposase genes were aligned with MAFFT [78]; alignments were visualized using BOXSHADE http://www.ch.embnet.org/software/BOX_form.html. For detection of direct repeats the nucleic acid alignments of the IS elements and their genomic neighborhood were

searched manually. We grouped and classified IS elements using Blast against the ISfinder website <http://www-is.biotoul.fr/> [43] and the following criteria: (i) a minimum amino acid sequence identity of 30% of the transposase to described transposases, (ii) the presence of flanking inverted repeats (exception: IS elements belonging to family IS110 and IS200/605, which do not have flanking inverted repeats), and (iii) the presence of at least two copies in the genome. IS element copies that shared more than 80% nucleic and amino acid sequence identity over at least 98% of their length were considered isoforms. The nomenclature suggested by the ISFinder website was used for naming of IS elements <http://www-is.biotoul.fr/> [43]. mRNA secondary structures were predicted using the Mfold web server [79]. For calculations of phylogenetic relationships of the transposases from selected IS elements, the amino acid sequences of overlapping ORFs were merged resulting in a single peptide sequence (in the case of IS elements with predicted ribosomal frameshifting), aligned with MAFFT [78] and imported into ARB [80]. Phylogenetic trees were constructed with the Phylip maximum parsimony, distance matrix (Fitch), ProML (using the JTT amino acid replacement model) methods and the TREE-PUZZLE algorithm (using the VT model of amino acid substitution) [81,82] implemented in ARB. Maximum parsimony bootstrap analysis was performed with 1000 resamplings. A filter considering only those alignment positions that were conserved in at least 10% of all sequences (resulting in a total number of 274 and 228 alignment columns for ISCa3 and ISCa4, respectively) was used for all treeing calculations. For each IS element analyzed, the overall tree topology between the different treeing methods applied was consistent, thus only trees calculated using the TREE-PUZZLE algorithm are shown.

Cultivation and isolation of amoebae

Amoebae harboring *A. asiaticus* 5a2 (ATCC no. PRA-228) and amoebae harboring *A. asiaticus* EIDS3 (ATCC no. PRA-221) were maintained as adherent culture in 25 cm² tissue culture flasks containing 10 ml peptone-yeast-glucose medium (PYG: 20 g/l proteose peptone, 2 g/l yeast extract, 90 mM glucose, 4 mM MgSO₄*7H₂O, 3.4 mM C₆H₅Na₃O₇*2H₂O, 2.5 mM KH₂PO₄, 1.3 mM Na₂HPO₄*2H₂O, 51 μM Fe(NH₄)₂(SO₄)₂*6H₂O). Cultures were incubated at 27°C and passaged at confluency by 1:10 dilution of the culture every five to ten days. Amoebae harboring *A. asiaticus* WR and amoebae harboring *A. asiaticus* US1 were isolated from soil and lake sediment (Alkaline lake "Unterer Stinker", Burgenland, Austria) samples, respectively, using non-nutrient agar plates seeded with live or heat-inactivated *Escherichia coli* as described previously [83]. Both isolates were

cultivated as described above using modified PYNFH (10 g/l bacteriological peptone, 10 g/l yeast extract, 1 g/l yeast nucleic acid, 15 mg/l folic acid, 1 mg/l hemin, 2.6 mM KH_2PO_4 , 2.8 mM $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$).

DNA isolation

Amoebae harboring *A. asiaticus* 5a2, EIDS3, WR and US1 were harvested by centrifugation ($5000 \times g$, 10 min). The cell pellet was resuspended in 250 μl 1 \times TE buffer (10 mM Tris, 1 mM EDTA, pH 8) and subsequently used for high molecular weight DNA isolation using a modified protocol from Zhou et al. [84]. Briefly, 675 μl DNA extraction buffer (100 mM Tris/HCl, 100 mM EDTA, 100 mM sodium-phosphate, 1.5 M NaCl, 1% (w/v) cetyltrimethylammonium bromide (CTAB), 200 $\mu\text{g}/\text{ml}$ proteinase K, pH 8.0) were added to the cell pellet and incubated for 30 min at 37°C. After addition of 75 μl 20% (w/v) SDS, the samples were incubated at 65°C for 1 h. To recover the aqueous phase, the lysate was mixed with an equal volume of chloroform/isoamylalcohol (24:1, v/v) and centrifuged ($11200 \times g$, 10 min). Nucleic acids were precipitated with 0.6 volume isopropanol at room temperature for 1 h. The resulting pellet from centrifugation ($16000 \times g$, 20 min) was washed with 70% ethanol, centrifuged again ($16000 \times g$, 5 min), resuspended in ddH_2O and stored at -20°C until use.

Transcription analysis

Amoebae harboring *A. asiaticus* 5a2 were harvested by centrifugation ($7000 \times g$, 3 min, 27°C). The resulting cell pellet was resuspended in 750 μl TRIzol (Invitrogen Life Technologies), transferred to a Lysing Matrix A tube (MP Biomedicals) and homogenized using a BIO101/Savant FastPrep FP120 instrument (speed: 4.5 m/sec, 30 sec). RNA was extracted by phase separation, precipitation, washing and redissolving according to the recommendations of the manufacturer (TRIzol, Invitrogen Life Technologies). Remaining DNA was removed using the TURBO DNA-free Kit (Ambion). After DNase treatment RNA was resuspended in $\text{ddH}_2\text{O}_{\text{DEPC}}$ and stored at -80°C until use. The absence of DNA contamination in the DNase-treated RNA was verified by performing a control PCR with 42 cycles using primers targeting the 16S rRNA gene of *A. asiaticus* 5a2 (Additional file 2, Table S1). DNA-free total RNA (containing host and symbiont RNA) was used to synthesize cDNA using the RevertAid™ First Strand cDNA Synthesis Kit (Fermentas) according to the recommendations of the manufacturer. cDNA was subsequently used as template in standard PCR reactions (35 cycles and annealing temperatures according to the optimal conditions for the primers listed in Additional file 2, Table S1). Negative controls (no cDNA added) and positive controls (genomic DNA) were included in all PCR reactions.

Amplification products were sequenced to ensure that amplification was specific. All experiments were performed in biologically independent triplicates.

PCR screening for IS elements in different *A. asiaticus* strains

A standard PCR cycling program with 35 cycles at low stringency (annealing temperature 45°C) with primers specific for different *A. asiaticus* 5a2 IS elements was used for the detection of IS elements in the *A. asiaticus* strains EIDS3, WR and US1 (see Additional file 2, Table S1 for primer sequences). Negative (no DNA added) and positive controls (genomic DNA from *A. asiaticus* 5a2) were included in all PCR reactions. The amplified fragments from *A. asiaticus* EIDS3 were cloned using the TOPO TA cloning kit and cloning vector pCRII (Invitrogen Life Technologies). Nucleotide sequences of the cloned DNA fragments were determined on an ABI 3130 XL genetic analyzer using the BigDye Terminator kit v3.1 (Applied Biosystems).

Southern hybridizations

Southern hybridization was performed using a modified protocol based on Sambrook et al. [71]. Two μg DNA (containing host amoeba and *A. asiaticus* DNA) were digested with Eco32I for all investigated IS elements, except for ISCa2, for which DNA was digested with HindIII and subsequently separated on a 0.7% TAE agarose gel (4°C, 17 h, 30 V). The gel was depurinated for 10 min in 0.25 M HCl, denatured for 30 min in 1.5 M NaCl/0.5 M NaOH and neutralized for 30 min in 1.5 M NaCl/1 M Tris-HCl (pH 7.5). Between each of these steps the gel was briefly rinsed in ddH_2O . DNA was transferred onto Hybond N⁺ nylon membranes (GE Healthcare) with a vacuum transfer system and 20 \times SSC (3M NaCl, 0.3M sodium citrate, pH 7.0) as transfer buffer for 30 min. After immobilizing the DNA by UV cross-linking (120000 $\mu\text{J cm}^{-2}$), the membrane was briefly rinsed in ddH_2O . Pre-hybridization was carried out for 2 h at 42°C in hybridization buffer (containing 50% formamide, 5 \times SSC, 2% blocking reagent (Roche), 0.1% N-lauroyl sarcosyl sodium salt, 0.02% SDS (v/v)) in a rotation hybridization chamber as the following steps. The blot was hybridized with digoxigenin (DIG)-labeled probes (synthesized using the PCR DIG Probe Synthesis Kit, Roche; each probe was specific for a single IS element; see Additional file 2, Table S1) and hybridization buffer over night at 42°C. The membrane was washed twice for 15 min each with 2 \times SSC/0.1% SDS at 25°C, and twice with 0.2 \times SSC/0.1% SDS at 60°C for 15 min, followed by 2 min with DIG washing buffer (0.5 M maleic acid, 0.75 M NaCl, 0.3% Tween 20, pH 7.5) at 25°C, 30 min with buffer 2 (0.5 M maleic acid, 0.75 M NaCl, 0.3% Tween 20, 20% blocking reagent) at 25°C, 30

min with buffer 2 and Anti-Digoxigenin-AP Fab fragments (1:10000) at 25°C, twice for 15 min with DIG washing buffer at 25°C and finally for 5 min in 100 mM Tris/100 mM NaCl/50 mM MgCl₂ (pH 9.5) at 25°C. The membrane was swayed for 1 min in 1% CSPD solution (Roche) and subsequently exposed to Amersham Hyperfilm™ ECL (GE Healthcare).

Amplification of 16S and 18S rRNA genes

Oligonucleotide primers targeting 16S rRNA or 18S rRNA gene signature regions were used for PCR to obtain near full-length bacterial 16S rRNA or amoeba 18S rRNA gene fragments of the novel isolates *Acanthamoeba* sp. WR (containing *A. asiaticus* WR) and *Acanthamoeba* sp. US1 (containing *A. asiaticus* US1); see Additional file 2, Table S1. Nucleotide sequences of DNA fragments were determined on an ABI 3130 XL genetic analyzer using the BigDye Terminator kit v3.1 (Applied Biosystems).

Nucleotide sequence accession numbers

Obtained nucleotide sequences of IS elements of *A. asiaticus* EIDS3 and 16S and 18S rRNA genes of the isolates *Acanthamoeba* sp. WR (containing *A. asiaticus* WR) and *Acanthamoeba* sp. US1 (containing *A. asiaticus* US1) were submitted to EMBL/DDBJ/GenBank under accession numbers HM159367 to HM159370. The sequences of the *A. asiaticus* IS elements were deposited at EMBL/DDBJ/GenBank under accession numbers HM159371 to HM159380 and the ISFinder database [http://www-is.biotoul.fr/\[43\]](http://www-is.biotoul.fr/[43]).

Additional material

Additional file 1: pdf-file containing Figures S1 to S12.

Additional file 2: pdf-file containing Tables S1 to S32.

Acknowledgements

We are grateful to Daniela Teichmann and Christa Schleper (Department for Genetics in Ecology, University of Vienna) for providing protocols and helpful discussions. Technical assistance by Gabriele Schwammel, Diana Perez-Lopez and Christian Baranyi is greatly acknowledged. This work was supported by grants from the Austrian Science Fund (FWF) to SSE (grant no. P22703-B17) and MH (Y277-B03).

Author details

¹Department of Microbial Ecology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria. ²Department of Genetics in Ecology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria. ³Institute for Milk Hygiene, University of Veterinary Medicine Vienna Veterinärplatz 1, 1210 Vienna, Austria.

Authors' contributions

SSE and MH designed the study. SSE, TP and AS performed sequence analyses; TP and AS carried out the molecular biology experiments. SSE and MH wrote the manuscript; all authors read, edited, and approved the final manuscript.

Received: 17 May 2011 Accepted: 26 September 2011

Published: 26 September 2011

References

- Frost LS, Leplae R, Summers AO, Toussaint A: **Mobile genetic elements: the agents of open source evolution.** *Nat Rev Microbiol* 2005, **3**(9):722-732.
- Siguier P, Filee J, Chandler M: **Insertion sequences in prokaryotic genomes.** *Curr Opin Microbiol* 2006, **9**(5):526-531.
- Touchon M, Rocha EP: **Causes of insertion sequences abundance in prokaryotic genomes.** *Mol Biol Evol* 2007, **24**(4):969-981.
- Wagner A, Lewis C, Bichsel M: **A survey of bacterial insertion sequences using IScan.** *Nucleic Acids Res* 2007, **35**(16):5284-5293.
- Aziz RK, Breitbart M, Edwards RA: **Transposases are the most abundant, most ubiquitous genes in nature.** *Nucleic Acids Res* 2010, **38**(13):4207-4217.
- Newton IL, Bordenstein SR: **Correlations Between Bacterial Ecology and Mobile DNA.** *Curr Microbiol* 2011, **62**(1):198-208.
- Rocha EP: **Order and disorder in bacterial genomes.** *Curr Opin Microbiol* 2004, **7**(5):519-527.
- Moran NA, Plague GR: **Genomic changes following host restriction in bacteria.** *Curr Opin Genet Dev* 2004, **14**(6):627-633.
- Klasson L, Westberg J, Sapountzis P, Naslund K, Lutnaes Y, Darby AC, Veneti Z, Chen L, Braig HR, Garrett R, et al: **The mosaic genome structure of the *Wolbachia* wRi strain infecting *Drosophila simulans*.** *Proc Natl Acad Sci USA* 2009, **106**(14):5725-5730.
- Tan HM: **Bacterial catabolic transposons.** *Appl Microbiol Biotechnol* 1999, **51**(1):1-12.
- Wagner A: **Cooperation is fleeting in the world of transposable elements.** *PLoS Comput Biol* 2006, **2**(12):e162.
- de Visser JA, Akkermans AD, Hoekstra RF, de Vos WM: **Insertion-sequence-mediated mutations isolated during adaptation to growth and starvation in *Lactococcus lactis*.** *Genetics* 2004, **168**(3):1145-1157.
- Riehle MM, Bennett AF, Long AD: **Genetic architecture of thermal adaptation in *Escherichia coli*.** *Proc Natl Acad Sci USA* 2001, **98**(2):525-530.
- Schneider D, Duperchy E, Coursange E, Lenski RE, Blot M: **Long-term experimental evolution in *Escherichia coli*. IX. Characterization of insertion sequence-mediated mutations and rearrangements.** *Genetics* 2000, **156**(2):477-488.
- Treves DS, Manning S, Adams J: **Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*.** *Mol Biol Evol* 1998, **15**(7):789-797.
- Mahillon J, Chandler M: **Insertion sequences.** *Microbiol Mol Biol Rev* 1998, **62**(3):725-774.
- Carlson PE Jr, Horzempa J, O'Dee DM, Robinson CM, Neophytou P, Labrinidis A, Nau GJ: **Global transcriptional response to spermine, a component of the intramacrophage environment, reveals regulation of *Francisella* gene expression through insertion sequence elements.** *J Bacteriol* 2009, **191**(22):6855-6864.
- Ciampi MS, Schmid MB, Roth JR: **Transposon Tn10 provides a promoter for transcription of adjacent sequences.** *Proc Natl Acad Sci USA* 1982, **79**(16):5016-5020.
- Kallastu A, Horak R, Kivisaar M: **Identification and characterization of IS1411, a new insertion sequence which causes transcriptional activation of the phenol degradation genes in *Pseudomonas putida*.** *J Bacteriol* 1998, **180**(20):5306-5312.
- Lin H, Li TY, Xie MH, Zhang Y: **Characterization of the variants, flanking genes, and promoter activity of the *Leifsonia xyli* subsp. *cynodontis* insertion sequence IS1237.** *J Bacteriol* 2007, **189**(8):3217-3227.
- Safi H, Barnes PF, Lakey DL, Shams H, Samten B, Vankayalapati R, Howard ST: **IS6110 functions as a mobile, monocyte-activated promoter in *Mycobacterium tuberculosis*.** *Mol Microbiol* 2004, **52**(4):999-1012.
- Han HJ, Kuwae A, Abe A, Arakawa Y, Kamachi K: **Differential Expression of Type III Effector BteA Protein Due to IS481 Insertion in *Bordetella pertussis*.** *PLoS One* 2011, **6**(3):e17797.
- Bordenstein SR, Reznikoff WS: **Mobile DNA in obligate intracellular bacteria.** *Nat Rev Microbiol* 2005, **3**(9):688-699.
- Moya A, Pereto J, Gil R, Latorre A: **Learning how to live together: genomic insights into prokaryote-animal symbioses.** *Nat Rev Genet* 2008, **9**(3):218-229.
- Song H, Hwang J, Yi H, Ulrich RL, Yu Y, Nierman WC, Kim HS: **The early stage of bacterial genome-reductive evolution in the host.** *PLoS Pathog* 2010, **6**(5):e1000922.

26. Plague GR, Dunbar HE, Tran PL, Moran NA: **Extensive proliferation of transposable elements in heritable bacterial symbionts.** *J Bacteriol* 2008, **190**(2):777-779.
27. Yang F, Yang J, Zhang X, Chen L, Jiang Y, Yan Y, Tang X, Wang J, Xiong Z, Dong J, *et al*: **Genome dynamics and diversity of *Shigella* species, the etiologic agents of bacillary dysentery.** *Nucleic Acids Res* 2005, **33**(19):6445-6458.
28. Cho NH, Kim HR, Lee JH, Kim SY, Kim J, Cha S, Kim SY, Darby AC, Fuxelius HH, Yin J, *et al*: **The *Orientia tsutsugamushi* genome reveals massive proliferation of conjugative type IV secretion system and host-cell interaction genes.** *Proc Natl Acad Sci USA* 2007, **104**(19):7981-7986.
29. Nakayama K, Yamashita A, Kurokawa K, Morimoto T, Ogawa M, Fukuhara M, Urakami H, Ohnishi M, Uchiyama I, Ogura Y, *et al*: **The Whole-genome sequencing of the obligate intracellular bacterium *Orientia tsutsugamushi* revealed massive gene amplification during reductive genome evolution.** *DNA Res* 2008, **15**(4):185-199.
30. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N, *et al*: **Phylogenomics of the reproductive parasite *Wolbachia pipiens* wMel: a streamlined genome overrun by mobile genetic elements.** *PLoS Biol* 2004, **2**(3):E69.
31. Klasson L, Walker T, Sebahia M, Sanders MJ, Quail MA, Lord A, Sanders S, Earl J, O'Neill SL, Thomson N, *et al*: **Genome Evolution of *Wolbachia* Strain wPip from the *Culex pipiens* Group.** *Mol Biol Evol* 2008, **25**(9):1877-1887.
32. Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO, Boffelli D, Anderson IJ, Barry KW, Shapiro HJ, *et al*: **Symbiosis insights through metagenomic analysis of a microbial consortium.** *Nature* 2006, **443**(7114):950-955.
33. Gil R, Belda E, Gosalbes MJ, Delaye L, Vallier A, Vincent-Monegat C, Heddi A, Silva FJ, Moya A, Latorre A: **Massive presence of insertion sequences in the genome of SOPE, the primary endosymbiont of the rice weevil *Sitophilus oryzae*.** *Int Microbiol* 2008, **11**(1):41-48.
34. Schmitz-Esser S, Tischler P, Arnold R, Montanaro J, Wagner M, Rattei T, Horn M: **The genome of the amoeba symbiont "*Candidatus Amoebophilus asiaticus*" reveals common mechanisms for host cell interaction among amoeba-associated bacteria.** *J Bacteriol* 2010, **192**(4):1045-1057.
35. Schmitz-Esser S, Toenshoff ER, Haider S, Heinz E, Hoenninger VM, Wagner M, Horn M: **Diversity of bacterial endosymbionts of environmental *Acanthamoeba* isolates.** *Appl Environ Microbiol* 2008, **74**(18):5822-5831.
36. Choi SH, Cho MK, Ahn SC, Lee JE, Lee JS, Kim DH, Xuan YH, Hong YC, Kong HH, Chung DI, *et al*: **Endosymbionts of *Acanthamoeba* isolated from domestic tap water in Korea.** *Korean J Parasitol* 2009, **47**(4):337-344.
37. Horn M, Harzenetter MD, Linner T, Schmid EN, Muller KD, Michel R, Wagner M: **Members of the *Cytophaga-Flavobacterium-Bacteroides* phylum as intracellular bacteria of acanthamoebae: proposal of '*Candidatus Amoebophilus asiaticus*'.** *Environ Microbiol* 2001, **3**(7):440-449.
38. Xuan YH, Yu HS, Jeong HJ, Seol SY, Chung DI, Kong HH: **Molecular characterization of bacterial endosymbionts of *Acanthamoeba* isolates from infected corneas of Korean patients.** *Korean J Parasitol* 2007, **45**(1):1-9.
39. Zchori-Fein E, Perlman SJ, Kelly SE, Katzir N, Hunter MS: **Characterization of a '*Bacteroidetes*' symbiont in *Encarsia* wasps (Hymenoptera: Aphelinidae): proposal of '*Candidatus Cardinium hertigii*'.** *Int J Syst Evol Microbiol* 2004, **54**(Pt 3):961-968.
40. Sunagawa S, Woodley CM, Medina M: **Threatened corals provide underexplored microbial habitats.** *PLoS ONE* 2010, **5**(3):e9554.
41. Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D: **Massive comparative genomic analysis reveals convergent evolution of specialized bacteria.** *Biol Direct* 2009, **4**:13.
42. Moran NA, McCutcheon JP, Nakabachi A: **Genomics and evolution of heritable bacterial symbionts.** *Annu Rev Genet* 2008, **42**:165-190.
43. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M: **ISfinder: the reference centre for bacterial insertion sequences.** *Nucleic Acids Res* 2006, **34** Database: D32-36.
44. Ohta S, Tsuchida K, Choi S, Sekine Y, Shiga Y, Ohtsubo E: **Presence of a characteristic D-D-E motif in IS1 transposase.** *J Bacteriol* 2002, **184**(22):6146-6154.
45. Baranov PV, Fayet O, Hendrix RW, Atkins JF: **Recoding in bacteriophages and bacterial IS elements.** *Trends Genet* 2006, **22**(3):174-181.
46. Namy O, Rousset JP, Naphtine S, Brierley I: **Reprogrammed genetic decoding in cellular gene expression.** *Mol Cell* 2004, **13**(2):157-168.
47. Cordaux R: **ISWpi1 from *Wolbachia pipiens* defines a novel group of insertion sequences within the IS5 family.** *Gene* 2008, **409**(1-2):20-27.
48. Cordaux R, Pichon S, Ling A, Perez P, Delaunay C, Vavre F, Bouchon D, Greve P: **Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont.** *Mol Biol Evol* 2008, **25**(9):1889-1896.
49. Cridge AG, Major LL, Mahagaonkar AA, Poole ES, Isaksson LA, Tate WP: **Comparison of characteristics and function of translation termination signals between and within prokaryotic and eukaryotic organisms.** *Nucleic Acids Res* 2006, **34**(7):1959-1973.
50. Poole ES, Brown CM, Tate WP: **The identity of the base following the stop codon determines the efficiency of in vivo translational termination in *Escherichia coli*.** *Embo J* 1995, **14**(1):151-158.
51. Barabas O, Ronning DR, Guynet C, Hickman AB, Ton-Hoang B, Chandler M, Dyda F: **Mechanism of IS200/IS605 family DNA transposases: activation and transposon-directed target site selection.** *Cell* 2008, **132**(2):208-220.
52. Guynet C, Hickman AB, Barabas O, Dyda F, Chandler M, Ton-Hoang B: **In vitro reconstitution of a single-stranded transposition mechanism of IS608.** *Mol Cell* 2008, **29**(3):302-312.
53. Beuzon CR, Chessa D, Casadesus J: **IS200: an old and still bacterial transposon.** *Int Microbiol* 2004, **7**(1):3-12.
54. Martusewitsch E, Sensen CW, Schleper C: **High spontaneous mutation rate in the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by transposable elements.** *J Bacteriol* 2000, **182**(9):2574-2581.
55. Ogata H, La Scola B, Audic S, Renesto P, Blanc G, Robert C, Fournier PE, Claverie JM, Raoult D: **Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens.** *PLoS Genet* 2006, **2**(5):e76.
56. Liu Y, Whitman WB: **Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea.** *Ann N Y Acad Sci* 2008, **1125**:171-189.
57. Khan NA: ***Acanthamoeba*: biology and increasing importance in human health.** *FEMS Microbiol Rev* 2006, **30**(4):564-595.
58. Rodríguez-Zaragoza S: **Ecology of free living amoebae.** *Crit Rev Microbiol* 1994, **20**(3):225-241.
59. Wagner A, de la Chaux N: **Distant horizontal gene transfer is rare for multiple families of prokaryotic insertion sequences.** *Mol Genet Genomics* 2008, **280**(5):397-408.
60. Hooper SD, Mavromatis K, Kyrpides NC: **Microbial co-habitation and lateral gene transfer: what transposases can tell us.** *Genome Biol* 2009, **10**(4):R45.
61. Wagner A: **Periodic extinctions of transposable elements in bacterial lineages: evidence from intragenomic variation in multiple genomes.** *Mol Biol Evol* 2006, **23**(4):723-733.
62. Lawrence JG, Ochman H: **Amelioration of bacterial genomes: rates of change and exchange.** *J Mol Evol* 1997, **44**(4):383-397.
63. Callanan M, Kaleta P, O'Callaghan J, O'Sullivan O, Jordan K, McAuliffe O, Sangrador-Vegas A, Slattery L, Fitzgerald GF, Beresford T, *et al*: **Genome sequence of *Lactobacillus helveticus*, an organism distinguished by selective gene loss and insertion sequence element expansion.** *J Bacteriol* 2008, **190**(2):727-735.
64. Cordaux R: **Gene conversion maintains nonfunctional transposable elements in an obligate mutualistic endosymbiont.** *Mol Biol Evol* 2009, **26**(8):1679-1682.
65. Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, Bhattacharyya A, Kapatral V, Kumar S, Posfai J, *et al*: **The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode.** *PLoS Biol* 2005, **3**(4):e121.
66. Nagy Z, Chandler M: **Regulation of transposition in bacteria.** *Res Microbiol* 2004, **155**(5):387-398.
67. Kleckner N: **Regulating Tn10 and IS10 Transposition.** *Genetics* 1990, **124**(3):449-454.
68. Jager D, Sharma CM, Thomsen J, Ehlers C, Vogel J, Schmitz RA: **Deep sequencing analysis of the *Methanosarcina mazei* Go1 transcriptome in response to nitrogen availability.** *Proc Natl Acad Sci USA* 2009, **106**(51):21878-21882.
69. Perkins TT, Kingsley RA, Fookes MC, Gardner PP, James KD, Yu L, Assefa SA, He M, Croucher NJ, Pickard DJ, *et al*: **A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*.** *PLoS Genet* 2009, **5**(7):e1000569.

70. Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons BA, Sorek R: **A single-base resolution map of an archaeal transcriptome.** *Genome Res* 2010, **20**(1):133-141.
71. Sambrook J, Russell DW: **Molecular cloning: a laboratory manual.** Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press,; 3 2001.
72. Twiss E, Coros AM, Tavakoli NP, Derbyshire KM: **Transposition is modulated by a diverse set of host factors in *Escherichia coli* and is stimulated by nutritional stress.** *Mol Microbiol* 2005, **57**(6):1593-1607.
73. Bayley DP, Rocha ER, Smith CJ: **Analysis of *cepA* and other *Bacteroides fragilis* genes reveals a unique promoter structure.** *FEMS Microbiol Lett* 2000, **193**(1):149-154.
74. Chen S, Bagdasarian M, Kaufman MG, Bates AK, Walker ED: **Mutational analysis of the *ompA* promoter from *Flavobacterium johnsoniae*.** *J Bacteriol* 2007, **189**(14):5108-5118.
75. Mittl PR, Schneider-Brachert W: **Sel1-like repeat proteins in signal transduction.** *Cell Signal* 2007, **19**(1):20-31.
76. Frishman D, Albermann K, Hani J, Heumann K, Metanomski A, Zollner A, Mewes HW: **Functional and structural genomics using PEDANT.** *Bioinformatics* 2001, **17**(1):44-57.
77. Rice P, Longden I, Bleasby A: **EMBOSS: the European Molecular Biology Open Software Suite.** *Trends Genet* 2000, **16**(6):276-277.
78. Katoh K, Toh H: **Recent developments in the MAFFT multiple sequence alignment program.** *Brief Bioinform* 2008, **9**(4):286-298.
79. Zuker M: **Mfold web server for nucleic acid folding and hybridization prediction.** *Nucleic Acids Res* 2003, **31**(13):3406-3415.
80. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Buchner A, Lai T, Steppi S, Jobb G, *et al*: **ARB: a software environment for sequence data.** *Nucleic Acids Res* 2004, **32**(4):1363-1371.
81. Felsenstein J: **PHYLIP - Phylogeny inference package (version 3.2).** *Cladistics* 1989, **5**:164-166.
82. Schmidt HA, Strimmer K, Vingron M, von Haeseler A: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing.** *Bioinformatics* 2002, **18**(3):502-504.
83. Heinz E, Kolarov I, Kastner C, Toenshoff ER, Wagner M, Horn M: **An *Acanthamoeba* sp. containing two phylogenetically different bacterial endosymbionts.** *Environ Microbiol* 2007, **9**(6):1604-1609.
84. Zhou J, Bruns MA, Tiedje JM: **DNA recovery from soils of diverse composition.** *Appl Environ Microbiol* 1996, **62**(2):316-322.

doi:10.1186/1471-2148-11-270

Cite this article as: Schmitz-Esser *et al*: A bacterial genome in transition - an exceptional enrichment of IS elements but lack of evidence for recent transposition in the symbiont *Amoebophilus asiaticus*. *BMC Evolutionary Biology* 2011 **11**:270.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



Chapter IV

The genome of the amoeba symbiont "*Candidatus Amoebophilus asiaticus*" encodes an afp-like prophage possibly used for protein secretion

published in Virulence. 2010 Nov-Dec;1(6):541-5. PMID: 21178499

The genome of the amoeba symbiont “*Candidatus Amoebophilus asiaticus*” encodes an afp-like prophage possibly used for protein secretion

Thomas Penz, Matthias Horn and Stephan Schmitz-Esser*[†]

Department of Microbial Ecology; University of Vienna; Vienna, Austria

[†]Current address: Institute for Milk Hygiene; University of Veterinary Medicine; Vienna, Austria

The recently sequenced genome of the obligate intracellular amoeba symbiont “*Candidatus Amoebophilus asiaticus*” is unique among prokaryotic genomes due to its extremely large fraction of genes encoding proteins harboring eukaryotic domains such as ankyrin-repeats, TPR/SEL1 repeats, leucine-rich repeats, as well as F- and U-box domains, most of which likely serve in the interaction with the amoeba host. Here we provide evidence for the presence of additional proteins, which are presumably presented extracellularly and should thus also be important for host cell interaction. Surprisingly, we did not find homologs of any of the well-known protein secretion systems required to translocate effector proteins into the host cell in the *A. asiaticus* genome, and the type six secretion system seems to be incomplete. Here we describe the presence of a putative prophage in the *A. asiaticus* genome, which shows similarity to the antifeeding prophage from the insect pathogen *Serratia entomophila*. In *S. entomophila* this system is used to deliver toxins into insect hosts. This putative antifeeding-like prophage might thus represent the missing protein secretion apparatus in *A. asiaticus*.

bacteria (including also many pathogens) have been shown to be able to survive and multiply within protozoa.³⁻⁶ Protozoa such as free-living amoebae may thus serve as environmental reservoir as well as vectors for the transmission of pathogenic bacteria to humans and might even represent evolutionary training grounds facilitating the adaptation of bacteria to survival within eukaryotic cells.^{5,7-9} Most of these amoeba-associated bacteria show a facultative intracellular lifestyle; however, bacteria with an obligate intracellular lifestyle also can be found thriving in protozoa: Bacteria belonging to three different phyla—the Proteobacteria, the Chlamydiae and the Bacteroidetes—have been described as obligate intracellular amoeba symbionts in the last few years.^{3-5,10}

The obligate intracellular Acanthamoeba symbiont “*Candidatus Amoebophilus asiaticus*” (in the following referred to as “*A. asiaticus*”) is a representative of obligate intracellular amoeba symbionts affiliating to the Bacteroidetes phylum.¹¹⁻¹³ The genome of *A. asiaticus* strain 5a2 has a size of 1.89 Mbp, a G+C content of 35%, and it encodes 1,557 proteins with a coding density of 81.8%.¹² Compared to the genome size of many other obligate intracellular bacteria, the genome of *A. asiaticus* is only moderately reduced in size but has extremely limited biosynthetic capabilities. In order to compensate for its reduced biosynthetic capabilities, *A. asiaticus* has to take up essential nutrients from its Acanthamoeba host cell using a variety (n = 82) of different transport proteins, including several transporters for uptake

Key words: *Amoebophilus asiaticus*, endosymbiont, bacteroidetes, acanthamoeba, *Serratia entomophila* antifeeding prophage, type six secretion system

Submitted: 09/17/10

Accepted: 09/24/10

Previously published online:
www.landesbioscience.com/journals/
virulence/article/13800

DOI: 10.4161/viru.1.6.13800

*Correspondence to: Stephan Schmitz-Esser;
Email: stephan.schmitz-esser@vetmeduni.ac.at

Addendum to: Schmitz-Esser S, Tischler P, Arnold R, Montanaro J, Wagner M, Rattei T, et al. The genome of the amoeba symbiont “*Candidatus Amoebophilus asiaticus*” reveals common mechanisms for host cell interaction among amoeba-associated bacteria. *J Bacteriol* 2010; 192:1045–57; PMID: 20023027; DOI: 10.1128/JB.01379-09.

Today numerous genomes of intracellular bacteria have been sequenced.^{1,2} Analyses of these genomes revealed both, common and different mechanisms for the interaction between intracellular bacteria and eukaryotic cells.^{1,2} Interestingly, an increasingly large number of intracellular

of oligopeptides and amino acids, cofactors and an ATP/ADP translocase.

A large fraction of the genome consists of a diverse arsenal of proteins most likely important for the interaction with its *Acanthamoeba* host cell. These proteins include—among others—five patatin-like proteins, two phospholipases D, a eukaryotic serine/threonine protein kinase, two proteins with similarity to insecticidal toxins (toxin complex) found in various *Photothabdus* spp. and other Gammaproteobacteria, a gene cluster encoding a putative lasso-peptide, and—compared to all other prokaryotic genomes—an extremely high number of proteins with domains predominantly found in eukaryotes (8% of all CDSs, $n = 129$). These eukaryotic domains—including ankyrin repeats, TPR/SEL1 repeats and leucine-rich repeats—are most likely targeted to the host cell and mediate protein-protein interaction.¹⁴⁻¹⁶ Proteins encoding such eukaryotic domains have been demonstrated to be important for the interaction of various intracellular bacterial pathogens with their eukaryotic host cells.¹⁶⁻²¹ *A. asiaticus* also encodes an extraordinary large number of proteins ($n = 26$) responsible for interference with the host ubiquitin system such as proteins with F-box and U-box domains and two putative ubiquitin-specific proteases. Proteins harboring F- and U-box domains transfer ubiquitin to target molecules, whereas ubiquitin proteases remove ubiquitin from target molecules—this combination of complementary proteins for ubiquitin interference has not been found among prokaryotic genomes so far.²² Interestingly, the two putative ubiquitin-specific proteases represent the first prokaryotic members of the CA clan C19 family of ubiquitin proteases, which is the most common family of eukaryotic ubiquitin proteases.^{22,23} Taken together, proteins with eukaryotic domains are probably exceptionally important for host cell manipulation in order to stimulate infection, intracellular survival and replication of *A. asiaticus* in amoebae.

In our previous analysis, we showed that proteins encoding eukaryotic domains are significantly enriched in *A. asiaticus* and in the genomes of other amoeba-associated bacteria compared to free-living

bacteria.¹² Bacteria that can exploit amoebae as hosts thus share a common set of eukaryotic protein domains important for host cell interaction despite their different life styles and their large phylogenetic diversity. Recently, additional genomes of amoeba-associated bacteria became available: several genomes of *Legionella longbeachae* strains^{24,25} and a draft genome of the obligate intracellular amoeba symbiont *Parachlamydia acanthamoebae* str. Hall's coccus.²⁶ Here we analyzed whether the trend that proteins with eukaryotic domains are overrepresented in amoeba-associated bacteria is also apparent in these recently sequenced genomes. An accumulation of proteins with eukaryotic domains is indeed also evident in these genomes, even though not to such a high degree as in *A. asiaticus* (Sup. Table 1)—corroborating our previous results. The actual numbers of proteins harboring eukaryotic domains might be higher in these genomes as only draft sequences are available for *L. longbeachae* D-4968 (and other *L. longbeachae* genomes)²⁵ and *P. acanthamoebae* str. Hall's coccus.²⁶

A crucial step for intracellular bacteria is the attachment prior to entry into a eukaryotic host cell. Proteins that are displayed in the outer membrane and exposed to the extracellular space are thus important for cell-recognition and adhesion. Some of these proteins show biased and repetitive amino acid composition and are frequently glycosylated.²⁷⁻³⁰ We systematically searched the *A. asiaticus* genome for predicted proteins with an elevated level of either one or more of following amino acids: Ser, Thr, Ala, Glu, Gln, Lys—compared to all other *A. asiaticus* proteins. In total we identified 14 proteins with an atypical amino acid composition. Based on sequence similarities to known proteins nine of these are likely displayed extracellularly. One of them (Aasi_0548) shows a highly enriched serine/threonine content (18% Ser and Thr combined). Aasi_0548 is a 1,448 amino acid protein containing a predicted signal peptide and no predicted transmembrane helices; it is thus most likely not an integral membrane protein. Immunogenic (extracellular) glycoproteins with enriched serine/threonine content have been described in many

bacterial pathogens such as *Ehrlichia* spp., *Streptococcus* spp., or *Staphylococcus aureus*.³¹⁻³³ Aasi_0548 has several predicted O-glycosylation sites (at Ser/Thr residues) and might thus also represent a glycoprotein. In addition, it shows weak (22%) amino acid identity to surface proteins from *Mycoplasma* spp. (Lmp1, surface-located membrane protein; Msp, massive surface protein family).^{34,35} Interestingly, seven Aasi_0548-homologs are present in the *A. asiaticus* genome—however they might represent pseudogenes as they are shorter or disrupted by IS elements.

Another protein that might be involved cell-recognition and adhesion is Aasi_1714, an 891 amino acid protein which shows highest amino acid identities to adhesins from eukaryotic and bacterial pathogens such as *vmcA* from *Mycoplasma* spp. (~30%),²⁹ a 200 kDa antigen from *Babesia bigemina* (~25%),³⁶ or to the MSP3 proteins (merozoite surface protein 3 family) from *Plasmodium vivax* (~30%).³⁰ The amino acid composition of Aasi_1714 is highly biased towards alanine (20.3%), glutamate (12.1%), glutamine (11.2%) and lysine (12%); these four amino acids together make up 55% of the whole amino acid residues of Aasi_1714. Also, 14 copies of a 26 amino acid repeat unit have been identified in Aasi_1714; in addition, two (identical) copies of 38 and 63 amino acid residue repeats are present. Furthermore, Aasi_1714 is predicted to form a similar secondary and tertiary structure as the *Plasmodium* MSP3 proteins: high numbers of α -helices and coiled-coil regions in an alanine-rich core region (29% alanine content in Aasi_1714; 31% in MSP3). In conclusion, the similarities between Aasi_0548 and Aasi_1714 to other extracellular or surface proteins—particularly to MSP3 proteins (for Aasi_1714)—suggest that Aasi_0548 as well as Aasi_1714 are also surface-exposed and therefore important for host cell interaction.

Bacteria associated with eukaryotic cells (showing either a mutualistic, commensal or parasitic lifestyle) generally encode various secretion systems to export bacterial effector proteins and to translocate them into their host cells.³⁷ Despite the high number of putative effector proteins mediating host cell interaction in *A. asiaticus*, we did not find homologs of

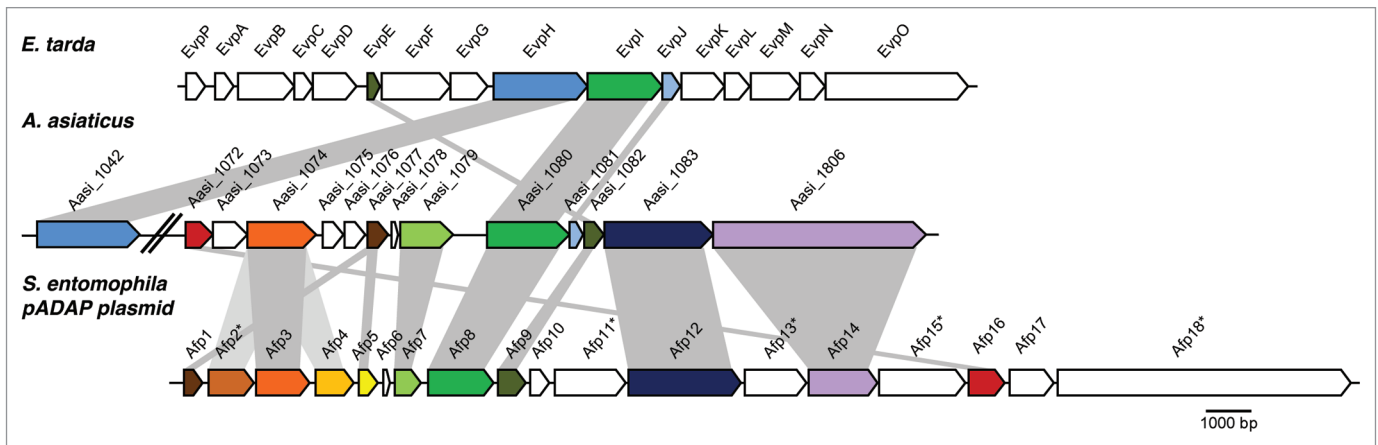


Figure 1. Schematic representation of the genomic organization of the *afp*-like *A. asiaticus* gene cluster compared to the *Serratia entomophila* anti-feeding prophage (*afp*) gene cluster on the pADAP plasmid and the type six secretion system gene cluster of *Edwardsiella tarda*. *A. asiaticus* locus tags and *S. entomophila* and *E. tarda* gene names are shown above the respective gene. Homologous proteins are depicted in the same color. An asterisk indicates proteins which have been shown to be essential for pathogenicity in *S. entomophila*. Afp17 and Afp18 most likely represent toxins and have no homologs outside *Serratia* spp.

these well known protein secretion systems in the *A. asiaticus* genome, except for the sec-dependent pathway for protein export across the inner membrane. We did also not identify a homolog of the recently described Bacteroidetes-specific Por secretion system, although *A. asiaticus* encodes 13 gliding motility genes, some of which are part of the Por secretion system.³⁸ The Por secretion system has been described in *Porphyromonas gingivalis*. In this microorganism 11 Por proteins are essential for the secretion of cell surface proteases (gingipains). These include six proteins with similarity to gliding motility proteins (only three are found in *A. asiaticus*) and five additional loci, which have no homologs in *A. asiaticus*. *A. asiaticus* thus does not encode a functional Por secretion system.

We previously identified a gene cluster consisting of 13 genes (Aasi_1072 to Aasi_1806) with predicted operon structure including one protein (Aasi_1081) showing highest similarity (63% amino acid identity) to EvpJ,¹² a protein from the type six secretion system (T6SS) of the human and fish pathogen *Edwardsiella tarda* (Suppl. Table 2).³⁹ T6SS have only relatively recently been described to be widely—and almost exclusively—distributed among the Proteobacteria and to be organized in operons of 15 to 20 genes.⁴⁰ Based on the high similarity of Aasi_1081 to the T6SS protein EvpJ, we speculated

that this operon found in *A. asiaticus* represents a putative—although divergent—T6SS.¹² To further test this hypothesis, we performed more detailed comparative sequence analyses of this gene cluster with known T6SS. Five of these proteins indeed show weak similarity to some core components of T6SS proteins (Fig. 1 and Sup. Table 2): In addition to Aasi_1081 (EvpJ, a putative effector protein³⁹), we detected a distant homolog of EvpI (VgrG; Aasi_1080), which could assemble into a membrane-penetrating device;⁴¹ and three phage-tail-associated proteins described as T6SS components are present (Aasi_1074: gp18, Aasi_1077: gp19, Aasi_1082: gp25).⁴² Furthermore, a homolog of EvpH (Aasi_1042), which is found in many but not all T6SS gene clusters,⁴³ is also present in the genome of *A. asiaticus*, but 32 kb upstream of the putative T6SS cluster. However, the Aasi_1081 homolog EvpJ is not specific for T6SS systems, and we did not identify homologs of other T6SS proteins known to be essential in functionally characterized T6SS (e.g., EvpA, EvpB, EvpN, EvpO).^{39,43}

We noted that eight of the 13 genes in this gene cluster show similarity to proteins from the anti-feeding prophage (*afp*) of the plasmid pADAP (“amber disease associated plasmid”) in the insect pathogen *Serratia entomophila* (Fig. 1 and Sup. Table 2).⁴⁴ In addition, the *A. asiaticus* gene cluster and the *afp* gene cluster are

largely syntenic (Fig. 1). In *S. entomophila* the *afp* gene cluster encodes 18 proteins, 16 of which have homologs outside *Serratia* spp., and it represents a prophage morphologically resembling phage-tail (R-type) bacteriocins.^{44–47} However, in contrast to known R-type pyocins, the *afp* prophage of *S. entomophila* does not show antibacterial activity but delivers toxins into the eukaryotic host of *S. entomophila*, larvae of the New Zealand grass grub *Costelytra zealandica* (Coleoptera). After entry of *S. entomophila* into the larval gut, which is normally dark in color, the gut clears, the insects turn amber and the level of major digestive enzymes of the gut decrease, the infected larvae stay in this state for up to three months before *S. entomophila* finally invades the hemocoel leading to rapid death of the insect.^{44–47} Similar eukaryotic toxin-encoding prophages have been identified in various *Photobacterium* spp. and, surprisingly, also in a number of marine bacteria with no known association with infectious disease or parasitic life style, including many Bacteroidetes.^{44,47,48} The *A. asiaticus* gene cluster lacks few genes reported to be essential for pathogenicity of *S. entomophila* (Afp15, Afp11 and Afp13). Nevertheless, based on the synteny of the *A. asiaticus* gene cluster with the *afp*-like prophage of *S. entomophila* and the higher sequence similarities of the encoded proteins to phage proteins compared to T6SS proteins, we conclude that

the *A. asiaticus* gene cluster resembles afp-like prophages rather than known T6SS.

Neither G+C content nor G+C skew of the afp-like gene cluster region of *A. asiaticus* show significant deviations from the surrounding genomic regions and thus give no hints for a recent acquisition of the afp-like gene cluster from a donor with a significantly different genomic G+C content than *A. asiaticus*. This might indicate that this region has been part of the *A. asiaticus* genome for evolutionary longer times. Recently, a structural relationship between T4 bacteriophages, R-type pyocins, afp-like prophages and T6SS was noted, and it has been suggested that the T6SS and phage tail-associated protein complexes share a common evolutionary origin.^{41,42} Taking into account the similarity of the afp-like gene cluster of *A. asiaticus* with the afp-like prophage of *S. entomophila*, which functions as a protein secretion apparatus, it seems likely that the afp-like gene cluster of *A. asiaticus* represents a protein secretion apparatus, which—similar to known T6SS—originated from a tailed bacteriophage. Taken together, we propose that the *A. asiaticus* afp-like gene cluster represents a prophage acquired early in evolution, which developed into a chromosomally encoded protein secretion apparatus used for the delivery of effector proteins into the amoeba host cell. This putative secretion apparatus might either be used for the delivery of specific effector(s) encoded within the afp-gene cluster (e.g., Aasi_1081), or as a more generally used secretion system for the numerous *A. asiaticus* proteins harboring eukaryotic domains.

Acknowledgements

This work was supported by grants of the Austrian Science Fund FWF to S.S.E. and M.H. (P22703-B17, Y277-B03).

Note

Supplementary materials can be found at: www.landesbioscience.com/supplement/PenzVIRUI-6-Sup.pdf

References

- Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biol Direct* 2009; 4:13.
- Toft C, Andersson SG. Evolutionary microbial genomics: Insights into bacterial host adaptation. *Nature Rev Genet* 2010; 11:465-75.
- Greub G, Raoult D. Microorganisms resistant to free-living amoebae. *Clin Microbiol Rev* 2004; 17:413-33.
- Horn M, Wagner M. Bacterial endosymbionts of free-living amoebae. *J Eukaryot Microbiol* 2004; 51:509-14.
- Molmeret M, Horn M, Wagner M, Santic M, Abu Kwaik Y. Amoebae as training grounds for intracellular bacterial pathogens. *Appl Environ Microbiol* 2005; 71:20-8.
- Thomas V, McDonnell G, Denyer SP, Maillard JY. Free-living amoebae and their intracellular pathogenic microorganisms: risks for water quality. *FEMS Microbiol Rev* 2009; PMID: 19744244.
- Albert-Weissenberger C, Cazalet C, Buchrieser C. *Legionella pneumophila*—a human pathogen that co-evolved with fresh water protozoa. *Cell Mol Life Sci* 2007; 64:432-48.
- Barker J, Brown MRW. Trojan horses of the microbial world: Protozoa and the survival of bacterial pathogens in the environment. *Microbiology* 1994; 140:1253-9.
- Harb OS, Gao LY, Kwaik YA. From protozoa to mammalian cells: a new paradigm in the life cycle of intracellular bacterial pathogens. *Environ Microbiol* 2000; 2:251-65.
- Horn M. Chlamydiae as symbionts in eukaryotes. *Annu Rev Microbiol* 2008; 62:113-31.
- Horn M, Harzenetter MD, Linner T, Schmid EN, Muller KD, Michel R, et al. Members of the Cytophaga-Flavobacterium-Bacteroides phylum as intracellular bacteria of acanthamoebae: proposal of *Candidatus Amoebophilus asiaticus*. *Environ Microbiol* 2001; 3:440-9.
- Schmitz-Esser S, Tischler P, Arnold R, Montanaro J, Wagner M, Rattai T, et al. The genome of the amoeba symbiont "*Candidatus Amoebophilus asiaticus*" reveals common mechanisms for host cell interaction among amoeba-associated bacteria. *J Bacteriol* 2010; 192:1045-57.
- Schmitz-Esser S, Toenshoff ER, Haider S, Heinz E, Hoenninger VM, Wagner M, et al. Diversity of bacterial endosymbionts of environmental Acanthamoeba isolates. *Appl Environ Microbiol* 2008; 74:5822-31.
- Bella J, Hindle KL, McEwan PA, Lovell SC. The leucine-rich repeat structure. *Cell Mol Life Sci* 2008; 65:2307-33.
- Mittl PR, Schneider-Brachert W. Sell-like repeat proteins in signal transduction. *Cell Signal* 2007; 19:20-31.
- Al-Khodor S, Price CT, Kalia A, Abu Kwaik Y. Functional diversity of ankyrin repeats in microbial proteins. *Trends Microbiol* 2010; 18:132-9.
- Al-Khodor S, Price CT, Habyarimana F, Kalia A, Abu Kwaik Y. A Dot/Icm-translocated ankyrin protein of *Legionella pneumophila* is required for intracellular proliferation within human macrophages and protozoa. *Mol Microbiol* 2008; 70:908-23.
- Habyarimana F, Al-Khodor S, Kalia A, Graham JE, Price CT, Garcia MT, et al. Role for the Ankyrin eukaryotic-like genes of *Legionella pneumophila* in parasitism of protozoan hosts and human macrophages. *Environ Microbiol* 2008; 10:1460-74.
- Kubori T, Hyakutake A, Nagai H. Legionella translocates an E3 ubiquitin ligase that has multiple U-boxes with distinct functions. *Mol Microbiol* 2008; 67:1307-19.
- Price CT, Al-Khodor S, Al-Quadan T, Santic M, Habyarimana F, Kalia A, et al. Molecular mimicry by an F-box effector of *Legionella pneumophila* hijacks a conserved polyubiquitination machinery within macrophages and protozoa. *PLoS Pathogens* 2009; 5:1000704.
- Pan X, Luhrmann A, Satoh A, Laskowski-Arce MA, Roy CR. Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* 2008; 320:1651-4.
- Rytönen A, Holden DW. Bacterial interference of ubiquitination and deubiquitination. *Cell Host Microbe* 2007; 1:13-22.
- Nijman SM, Luna-Vargas MP, Velds A, Brummelkamp TR, Dirac AM, Sixma TK, et al. A genomic and functional inventory of deubiquitinating enzymes. *Cell* 2005; 123:773-86.
- Cazalet C, Gomez-Valero L, Rusniok C, Lomma M, Dervins-Ravault D, Newton HJ, et al. Analysis of the *Legionella longbeachae* genome and transcriptome uncovers unique strategies to cause Legionnaires' disease. *PLoS Genetics* 2010; 6:1000851.
- Kozak NA, Buss M, Lucas CE, Frace M, Govil D, Travis T, et al. Virulence factors encoded by *Legionella longbeachae* identified on the basis of the genome sequence analysis of clinical isolate D-4968. *J Bacteriol* 2010; 192:1030-44.
- Greub G, Kebbi-Beghdadi C, Bertelli C, Collyn F, Riederer BM, Yersin C, et al. High throughput sequencing and proteomics to identify immunogenic proteins of a new pathogen: the dirty genome approach. *PLoS ONE* 2009; 4:8423.
- Upreti RK, Kumar M, Shankar V. Bacterial glycoproteins: functions, biosynthesis and applications. *Proteomics* 2003; 3:363-79.
- Zhou M, Wu H. Glycosylation and biogenesis of a family of serine-rich bacterial adhesins. *Microbiology* 2009; 155:317-27.
- Wise KS, Foecking MF, Roske K, Lee YJ, Lee YM, Madan A, et al. Distinctive repertoire of contingency genes conferring mutation-based phase variation and combinatorial expression of surface lipoproteins in *Mycoplasma capricolum* subsp. *capricolum* of the *Mycoplasma mycoides* phylogenetic cluster. *J Bacteriol* 2006; 188:4926-41.
- Galinski MR, Ingravallo P, Corredor-Medina C, Al-Khedery B, Pova M, Barnwell JW. *Plasmodium vivax* merozoite surface proteins-3β and -3γ share structural similarities with *P. vivax* merozoite surface protein-3α and define a new gene family. *Mol Biochem Parasitol* 2001; 115:41-53.
- Li Y, Chen Y, Huang X, Zhou M, Wu R, Dong S, et al. A conserved domain of previously unknown function in Gap1 mediates protein-protein interaction and is required for biogenesis of a serine-rich streptococcal adhesin. *Mol Microbiol* 2008; 70:1094-104.
- McBride JW, Doyle CK, Zhang X, Cardenas AM, Popov VL, Nethery KA, et al. Identification of a glycosylated *Ehrlichia canis* 19-kilodalton major immunoreactive protein with a species-specific serine-rich glycopeptide epitope. *Infect Immun* 2007; 75:74-82.
- Seifert KN, Adderson EE, Whiting AA, Bohnsack JF, Crowley PJ, Brady LJ. A unique serine-rich repeat protein (Str-2) and novel surface antigen (epsilon) associated with a virulent lineage of serotype III *Streptococcus agalactiae*. *Microbiology* 2006; 152:1029-40.
- Ladefoged SA, Birkelund S, Hauge S, Brock B, Jensen LT, Christiansen G. A 135-kilodalton surface antigen of *Mycoplasma hominis* PG21 contains multiple directly repeated sequences. *Infect Immun* 1995; 63:212-23.
- Dybvig K, Zuhua C, Lao P, Jordan DS, French CT, Tu AH, et al. Genome of *Mycoplasma arthritidis*. *Infect Immun* 2008; 76:4000-8.
- Tebele N, Skilton RA, Katende J, Wells CW, Nene V, McElwain T, et al. Cloning, characterization and expression of a 200-kilodalton diagnostic antigen of *Babesia bigemina*. *J Clin Microbiol* 2000; 38:2240-7.
- Tseng TT, Tyler BM, Setubal JC. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol* 2009; 9:2.

38. Sato K, Naito M, Yukitake H, Hirakawa H, Shoji M, McBride MJ, et al. A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proc Natl Acad Sci USA* 2010; 107:276-81.
39. Zheng J, Leung KY. Dissection of a type VI secretion system in *Edwardsiella tarda*. *Mol Microbiol* 2007; 66:1192-206.
40. Pukatzki S, McAuley SB, Miyata ST. The type VI secretion system: translocation of effectors and effector-domains. *Curr Opin Microbiol* 2009; 12:11-7.
41. Bonemann G, Pietrosiuk A, Mogk A. Tubules and donuts: a type VI secretion story. *Mol Microbiol* 2010; 76:815-21.
42. Leiman PG, Basler M, Ramagopal UA, Bonanno JB, Sauder JM, Pukatzki S, et al. Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc Natl Acad Sci USA* 2009; 106:4154-9.
43. Boyer F, Fichant G, Berthod J, Vandenbrouck Y, Attree I. Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? *BMC Genomics* 2009; 10:104.
44. Hurst MR, Glare TR, Jackson TA. Cloning *Serratia entomophila* antifeeding genes—a putative defective prophage active against the grass grub *Costelytra zealandica*. *J Bacteriol* 2004; 186:5116-28.
45. Hurst MR, Beard SS, Jackson TA, Jones SM. Isolation and characterization of the *Serratia entomophila* antifeeding prophage. *FEMS Microbiol Lett* 2007; 270:42-8.
46. Sen A, Rybakova D, Hurst MR, Mitra AK. Structural study of *Serratia entomophila* antifeeding prophage: 3D structure of the helical sheath. *J Bacteriol* 2010; 192:4522-5.
47. Yang G, Dowling AJ, Gerike U, French-Constant RH, Waterfield NR. Photorhabdus virulence cassettes confer injectable insecticidal activity against the wax moth. *J Bacteriol* 2006; 188:2254-61.
48. Persson OP, Pinhassi J, Riemann L, Marklund BI, Rhen M, Normark S, et al. High abundance of virulence gene homologues in marine bacteria. *Environ Microbiol* 2009; 11:1348-57.

©2011 Landes Bioscience.
Do not distribute.

Chapter V

Host adaptation of a symbiont: The biphasic life cycle of *Amoebophilus asiaticus* and its phage derived protein secretion system

Manuscript in preparation

Host adaptation of a symbiont: The biphasic life cycle of *Amoebophilus asiaticus* and its phage derived protein secretion system

**Thomas Penz^a, Agnes Harreither^a, Han-Fei Tsao^a, Karin Aistleitner^a, Rok Kostanjsek^b,
Martin Pilhofer^c, Grant J. Jensen^c, Stephan Schmitz-Esser^d, Matthias Horn^{a,1}**

^aDepartment of Microbial Ecology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria; ^bDepartment of Biology, University of Ljubljana, Večna pot 111, 1000 Ljubljana, Slovenia; ^cDivision of Biology, California Institute of Technology, 1200 East California Boulevard, Pasadena, 91125 California, USA; ^dInstitute for Milk Hygiene, University of Veterinary Medicine Vienna, Veterinärplatz 1, 1210 Vienna, Austria

¹corresponding author:

Matthias Horn

Department of Microbial Ecology
University of Vienna
Althanstrasse 14
1090 Vienna
Austria

Phone: +43 1 4277 54393

Fax: +43 1 4277 54389

E-mail: horn@microbial-ecology.net

Abstract

Acanthamoebae are ubiquitous protozoa and predators of bacteria. However, some bacteria have learned to survive phagocytosis and are able to use the intracellular environment of amoebae as a niche for survival, multiplication and spreading. In this study we describe the complex life cycle of the obligate intracellular *Bacteroidetes* symbiont of *Acanthamoebae* *Amoebophilus asiaticus* with transmission and cryo-electron microscopy. The life cycle of *Amoebophilus* starts with an extracellular infectious stage and continues with the entry into an *Acanthamoeba*. After intracellular establishment, *Amoebophilus* differentiates into its replicative stage and continues with intracellular replication in the *Acanthamoeba* host until the host is completely packed with the symbiont. At this life stage, *Amoebophilus* lyses its *Acanthamoeba* host and the life cycle starts again. Interestingly, not only the shape of *Amoebophilus* changes from a rod-shaped extracellular infectious form to a coccoid non-infectious form during the life cycle. Also, the expression of a phage-derived secretion apparatus alters during the infection process. The high expression level of the secretion apparatus in the extracellular *Amoebophilus* stage, which was verified with qPCR on RNA level and with mass spectrometry on protein level, confirms the hypothesis that the phage-derived apparatus might be used in delivering effector proteins to their location of action promoting infection, intracellular survival and replication. With comparative sequence analysis, we could show that the putative secretion apparatus is also present in the genomes of other *Bacteroidetes* classes and that the genetic organization is similar to a defective prophage, identified in the entomopathogen *Serratia entomophila*.

Introduction

The obligate intracellular symbiont of *Acanthamoebae* *Amoebophilus asiaticus* belongs to the diverse phylum of *Bacteroidetes* (Horn et al, 2001; Thomas et al, 2011). Within the *Bacteroidetes* different symbiotic lifestyles are observed and vary from a free living to an endosymbiotic one. The host range of *Bacteroidetes* is broad and spans large parts of the eukaryotic domain. For instance the human gut microbiota is dominated by *Bacteroidetes* (Marchesi, 2010). Some members of *Bacteroidetes* are essential nutrient providers for insects (Sabree et al, 2009) while others are able to influence the reproducibility of insects (Hunter et al, 2003). However, some *Bacteroidetes* are associated with Eukaryotes in another way than a symbiosis. These marine *Bacteroidetes* appear at algal blooms have been reported to be algicidal (Lee et al, 2011; Saw et al, 2012; Zhou et al, 2012).

Phylogentic analyses place *Amoebophilus* together with insect endosymbiont *Cardinium hertigii* (Gruwell et al, 2007), the nematode symbiont ‘*Candidatus* Paenicardinium endonii’ (Noel & Atibalentja, 2006) within one phylogentic clade. Other sequences clustering into this *Amoebophilus/Cardinium* clade are retrieved from coral samples (Sunagawa et al, 2010). The well known obligate *Bacteroidetes* insect endosymbionts *Blattabacterium* sp. and *Sulcia* sp. are only distantly related to this clade.

The natural hosts of *Amoebophilus* are *Acanthamoebae* (Schmitz-Esser et al, 2008), ubiquitous protozoa found in various natural environments. Besides being opportunistic human pathogens, causing blinding keratitis or fatal encephalitis (Khan, 2006), amoebae are predators of bacteria (Rodriguez-Zaragoza, 1994) and have thus a great impact on microbial community composition. However, some bacteria have developed mechanisms to resist amoebal phagocytosis and are able to use the intracellular environment of an amoeba as a

niche for survival, multiplication and spreading. There are several lines of evidence that protozoa contribute to the adaption of bacteria to new intracellular environments such as higher eukaryotic organisms like insects and mammals (Penz et al, 2012; Toft & Andersson, 2010). Therefore it has been suggested that amoebae are “training grounds” for intracellular pathogenic bacteria (Molmeret et al, 2005).

Besides harboring obligate intracellular bacteria amoebae harbor also facultative intracellular bacteria including the human pathogens *Legionella pneumophila* and *Chlamydomphila pneumoniae* (Albert-Weissenberger et al, 2007; Greub & Raoult, 2004). For some bacterial symbionts a biphasic lifestyle with morphological and physiological distinct stages has been described. The biphasic lifestyle of these bacteria consist of an extracellular infectious and an intracellular replicating stage (Harb et al, 2000; Horn, 2008; Molofsky & Swanson, 2004). To our knowledge for none of the bacteria belonging to the phylum of *Bacteroidetes* such a biphasic lifestyle has been described.

Similar to most obligate intracellular symbionts of amoebae, *Amoebophilus* possess reduced biosynthetic capabilities compared with free-living bacteria. Under environmental conditions *Amoebophilus* is not able to replicate outside the host and is thus highly dependent on the host with regard to the provision of essential compounds such as amino acids, vitamins, nucleotides and cofactors. To compensate for its reduced metabolic capabilities the genome of *Amoebophilus* encodes an arsenal of transport proteins used to parasitize energy from its *Acanthamoeba* host (Schmitz-Esser et al, 2010).

However, the genome of *Amoebophilus* is unique among prokaryotic genomes due to its large fraction of genes encoding putative host cell interaction proteins (n=129; 8% of all coding sequences) (Schmitz-Esser et al, 2010). This fraction of proteins includes proteins with typical eukaryotic protein-protein interaction motifs such as ankyrin-repeats (ANKs), tetratricopeptide-repeats (TPRs), leucine-rich repeats (LRRs) as well as proteins with F- and U-box domains

and an ubiquitin specific protease that is able to interfere with the host's ubiquitin system. The secretion of such effector proteins into the bacterial environment and into the host cell plays a crucial role in modulating the interaction between symbiotic bacteria with their hosts. Surprisingly, no homologues of the well-known Gram-negative bacterial secretion systems (type three, four and six) (Tseng et al, 2009), which would be required to translocate effector proteins into the *Acanthamoeba* host, were recognized in the genome of *Amoebophilus*. Instead of using any of the known secretion systems, *Amoebophilus* effector proteins might be translocated by any other secretion system. The only gene cluster present in the genome of *Amoebophilus* that could encode for a putative secretion apparatus might be a gene cluster derived from a prophage. This encoded prophage is not functional anymore because important parts such as lysis cassettes for the lytic life cycle of a prophage are missing. The prophage gene cluster still contains all proteins to potentially form a functional secretion machinery. This putative secretion apparatus of *Amoebophilus* is distantly related to the antifeeding prophage from the entomopathogen *Serratia entomophila* (Penz et al, 2011). In *Serratia entomophila* and *Photorhabdus* sp. the antifeeding prophage is used to deliver insecticidal toxins to its insect host causing the so called amber disease which leads to the death of the insect (Hurst et al, 2007; Hurst et al, 2004; Yang et al, 2006).

Here we describe for the first time a life cycle of a *Bacteroidetes* endosymbiont by using fluorescence in situ hybridization (FISH) and transmission electron microscopy (TEM) imaging. Infection studies suggest an extracellular infectious stage and an intracellular replicating stage of the *Acanthamoeba* endosymbiont *Amoebophilus asiaticus*. Interestingly, we could show by using real time quantitative PCR (RT q-PCR) that the defective prophage of *Amoebophilus* which shows similarities to secretion systems is differentially expressed during the biphasic lifecycle of *Amoebophilus*. This suggests that the defective prophage of *Amoebophilus* has an important function during the infection process. In this host-symbiont system the unusual secretion apparatus might be used to deliver effector proteins into the host

cell and plays thus a crucial role in this symbiosis. In addition we could also verify the expression of the prophage proteins with mass spectrometry and give insights into the cellular localization and structure of this protein complex with cryo-electron microscopy.

Materials and Methods

Cultivation of *Acanthamoeba* sp.

Uninfected *Acanthamoeba* sp. and *Acanthamoeba* sp. infected with *Amoebophilus asiaticus* strain 5a2 (ATCC no. PRA-228), which were isolated from a lake sediment in Austria (Schmitz-Esser et al, 2008), were maintained as adherent culture in 25 cm² tissue culture flasks containing 10 ml trypticase soy broth with yeast extract (TSY; 30 g/l trypticase soy broth, 10 g/l yeast extract, pH 7.3). Cultures were incubated at 27°C and passaged at confluency by 1:10 dilution of the culture every five to ten days.

Infection experiments

Two different methods were used to isolate once intracellular bacteria that are inside of an *Acanthamoeba* and once extracellular bacteria that were released from an infected *Acanthamoeba* culture. To gain intracellular bacteria, *Acanthamoeba* sp. infected with *Amoebophilus asiaticus* strain 5a2 were harvested by centrifugation (7000 rpm, 3 min, 27°C) and lysed in TSY media using a Dounce tissue grinder (Wheaton). To separate the bacterial cells from lysed *Acanthamoebae*, the homogenized suspension was passed through a five µm pore-size cellulose filter (Sartorius). The harvested bacteria were immediately used for following infection experiments.

To gain extracellular bacteria, freshly released *Amoebophilus* stages were harvested from well grown *Acanthamoeba* sp. cultures infected with *Amoebophilus*, concentrated by centrifugation (8000 rpm, 5 min, 27°C) and passed through a five µm pore-size cellulose filter (Sartorius). The harvested bacteria were immediately used for following infection experiments.

The number of *Amoebophilus* cells was determined by counting of DAPI (4',6-diamidino-2-phenylindole) stained bacterial cells on a 0.22 µm polycarbonate membrane filter (Millipore). The number of *Acanthamoebae* was determined using a Neubauer haemocytometer. To remove bacteria that have not been taken up by *Acanthamoebae*, infected *Acanthamoebae* ssp. cultures were washed with TSY media prior further proceeding for early infection time points or at latest after 24 hours post infection (h p.i). Infection experiments were monitored by fluorescence in situ hybridization (FISH) in combination with confocal laser scanning microscopy (Schmitz-Esser et al, 2008).

Symbiont-free *Acanthamoeba* sp. were infected with either intracellular *Amoebophilus* cells from mechanically lysed *Acanthamoebae* or freshly released *Amoebophilus* cells by adding them to uninfected *Acanthamoebae* ssp. with a multiplicity of infection (MOI) of 500. The high MOI was required to gain a saturation of the infection.

Prophage tail sheath preparation

Freshly released extracellular *Amoebophilus* cells were harvested as described above. Prophage tail sheath preparation was done as described elsewhere (Basler et al, 2012). Briefly, *Amoebophilus* cells were lysed and cell debris was removed by centrifugation. Cleared lysates were subjected to ultraspeed centrifugation three times. Purified prophage tail sheaths were analyzed with SDS-PAGE and stored at -20°C for further analysis.

Mass spectrometry analysis

For mass spectrometry analysis isolated phage tail sheaths were loaded on a SDS-PAGE gel which was separated for 2 cm. The coomassie-stained gel copped in three equal parts. Coomassie-stained gel sections were washed with 50 mM ammonium bicarbonate (ABC) (pH 8.5) and dried with acetonitrile (ACN). Disulfide bonds were reduced by DTT (200 μ l of 10 mM dithiothreitol for 30 min at 56°C). DTT was washed off and cysteins were alkylated by incubation with 100 μ l of 54 mM iodoacetamide for 20 min at RT in the dark. Gel pieces were dried with ACN, then swollen in 10 ng/ μ l trypsin (recombinant, proteomics grade, Roche) in 50 mM ABC and incubated over night at 37°C. Reaction was stopped by adding formic acid to a final concentration of approximately 1% and peptides were extracted by sonication. Peptides were separated on an UltiMate 3000 HPLC system (Dionex, Thermo Fisher Scientific). Digests were loaded on a trapping column (PepMap C18, 5 μ m particle size, 300 μ m i.d. x 5 mm, Thermo Fisher Scientific) equilibrated with 0.1% TFA (trifluoroacetic acid) and separated on an analytical column (PepMap C18, 3 μ m, 75 μ m i.d. x 150 mm, Thermo Fisher Scientific) applying a 60 min linear gradient from 2.5% up to 40% ACN with 0.1% formic acid followed by a washing step with 80% ACN and 10% TFE (trifluoroethanol). The HPLC was directly coupled to an LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) equipped with a nanoelectrospray ionization source (Proxeon, Thermo Fisher Scientific). The electrospray voltage was set to 1500 V. The mass spectrometer was operated in the data-dependent mode: 1 full scan (m/z: 350-1800, resolution 60000) with lock mass enabled was followed by maximal 10 MS/MS scans. The lock mass was set at the signal of polydimethylcyclsiloxane at m/z 445.120025. Monoisotopic precursor selection was enabled, singly charged signals were excluded from fragmentation. The collision energy was set at 35%, Q-value at 0.25 and the activation time at 10 msec. Fragmented ions were set onto an exclusion list for 60 sec. Raw spectra were interpreted by Mascot 2.2.04 (Matrix Science) using Mascot Daemon 2.2.2. Spectra were searched against

the bacterial nr-database with the following parameters: the peptide tolerance was set to 2 ppm, MS/MS tolerance was set to 0.8 Da, carbamidomethylcysteine was set as static modification, oxidation of methionine as a variable modification. Trypsin was selected as the protease and two missed cleavages were allowed. MASCOT results were loaded into Scaffold (Ver. 3.00.02; Proteome Software). Peptide identifications were accepted if they could be established at a probability greater than 95% as assigned by the Protein Prophet algorithm. Protein identifications were accepted if they could be established at a probability greater than 99%. Additionally, at least two identified peptides per protein were required.

Transmission electron microscopy

Acanthamoebae ssp. infected with *Amoebophilus* and extracellular bacteria obtained from culture supernatants were harvested as described above. Samples for transmission electron microscopy were preserved in 3.5% glutaraldehyde in 0.1 M phosphate buffer (pH 7.2), washed in 0.1 M phosphate buffer (pH 7.2), post-fixed in 1% OsO₄ for 1 h and dehydrated in an increasing ethanol series. Dehydrated samples were embedded in Agar100 resin and cut. Ultrathin sections were stained with uranyl acetate and Reynold's lead citrate prior to examination with a Philips CM 100 transmission electron microscope operating at 80 kV.

For negative staining, samples were applied to a Formvar-coated, carbon-coated, glow-discharged copper EM grid (Electron Microscopy Sciences). Samples were aspirated washed with water, stained with 2% uranylacetate and imaged with a Tecnai T12 transmission electron microscope (FEI).

Cryo-electron microscopy

For plunge freezing copper/rhodium EM grids (R2/1, Quantifoil) were glow-discharged for 1 min. A 20x concentrated bovine serum albumin-treated solution of 10 nm colloidal gold (Sigma) was added to purified bacterial cells (1:4 v/v) immediately before plunge freezing. A 4 μ l droplet of the mixture was applied to the EM grid, then automatically blotted and plunge-frozen into a liquid ethane-propane mixture (Tivol et al, 2008) using a Vitrobot (FEI Company) (Iancu et al, 2006). Cryo-EM images of plunge frozen cells were collected using a Polara 300 kV FEG transmission electron microscope (FEI Company) equipped with an energy filter (slit width 20 eV; Gatan) on a lens-coupled 4k by 4k UltraCam (Gatan). Pixels on the CCD represented 0.95 nm (22,500x) or 0.63 nm (34,000x) at the specimen level. Typically, tilt series were recorded from -60° to $+60^\circ$ with an increment of 1° at 10 μ m under-focus. The cumulative dose of a tilt-series was 180-220 $e^-/\text{\AA}^2$ (for whole cells) or 100 $e^-/\text{\AA}^2$ (for cryo-sections and sheath preparations). UCSF Tomo (Zheng et al, 2007) was used for automatic tilt-series acquisition. Three-dimensional reconstructions were calculated using the IMOD software package (Mastronarde, 2008) or Raptor (Amat et al, 2008).

RNA isolation

RNA was isolated from freshly released extracellular *Amoebophilus* cells after harvesting as described above. RNA from *Acanthamoebae* ssp. infected with *Amoebophilus* was isolated from the time points 12 hours post infection (h p.i.), 24 h p.i., 68 h p.i., 140 h p.i., and from uninfected *Acanthamoebae* ssp. after harvesting cells by centrifugation (8000 x g, 2 min, 27°C). The resulting cell pellets were resuspended in 750 μ l TRIzol (Invitrogen Life Technologies), transferred to a Lysing Matrix A tube (*MP* Biomedicals) and homogenized

using a BIO101/Savant FastPrep FP120 instrument (speed: 4.5 m/sec, 30 sec). RNA was extracted by phase separation, precipitation, washing and dissolving according to the recommendations of the manufacturer (TRIzol, Invitrogen Life Technologies). Remaining DNA was removed using the TURBO DNA-*free* Kit (Ambion). After DNase treatment RNA was dissolved in ddH₂O_{DEPC} and stored at -80°C until use. The absence of DNA contamination in the DNase-treated RNA was verified by performing a control PCR with 42 cycles using primers targeting a 141 bp fragment of the gene Aasi_1074 (encoding the prophage tail sheath protein; see below). Quality control of purified RNA was performed using the Experion Automated Electrophoresis Station (Bio-Rad). All RNA isolations were performed in three biological triplicates: RNA was isolated simultaneously from three infection experiments.

Quantitative reverse transcriptase PCR

RT q-PCR was performed according to the MIQE guidelines (Table S1) (Bustin et al, 2009). 500 ng DNase treated total RNA was used for first strand cDNA synthesis per reaction. Reverse transcription was performed using SuperScript[®] III Reverse Transcriptase (Life Technologies) with random hexamers. Reverse transcription of each infection time point was done for 60 min at 50°C in technical triplicates (three times RNA from the same infection timepoint) in a total volume of 25 µl per reaction. RT-qPCR was performed using a CFX96 Touch[™] Real-Time PCR Detection System (Bio-Rad, Hercules, CA) with primers targeting a 141 bp fragment of the prophage tail sheath gene of *Amoebophilus* Aasi_1074 (Aasi_1074 qPCR F1: 5'-GTGGTGCAGATTGCTACATCAT-3'; Aasi_1074 qPCR R1: 5'-AGTCGGGCATAAGCAACATAGT-3'), and primers targeting a 167 bp fragment of the beta subunit of the RNA polymerase gene Aasi_1396 as reference gene (Aasi_1396 qPCR F2: 5'-ACTAGGTACGCCACCTGAAAAA-3'; Aasi_1396 qPCR R2: 5'-

AAGTTACTCCCCTTTCCACACA-3'). All primers were used in a final concentration of 200 nM. RT-qPCR reactions were prepared with the iQTM SYBR[®] Green Supermix (Bio-Rad). qPCR conditions including annealing temperature and primer concentrations were optimized with genomic DNA isolated from a continuous culture of *Acanthamoebae* ssp. infected with *Amoebophilus*. For all RT-qPCR reactions a thermal cycling was used as follows: initial denaturation step at 95°C for 3 min, followed by 45 cycles of denaturation at 95°C for 30 sec, annealing at 65.7°C for 30 sec, and elongation at 72°C for 30 sec. To assess the specificity of the amplification, a melting curve was subsequently performed, starting at 55 °C and increasing the temperature by 0.5°C each 10 sec up to a temperature of 95°C. For both genes, standard curves were obtained with TOPO[®] XL plasmids (Life Technologies) containing a 1507 bp insert with the complete phage tail sheath gene and an 851 bp fragment of the reference gene, respectively. RT q-PCR standards were quantified using Quant-IT PicoGreen ds Assay Kit (Life Technologies). RT q-PCR controls were done as follows: To determine the amount of DNA in DNase treated RNA RT q-PCR reactions were performed with DNase treated RNA from different infection samples. In addition to a non target control (NTC) (cDNA from uninfected *Acanthamoabae*) to ensure that the RT q-PCR assay is specific a no template control (NTaC) (reverse transcription with water) was included in the RT q-PCR assay. RT q-PCR data were analyzed using the CFX Manager (v 2.1, Bio-Rad). Fold increase ratios of relative prophage tail sheath mRNA level were calculated using mean starting quantity (SQ) calculated from RT q-PCR standards as percentage of the ratio of the sample 68 h p.i. and divided by 100.

Phylogentic analysis

Nucleotide sequence alignments of 16S rRNA genes of selected *Bacteroidetes* (complete genomes with exception of *Cardinium hertigii* cEper1) were done using MUSCLE (Edgar, 2004), and a phylogenetic tree was reconstructed using the software MEGA5 (Kumar et al, 2008). The phylogentic tree was calculated using the maximum likelihood algorithm (1000 bootstrap resamplings). Graphical manipulation of phylogentic trees was performed using the online tool Interactive Tree Of Life (iTOL) version v2.2 (Letunic & Bork, 2011).

Antifeeding prophage (AFP)-like clusters in complete genomes of *Bacteroidetes* were identified as described elsewhere (Penz et al, 2012).

Results

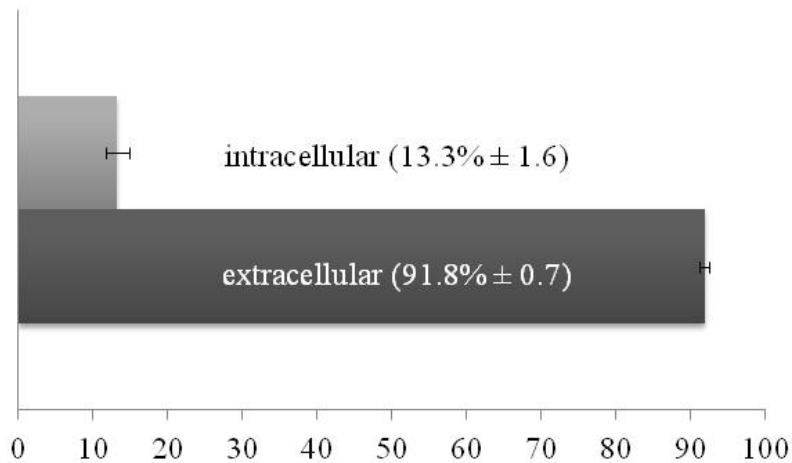
The life cycle of *Amoebophilus*

Many bacteria associated with eukaryotic hosts harbor a complex life cycle which usually starts with an extracellular infective form, continues with an intracellular replicative stage and ends with a lytic form that is able to spread and to infect again new eukaryotic hosts. To investigate if *Amoebophilus* harbors a life cycle similar to other bacteria associated with Eukaryotes we performed infection experiments in *Acanthamoebae* cultures.

Increased infectivity of extracellular vs. intracellular stages of *Amoebophilus*

From previous infection experiments we observed that *Amoebophilus* cells that were isolated from lysates from long incubated *Acanthamoebae* cultures had a much higher infectivity

compared to others isolated from short incubated *Acanthamoebae* cultures. Long incubated *Acanthamoebae* cultures harbor possibly more extracellular *Amoebophilus* stages than short incubated ones. This prompted us to investigate the infectivity with infection experiments of two different *Amoebophilus* stages, the extracellular and the intracellular stage of *Amoebophilus*. For this infection experiment, uninfected and *Acanthamoebae* infected with *Amoebophilus* were counted after 48 h p.i. using FISH. The infection stages 48 h p.i. represent intracellular stages of *Amoebophilus* (see section life cycle of *Amoebophilus*). *Amoebophilus* cells of the extracellular intermediates showed a significantly increased infectivity by seven times ($91.8\% \pm 0.7$ for extracellular versus $13.3\% \pm 1.6$ for intracellular intermediates of *Amoebophilus*) in comparison with intracellular *Amoebophilus* intermediates (Figure 1, Table S2). To exclude that the mechanical harvest treatment of intracellular *Amoebophilus* cells with the Dounce tissue grinder (Wheaton) influences the infectivity in infection experiments, control infection experiments with extracellular *Amoebophilus* cells that were treated with the same harvest method and then subsequently used for infecting *Acanthamoebae* were performed. Here no differences in infectivity between untreated extracellular and extracellular *Amoebophilus* cells treated with the Dounce tissue grinder (Wheaton) were observed (data not shown). To test if a factor in the amoebal lysate influences the infectivity of *Amoebophilus*, extracellular *Amoebophilus* intermediates with and without amoebal lysate were used to infect empty *Acanthamoebae*. Again no differences in infectivity between extracellular *Amoebophilus* cells treated with and without amoebal lysate were observed.



infected *Acanthamoebae* with *Amoebophilus* in % in one infection experiment

Figure 1. Infectivity of intracellular and extracellular stages of *Amoebophilus*. Infectivity of intracellular and extracellular stages of *Amoebophilus* was compared by adding once *Amoebophilus* cells isolated from infected *Acanthamoebae* and once extracellular *Amoebophilus* cells harvested from the supernatant of an infected amoebae culture. The extracellular stage of *Amoebophilus* shows a significantly increased infection rate to *Acanthamoebae* compared to the intracellular intermediate at 48 h p.i. Arrow bars of standard deviation are indicated.

The life cycle of *Amoebophilus* starts with an extracellular infective stage

With the observations from the previous experiment we performed following infection experiments starting with extracellular *Amoebophilus* cells isolated from long incubated *Acanthamoebae* cultures. The extracellular stage of *Amoebophilus* is different in shape compared to its intracellular stage. In addition this extracellular coccoid shaped *Amoebophilus* stage showed in classical transmission electron micrographs (TEM) (Figure 4A3) cytoplasmic fibril like structures similar to phage tail sheath bundles. These fibril like structures, with a diameter of 4-6 nm are always organized in bundles of five to seven (indicated with black arrows in Figure 4A3). The next stage in the life cycle of *Amoebophilus* is the attachment to its *Acanthamoeba* host (Figure 4B1-3) which is observed 12 h p.i. Here coccoid *Amoebophilus* cells are attached to the cell membrane of *Acanthamoebae*. After 17 to 24 h p.i. coccoid *Amoebophilus* cells invade its host cell (Figure 4C1-3). Again, cytoplasmic phage tail

sheath like structures, already observed in the extracellular *Amoebophilus* intermediate, are present immediately after internalization (indicated with black arrows in Figure 4C3). Subsequent to the internalization and following intracellular establishment, *Amoebophilus* changes its coccoid shape to a rod shaped one at 68 h p.i. (Figure 4D1-3). The replicative intermediate of *Amoebophilus* is associated with membranes of the rough endoplasmatic reticulum (Horn et al, 2001; Schmitz-Esser et al, 2010) and replicates within the cytoplasm of *Acanthamoebae* until the *Acanthamoeba* is densely packed with *Amoebophilus*. Compared to the extracellular *Amoebophilus* cells no fibril like structures were observed at this bacterial life stage. Again, *Amoebophilus* changes from a rod shaped bacterium to a coccoid one and lyses the host cell after 140 h p.i. (Figure 4E1-3). Here the life cycle of *Amoebophilus* starts again with its extracellular intermediate able to infect new *Acanthamoeba*.

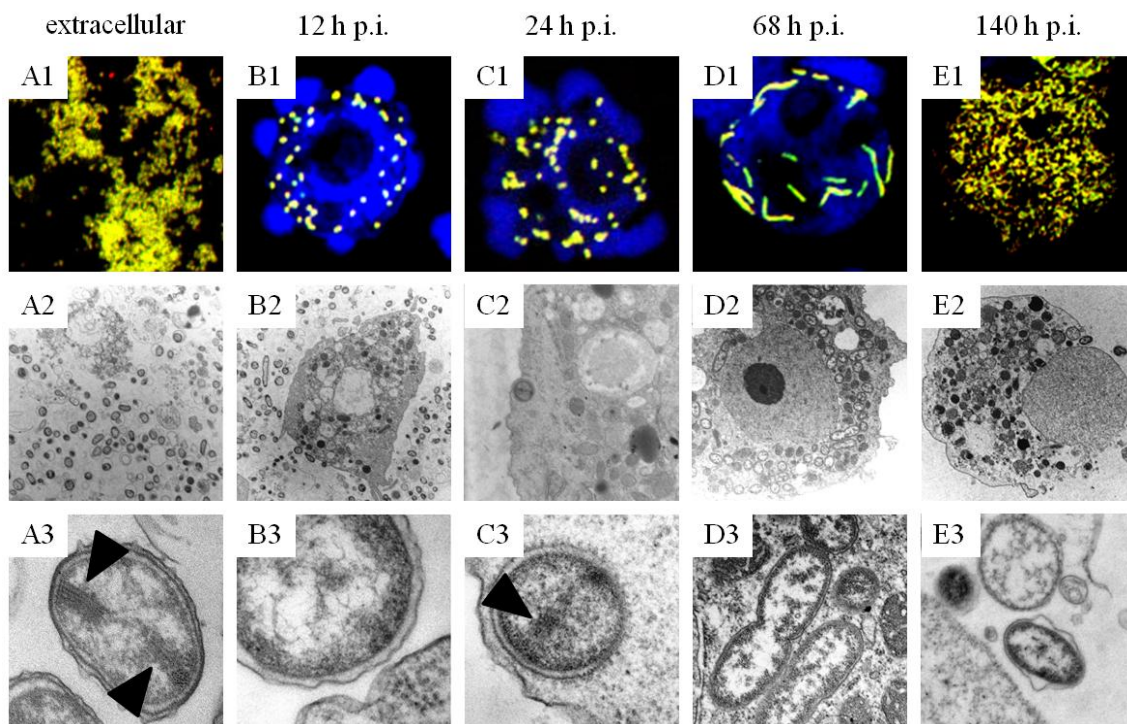


Figure 4: The infection cycle of *Amoebophilus* in *Acanthamoebae*. (1) FISH images. (2-3) Transmission electron micrographs. (A) extracellular infectious *Amoebophilus* stages with fibril like structures. (B) *Amoebophilus* attaching to the membrane of its *Acanthamoeba* host at 12 h p.i. (C) *Amoebophilus* at an early intracellular stage inside its *Acanthamoeba* host at 24 h p.i., again fibril like structures are visible. (D) *Amoebophilus* in its replicative dividing stage inside its *Acanthamoeba* host at 68 h p.i., *Amoebophilus* appears in rod shaped forms. (E) extracellular *Amoebophilus* after the lysis of its *Acanthamoeba* host 140 h p.i. in its coccoid infectious intermediate.

Ultrastructure of the prophage tail sheath in extracellular stages of *Amoebophilus*

As already demonstrated with TEM imaging prophage tail sheath structures are present in extracellular infectious stages of *Amoebophilus*. Cryo-EM imaging of extracellular infectious stages of *Amoebophilus* intermediates confirmed these findings (Figure 5A-F). Here again we could show that these fibril like structures are organized in bundles of five to seven. Interestingly, in most extracellular *Amoebophilus* stages those bundles appeared at two different locations within the bacterial cell. Those observed phage tail sheath like structures were isolated from extracellular infectious intermediates of *Amoebophilus* with ultracentrifugation and analyzed with SDS-PAGE and negative stain electron microscopy. Phage tail sheath like structures were also characterized with mass spectrometry. Negative stain electron microscopy revealed straight, hollow and helical tubular structures (Figure 6A-D). SDS-PAGE gel analysis revealed two major bands between 55 and 35 kDa (Figure 6E). Mass spectrometry analysis of the purified helical tubular structures revealed that the prophage tail sheath protein Aasi_1074 is with 66 detected peptides one of the most prominent measured proteins in the purification (Table 1, Table S3). Peptides of nine additional proteins encoded in the prophage gene cluster of *Amoebophilus* (Penz et al, 2011) were detected with mass spectrometry (Table 1, Table S3 and 4). These proteins include proteins related to a virus fibre protein (Aasi_0556), to phage baseplate protein (Aasi_0557), to a VgrG-like protein (Aasi_1080) and other proteins without any obvious function.

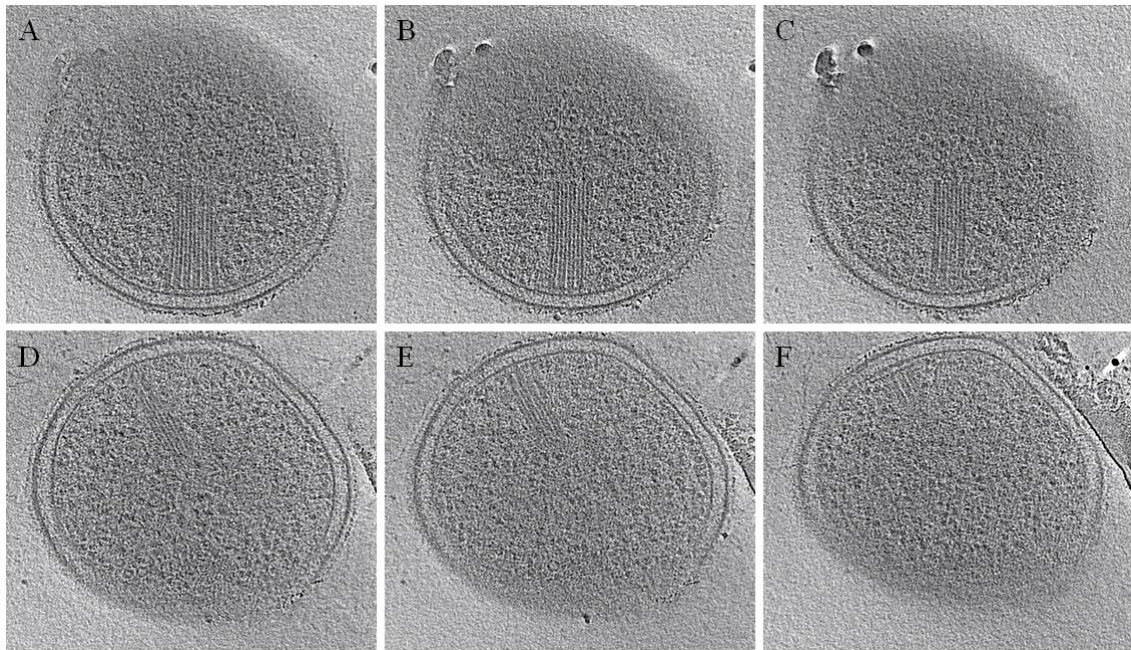


Figure 5: Cryo-electron imaging of an extracellular infectious *Amoebophilus* cell. Different sections (A)-(F) of a single extracellular infectious *Amoebophilus* cell showing cytoplasmic phage tail sheath like structures organized in bundles at two different areas within the cell.

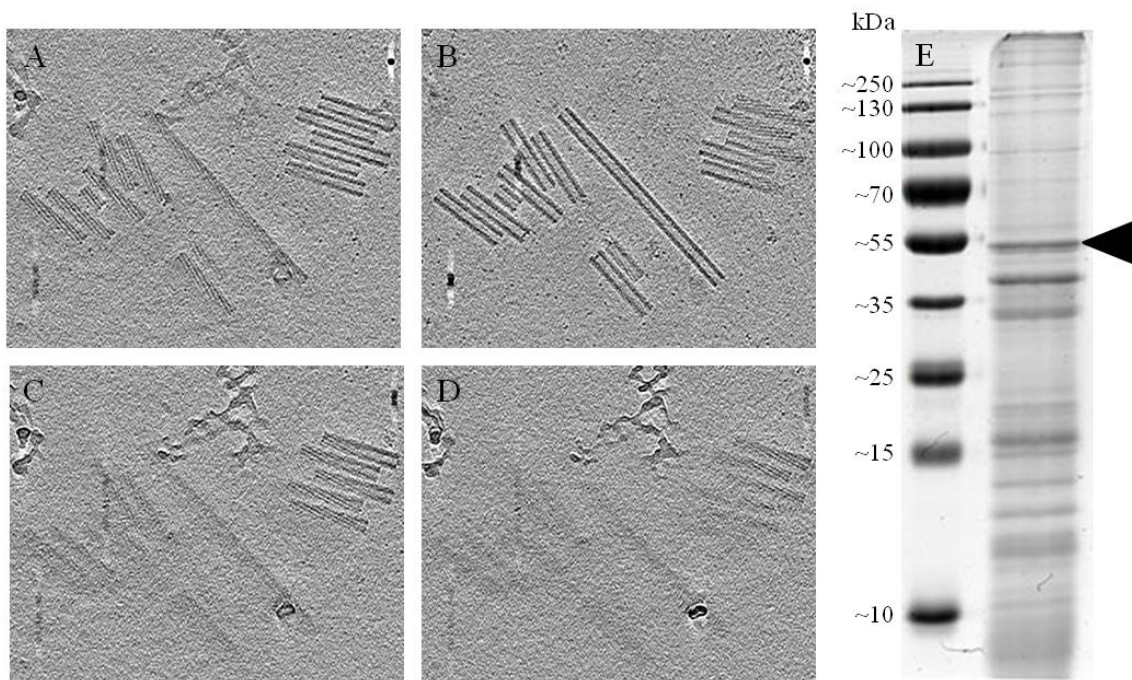


Figure 6: Purified phage tail sheath like structures of extracellular infectious *Amoebophilus* stages. (A)-(D) Different sections of negative stained phage tail sheath like structures showing straight, hollow and helical tubular organized fibril like structures. (E) 12,5 % SDS-PAGE of purified phage tail sheath like structures revealing two major bands between 35 and 55 kDa .

Table 1. **Identified AFP-like proteins with mass spectrometry**

<i>Amoebophilus</i> AFP-like protein	molecular weight in kDa	detected with mass spectrometry	number of detected peptides overall measurement	putative function
Aasi_0556	28	yes	13	virus fibre protein
Aasi_0557	145	yes	27	phage baseplate
Aasi_1072	22	no	-	n.d.
Aasi_1073	33	yes	4	n.d.
Aasi_1074	55	yes	66	phage tail sheath
Aasi_1075	18	yes	9	n.d.
Aasi_1076	16	yes	6	n.d.
Aasi_1077	17	no	-	phage tail tube
Aasi_1078	7	no	-	n.d.
Aasi_1079	38	yes	10	n.d.
Aasi_1080	66	yes	18	VgrG
Aasi_1081	10	no	-	n.d.
Aasi_1082	15	no	-	lysozyme
Aasi_1083	97	yes	15	n.d.
Aasi_1806	180	yes	5	n.d.

n.d. not detected

Differential expression of the prophage during infection

To further investigate the prophage structures that were predominantly observed in the extracellular stage of *Amoebophilus*, we analyzed the expression of one of its core components (phage tail sheath, Aasi_1074) by quantitative reverse transcriptase PCR. For this we isolated RNA at different time points of the infection cycle of *Amoebophilus* (12 h p.i., 24 h p.i., 68 h p.i., 140 h p.i, and extracellular *Amoebophilus* cells). To determine the relative mRNA level of Aasi_1074 we performed quantitative reverse transcriptase PCR (Figure S1). The mRNA level of Aasi_1074 was normalized against the mRNA level of the housekeeping gene Aasi_1396 (beta subunit of the RNA polymerase). The efficiency of the RT q-PCR assay was 88.1% ($R^2=0.999$) for Aasi_1074 and 87.6% ($R^2=0.998$) for Aasi_1396 (Tables S5 and S6, Figure S2 and S3). The DNA amount in the DNase treated RNA for the RT reactions contained amounts of DNA within the range of the last detectable standard or even below the

limit of detection (Figures S4 to S6). Including a no template control (water instead of template in the RT reaction) confirmed purity of RT q-PCR components (Figure S7). The specificity of the RT q-PCR assay was confirmed by using cDNA from uninfected *Acanthamoeba*. (Figure S8)

The lowest relative mRNA level of the prophage tail sheath gene Aasi_1074 was observed at 68 h p.i. which represents the replicating intracellular *Amoebophilus* stage. Here the relative prophage tail sheath mRNA level was assumed to be 1 (Figure 7, Table 2, Tables S7 and 9, Figure S9 and S10). The highest relative prophage tail sheath mRNA level was observed at the extracellular stage of *Amoebophilus*. Here the relative prophage tail sheath mRNA level is increased 230 times compared to 68 h p.i.. During the attachment of *Amoebophilus* to its *Acanthamoeba* host at 12 h p.i. the relative prophage tail sheath mRNA level increases 4 times compared to 68 h p.i.. After 140 h p.i. *Amoebophilus* starts to lyse the *Acanthamoeba* host and the relative prophage tail sheath mRNA level was increased 30 times compared to 68 h p.i..

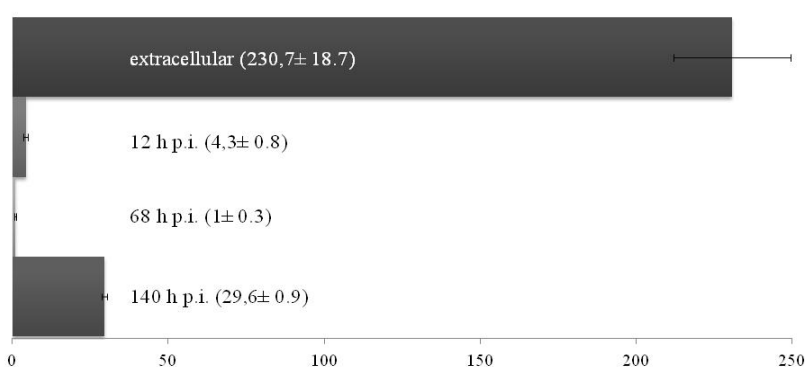


Figure 7: Differential expression of the prophage tail sheath gene relative to the beta subunit of the RNA polymerase during the infection cycle of *Amoebophilus*. Relative mRNA levels were normalized relative to 68 h p.i. timepoint. mRNA level at 68 h p.i. was assumed to be one. mRNA level of the phage tail was increased 230 times in extracellular intermediates of *Amoebophilus*. mRNA level of the phage tail was increased 4 times at 12 h p.i. (attachment of *Amoebophilus* to *Acanthamoeba*) and 30 times at 140 h p.i. (lysis of *Acanthamoeba* and release of infectious *Amoebophilus* cells).

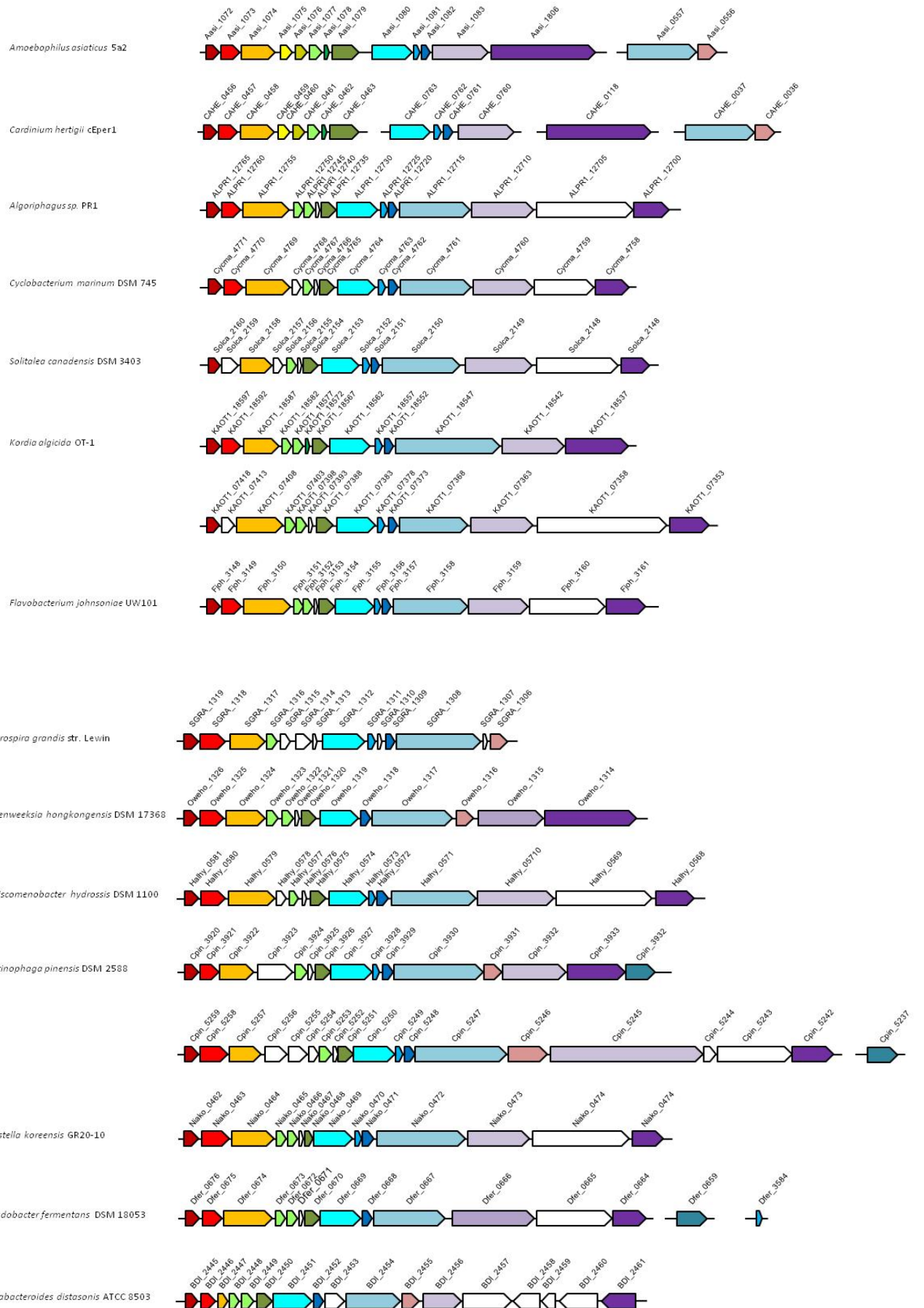
Table 2. **Relative quantification of phage tail sheath mRNAs in different life cycle stages of *Amoebophilus asiaticus***

<i>Amoebophilus</i> intermediate	mean ratio of Cq value Aasi_1074 to Aasi_1396*	standard deviation	fold increase	standard deviation
extracellular	10.72	0.87	230.7	18.7
12 h p.i.	0.20	0.04	4.4	0.8
68 h p.i.	0.05	0.01	1.0	0.3
140 h p.i.	1.38	0.04	29.6	0.8

*data are derived from 3 independent biological triplicates and 3 technical triplicates

The presence of the *Amoebophilus* prophage gene cluster in other *Bacteroidetes*

The prophage gene cluster encoded in the genome of *Amoebophilus* is also present in few other bacteria and some bacteria belonging to the phylum *Bacteroidetes*. Most of the 15 proteins encoded in the prophage derived gene cluster are present in 18 other *Bacteroidetes* (Figure 8, Table S9). The components of the prophage gene cluster in *Amoebophilus* that show similarities to characterized proteins or similarities to functional characterized domains are a phage baseplate-like protein (Aasi_0557) which might be the connection to the inner bacterial membrane. Further there are a phage tail sheath (Aasi_1074) and a phage tail tube (Aasi_1077) proteins present which might span through the inner bacterial membrane, the periplasm and the outer bacterial membrane. Also a VgrG-like protein (Aasi_1080) which might be able to puncture holes (Pukatzki et al, 2007) into the target cell is present. Most of the other proteins encoded in this prophage gene cluster do not allow prediction of a clear function. Besides to the prophage core components (Aasi_0557, Aasi_1074, Aasi_1077 and Aasi_1080) there are also the proteins Aasi_1072, Aasi_1082 and Aasi_1083 present and conserved in the 18 other *Bacteroidetes*. The prophage gene cluster is present in most of the so far known classes of *Bacteroidetes* (Figure 9). In some *Bacteroidetes* the prophage gene cluster is present twice (*Kordia algicida*, *Chitinophaga pinensis*, *Microscilla marina*).



Such a biphasic lifestyle was observed for the first time in a bacterium belonging to the phylum *Bacteroidetes*: the obligate intracellular *Acanthamoeba* endosymbiont *Amoebophilus asiaticus*. Developmental cycles and differentiations between intracellular and extracellular stages of some endosymbionts play central and crucial roles in their ecology. The biphasic life cycle of *Amoebophilus* (Figure 10) starts with its extracellular infectious coccoid shaped intermediate which is the only infectious form. This extracellular form, which is able to maintain its infectivity in host-free media for up to six days (Figure 2), is essential for the horizontal transmission of *Amoebophilus* to new *Acanthamoeba* hosts. In this stage the obligate intracellular *Amoebophilus* is exposed to the environment and not protected by its *Acanthamoeba* host. The genome of *Amoebophilus* does not have the potential to synthesize most cofactors, vitamins and amino acids *de novo* (Schmitz-Esser et al, 2010). Thus *Amoebophilus* is only able to survive a few days outside an *Acanthamoeba* and therefore highly dependent on its host. The attachment to the *Acanthamoeba* is the next step in the life cycle of *Amoebophilus*. Here membrane anchored proteins that are exposed to the extracellular environment are important for cell-recognition and adhesion. In the genome of *Amoebophilus* there are some glycosylated proteins encoded with atypical amino acid composition (Penz et al, 2011) which are often found to be involved in cell-recognition and adhesion (Galinski et al, 2001; Upreti et al, 2003; Wise et al, 2006; Zhou & Wu, 2009). The fibril like structures in the infectious extracellular intermediate of *Amoebophilus* that are derived from the prophage and the presence of genes encoding for proteins important for cell-recognition and adhesion suggests that *Amoebophilus* plays an active role in invasion of an *Acanthamoeba*. With its extremely high number of proteins with domains predominantly found in Eukaryotes (n=129; 8% of all coding sequences) like proteins containing common protein-protein interaction motifs such as ankryin repeats, TPR/Sell repeats, leucin rich repeats and proteins possibly interfering with the host ubiquitin system (Schmitz-Esser et al, 2010). *Amoebophilus* is very well equipped to interfere actively with the eukaryotic

Acanthamoeba host. Once internalized and escaped from the *Acanthamoeba* phagosome which might be achieved by two phospholipases D (Aasi_0130 and Aasi_0227) encoded in the *Amoebophilus* genome, *Amoebophilus* resides within the host cytoplasm like rickettsiae (Hackstadt, 1998) and thus has direct access to nutrients in this intracellular niche. From ultrastructure no indications for the presence of membranes that surround *Amoebophilus* are evident. This is in contrast to many other intracellular bacteria such as *Chlamydiae* (Dumoux et al, 2012), *Legionella* (AbuKwaik, 1996), and *Ehrlichia* (Weiss, 1991). During its intracellular intermediate *Amoebophilus* differentiates from a coccus to a replicative rod. The rod shaped form of *Amoebophilus* is associated with the rough endoplasmic reticulum (ER) of the *Acanthamoeba*. The precise role of the host ER in the infection cycle of intracellular bacteria remains largely unknown (Roy et al, 2006). This rod shaped form is the replicative noninfectious intermediate of *Amoebophilus*. This rod shaped form is visible at 68 h p.i. by ultrastructure. At this time point many dividing cells are visible. Extracted *Amoebophilus* cells of this replicative stage are not able to infect *Acanthamoebae*. It seems to be that this noninfectious stage cannot respond to new environmental conditions before *Amoebophilus* develops intracellular into its coccid shaped infectious extracellular stage. The development from a replicative non-infectious stage to an extracellular infectious stage can only be achieved intracellular in the *Acanthamoeba* host. This strongly suggests a biphasic lifestyle of the *Bacteroidetes* endosymbiont *Amoebophilus*. The biphasic developmental cycle of *Amoebophilus* ends after approximately 6 days by host cell lysis and the release of coccid shaped *Amoebophilus* intermediates to infect new *Acanthamoeba* hosts.

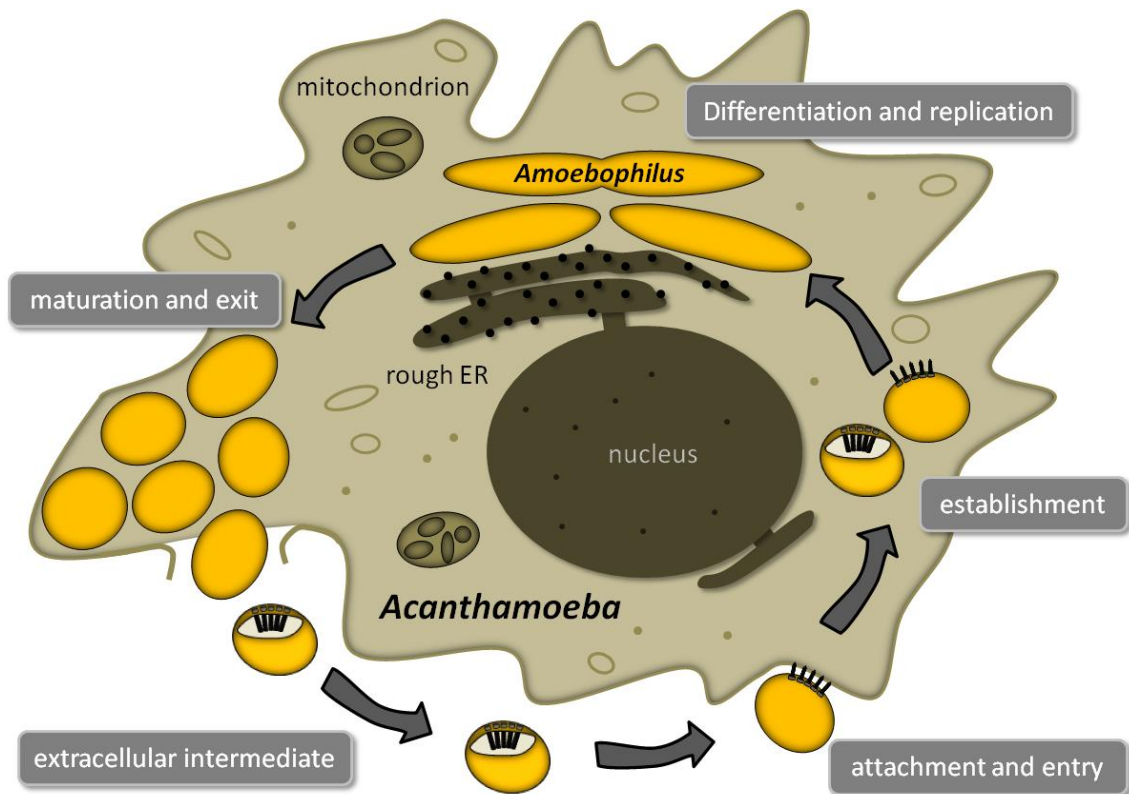


Figure 10: Schematic life cycle of *Amoebophilus*. The life cycle of *Amoebophilus* starts with its extracellular infectious coccoid shaped stage and continues with the attachment and entry into an *Acanthamoeba*. After the intracellular establishment, *Amoebophilus* differentiates in its replicative rod shaped form which is associated with the rough endoplasmic reticulum. Again after successful replication the *Acanthamoeba* host is completely packed with *Amoebophilus*. At this life stage *Amoebophilus* differentiates to a coccoid shaped bacterium and lyses its *Acanthamoeba* host. After lysis the life cycle of *Amoebophilus* starts again with the attachment to an *Acanthamoeba*.

The presence of fibril like structures derived from the prophage encoded in the *Amoebophilus* genome seems to play an important role during the extracellular intermediate. The relative expression of the prophage derived genes on transcriptional level is at highest in the infectious extracellular intermediate. This is a strong hint that the defective prophage is important in the infection process. Genes encoding for the defective prophage are also present in genomes of other *Bacteroidetes* belonging to different phylogenetic classes (Persson et al, 2009). With the exception of *Amoebophilus* and *Cardinium* (an endosymbiont of arthropods), all other *Bacteroidetes* harboring the defective prophage have a non obligate intracellular lifestyle and are able to replicate without direct dependency on a host. Interestingly, transmission electron

microscopy shows similar subcellular structures derived from the defective prophage in *Cardinium* (Penz et al, 2012). Many *Bacteroidetes* harboring the defective prophage are associated with Eukaryotes: *Algoriphagus* sp. PR1 with choanoflagellates (Alegado et al, 2011), *Dyadobacter fermentans* with nematodes (Nour et al, 2003), *Flavobacterium johnsoniae* with fishes (Flemming et al, 2007), *Bacteroides cellulosilyticus* (Robert et al, 2007) and *Parabacteroides distasonis* (Eggerth & Gagnon, 1933) with the human gut. As many members of the phylum *Bacteroidetes* (Abell & Bowman, 2005; Gomez-Consarnau et al, 2007; Pinhassi et al, 2004; Riemann et al, 2000), some *Bacteroidetes* harboring the defective prophage are important in the degradation of organic matter during and after algal blooms in the sea and are thus associated with algae or reported as algicidal: *Kordia algicida* (Lee et al, 2011), *Owenweeksia hongkongensis* (Zhou et al, 2012) and *Saprospira grandis* (Saw et al, 2012). In some genomes of these *Bacteroidetes* such as *Kordia algicida* OT-1, *Chitinophaga pinensis* DSM 2588 and *Microscilla marina* ATCC 23134 the prophage is present twice. Interestingly, the conservation of the prophages in those genomes is weaker compared to some other prophages from different *Bacteroidetes* genomes. The presence of more distantly related T6SSs in the case of *Salmonella* (Blondel et al, 2009) or prophage clusters in *Bacteroidetes* in a single genome reflects specific roles for each of them during the infection of different hosts.

One of the few characterized components of the defective *Bacteroidetes* prophage besides the proteins of this study is the phage tail sheath protein SCFP from the algicidal bacterium *Saprospira* sp. The SCFP is only present in gliding *Saprospira* sp. cells which are hunting for algae. In non-gliding/non-hunting *Saprospira* sp. cells none of these proteins were observed (Furusawa et al, 2005; Furusawa et al, 2003). This suggests again a strong evidence that the defective prophage is important in the interaction of those algicidal bacteria with their prey by direct attack (using the defective prophage to puncture algae and inject effector proteins) or by indirect interaction (secretion of algicides into the environment) (Azam, 1998).

In both cases a secretion system would be needed to translocate algicidal effector proteins outside the bacterial cell. So far none of the described Gram-negative secretion systems (type one to six) has been discovered in the genomes of those algicidal *Bacteroidetes*.

Furthermore based on weak amino acid similarities (Penz et al, 2011) (Table S6), the defective *Bacteroidetes* prophage is somewhat similar to the antifeeding prophage of *Serratia entomophila* which delivers toxins into the hemocytes of its insect host (Hurst et al, 2007). Interestingly, the purified *Amoebophilus* phage tail sheath particles of the defective prophage show structural similarities to the contractile phage tail-like structures of the type six secretion system (T6SS) of *Vibrio cholerae* (Basler et al, 2012). Similar to the T4 phage sheath (Aksyuk et al, 2009) and the T6SS of *Vibrio cholerae* the purified sheath particles appear as straight, hollow and helical tubular structures. Also the cellular localization and the appearance seem to be comparable to the T6SS of *Vibrio cholerae*. Analogous to T6SSs the defective prophage of the *Bacteroidetes* might have originated from a tailed bacteriophage. Structural relationships between T4 bacteriophages, R-type pyocins, antifeeding prophages and T6SS were observed and it has been suggested that the T6SS and phage tail-associated protein complexes share a common evolutionary origin (Bonemann et al, 2010; Leiman et al, 2009).

In addition the purified protein fraction of the *Amoebophilus* phage tail sheath particles is highly enriched in putative effector proteins. Surprisingly, the most abundant peptides in the mass spectrometry measurement are derived from two proteins (Aasi_1414 and Aasi_1417) which are similar to insecticidal toxins from various *Photorhabdus* spp. and other members of the *Gammaproteobacteria* (Schmitz-Esser et al, 2010). These insecticidal toxins consist of up to three subunits, one of which is the toxin itself and one or two potentiators (Ffrench-Constant & Waterfield, 2006). The high abundance of these effector proteins in the purified phage tail sheath particles suggests that the phage tail sheath proteins could have a potential

role in the translocation of the putative effector proteins. The phage tail sheaths of *Amoebophilus* might be already loaded with some effector proteins and are waiting in a 'ready to fire' confirmation. The proposed model of the secretion mechanism of the unusual *Amoebophilus* secretion apparatus could be similar as proposed for the T6SS of *Vibrio cholerae* (Basler et al, 2012). In short: The first step could be the assembly of the phage base plate complex (Aasi_0557) and the phage tube (Aasi_1077) polymerization. The second step is probably the polymerization of the phage tail sheath (Aasi_1074) around the phage tube. The final step in the assembly is the adding of a VgrG-like protein (Aasi_1080) that is able to puncture holes in to the target cell. The now in 'ready to fire' confirmation residing secretion apparatus resides now in its contracted confirmation and is ready for contraction upon an extracellular signal to deliver effector proteins into the new environment.

Taken together, the expression of the defective prophage in infectious and early intracellular *Amoebophilus* intermediates, the expression of phage tail sheath proteins in hunting algicidal *Saprospira* spp., the amino acid similarities of the prophage gene cluster to the antifeeding prophage of *Serratia entomophila* (which delivers insecticidal toxins to its hosts), the structural and organizational similarities to the T6SS of *Vibrio cholerae*, the high co-abundance of purified *Amoebophilus* prophage particles with putative effector proteins in the mass spectrometry, the fact that some secretion systems and phage tail-associated protein complexes share a common evolutionary origin strongly suggests that the prophage cluster encoded in the genomes of several Eukaryote associated *Bacteroidetes* is used as an unusual secretion apparatus and might thus compensate for the missing secretion systems in the genomes of *Bacteroidetes*.

Acknowledgement

This work was funded by a grant from the Austrian Science Fund FWF (project numbers Y277-B03 and P22703-B17) to MH and SSE.

References

Abell GC, Bowman JP (2005) Ecological and biogeographic relationships of class *Flavobacteria* in the Southern Ocean. *FEMS Microbiol Ecol* **51**: 265-277

AbuKwaik Y (1996) The phagosome containing *Legionella pneumophila* within the protozoan *Hartmannella vermiformis* is surrounded by the rough endoplasmic reticulum. *Applied and Environmental Microbiology* **62**: 2022-2028

Aksyuk AA, Leiman PG, Kurochkina LP, Shneider MM, Kostyuchenko VA, Mesyanzhinov VV, Rossmann MG (2009) The tail sheath structure of bacteriophage T4: a molecular machine for infecting bacteria. *EMBO J* **28**: 821-829

Albert-Weissenberger C, Cazalet C, Buchrieser C (2007) *Legionella pneumophila* - a human pathogen that co-evolved with fresh water protozoa. *Cellular and Molecular Life Sciences* **64**: 432-448

Alegado RA, Ferriera S, Nusbaum C, Young SK, Zeng Q, Imamovic A, Fairclough SR, King N (2011) Complete genome sequence of *Algoriphagus* sp. PR1, bacterial prey of a colony-forming choanoflagellate. *Journal of Bacteriology* **193**: 1485-1486

Amat F, Moussavi F, Comolli LR, Elidan G, Downing KH, Horowitz M (2008) Markov random field based automatic image alignment for electron tomography. *J Struct Biol* **161**: 260-275

Azam F (1998) Microbial control of oceanic carbon flux: The plot thickens. *Science* **280**: 694-696

Basler M, Pilhofer M, Henderson GP, Jensen GJ, Mekalanos JJ (2012) Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* **483**: 182-186

Blondel CJ, Jimenez JC, Contreras I, Santiviago CA (2009) Comparative genomic analysis uncovers 3 novel loci encoding type six secretion systems differentially distributed in *Salmonella* serotypes. *Bmc Genomics* **10**

Bonemann G, Pietrosiuk A, Mogk A (2010) Tubules and donuts: a type VI secretion story. *Mol Microbiol* **76**: 815-821

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* **55**: 611-622

Dumoux M, Clare DK, Saibil HR, Hayward RD (2012) *Chlamydiae* assemble a pathogen synapse to hijack the host endoplasmic reticulum. *Traffic*

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797

Eggerth AH, Gagnon BH (1933) The *Bacteroides* of Human Feces. *Journal of Bacteriology* **25**: 389-413

Ffrench-Constant R, Waterfield N (2006) An ABC guide to the bacterial toxin complexes. *Adv Appl Microbiol* **58**: 169-183

Flemming L, Rawlings D, Chenia H (2007) Phenotypic and molecular characterisation of fish-borne *Flavobacterium johnsoniae*-like isolates from aquaculture systems in South Africa. *Res Microbiol* **158**: 18-30

Furusawa G, Yoshikawa T, Takano Y, Mise K, Furusawa I, Okuno T, Sakata T (2005) Characterization of cytoplasmic fibril structures found in gliding cells of *Saprospira* sp. *Can J Microbiol* **51**: 875-880

Furusawa G, Yoshikawa T, Yasuda A, Sakata T (2003) Algicidal activity and gliding motility of *Saprospira* sp. SS98-5. *Can J Microbiol* **49**: 92-100

Galinski MR, Ingravallo P, Corredor-Medina C, Al-Khedery B, Povoia M, Barnwell JW (2001) *Plasmodium vivax* merozoite surface proteins-3beta and-3gamma share structural similarities with *P. vivax* merozoite surface protein-3alpha and define a new gene family. *Mol Biochem Parasitol* **115**: 41-53

Gomez-Consarnau L, Gonzalez JM, Coll-Llado M, Gourdon P, Pascher T, Neutze R, Pedros-Alio C, Pinhassi J (2007) Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. *Nature* **445**: 210-213

Greub G, Raoult D (2004) Microorganisms resistant to free-living amoebae. *Clin Microbiol Rev* **17**: 413-433

Gruwell ME, Morse GE, Normark BB (2007) Phylogenetic congruence of armored scale insects (Hemiptera : Diaspididae) and their primary endosymbionts from the phylum *Bacteroidetes*. *Molecular Phylogenetics and Evolution* **44**: 267-280

Hackstadt T (1998) The diverse habitats of obligate intracellular parasites. *Curr Opin Microbiol* **1**: 82-87

Harb OS, Gao LY, Abu Kwaik Y (2000) From protozoa to mammalian cells: a new paradigm in the life cycle of intracellular bacterial pathogens. *Environ Microbiol* **2**: 251-265

Horn M (2008) *Chlamydiae* as Symbionts in Eukaryotes. *Annual Review of Microbiology* **62**: 113-131

Horn M, Harzenetter MD, Linner T, Schmid EN, Muller KD, Michel R, Wagner M (2001) Members of the *Cytophaga-Flavobacterium-Bacteroides* phylum as intracellular bacteria of acanthamoebae: proposal of '*Candidatus* Amoebohilus asiaticus'. *Environ Microbiol* **3**: 440-449

Hunter MS, Perlman SJ, Kelly SE (2003) A bacterial symbiont in the *Bacteroidetes* induces cytoplasmic incompatibility in the parasitoid wasp *Encarsia pergandiella*. *Proc Biol Sci* **270**: 2185-2190

Hurst MR, Beard SS, Jackson TA, Jones SM (2007) Isolation and characterization of the *Serratia entomophila* antifeeding prophage. *FEMS Microbiol Lett* **270**: 42-48

Hurst MR, Glare TR, Jackson TA (2004) Cloning *Serratia entomophila* antifeeding genes--a putative defective prophage active against the grass grub *Costelytra zealandica*. *J Bacteriol* **186**: 5116-5128

Iancu CV, Tivol WF, Schooler JB, Dias DP, Henderson GP, Murphy GE, Wright ER, Li Z, Yu Z, Briegel A, Gan L, He Y, Jensen GJ (2006) Electron cryotomography sample preparation using the Vitrobot. *Nat Protocols* **1**: 2813-2819

Khan NA (2006) *Acanthamoeba*: biology and increasing importance in human health. *FEMS Microbiol Rev* **30**: 564-595

Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* **9**: 299-306

Lee HS, Kang SG, Kwon KK, Lee JH, Kim SJ (2011) Genome sequence of the algicidal bacterium *Kordia algicida* OT-1. *Journal of Bacteriology* **193**: 4031-4032

Leiman PG, Basler M, Ramagopal UA, Bonanno JB, Sauder JM, Pukatzki S, Burley SK, Almo SC, Mekalanos JJ (2009) Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc Natl Acad Sci U S A* **106**: 4154-4159

Letunic I, Bork P (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* **39**: W475-478

Marchesi JR (2010) Prokaryotic and eukaryotic diversity of the human gut. *Adv Appl Microbiol* **72**: 43-62

Mastrorade DN (2008) Correction for non-perpendicularity of beam and tilt axis in tomographic reconstructions with the IMOD package. *J Microsc* **230**: 212-217

Molmeret M, Horn M, Wagner M, Santic M, Abu Kwaik Y (2005) Amoebae as training grounds for intracellular bacterial pathogens. *Appl Environ Microbiol* **71**: 20-28

Molofsky AB, Swanson MS (2004) Differentiate to thrive: lessons from the *Legionella pneumophila* life cycle. *Mol Microbiol* **53**: 29-40

Noel GR, Atibalentja N (2006) '*Candidatus* Paenicardinium endonii', an endosymbiont of the plant-parasitic nematode *Heterodera glycines* (Nemata: *Tylenchida*), affiliated to the phylum *Bacteroidetes*. *Int J Syst Evol Microbiol* **56**: 1697-1702

Nour SM, Lawrence JR, Zhu H, Swerhone GD, Welsh M, Welacky TW, Topp E (2003) Bacteria associated with cysts of the soybean cyst nematode (*Heterodera glycines*). *Appl Environ Microbiol* **69**: 607-615

Penz T, Horn M, Schmitz-Esser S (2011) The genome of the amoeba symbiont "*Candidatus* Amoebophilus asiaticus" encodes an afp-like prophage possibly used for protein secretion. *Virulence* **1**: 541-545

Penz T, Schmitz-Esser S, Kelly SE, Cass BN, Muller A, Woyke T, Malfatti SA, Hunter MS, Horn M (2012) Comparative genomics suggests an independent origin of cytoplasmic incompatibility in *Cardinium hertigii*. *PLoS Genet* **8**: e1003012

Persson OP, Pinhassi J, Riemann L, Marklund BI, Rhen M, Normark S, Gonzalez JM, Hagstrom A (2009) High abundance of virulence gene homologues in marine bacteria. *Environ Microbiol* **11**: 1348-1357

Pinhassi J, Sala MM, Havskum H, Peters F, Guadayol O, Malits A, Marrase C (2004) Changes in bacterioplankton composition under different phytoplankton regimens. *Appl Environ Microbiol* **70**: 6753-6766

Pukatzki S, Ma AT, Revel AT, Sturtevant D, Mekalanos JJ (2007) Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 15508-15513

Riemann L, Steward GF, Azam F (2000) Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. *Appl Environ Microbiol* **66**: 578-587

Robert C, Chassard C, Lawson PA, Bernalier-Donadille A (2007) *Bacteroides cellulosilyticus* sp. nov., a cellulolytic bacterium from the human gut microbial community. *Int J Syst Evol Microbiol* **57**: 1516-1520

Rodriguez-Zaragoza S (1994) Ecology of free-living amoebae. *Crit Rev Microbiol* **20**: 225-241

Roy CR, Salcedo SP, Gorvel JPE (2006) Pathogen-endoplasmic-reticulum interactions: in through the out door. *Nature Reviews Immunology* **6**: 136-147

Sabree ZL, Kambhampati S, Moran NA (2009) Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. *Proc Natl Acad Sci U S A* **106**: 19521-19526

Saw JH, Yuryev A, Kanbe M, Hou S, Young AG, Aizawa S, Alam M (2012) Complete genome sequencing and analysis of *Saprospira grandis* str. Lewin, a predatory marine bacterium. *Stand Genomic Sci* **6**: 84-93

Schmitz-Esser S, Tischler P, Arnold R, Montanaro J, Wagner M, Rattei T, Horn M (2010) The genome of the amoeba symbiont "*Candidatus Amoebophilus asiaticus*" reveals common mechanisms for host cell interaction among amoeba-associated bacteria. *J Bacteriol* **192**: 1045-1057

Schmitz-Esser S, Toenshoff ER, Haider S, Heinz E, Hoenninger VM, Wagner M, Horn M (2008) Diversity of bacterial endosymbionts of environmental acanthamoeba isolates. *Appl Environ Microbiol* **74**: 5822-5831

Sunagawa S, Woodley CM, Medina M (2010) Threatened corals provide underexplored microbial habitats. *PLoS One* **5**: e9554

Thomas F, Hehemann JH, Rebuffet E, Czjzek M, Michel G (2011) Environmental and gut bacteroidetes: the food connection. *Front Microbiol* **2**: 93

Tivol WF, Briegel A, Jensen GJ (2008) An improved cryogen for plunge freezing. *Microsc Microanal* **14**: 375-379

Toft C, Andersson SG (2010) Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat Rev Genet* **11**: 465-475

Tseng TT, Tyler BM, Setubal JC (2009) Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol* **9 Suppl 1**: S2

Upreti RK, Kumar M, Shankar V (2003) Bacterial glycoproteins: functions, biosynthesis and applications. *Proteomics* **3**: 363-379

Weiss E (1991) Biology of ehrlichiae. *Eur J Epidemiol* **7**: 253-258

Wise KS, Foecking MF, Roske K, Lee YJ, Lee YM, Madan A, Calcutt MJ (2006) Distinctive repertoire of contingency genes conferring mutation- based phase variation and combinatorial expression of surface lipoproteins in *Mycoplasma capricolum* subsp. *capricolum* of the *Mycoplasma mycoides* phylogenetic cluster. *Journal of Bacteriology* **188**: 4926-4941

Yang G, Dowling AJ, Gerike U, ffrench-Constant RH, Waterfield NR (2006) *Photorhabdus* virulence cassettes confer injectable insecticidal activity against the wax moth. *Journal of Bacteriology* **188**: 2254-2261

Zheng SQ, Keszthelyi B, Branlund E, Lyle JM, Braunfeld MB, Sedat JW, Agard DA (2007) UCSF tomography: an integrated software suite for real-time electron microscopic tomographic data collection, alignment, and reconstruction. *J Struct Biol* **157**: 138-147

Zhou M, Wu H (2009) Glycosylation and biogenesis of a family of serine-rich bacterial adhesins. *Microbiology* **155**: 317-327

Zhou Y, Su J, Lai Q, Li X, Yang X, Dong P, Zheng T (2012) *Phaeocystidibacter luteus* gen. nov. sp. nov., a new member of the family *Cryomorphaceae* isolated from the marine alga *Phaeocystis globosa* and emended description of *Owenweeksia hongkongensis*. *Int J Syst Evol Microbiol*

Table S1. MIQE guidelines for RT q-PCR assay

item	importance	status	remarks
experimental design			
definition of experimental and control groups	E	OK	uninfected <i>Acanthamoebae</i> , <i>Acanthamoebae</i> infected with <i>Amoebophilus asiaticus</i> , extracellular <i>Amoebophilus asiaticus</i>
number within each group	E	OK	three biological replicates per sample, three technical replicates per biological replicates
assay carried out by core lab or investigator's lab?	D		
acknowledgement of authors' contributions	D		
sample			
description	E	OK	uninfected <i>Acanthamoebae</i> , <i>Acanthamoebae</i> infected with <i>Amoebophilus asiaticus</i> , extracellular <i>Amoebophilus asiaticus</i>
volume/mass of sample processed	D	OK	varying amounts of extracellular <i>Amoebophilus asiaticus</i> , infected and uninfected <i>Acanthamoeba</i> culture
microdissection or macrodissection	E	OK	neither (no animal/organ/tissue to dissect anything)
processing procedure	E	OK	samples were not processed
if frozen - how and how quickly?	E	OK	samples were not frozen
if fixed - with what, how quickly?	E	OK	samples were not fixed, TRIzol reagent added immediately after harvesting cells
sample storage conditions and duration (especially for ffpe samples)	E	OK	from opening cell culture flasks until addition of TRIzol approximately 5min
nucleic acid extraction			
procedure and/or instrumentation	E	OK	TRIzol-based total RNA extraction, bead beating
name of kit and details of any modifications	E	OK	TRIzol reagent (Life Technologies)
source of additional reagents used	D	OK	Chloroform (Roth), Isopropanol (Roth), Ethanol (Merck), water (life technologies)
details of DNase or RNase treatment	E	OK	TURBO DNA-free™ kit (Life Technologies)

contamination assessment (DNA or RNA)	E	OK	PCR and agarose gelelectrophoresis; -RT controls (reverse transcription reaction without addition of enzyme) analysed with qPCR
nucleic acid quantification	E	OK	NanoDrop for RNA, cDNA, gDNA, PicoGreen for quantification of standards
instrument and method	E	OK	NanoDrop ND-1000 UV-vis spectrophotometer (Thermo Scientific), tecan infinite m200 (Tecan)
purity (a260/a280)	D		
yield	D		
RNA integrity method/instrument	E	OK	Experion™ Automated Electrophoresis System (Bio-Rad)
RIN/RQI or Cq of 3' and 5' transcripts	E	OK	RIN > 8
electrophoresis traces	D		
inhibition testing (Cq dilutions, spike or other)	E	OK	no inhibition
reverse transcription			
complete reaction conditions	E	OK	
amount of RNA and reaction volume	E	OK	5 µl, corresponding to 500 ng, DNase treated RNA in a total reaction volume of 25 µl
priming oligonucleotide (if using GSP) and concentration	E	OK	random hexamer primers (final concentration 8 ng µl ⁻¹)
reverse transcriptase and concentration	E	OK	SuperScript™ III Reverse Transcriptase (final concentration 8 U µl ⁻¹)
temperature and time	E	OK	10 min 25 °C, 60 min 50 °C, 15 min 70 °C
manufacturer of reagents and catalogue numbers	D	OK	Invitrogen
Cqs with and without RT	D	OK	below limit of detection
storage conditions of cDNA	D	OK	-20 °C
qPCR target information			
if multiplex, efficiency and LOD of each assay.	E	OK	no multiplexing
sequence accession number	E		Aasi_1074 (Gene ID: 6377405), Aasi_1396 (GeneID:6377582)

location of amplicon	D	
amplicon length	E	OK Aasi_1074 qPCR F1/R1: 141 bp; Aasi_1396 qPCR F2/R2: 167 bp
<i>in silico</i> specificity screen (blast, etc)	E	OK done
pseudogenes, retrospseudogenes or other homologs?	D	
sequence alignment	D	
secondary structure analysis of amplicon	D	
location of each primer by exon or intron (if applicable)	E	OK not relevant
what splice variants are targeted?	E	OK not relevant
qPCR oligonucleotides		
primer sequences	E	OK Aasi_1074 qPCR F1 (5'-GTGGTGCAGATTGCTACATCAT-3'); Aasi_1074 qPCR R1 (5'-AGTCGGGCATAAGCAACATAGT-3'); Aasi_1396 qPCR F2 (5'-ACTAGGTACGCCACCTGAAAAA-3'); Aasi_1396 qPCR R2 (5'-AAGTTACTCCCTTCCACACA-3')
RTPPrimerdb identification number	D	OK not relevant, as all of them are unpublished newly designed primers
probe sequences	D	OK not relevant, as no probes were used
location and identity of any modifications	E	OK no modifications
manufacturer of oligonucleotides	D	OK Thermo Fisher Scientific
purification method	D	OK HPLC
qPCR protocol		
complete reaction conditions	E	OK
reaction volume and amount of cDNA/DNA	E	OK reaction volume = 50 µl; amount of cDNA = 5 µl, 1:10 dilution, hypothetically 10 ng DNA
primer, (probe), Mg ⁺⁺ and dNTP concentrations	E	OK Primer concentration = 200 pmol µl ⁻¹ (for all), Magnesium is part of the Bio-Rad supermix in unknown concentration, same for dNTPs
polymerase identity and concentration	E	OK Bio-Rad supermix
buffer/kit identity and manufacturer	E	OK Bio-Rad supermix

exact chemical constitution of the buffer	D	OK	Bio-Rad supermix
additives (SYBR green I, DMSO, etc.)	E	OK	SYBR Green I is contained in the supermix; no further additives
manufacturer of plates/tubes and catalog number	D	OK	Hard-Shell® Thin-Wall 96-Well Skirted PCR Plates (Bio-Rad) with white shell and clear wells
complete thermocycling parameters	E	OK	3 min 95 °C, 45× (30 sec 95 °C, 30 sec 65.7 °C, 30 sec 72 °C), 1 min 95 °C, melting curve (10 sec 55 °C, + 0.5 °C after each cycle up to 95 °C)
reaction setup (manual/robotic)	D	OK	manual
manufacturer of qpcr instrument	E	OK	Bio-Rad
qPCR validation			
evidence of optimisation (from gradients)	D	OK	done
specificity (gel, sequence, melt, or digest)	E	OK	done
for SYBR green I, C _q of the NTC	E	OK	no amplification
standard curves with slope and y-intercept	E	OK	done
PCR efficiency calculated from slope	E	OK	88.1 % (Aasi_1074) and 87.6 % (Aasi_1396)
confidence interval for PCR efficiency or standard error	D	OK	not determined
R ² of standard curve	E	OK	0.999 (Aasi_1074) and 0.998 (Aasi_1396)
linear dynamic range	E	OK	$1.2 \times 10^8 - 1.2 \times 10^1$ (Aasi_1074) and $1.5 \times 10^8 - 1.5 \times 10^1$ (Aasi_1396) copies
C _q variation at lower limit	E	OK	36.88 (Aasi_1074) and 36.87 (Aasi_1396)
confidence intervals throughout range	D	OK	not determined
evidence for limit of detection	E	OK	12 copies for Aasi_1074, 15 copies for Aasi_1396
if multiplex, efficiency and LOD of each assay.	E	OK	not relevant, no multiplexing
data analysis			
qPCR analysis program (source, version)	E	OK	CFX Manager v2.1 (Bio-Rad)

C _q method determination	E	OK	CFX Manager default settings (baseline subtracted curve fit, single threshold, automatically calculated). Threshold manually curated for maximum efficiency within linear range for each plate
outlier identification and disposition	E	OK	done
results of NTCs	E	OK	no amplificate
justification of number and choice of reference genes	E	OK	assumption that β -SU of RNA polymerase is constant expressed
description of normalisation method	E	OK	relative quantification – copy number of phage tail sheath mRNAs per sample normalised against copy number of reference mRNAs within the same sample
number and concordance of biological replicates	D	OK	3 biological replicates
number and stage (RT or qPCR) of technical replicates	E	OK	3 technical (RT) replicates for all samples; samples were analysed by qPCR without additional (qPCR) replicates; standards were applied in 3 technical (qPCR) replicates
repeatability (intra-assay variation)	E	OK	repeatable
reproducibility (inter-assay variation, %CV)	D	OK	not determined (strongly recommended for clinical/diagnostic applications, but not other assays)
power analysis	D	OK	not done
statistical methods for result significance	E	OK	done
software (source, version)	E	OK	Bio-Rad, CFX Manager v2.1
C _q or raw data submission using RDML	D	OK	done

E essential, D recommended

Table S2. Infectivity of intracellular and extracellular intermediates of *Amoebophilus*

	total number <i>Acanthamoebae</i>	infected <i>Acanthamoebae</i>	% infected <i>Acanthamoebae</i>	standard deviation % infected <i>Acanthamoebae</i>
intracellular <i>Amoebophilus</i> count 1	257	33	12.8	
intracellular <i>Amoebophilus</i> count 2	232	35	15.1	
intracellular <i>Amoebophilus</i> count 3	252	30	11.9	
mean % infected <i>Acanthamoebae</i> with intracellular <i>Amoebophilus</i>			13.3	1.6
extracellular <i>Amoebophilus</i> count 1	299	272	91.0	
extracellular <i>Amoebophilus</i> count 2	247	228	92.3	
extracellular <i>Amoebophilus</i> count 3	242	223	92.1	
mean % infected <i>Acanthamoebae</i> with extracellular <i>Amoebophilus</i>			91.8	0.7

Table S3. Mass spectrometry analysis results from purified phage tail sheath like structures

rank	identified <i>Ameobophilus</i> proteins (347)	accession number	molecular weight in kDa	number of detected peptides overall measurement	number of detected peptides in fraction 1	number of detected peptides in fraction 2	number of detected peptides in fraction 3	description
1	Aasi_1417	gi 189502722	370	90	25	7	18	putative insecticidal toxin
2	Aasi_1414	gi 189502719	289	77	0	0	10	putative insecticidal toxin
3	Aasi_1944	gi 294661399	138	77	24	24	17	putative uncharacterized protein
4	Aasi_1395	gi 189502703	161	71	23	32	25	DNA-directed RNA polymerase subunit beta
5	Aasi_1074	gi 189502431	55	66	8	8	3	Afp3-like phage tail sheath protein
6	Aasi_0378	gi 189501812	40	60	79	11	14	lysyl aminopeptidase
7	Aasi_0808	gi 189502196	45	60	14	13	20	gliding motility-associated lipoprotein GldJ
8	Aasi_0258	gi 189501713	45	59	41	19	35	outer membrane protein (porin)
9	Aasi_0008	gi 189501477	93	48	15	40	2	putative uncharacterized protein
10	Aasi_0013	gi 189501482	79	45	11	8	16	polyribonucleotide nucleotidyltransferase
11	Aasi_1396	gi 189502704	144	39	11	11	10	DNA-directed RNA polymerase subunit beta
12	Aasi_0153	gi 189501611	58	38	13	10	15	Rne/Rng family ribonuclease
13	Aasi_0380	gi 189501814	20	38	0	0	3	probable ferritin-1
14	Aasi_1916	gi 294661382	65	36	27	12	15	DEAD-box ATP-dependent RNA helicase CshA
15	Aasi_1513	gi 294661128	42	35	14	9	8	gliding motility-associated lipoprotein GldK
16	Aasi_0266	gi 189501718	22	34	9	6	10	adenylate kinase
17	Aasi_0340	gi 189501781	273	32	12	9	15	ANK
18	Aasi_1773	gi 294661294	49	28	15	13	16	signal recognition particle protein
19	Aasi_0557	gi 189501970	145	27	12	7	9	Afp11-like phage baseplate protein
20	Aasi_0590	gi 189502003	39	27	0	0	0	tetrahedral aminopeptidase
21	Aasi_0925	gi 189502302	55	27	5	4	14	F0F1 ATP synthase subunit beta
22	Aasi_0308	gi 189501756	58	26	4	6	10	chaperonin GroEL
23	Aasi_0481	gi 189501902	68	26	6	3	7	molecular chaperone DnaK

24	Aasi_0583	gi 189501996	25	25	3	0	6	putative uncharacterized protein
25	Aasi_1249	gi 189502583	53	25	17	13	15	cell division protein FtsZ
26	Aasi_1403	gi 189502711	44	24	10	8	15	elongation factor Tu
27	Aasi_1498	gi 294661116	67	24	18	4	8	peptide/opine/nickel uptake transporter
28	Aasi_0430	gi 189501858	29	23	10	10	12	ribosomal large subunit pseudouridine synthase B
29	Aasi_0038	gi 189501505	86	22	12	7	8	UvrD/REP helicase
30	Aasi_0191	gi 189501648	27	22	6	4	5	30S ribosomal protein S3
31	Aasi_1449	gi 189502750	64	22	8	2	0	putative uncharacterized protein
32	Aasi_0274	gi 189501725	70	21	3	2	2	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA
33	Aasi_0638	gi 189502047	35	21	9	14	4	putative uncharacterized protein
34	Aasi_0249	gi 189501704	37	20	8	9	9	rod shape-determining protein MreB
35	Aasi_0267	gi 189501719	36	19	9	8	8	GTPase obg
36	Aasi_0782	gi 189502172	102	19	7	7	8	putative uncharacterized protein
37	Aasi_0069	gi 189501531	56	18	0	8	0	ADP/ATP carrier protein
38	Aasi_0839	gi 189502222	125	18	8	5	10	solute sodium symporter
39	Aasi_1080	gi 189502437	66	18	6	8	13	Alp8-like VgrG protein
40	Aasi_1399	gi 189502707	25	18	10	11	9	50S ribosomal protein L1
41	Aasi_0694	gi 189502096	108	17	4	5	4	solute sodium symporter
42	Aasi_0806	gi 189502194	128	17	6	8	11	preprotein translocase subunit SecA
43	Aasi_1179	gi 189502524	45	17	10	10	12	ATP-dependent protease ATP-binding subunit ClpX
44	Aasi_0071	gi 189501532	24	16	5	4	3	alkyl hydroperoxide reductase/thiol specific antioxidant/Mal allergen
45	Aasi_1324	gi 189502646	39	16	9	8	9	recombinase A
46	Aasi_0559	gi 189501972	74	15	11	7	9	threonyl-tRNA synthetase
47	Aasi_1083	gi 189502440	97	15	14	6	10	Alp12-like protein
48	Aasi_1512	gi 294661127	37	15	0	4	2	putative uncharacterized protein
49	Aasi_0289	gi 189501739	52	14	0	3	5	bifunctional UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase/(3R)-hydroxymyristoyl-[acyl-carrier-protein] dehydratase
50	Aasi_1169	gi 189502515	48	14	7	6	7	GTP-binding protein EngA
51	Aasi_1447	gi 189502748	39	14	0	6	10	putative uncharacterized protein
52	Aasi_0125	gi 189501584	79	13	5	6	7	translation elongation factor G

53	Aasi_0226	gi 189501682	105	13	8	4	9	UvrABC system protein A
54	Aasi_0556	gi 189501969	28	13	2	2	8	Afp13-like virus fibre protein
55	Aasi_1018	gi 189502380	54	13	4	2	5	putative uncharacterized protein
56	Aasi_1398	gi 189502706	20	13	4	4	5	50S ribosomal protein L10
57	Aasi_0084	gi 189501545	53	12	0	0	2	peptidase S14 ClpP
58	Aasi_0173	gi 189501630	14	12	3	4	4	30S ribosomal protein S13
59	Aasi_0180	gi 189501637	18	12	4	6	7	30S ribosomal protein S5
60	Aasi_0793	gi 189502183	79	12	13	2	7	glutamate dehydrogenase
61	Aasi_0828	gi 189502213	91	12	7	2	2	DNA gyrase subunit A
62	Aasi_0979	gi 189502347	42	12	18	7	3	chaperone protein DnaJ
63	Aasi_1168	gi 189502514	51	12	10	6	10	DNA repair protein RadA
64	Aasi_1254	gi 189502588	42	12	0	0	14	putative uncharacterized protein
65	Aasi_0102	gi 189501562	90	11	7	3	3	phenylalanyl-tRNA synthetase subunit beta
66	Aasi_0283	gi 189501733	63	11	3	2	3	putative uncharacterized protein
67	Aasi_0825	gi 189502210	94	11	7	2	4	DNA topoisomerase IV subunit A
68	Aasi_1024	gi 189502384	61	11	6	5	7	uncharacterized ABC transporter ATP-binding protein YheS
69	Aasi_1181	gi 189502526	54	11	10	5	7	putative uncharacterized protein
70	Aasi_1182	gi 189502527	49	11	4	0	0	GTP pyrophosphokinase
71	Aasi_1248	gi 189502582	48	11	10	7	12	cell division protein FtsA
72	Aasi_1909	gi 294661376	30	11	0	2	7	putative uncharacterized protein
73	Aasi_0051	gi 189501516	15	10	2	6	4	30S ribosomal protein S9
74	Aasi_0066	gi 189501530	33	10	0	3	6	ATP synthase F1 subunit gamma
75	Aasi_0115	gi 189501575	38	10	3	4	0	3-oxoacyl-(acyl-carrier-protein) synthase III domain-containing protein
76	Aasi_0139	gi 189501597	101	10	2	2	3	translation initiation factor IF-2
77	Aasi_0182	gi 189501639	20	10	0	0	5	50S ribosomal protein L6
78	Aasi_0423	gi 189501851	40	10	3	0	4	leucine dehydrogenase
79	Aasi_0428	gi 189501856	29	10	4	4	6	putative uncharacterized protein
80	Aasi_1079	gi 189502436	38	10	2	4	6	Afp7-like protein
81	Aasi_1201	gi 189502543	50	10	4	3	4	tRNA modification GTPase MnmE

82	Aasi_1321	gi 189502643	49	10	9	0	4	4	DNA recombination protein rnuC homolog
83	Aasi_1391	gi 189502700	38	10	3	5	5	5	putative uncharacterized protein
84	Aasi_1697	gi 294661240	17	10	5	4	7	7	small heat shock protein C2
85	Aasi_1878	gi 294661362	89	10	2	0	0	0	putative uncharacterized protein
86	Aasi_0001	gi 189501471	55	9	4	4	4	4	chromosomal replication initiation protein
87	Aasi_0028	gi 189501497	107	9	6	0	2	2	putative uncharacterized protein
88	Aasi_0170	gi 189501627	37	9	0	5	5	5	DNA-directed RNA polymerase subunit alpha
89	Aasi_0222	gi 189501678	35	9	3	0	2	2	gliding motility ABC Transporter
90	Aasi_0325	gi 189501771	55	9	0	0	3	3	putative uncharacterized protein
91	Aasi_0442	gi 189501865	37	9	3	3	4	4	putative uncharacterized protein
92	Aasi_0614	gi 189502025	56	9	4	5	3	3	(Dimethylallyl)adenosine tRNA methyltransferase MiaB
93	Aasi_0705	gi 189502105	153	9	0	2	11	11	putative uncharacterized protein
94	Aasi_0916	gi 189502294	79	9	4	0	0	0	putative uncharacterized protein
95	Aasi_1040	gi 189502400	74	9	7	7	7	7	putative uncharacterized protein
96	Aasi_1075	gi 189502432	18	9	6	5	5	5	Atp-like protein
97	Aasi_1200	gi 189502542	42	9	7	4	5	5	tRNA-specific 2-thiouridylyase MnmA
98	Aasi_1319	gi 189502641	82	9	8	5	8	8	ribonuclease R
99	Aasi_1426	gi 189502729	27	9	0	3	4	4	putative uncharacterized protein
100	Aasi_0044	gi 189501510	52	8	2	2	3	3	putative uncharacterized protein
101	Aasi_0172	gi 189501629	14	8	3	5	4	4	30S ribosomal protein S11
102	Aasi_0188	gi 189501645	10	8	0	0	0	0	30S ribosomal protein S17
103	Aasi_0190	gi 189501647	16	8	0	0	2	2	50S ribosomal protein L16
104	Aasi_0192	gi 226733425	16	8	3	3	3	3	50S ribosomal protein L22
105	Aasi_0194	gi 189501651	30	8	5	2	8	8	50S ribosomal protein L2
106	Aasi_0379	gi 189501813	43	8	2	8	4	4	putative uncharacterized protein
107	Aasi_0457	gi 189501879	49	8	6	5	7	7	uncharacterized transporter yclF
108	Aasi_0653	gi 189502061	36	8	5	4	3	3	tRNA dimethylallyltransferase
109	Aasi_0710	gi 189502109	17	8	3	4	3	3	putative uncharacterized protein
110	Aasi_0731	gi 189502126	67	8	4	0	6	6	uncharacterized ABC transporter ATP-binding protein HI_1051

111	Aasi_0759	gi 189502152	25	8	6	5	5	lipoprotein Translocase (LPT)
112	Aasi_0794	gi 189502184	26	8	0	3	3	phosphatidylserine decarboxylase
113	Aasi_1039	gi 189502399	52	8	5	0	5	putative uncharacterized protein
114	Aasi_1049	gi 189502409	58	8	3	3	7	phosphoenolpyruvate carboxykinase
115	Aasi_0050	gi 189501515	34	7	0	0	6	30S ribosomal protein S2
116	Aasi_0099	gi 189501559	52	7	5	0	0	putative uncharacterized protein
117	Aasi_0126	gi 189501585	18	7	0	2	2	30S ribosomal protein S7
118	Aasi_0151	gi 189501609	47	7	2	2	3	enolase
119	Aasi_0197	gi 189501654	23	7	5	4	4	50S ribosomal protein L3
120	Aasi_0253	gi 189501708	58	7	0	2	0	putative uncharacterized protein
121	Aasi_0324	gi 189501770	89	7	0	0	0	putative uncharacterized protein
122	Aasi_0474	gi 189501896	15	7	0	0	0	nucleoside diphosphate kinase
123	Aasi_0482	gi 189501903	22	7	0	0	2	50S ribosomal protein L25/general stress protein Ctc
124	Aasi_0562	gi 189501975	13	7	0	0	0	50S ribosomal protein L20
125	Aasi_0772	gi 189502163	92	7	0	0	0	putative uncharacterized protein
126	Aasi_0942	gi 189502314	106	7	8	4	10	UvrABC system protein A
127	Aasi_1150	gi 189502498	88	7	0	3	7	DNA topoisomerase I
128	Aasi_1310	gi 189502633	23	7	0	0	2	putative uncharacterized protein
129	Aasi_1328	gi 189502650	56	7	0	5	0	putative uncharacterized protein
130	Aasi_1338	gi 189502657	58	7	2	0	0	putative uncharacterized protein
131	Aasi_0057	gi 189501522	23	6	4	2	3	methyladenine glycosylase
132	Aasi_0171	gi 189501628	23	6	0	0	0	RNA-binding S4 domain-containing protein
133	Aasi_0281	gi 189501731	35	6	0	0	2	putative uncharacterized protein
134	Aasi_0290	gi 189501740	36	6	3	3	4	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase
135	Aasi_0435	gi 189501860	51	6	0	0	0	dihydrolipoyl dehydrogenase
136	Aasi_0564	gi 189501977	35	6	3	0	6	PhoH-like protein
137	Aasi_0617	gi 189502027	24	6	2	4	2	probable GTP-binding protein EngB
138	Aasi_0635	gi 189502044	45	6	7	0	3	bifunctional protein GImU
139	Aasi_0651	gi 189502059	36	6	2	6	0	pyruvate dehydrogenase E1 component subunit beta

140	Aasi_0785	gi 189502175	50	6	8	0	0	0	putative uncharacterized protein
141	Aasi_0861	gi 189502242	57	6	0	2	0	0	putative metalloprotease ypwA
142	Aasi_0981	gi 189502349	89	6	0	0	2	0	penicillin-binding protein 1A
143	Aasi_1076	gi 189502433	16	6	4	4	4	4	Afp-like protein
144	Aasi_1164	gi 189502510	26	6	3	2	3	0	putative TrmH family tRNA/rRNA methyltransferase
145	Aasi_1354	gi 189502671	32	6	0	0	0	0	nucleoside triphosphate pyrophosphohydrolase
146	Aasi_1427	gi 189502730	44	6	6	0	0	0	aspartate aminotransferase A
147	Aasi_1439	gi 189502741	113	6	7	0	2	0	putative uncharacterized protein
148	Aasi_1645	gi 294661206	103	6	3	2	0	0	putative uncharacterized protein
149	Aasi_0012	gi 189501481	11	5	0	0	0	0	30S ribosomal protein S15
150	Aasi_0232	gi 189501688	77	5	4	4	7	0	Afp15-like ATPase
151	Aasi_0388	gi 189501821	26	5	0	2	2	0	tRNA pseudouridine synthase B
152	Aasi_0401	gi 189501831	51	5	3	4	0	0	cytosolic non-specific dipeptidase
153	Aasi_0420	gi 189501848	12	5	0	0	0	0	UPF0092 membrane protein aq_1254
154	Aasi_0422	gi 189501850	44	5	0	3	0	0	putative uncharacterized protein
155	Aasi_0436	gi 189501861	127	5	2	4	3	0	isoleucyl-tRNA synthetase
156	Aasi_0467	gi 189501889	83	5	0	0	7	0	putative uncharacterized protein
157	Aasi_0475	gi 189501897	39	5	7	0	0	0	phenylalanine--tRNA ligase alpha subunit
158	Aasi_0478	gi 189501899	69	5	2	0	0	0	putative uncharacterized protein
159	Aasi_0519	gi 189501934	30	5	0	0	0	0	putative uncharacterized protein
160	Aasi_0597	gi 189502010	107	5	2	0	0	0	putative uncharacterized protein
161	Aasi_0601	gi 189502014	37	5	2	2	4	0	uncharacterized protein MG371 homolog
162	Aasi_0618	gi 189502028	54	5	6	3	4	0	probable periplasmic serine endoprotease DegP-like
163	Aasi_0665	gi 189502073	39	5	6	0	3	0	uncharacterized RNA pseudouridine synthase Cpar_0723
164	Aasi_0698	gi 189502099	23	5	3	2	2	0	putative uncharacterized protein
165	Aasi_0930	gi 189502307	14	5	0	0	2	0	50S ribosomal protein L19
166	Aasi_0952	gi 189502323	37	5	5	5	4	0	putative uncharacterized protein
167	Aasi_1055	gi 189502414	31	5	0	0	0	0	probable transposase for insertion sequence element IS702
168	Aasi_1096	gi 189502451	44	5	13	0	0	0	yxeP

169	Aasi_1165	gi 189502511	83	5	2	2	3	3	ATP-dependent DNA helicase recQ
170	Aasi_1326	gi 189502648	22	5	3	0	3	3	peptide deformylase
171	Aasi_1353	gi 189502670	26	5	6	3	9	9	cell division ATP-binding protein FtsE
172	Aasi_1366	gi 189502682	49	5	0	0	0	0	putative uncharacterized protein
173	Aasi_1530	gi 294661138	33	5	0	2	2	2	probable chromosome-partitioning protein parB
174	Aasi_1806	gi 294661314	180	5	3	3	5	5	Afp14-like protein
175	Aasi_0027	gi 189501496	41	4	0	2	11	11	permease YjgP/YjgQ family protein
176	Aasi_0036	gi 189501503	48	4	4	0	0	0	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
177	Aasi_0092	gi 189501552	110	4	2	4	3	3	bifunctional preprotein translocase subunit SecD/SecE
178	Aasi_0101	gi 189501561	27	4	0	0	8	8	DNA repair protein RadC
179	Aasi_0166	gi 189501623	31	4	2	2	0	0	starch synthase catalytic domain-containing protein
180	Aasi_0185	gi 189501642	20	4	0	2	4	4	50S ribosomal protein L5
181	Aasi_0195	gi 189501652	11	4	0	0	0	0	50S ribosomal protein L23
182	Aasi_0196	gi 189501653	23	4	0	0	0	0	50S ribosomal protein L4
183	Aasi_0220	gi 189501676	64	4	4	4	0	0	gliding motility ABC Transporter
184	Aasi_0292	gi 189501742	17	4	0	0	2	2	putative uncharacterized protein
185	Aasi_0301	gi 189501750	45	4	3	3	0	0	CinA-like protein
186	Aasi_0399	gi 189501829	15	4	0	0	0	0	30S ribosomal protein S6
187	Aasi_0535	gi 189501948	41	4	5	4	3	3	ribonucleoside-diphosphate reductase subunit M2
188	Aasi_0578	gi 189501991	100	4	6	0	0	0	alanyl-tRNA synthetase
189	Aasi_0588	gi 189502001	56	4	0	0	0	0	uncharacterized protein YifB
190	Aasi_0595	gi 189502008	79	4	3	2	2	2	putative uncharacterized protein
191	Aasi_0644	gi 189502053	69	4	3	3	7	7	30S ribosomal protein S1
192	Aasi_0867	gi 189502248	39	4	0	2	3	3	drug exporter-1 (DrugE1)
193	Aasi_1073	gi 189502430	33	4	3	6	5	5	Afp-like protein
194	Aasi_1180	gi 189502525	25	4	0	0	0	0	ATP-dependent Clp protease proteolytic subunit
195	Aasi_1207	gi 189502547	45	4	4	4	0	0	3-oxoacyl-[acyl-carrier-protein] synthase 2
196	Aasi_1209	gi 189502549	53	4	0	0	3	3	pyruvate kinase 2
197	Aasi_1231	gi 189502568	48	4	4	0	0	0	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase

198	Aasi_1247	gi 189502581	30	4	0	0	0	0	0	putative uncharacterized protein
199	Aasi_1428	gi 189502731	16	4	0	0	0	0	0	putative uncharacterized protein
200	Aasi_1634	gi 294661199	56	4	0	0	0	2	0	putative uncharacterized protein
201	Aasi_0058	gi 189501523	30	3	0	0	0	0	0	enoyl-(acyl-carrier-protein) reductase
202	Aasi_0064	gi 189501528	23	3	2	0	0	3	0	ATP synthase F1 subunit delta
203	Aasi_0079	gi 189501540	48	3	0	0	0	6	0	peptidase M23 (membrane bound)
204	Aasi_0105	gi 189501565	26	3	0	0	0	0	0	3-oxoacyl-(acyl-carrier-protein) reductase
205	Aasi_0107	gi 189501567	12	3	0	0	0	0	0	50S ribosomal protein L21
206	Aasi_0109	gi 189501569	13	3	0	0	0	2	0	30S ribosomal protein S16
207	Aasi_0193	gi 189501650	10	3	0	0	0	0	0	30S ribosomal protein S19
208	Aasi_0203	gi 189501660	37	3	0	0	4	0	0	probable peptide ABC transporter ATP-binding protein y4R
209	Aasi_0223	gi 189501679	50	3	0	0	2	0	0	putative methylthiotransferase yqeV
210	Aasi_0275	gi 189501726	46	3	3	0	0	0	0	serine hydroxymethyltransferase
211	Aasi_0276	gi 189501727	46	3	3	3	3	3	0	uncharacterized zinc protease y4wA
212	Aasi_0311	gi 189501758	79	3	2	0	0	3	0	ATP-dependent DNA helicase RecG
213	Aasi_0424	gi 189501852	37	3	2	0	0	0	0	aspartate-semialdehyde dehydrogenase
214	Aasi_0425	gi 189501853	74	3	2	2	2	0	0	putative uncharacterized protein
215	Aasi_0468	gi 189501890	57	3	2	0	0	2	0	putative uncharacterized protein
216	Aasi_0506	gi 189501923	46	3	0	0	0	0	0	putative uncharacterized protein (TMH)
217	Aasi_0524	gi 189501939	95	3	2	2	3	0	0	primosomal protein N'
218	Aasi_0570	gi 189501983	10	3	0	0	0	0	0	50S ribosomal protein L28
219	Aasi_0573	gi 189501986	34	3	0	0	6	2	0	cell division protein FtsY homolog
220	Aasi_0624	gi 189502034	27	3	0	0	0	0	0	probable transcriptional regulatory protein
221	Aasi_0632	gi 189502041	46	3	3	0	0	0	0	putative uncharacterized protein
222	Aasi_0646	gi 189502055	79	3	0	0	0	0	0	putative uncharacterized protein
223	Aasi_0691	gi 189502093	17	3	0	0	0	0	0	putative uncharacterized protein
224	Aasi_0755	gi 189502148	67	3	0	0	0	0	2	putative uncharacterized protein
225	Aasi_0780	gi 189502170	39	3	3	0	0	0	0	S-adenosylmethionine:tRNA ribosyltransferase-isomerase
226	Aasi_0918	gi 189502296	36	3	2	0	3	2	0	UPF0176 protein CHU_2691

255	Aasi_1141	gi 189502490	34	2	6	2	3	GTP-binding protein Era
256	Aasi_1361	gi 189502677	63	2	0	0	0	putative uncharacterized protein
257	Aasi_1400	gi 189502708	15	2	0	2	0	50S ribosomal protein L11
258	Aasi_1429	gi 189502732	68	2	0	0	0	putative uncharacterized protein
259	Aasi_1430	gi 189502733	38	2	0	0	0	fructose-bisphosphate aldolase
260	Aasi_1494	gi 294661112	13	2	0	0	0	50S ribosomal protein L18
261	Aasi_1504	gi 294661121	31	2	0	0	0	putative uncharacterized protein
262	Aasi_1612	gi 294661180	13	2	0	0	0	putative uncharacterized protein
263	Aasi_1648	gi 294661209	90	2	0	0	2	putative uncharacterized protein
264	Aasi_0011	gi 189501480	79	0	0	2	0	permease YigP/YigQ family protein
265	Aasi_0014	gi 189501483	33	0	0	2	0	RNA polymerase sigma factor
266	Aasi_0034	gi 189501501	86	0	0	0	2	Sel1 domain-containing protein
267	Aasi_0040	gi 189501507	124	0	2	2	0	metal-dependent phosphohydrolase
268	Aasi_0048	gi 189501513	29	0	3	2	3	tRNA pseudouridine synthase A
269	Aasi_0052	gi 189501517	17	0	0	2	4	50S ribosomal protein L13
270	Aasi_0065	gi 189501529	57	0	3	3	3	ATP synthase F1 subunit alpha
271	Aasi_0120	gi 189501579	62	0	4	6	3	putative uncharacterized protein
272	Aasi_0155	gi 189501613	43	0	2	3	0	A/G-specific adenine glycosylase
273	Aasi_0168	gi 189501625	81	0	0	3	3	putative uncharacterized protein
274	Aasi_0169	gi 189501626	22	0	0	0	2	50S ribosomal protein L17
275	Aasi_0176	gi 189501633	28	0	0	7	6	methionine aminopeptidase
276	Aasi_0177	gi 189501634	48	0	0	4	3	preprotein translocase subunit SecY
277	Aasi_0183	gi 189501640	15	0	2	3	3	30S ribosomal protein S8
278	Aasi_0201	gi 189501658	47	0	0	4	0	uncharacterized metallophosphoesterase ykuE
279	Aasi_0225	gi 189501681	46	0	2	0	0	putative uncharacterized protein
280	Aasi_0229	gi 189501685	44	0	2	0	0	uncharacterized protein ybbC
281	Aasi_0264	gi 189501716	57	0	2	0	0	transcription termination factor Rho
282	Aasi_0299	gi 189501748	38	0	0	0	3	holliday junction DNA helicase RuvB
283	Aasi_0300	gi 189501749	18	0	3	3	2	dihydrofolate reductase type 3

284	Aasi_0303	gi 189501751	42	0	0	6	2	6	putative uncharacterized protein
285	Aasi_0341	gi 189501782	97	0	0	0	0	2	putative uncharacterized protein
286	Aasi_0351	gi 189501788	69	0	0	0	0	2	lipid A export ATP-binding/permease protein MsbA
287	Aasi_0355	gi 189501791	125	0	0	0	0	4	putative uncharacterized protein
288	Aasi_0359	gi 189501795	38	0	0	0	2	0	putative uncharacterized protein
289	Aasi_0370	gi 189501805	37	0	0	0	0	3	fructose-1,6-bisphosphatase class 1
290	Aasi_0426	gi 189501854	43	0	0	0	2	2	tetraacyl/disaccharide 4'-kinase
291	Aasi_0429	gi 189501857	15	0	0	0	2	2	putative uncharacterized protein
292	Aasi_0449	gi 189501872	127	0	2	2	3	2	putative uncharacterized protein
293	Aasi_0465	gi 189501887	94	0	0	3	2	0	negative regulator of genetic competence ClpC/MecB
294	Aasi_0471	gi 189501893	52	0	0	0	0	2	ATP-dependent protease ATP-binding subunit HslU
295	Aasi_0507	gi 189501924	52	0	0	0	0	2	magnesium transporter mgfE
296	Aasi_0534	gi 189501947	48	0	0	2	0	0	putative uncharacterized protein
297	Aasi_0541	gi 189501954	64	0	0	2	0	2	dimethylamine monooxygenase [N-oxide-forming] 5
298	Aasi_0566	gi 189501979	25	0	0	0	0	2	O-methyltransferase mdmC
299	Aasi_0577	gi 189501990	54	0	0	0	2	0	aspartyl/glutamyl-tRNA amidotransferase subunit B
300	Aasi_0628	gi 189502038	14	0	0	0	0	2	putative uncharacterized protein
301	Aasi_0636	gi 189502045	60	0	0	0	0	5	putative uncharacterized protein
302	Aasi_0648	gi 189502056	71	0	0	3	0	0	DNA topoisomerase IV subunit B
303	Aasi_0671	gi 189502078	40	0	0	0	3	0	uncharacterized protein RC0012
304	Aasi_0674	gi 189502079	66	0	0	0	0	2	uncharacterized ABC transporter ATP-binding protein HI_1051
305	Aasi_0708	gi 189502107	48	0	0	2	0	0	prolipoprotein diacylglycerol transferase
306	Aasi_0730	gi 189502125	68	0	0	0	0	2	GTP-binding protein TypA/BipA homolog
307	Aasi_0765	gi 189502157	31	0	0	0	0	4	putative uncharacterized protein
308	Aasi_0788	gi 189502178	54	0	0	0	0	3	asparaginyl-tRNA synthetase
309	Aasi_0804	gi 189502192	56	0	0	2	3	0	putative uncharacterized protein
310	Aasi_0824	gi 189502209	29	0	0	0	0	2	putative uncharacterized protein
311	Aasi_0838	gi 189502221	27	0	0	0	0	2	uncharacterized HTH-type transcriptional regulator yobV
312	Aasi_0858	gi 189502239	36	0	0	0	2	0	putative uncharacterized protein

313	Aasi_0900	gi 189502279	29	0	0	0	2	putative uncharacterized protein
314	Aasi_0939	gi 189502312	129	0	4	2	2	putative uncharacterized protein
315	Aasi_0944	gi 189502316	91	0	0	2	2	putative uncharacterized protein
316	Aasi_0948	gi 189502320	57	0	0	0	2	replicative DNA helicase
317	Aasi_0955	gi 189502325	203	0	0	0	2	putative uncharacterized protein
318	Aasi_0992	gi 189502358	31	0	2	4	6	probable transposase for insertion sequence element IS702
319	Aasi_0997	gi 189502363	16	0	4	3	3	uncharacterized sufe-like protein RHE_CH01250
320	Aasi_1006	gi 189502370	17	0	0	0	2	putative peroxiredoxin bep
321	Aasi_1036	gi 189502396	30	0	0	0	2	2,3,4,5-tetrahydropyridine-2,6-carboxylate N-succinyltransferase
322	Aasi_1042	gi 189502402	98	0	0	0	3	chaperone protein C1pB
323	Aasi_1048	gi 189502408	38	0	3	0	0	peptide chain release factor 2
324	Aasi_1155	gi 189502503	27	0	3	0	8	putative uncharacterized protein
325	Aasi_1156	gi 189502504	34	0	0	0	2	putative uncharacterized protein
326	Aasi_1170	gi 189502516	54	0	4	5	2	cysteine desulfurase activator complex subunit SufB
327	Aasi_1174	gi 189502520	71	0	2	2	2	DNA mismatch repair protein mutL
328	Aasi_1183	gi 189502528	50	0	2	2	0	proline/betaine transporter
329	Aasi_1196	gi 189502538	31	0	0	0	2	putative uncharacterized protein
330	Aasi_1230	gi 189502567	41	0	0	0	2	protein smf
331	Aasi_1232	gi 189502569	45	0	0	0	4	GTPase HflX
332	Aasi_1257	gi 189502591	21	0	2	0	0	uncharacterized protein ywjB
333	Aasi_1282	gi 189502610	27	0	0	0	3	insertion element IS1 3 protein insB
334	Aasi_1365	gi 189502681	36	0	0	2	0	tryptophan-tRNA ligase
335	Aasi_1410	gi 189502717	30	0	0	0	3	thymidylate synthase
336	Aasi_1421	gi 189502725	28	0	0	0	2	lipopolysaccharide export system ATP-binding protein LptB
337	Aasi_1424	gi 189502727	94	0	0	0	3	DNA translocase FtsK
338	Aasi_1444	gi 189502745	130	0	2	2	0	putative uncharacterized protein
339	Aasi_1519	gi 294661131	50	0	0	0	3	serine-tRNA ligase
340	Aasi_1524	gi 294661134	48	0	2	0	2	Xaa-Pro aminopeptidase
341	Aasi_1597	gi 294661171	63	0	2	0	0	DNA polymerase/3'-5' exonuclease PolX

342	Aasi_1615	gi 294661182	68	0	0	2	2	putative uncharacterized protein
343	Aasi_1658	gi 294661217	70	0	0	2	0	uncharacterized protein yyaL
344	Aasi_1766	gi 294661288	53	0	0	3	0	putative uncharacterized protein
345	Aasi_1922	gi 294661386	29	0	0	0	3	uncharacterized deoxyribonuclease YabD
346	Aasi_1958	gi 294661408	34	0	0	0	2	putative uncharacterized protein
347	Aasi_1962	gi 294661410	120	0	2	0	2	putative uncharacterized protein

Table S4. Antifeeding prophage (AFP) like proteins encoded in the genome of *Amoebophilus* and their amino acid similarities to the AFP of *Serratia entomophila* encoded on the pADAP plasmid

<i>Amoebophilus</i> AFP-like protein	<i>Serratia</i> (AFP_tag, aa identities to AFP, E-value, PFAM domain)	putative function
Aasi_0557	Afp11, I=24%/35%, E=1e-15/2e-8, -	phage baseplate
Aasi_0556	Afp13, I=31%, E=0.8, <i>Adeno_shaft</i> (PF00608)	virus fibre protein
Aasi_1072	Afp16, I=9%, E=0.069, -	n.d.
Aasi_1073	-	n.d.
Aasi_1074	Afp3, I=49%, E=7e-42, <i>Phage_sheath_I</i> (PF04984) Afp2, I=48%, E=3e-41, <i>Phage_sheath_I</i> (PF04984) Afp4, I=26%, E=3e-31,	phage tail sheath
Aasi_1075	-	n.d.
Aasi_1076	-	n.d.
Aasi_1077	Afp1, I=23%, E=9.9, <i>Phage_T4_gp19</i> (PF06841) Afp5, I=22%, E=3e-21, <i>Phage_T4_gp19</i> (PF06841)	phage tail tube
Aasi_1078	-	n.d.
Aasi_1079	Afp7, I=27%, E=3e-09, -	n.d.
Aasi_1080	Afp8, I=22%, E=2e-26, <i>Phage_GPD</i> (PF05954)/ <i>Phage_base_V</i> (PF04717)	VgrG
Aasi_1081	-	n.d.
Aasi_1082	Afp9, I=29%, E=5e-29, <i>GPW_gp25</i> (PF04965)	lysozyme
Aasi_1083	Afp12, I=22%, E=2e-15, -	n.d.
Aasi_1806	Afp14, I=15%/12%, E=8e-30/1, -	n.d.

n.d. not determined

Table S5. **RT q-PCR values standard curve Aasi_1074**

target gene	standard	Cq	Cq mean	Cq SD	SQ
Aasi_1074	1	10,99	10,97	0,023	115212773,00000
Aasi_1074	2	14,93	14,79	0,149	11521277,30000
Aasi_1074	3	18,48	18,36	0,139	1152127,73000
Aasi_1074	4	22,13	21,96	0,145	115212,77300
Aasi_1074	5	25,56	25,44	0,105	11521,27730
Aasi_1074	6	29,06	29,04	0,017	1152,12773
Aasi_1074	7	32,68	32,63	0,263	115,21277
Aasi_1074	8	36,22	36,88	0,604	11,52128
Aasi_1074	1	10,98	10,97		
Aasi_1074	2	14,63	14,79		
Aasi_1074	3	18,21	18,36		
Aasi_1074	4	21,87	21,96		
Aasi_1074	5	25,36	25,44		
Aasi_1074	6	29,02	29,04		
Aasi_1074	7	32,35	32,63		
Aasi_1074	8	37,41	36,88		
Aasi_1074	1	10,95	10,97		
Aasi_1074	2	14,81	14,79		
Aasi_1074	3	18,40	18,36		
Aasi_1074	4	21,89	21,96		
Aasi_1074	5	25,40	25,44		
Aasi_1074	6	29,04	29,04		
Aasi_1074	7	32,87	32,63		
Aasi_1074	8	37,00	36,88		

SQ...starting quantity of plasmid standards estimated with PicoGreen

Table S6. **RT q-PCR values standard curve Aasi_1396**

target gene	standard	Cq	Cq mean	Cq SD	SQ
Aasi_1396	1	10,52	10,75	0,236	145741066,00000
Aasi_1396	2	15,55	15,30	0,230	14574106,60000
Aasi_1396	3	18,96	18,87	0,185	1457410,66000
Aasi_1396	4	22,76	22,50	0,238	145741,06600
Aasi_1396	5	26,39	26,03	0,331	14574,10660
Aasi_1396	6	29,87	29,83	0,065	1457,41066
Aasi_1396	7	32,74	32,95	0,256	145,74107
Aasi_1396	8	36,51	36,87	0,598	14,57411
Aasi_1396	1	10,74	10,75		
Aasi_1396	2	15,10	15,30		
Aasi_1396	3	18,65	18,87		
Aasi_1396	4	22,30	22,50		
Aasi_1396	5	25,73	26,03		
Aasi_1396	6	29,76	29,83		
Aasi_1396	7	33,23	32,95		
Aasi_1396	8	37,56	36,87		
Aasi_1396	1	11,00	10,75		
Aasi_1396	2	15,26	15,30		
Aasi_1396	3	18,99	18,87		
Aasi_1396	4	22,43	22,50		
Aasi_1396	5	25,98	26,03		
Aasi_1396	6	29,87	29,83		
Aasi_1396	7	32,87	32,95		
Aasi_1396	8	36,55	36,87		

SQ...starting quantity of plasmid standards estimated with PicoGreen

Table S7. RT q-PCR values Aasi_1074

target	sample	Cq	Cq mean	Cq SD	SQ	SQ mean	SQ SD
Aasi_1074	12 h p.i. 1.1	28,26	28,79	0,494	2125,15890	1570,44870	502,55122
Aasi_1074	12 h p.i. 1.2	28,87			1440,67469		
Aasi_1074	12 h p.i. 1.3	29,24			1145,51252		
Aasi_1074	12 h p.i. 2.1	27,50	28,13	0,547	3428,06175	2398,67496	891,67409
Aasi_1074	12 h p.i. 2.2	28,43			1902,81751		
Aasi_1074	12 h p.i. 2.3	28,46			1865,14560		
Aasi_1074	12 h p.i. 3.1	28,48	28,66	0,166	1849,95876	1657,45346	175,85153
Aasi_1074	12 h p.i. 3.2	28,69			1617,14760		
Aasi_1074	12 h p.i. 3.3	28,80			1505,25403		
Aasi_1074	68 h p.i. 1.1	31,93	32,26	0,360	208,14967	172,66023	37,87909
Aasi_1074	68 h p.i. 1.2	32,19			177,05597		
Aasi_1074	68 h p.i. 1.3	32,65			132,77505		
Aasi_1074	68 h p.i. 2.1	32,33	33,42	1,001	162,19528	93,21062	60,90129
Aasi_1074	68 h p.i. 2.2	33,65			70,54215		
Aasi_1074	68 h p.i. 2.3	34,29			46,89441		
Aasi_1074	68 h p.i. 3.1	32,48	32,66	0,155	147,68448	132,38362	13,28630
Aasi_1074	68 h p.i. 3.2	32,73			125,70193		
Aasi_1074	68 h p.i. 3.3	32,76			123,76445		
Aasi_1074	140 h p.i. 1.1	24,09	24,19	0,093	29533,00595	27748,22148	1636,58978
Aasi_1074	140 h p.i. 1.2	24,21			27393,72741		
Aasi_1074	140 h p.i. 1.3	24,28			26317,93109		
Aasi_1074	140 h p.i. 2.1	23,19	23,28	0,071	52155,73965	49540,13696	2265,45333
Aasi_1074	140 h p.i. 2.2	23,32			48267,63074		
Aasi_1074	140 h p.i. 2.3	23,32			48197,04051		
Aasi_1074	140 h p.i. 3.1	23,85	24,00	0,164	34456,44314	31354,70636	3217,64515
Aasi_1074	140 h p.i. 3.2	23,99			31575,18266		
Aasi_1074	140 h p.i. 3.3	24,18			28032,49328		
Aasi_1074	extracellular 1.1	18,66	18,73	0,106	915398,80920	878766,82111	57627,76118
Aasi_1074	extracellular 1.2	18,67			908560,40972		
Aasi_1074	extracellular 1.3	18,85			812341,24439		
Aasi_1074	extracellular 2.1	17,75	17,91	0,248	1627572,25582	1480377,38527	220567,90932
Aasi_1074	extracellular 2.2	17,79			1586781,21439		
Aasi_1074	extracellular 2.3	18,20			1226778,68561		
Aasi_1074	extracellular 3.1	17,28	17,63	0,306	2192545,71877	1782178,16634	359234,10307
Aasi_1074	extracellular 3.2	17,75			1629415,70969		
Aasi_1074	extracellular 3.3	17,85			1524573,07057		

Table S8. RT q-PCR values Aasi_1396

target	sample	Cq	Cq mean	Cq SD	SQ	SQ mean	SQ SD
Aasi_1396	12 h p.i. 1.1	26,23	26,76	0,458	12338,48951	9116,29711	2795,20861
Aasi_1396	12 h p.i. 1.2	26,99			7667,36830		
Aasi_1396	12 h p.i. 1.3	27,06			7343,03352		
Aasi_1396	12 h p.i. 2.1	25,73	26,30	0,488	16891,40486	12263,01796	4015,93814
Aasi_1396	12 h p.i. 2.2	26,54			10196,38297		
Aasi_1396	12 h p.i. 2.3	26,62			9701,26606		
Aasi_1396	12 h p.i. 3.1	27,05	27,18	0,159	7370,57661	6826,02456	666,76122
Aasi_1396	12 h p.i. 3.2	27,13			7025,09537		
Aasi_1396	12 h p.i. 3.3	27,36			6082,40171		
Aasi_1396	68 h p.i. 1.1	27,75	28,07	0,281	4759,16210	3923,32102	725,02032
Aasi_1396	68 h p.i. 1.2	28,22			3546,40943		
Aasi_1396	68 h p.i. 1.3	28,25			3464,39153		
Aasi_1396	68 h p.i. 2.1	27,99	28,83	0,732	4098,59810	2602,93533	1298,23755
Aasi_1396	68 h p.i. 2.2	29,17			1942,65622		
Aasi_1396	68 h p.i. 2.3	29,32			1767,55166		
Aasi_1396	68 h p.i. 3.1	28,75	28,97	0,255	2540,54835	2221,62845	346,65113
Aasi_1396	68 h p.i. 3.2	28,92			2271,65499		
Aasi_1396	68 h p.i. 3.3	29,25			1852,68203		
Aasi_1396	140 h p.i. 1.1	25,42	25,51	0,080	20614,13471	19521,55519	986,21485
Aasi_1396	140 h p.i. 1.2	25,53			19253,33416		
Aasi_1396	140 h p.i. 1.3	25,57			18697,19670		
Aasi_1396	140 h p.i. 2.1	24,36	24,50	0,125	40134,97072	36773,63276	2947,97949
Aasi_1396	140 h p.i. 2.2	24,55			35558,40807		
Aasi_1396	140 h p.i. 2.3	24,59			34627,51948		
Aasi_1396	140 h p.i. 3.1	25,06	25,25	0,252	25780,50188	23027,07916	3500,36946
Aasi_1396	140 h p.i. 3.2	25,16			24212,90245		
Aasi_1396	140 h p.i. 3.3	25,54			19087,83314		
Aasi_1396	extracellular 1.1	23,14	23,26	0,183	86382,30239	80600,02458	8955,66972
Aasi_1396	extracellular 1.2	23,16			85133,70673		
Aasi_1396	extracellular 1.3	23,47			70284,06463		
Aasi_1396	extracellular 2.1	22,02	22,26	0,248	174946,44708	151475,91603	23419,18819
Aasi_1396	extracellular 2.2	22,25			151372,89037		
Aasi_1396	extracellular 2.3	22,51			128108,41062		
Aasi_1396	extracellular 3.1	21,91	22,23	0,310	186937,05795	155161,94983	30320,56523
Aasi_1396	extracellular 3.2	22,24			152005,99244		
Aasi_1396	extracellular 3.3	22,53			126542,79909		

Table S9. *Amoebophilus* antifeeding prophage (AFP) like proteins encoded in *Bacteroidetes* genomes. I...amino acid identity, E...e-value, n.d....not detected

	Aasi_1072	Aasi_1073	Aasi_1074	Aasi_1075	Aasi_1076	Aasi_1077	Aasi_1078	Aasi_1079	Aasi_1080	Aasi_1081	Aasi_1082	Aasi_1083	Aasi_1806	Aasi_0556	Aasi_0557
<i>Cardinium hertigii</i> cHepr1	CAHE_0456, I=37%, E=9e-30	CAHE_0457, I=35%, E=5e-41	CAHE_0458, I=76%, E=4e-228	CAHE_0459, I=29%, E=3e-20	CAHE_0460, I=30%, E=1e-10	CAHE_0461, I=41%, E=1e-26	CAHE_0462, I=48%, E=0.011	CAHE_0463, I=33%, E=2e-36	CAHE_0763, I=45%, E=3e-153	CAHE_0762, I=64%, E=9e-39	CAHE_0761, I=48%, E=1e-32	CAHE_0760, I=44%, E=5e-202	CAHE_0118, I=24%, E=7e-35 E=3e-67/8e-35	CAHE_0036, I=64%, E=3e-96	CAHE_0037, I=42%, E=1e-283
<i>Algoriphagus</i> sp. PRI	ALPRI_12765, I=32%, E=2e-10	*ALPRI_1276 0, I=14%, E=6e-29	ALPRI_12755, I=53%, E=1e-58	n.d.	n.d.	ALPRI_12745, I=25%, E=3e-09; ALPRI_12750, I=24%, E=0.11	n.d.	ALPRI_12735, I=30%, E=3e-18	ALPRI_12730, I=35%, E=3e-115	ALPRI_12725, I=5%, E=7e-23	ALPRI_12720, I=39%, E=2e-26	ALPRI_12710, I=27%, E=0.25	ALPRI_12700, I=21%, E=0.25	n.d.	ALPRI_12715, I=29%, E=2e-94
<i>Bacteroides cellulositricus</i> DSM 14838	BACCELL_03 032, I=23%, E=0.47	*BACCELL_0 3033, I=11%, E=7e-34	BACCELL_03 035, I=40%, E=44%, E=5e-132; BACCELL_03 034, I=50%, E=1e-59	n.d.	n.d.	BACCELL_03 036, I=22%, E=0.023	n.d.	039, I=33%, E=1e-10	BACCELL_03 040, I=31%, E=5e-94	BACCELL_03 041, I=57%, E=2e-24	BACCELL_03 042, I=43%, E=1e-22	BACCELL_03 046, I=22%, E=3e-18	BACCELL_03 941, I=22%, E=0.44	BACCELL_03 045, I=26%, E=2e-10	BACCELL_03 044, I=26%, E=1e-58
<i>Chitinophaga pinensis</i> DSM 2588	Cpin_3920, I=28%, E=3e-16; Cpin_5259, I=26%, E=4e-14	*Cpin_3921, I=20%, E=6e-46; *Cpin_5258, I=15%, E=4e-14	Cpin_3922, I=44%, E=5e-132; Cpin_5257, I=37%, E=1e-88;	n.d.	n.d.	Cpin_3924, I=32%, E=5e-09; Cpin_5253, I=26%, E=8e-08	n.d.	Cpin_3926, I=30%, E=8e-16; Cpin_5251, I=30%, E=4e-15	Cpin_3927, I=35%, E=1e-93; Cpin_5250, I=36%, E=1e-119	Cpin_3928, I=58%, E=5e-26; Cpin_5249, I=57%, E=1e-23	Cpin_3929, I=37%, E=1e-23; Cpin_5248, I=36%, E=1e-18	Cpin_3932, I=27%, E=2e-58; Cpin_5245, I=30%, E=6e-55	Cpin_3933, I=23%, E=2e-07; Cpin_5242, I=19%, E=2e-05	Cpin_3931, I=39%, E=2e-51; Cpin_5246, I=54%, E=8e-05	Cpin_3930, I=25%, E=2e-99; Cpin_5247, I=32%, E=7e-59
<i>Chlorobaculum parvum</i> NCIB 8327	Cpar_0903, I=26%, E=5e-08	*Cpar_0902, I=7%, E=0.007	Cpar_0901, I=54%, E=8e-67	n.d.	n.d.	Cpar_0899, I=22%, E=6e-04	n.d.	Cpar_0897, I=32%, E=9e-14	Cpar_0896, I=36%, E=1e-111	Cpar_0895, I=58%, E=2e-09	Cpar_0894, I=42%, E=2e-28	Cpar_0892, I=25%, E=1e-32	Cpar_0889, I=23%, E=5e-09	n.d.	Cpar_0893, I=29%, E=4e-96
<i>Cyclobacterium marinum</i> DSM 745	Cycma_4771, I=30%, E=8e-09	*Cycma_4770, I=12%, E=2e-22	Cycma_4769, I=49%, E=4e-58	n.d.	n.d.	Cycma_4767, I=22%, E=6e-06	n.d.	Cycma_4765, I=31%, E=6e-18	Cycma_4764, I=31%, E=2e-97	Cycma_4763, I=57%, E=4e-24	Cycma_4762, I=43%, E=3e-29	Cycma_4760, I=28%, E=6e-77	Cycma_4758, I=20%, E=0.34	n.d.	Cycma_4761, I=32%, E=1e-61
<i>Dyadobacter fermentans</i> DSM 18053	Dfer_0676, I=31%, E=1e-10	Dfer_0675, I=1 0%, E=0.005	Dfer_0674, I=52%, E=1e-57	n.d.	n.d.	Dfer_0672, I=27%, E=3e-12; Dfer_0673, I=24%, E=0.039	n.d.	Dfer_0670, I=32%, E=6e-21	Dfer_0669, I=35%, E=2e-110	n.d.	Dfer_0668, I=44%, E=7e-29	Dfer_0666, I=26%, E=4e-47	Dfer_0664, I=27%, E=3e-10	n.d.	Dfer_0667, I=27%, E=1e-115
<i>Flavobacterium johnsoniae</i> UW101	Fjoh_3148, I=30%, E=1e-10	*Fjoh_3149, I=15%, E=9e-39	Fjoh_3150, I=55%, E=8e-61	n.d.	n.d.	Fjoh_3152, I=27%, E=2e-10; Fjoh_3151, I=25%, E=0.18	n.d.	Fjoh_3154, I=29%, E=1e-15	Fjoh_3155, I=35%, E=2e-120	Fjoh_3156, I=55%, E=1e-22	Fjoh_3157, I=43%, E=1e-29	Fjoh_3159, I=28%, E=9e-91	Fjoh_3161, I=56%, E=0.005	n.d.	Fjoh_3158, I=30%, E=7e-99
<i>Haliscamenobacter hydrossus</i> DSM 1100	Hahly_0581, I=26%, E=1e-10	*Hahly_0580, I=12%, E=2e-11	Hahly_0579, I=50%, E=7e-58	n.d.	n.d.	Hahly_0577, I=26%, E=2e-10 KAOTI_07403, I=22%, E=0.066;	n.d.	Hahly_0575, I=28%, E=7e-12	Hahly_0574, I=36%, E=2e-113	Hahly_0573, I=50%, E=6e-11	Hahly_0572, I=39%, E=6e-25	Hahly_0570, I=28%, E=2e-35	Hahly_0568, I=22%, E=3e-05	n.d.	Hahly_0571, I=30%, E=5e-97
<i>Kordia algicida</i> OT-1	KAOTI_07418, I=23%, E=3.4; KAOTI_18597, I=26, E=5e-10	*KAOTI_1859 2, I=18%, E=3e-49	KAOTI_07408, I=53%, E=2e-61; KAOTI_18587, I=39%, E=6e-109	n.d.	n.d.	KAOTI_18572, I=31%, E=4.9	KAOTI_07388, I=30%, E=1e-13; KAOTI_18567, I=31%, E=4e-21	KAOTI_07383, I=34%, E=1e-105; KAOTI_18562, I=33%, E=2e-99	KAOTI_07378, I=56%, E=2e-25; KAOTI_18557, I=63%, E=8e-31	KAOTI_07373, I=38%, E=8e-25; KAOTI_18552, I=31%, E=2e-15	KAOTI_07363, I=27%, E=4e-70; KAOTI_18542, I=56%, E=2e-40	KAOTI_07353, I=26%, E=1e-05; KAOTI_18537, I=23%, E=0.003	n.d.	KAOTI_07368, I=27%, E=4e-87; KAOTI_18547, I=29%, E=5e-27	
<i>Leewenhoekiella blandensis</i> MED217	MED217_0865 0, I=30%, E=6e-12	*MED217_086 45, I=16%, E=7e-32	MED217_0864 0, I=55%, E=5e-64	n.d.	n.d.	MED217_0863 0, I=31%, E=3e-13; *MED217_086 35, I=20%, E=2e-36	n.d.	MED217_0862 0, I=29%, E=1e-17	MED217_0861 5, I=34%, E=6e-102	MED217_0861 0, I=59%, E=7e-29	MED217_0860 5, I=39%, E=2e-24	MED217_0859 5, I=28%, E=2e-87	*MED217_085 85, I=17%, E=9e-28	n.d.	MED217_0860 0, I=27%, E=7e-97

<i>Microscilla marina</i> ATCC 23134	M23134_01833 I=53%, E=2e-08; M23134_06080 I=22%, E=8e-05	*M23134_01834 I=16%, E=1e-26 *M23134_06081 I=15%, E=2e-38	M23134_01835 I=54%, E=2e-66; M23134_06082 I=40%, E=1e-129	n.d.	n.d.	M23134_01840 I=31%, E=3e-04; M23134_06076 I=53%, E=7e-16	M23134_01841 I=30%, E=5e-62; M23134_06086 I=24%, E=7e-19	n.d.	Niako_0470, I=55%, E=1e-24	Niako_0471, I=43%, E=1e-28	Oweho_1318, I=39%, E=1e-20	BDI_2452, I=34%, E=3e-17	HMPREF0619_03424, I=34%, E=6e-19	SGRA_1309, I=42%, E=3e-27	Solca_2151, I=41%, E=3e-27	M23134_01846 I=30%, E=1e-08; M23134_01847 I=28%, E=8e-06; M23134_06089 I=28%, E=8e-136	M23134_01845 I=28%, E=2e-82; M23134_06089 I=28%, E=8e-136	Niako_0472, I=51%, E=4e-119	n.d.	Oweho_1316, I=38%, E=1e-48	Oweho_1317, I=25%, E=1e-75	BDI_2454, I=26%, E=4e-56	HMPREF0619_03425, I=25%, E=6e-61	SGRA_1308, I=31%, E=6e-148	Solca_2150, I=28%, E=9e-98	
<i>Nastella koreensis</i> GR20-10	Niako_0462, I=29%, E=3e-13	*Niako_0463, I=20%, E=6e-17	Niako_0464, I=53%, E=5e-62	n.d.	n.d.	Niako_0468, I=53%, E=3e-22	Niako_0469, I=57%, E=9e-126	n.d.	Niako_0470, I=55%, E=1e-24	Niako_0471, I=43%, E=1e-28	Oweho_1320, I=53%, E=1e-21	BDI_2450, I=32%, E=2e-19	HMPREF0619_03422, I=32%, E=1e-19	SGRA_1316, I=21%, E=8e-34	Solca_2156, I=24%, E=5e-07	Niako_0465, I=26%, E=1e-04; Niako_0466, I=26%, E=2e-54	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
<i>Owenweeksia hongkongensis</i> DSM 17368	Oweho_1326, I=26%, E=1e-12	*Oweho_1325, I=14%, E=1e-31	Oweho_1324, I=40%, E=3e-125	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	Oweho_1322, I=23%, E=5e-11; Oweho_1323, I=25%, E=0.004	Oweho_1319, I=33%, E=5e-103	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
<i>Parabacteroides distasonis</i> ATCC 8503	BDI_2445, I=26%, E=0.016	*BDI_2446, I=14%, E=8e-28	BDI_2447, I=53%, E=3e-43	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	BDI_2448, I=23%, E=2e-38; BDI_2449, I=26%, E=7e-07	BDI_2451, I=31%, E=2e-92	BDI_3451, I=53%, E=4e-22	BDI_2452, I=34%, E=3e-17	HMPREF0619_00657, I=53%, E=4e-22	HMPREF0619_03423, I=32%, E=2e-78	SGRA_1312, I=31%, E=4e-93	Solca_2153, I=30%, E=3e-76	n.d.		
<i>Parabacteroides</i> sp. D13	HMPREF0619_03416, I=26%, E=0.025	*HMPREF0619_03417, I=14%, E=3e-29	HMPREF0619_03418, I=54%, E=8e-44	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	HMPREF0619_03419, I=23%, E=2e-38; HMPREF0619_03420, I=26%, E=7e-07	HMPREF0619_03423, I=32%, E=2e-78	HMPREF0619_00657, I=53%, E=4e-22	HMPREF0619_03424, I=34%, E=6e-19	HMPREF0619_03427, I=22%, E=2e-10	HMPREF0619_03431, I=19%, E=9.0	SGRA_1317, I=41%, E=5e-117	Solca_2158, I=38%, E=8e-95	n.d.	n.d.	
<i>Saprospira grandis</i> str. Levitt	SGRA_1319, I=31%, E=9e-23	*SGRA_1318, I=11%, E=9e-33	SGRA_1317, I=41%, E=5e-117	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	SGRA_1311, I=49%, E=3e-15	SGRA_1312, I=31%, E=4e-93	SGRA_1311, I=49%, E=3e-15	SGRA_1309, I=42%, E=3e-27	SGRA_1346, I=31%, E=7e-65	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
<i>Solitalea canadensis</i> DSM 3403	Solca_2160, I=34%, E=1e-13	n.d.	Solca_2158, I=38%, E=8e-95	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	Solca_2152, I=55%, E=2e-22	Solca_2153, I=30%, E=3e-76	Solca_2152, I=55%, E=2e-22	Solca_2151, I=41%, E=3e-27	Solca_2149, I=28%, E=2e-69	Solca_2147, I=2.6%, E=2e-12	n.d.	n.d.	n.d.	n.d.	n.d.

*PSI blast (6 iterations)

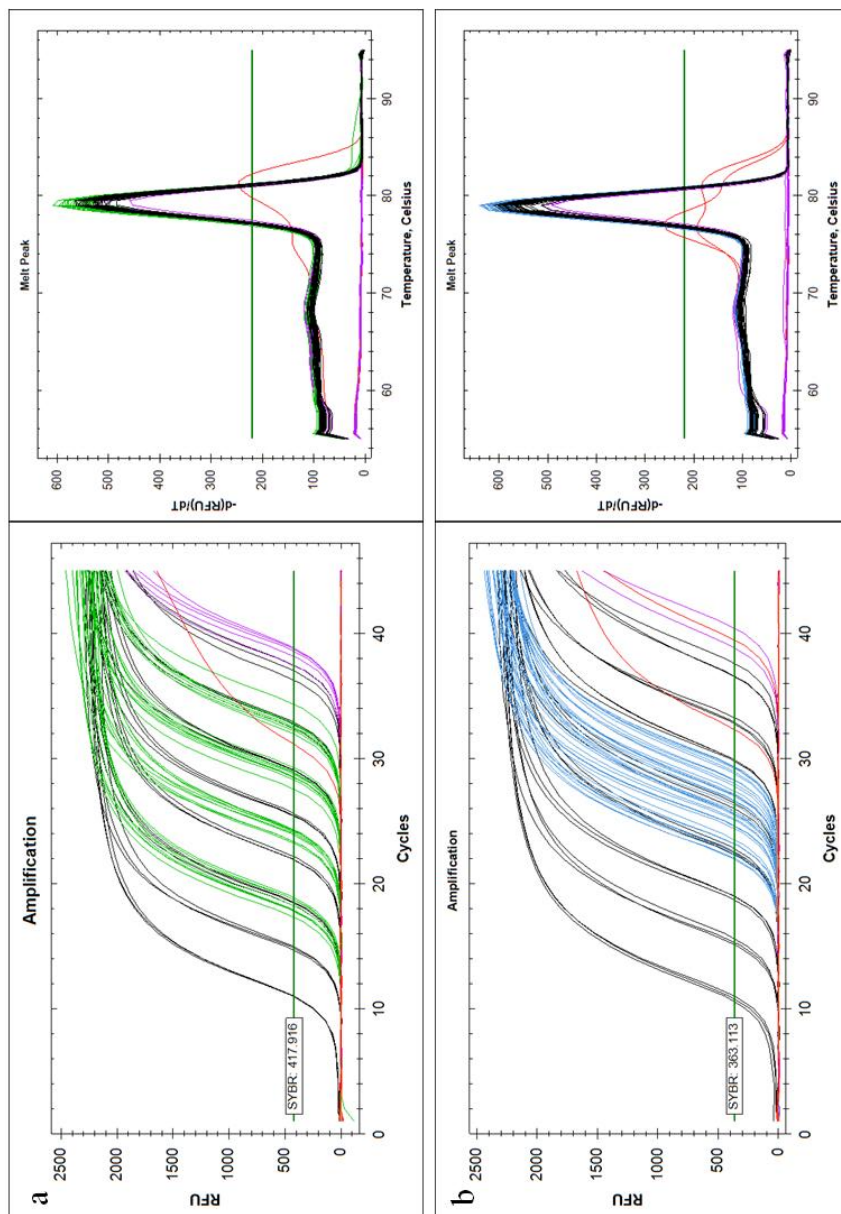


Figure S1. **Complete RT-qPCR assay.** (a) Amplification with primers Aasi_1074 qPCR F1/R1, biological replicate 2 (b) Amplification with Aasi_1396 qPCR F2/R2. Standards in black, cDNA samples in green and blue, respectively, NRT controls in purple, NTC in orange, NTC in red.

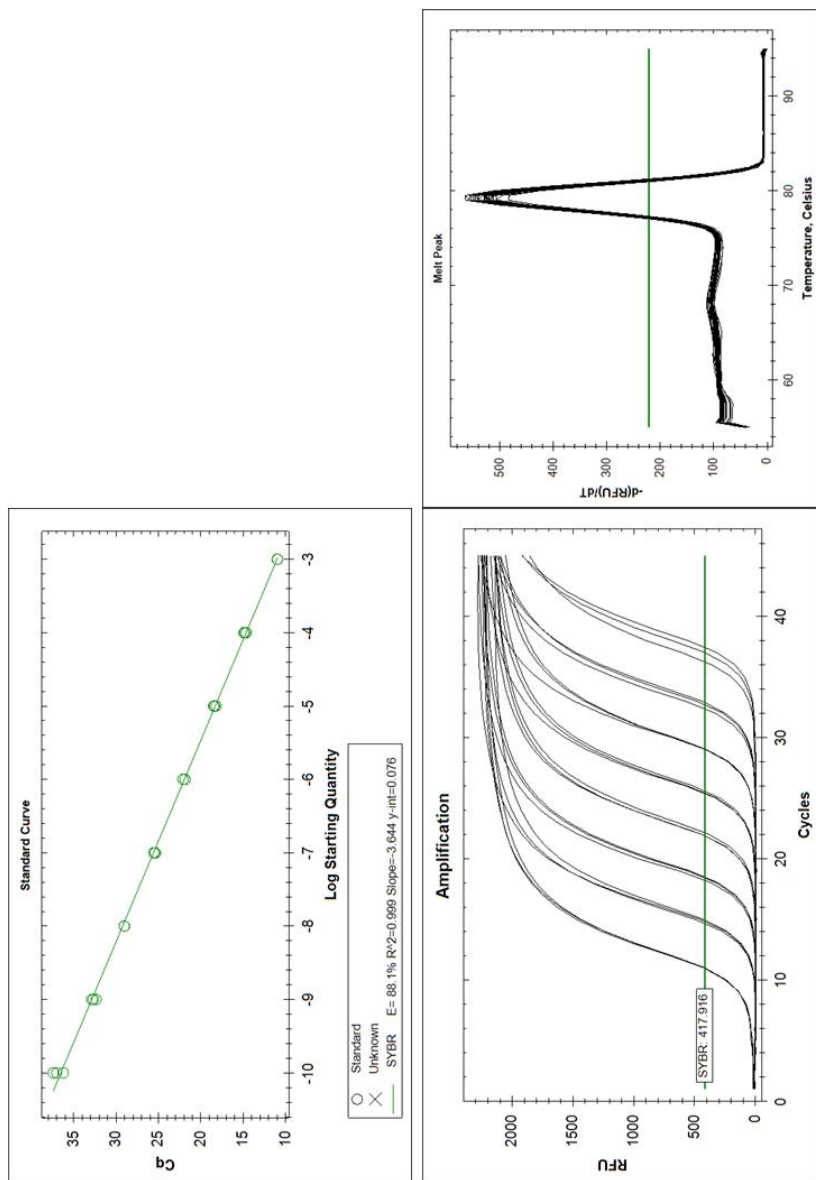


Figure S2. **Standard curve for phage tail sheath gene Aasi_1074.** A TOPO[®] XL plasmid (Life Technologies) containing a 1507 bp insert with the phage tail sheath gene was used to generate the standard curve. qPCR reactions were performed in triplicates. Dynamic range: $1.2 \times 10^8 - 1.2 \times 10^1$ copies. Mean Cq at limit of detection: 36.88 ± 0.604 .

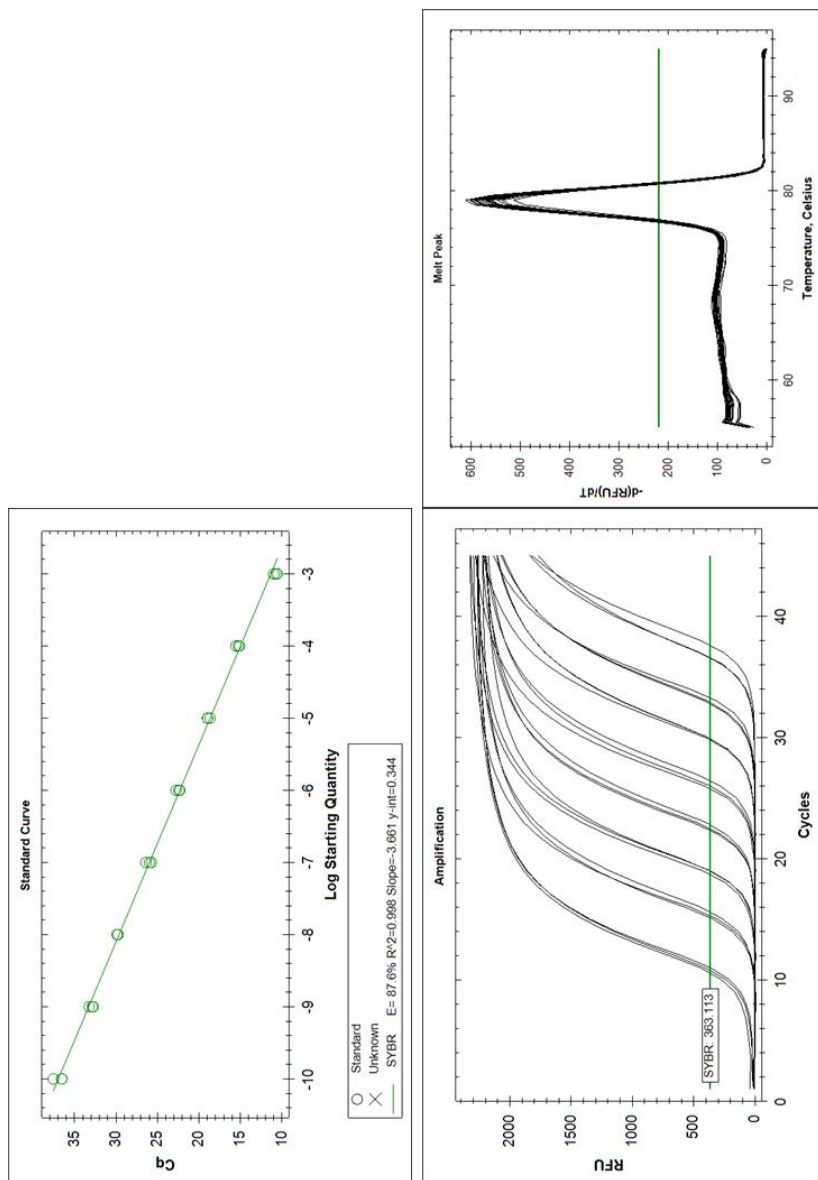


Figure S3. Standard curve for β subunit of RNA polymerase gene Aasi_1396. A TOPO[®] XL plasmid (Life Technologies) containing an 851 bp fragment of the reference gene was used to generate the standard curve. qPCR reactions were performed in triplicates. Dynamic range: $1.5 \times 10^8 - 1.5 \times 10^1$ copies. Mean Cq at limit of detection: 36.87 ± 0.598 .

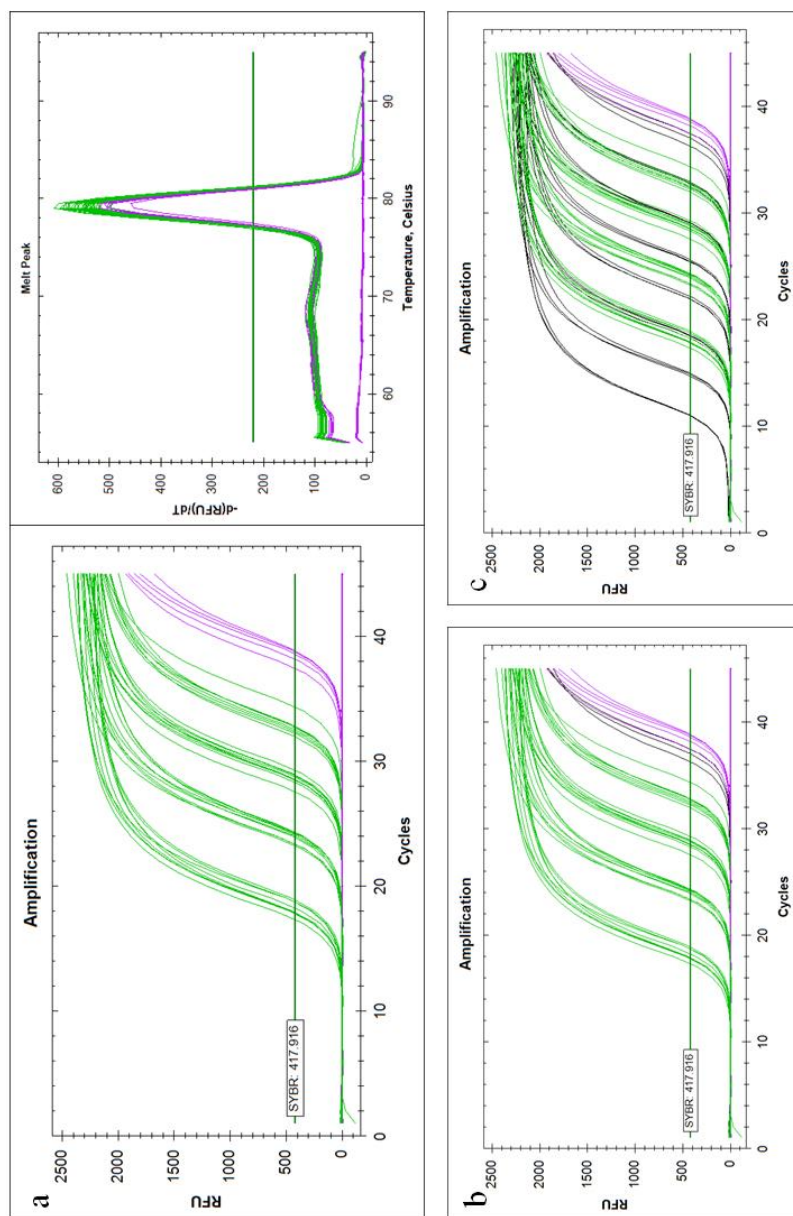


Figure S4. cDNA samples and no-reverse transcriptase controls for phage tail sheath gene *Aasi_1074*. (a) Without standards, (b) with last detectable standard (12 copies per reaction), (c) with all standards. Most NRT controls contain no detectable amounts of DNA. If amplification occurs, it is within the range of the last detectable standard or even beyond limit of detection.

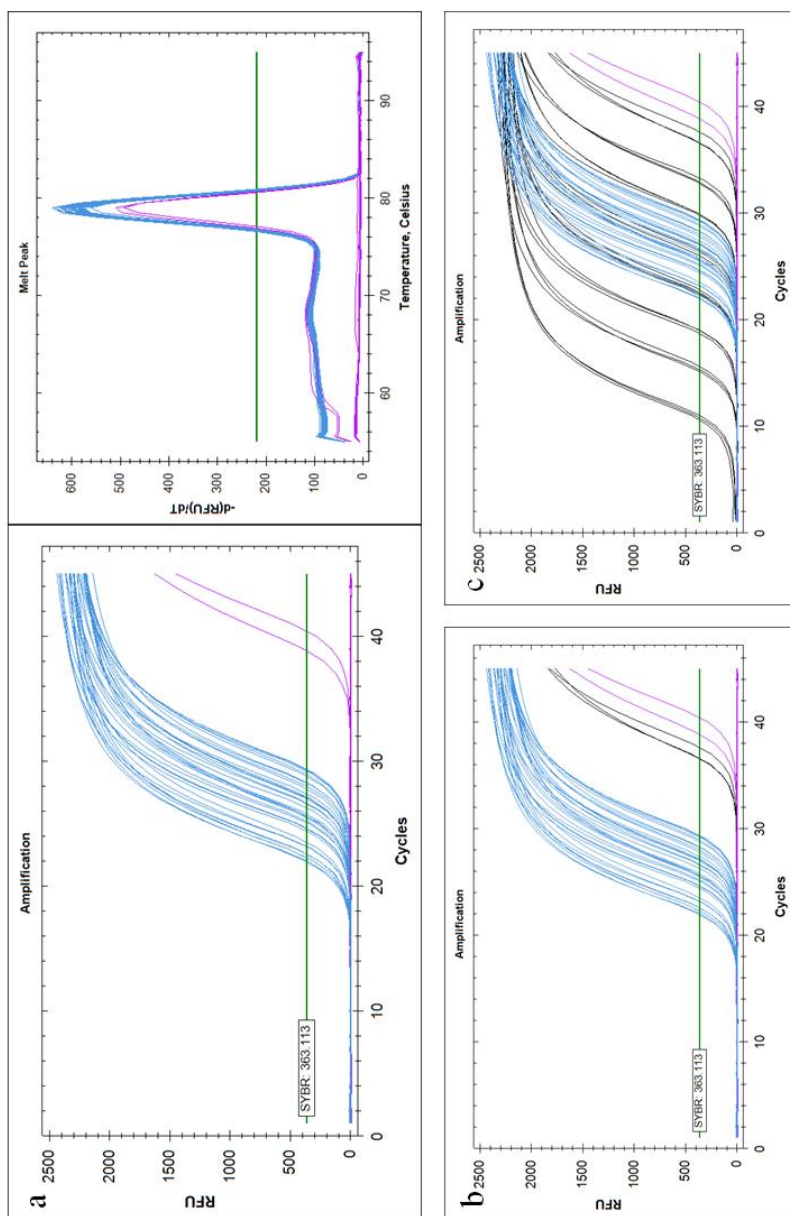


Figure S5. cDNA samples and no-reverse transcriptase controls for β subunit of RNA polymerase gene Aasi_1396. (a) Without standards, (b) with last detectable standard (15 copies per reaction), (c) with all standards. Most NRT controls contain no detectable amounts of DNA. If amplification occurs, it is within the range of the last detectable standard or even beyond limit of detection.

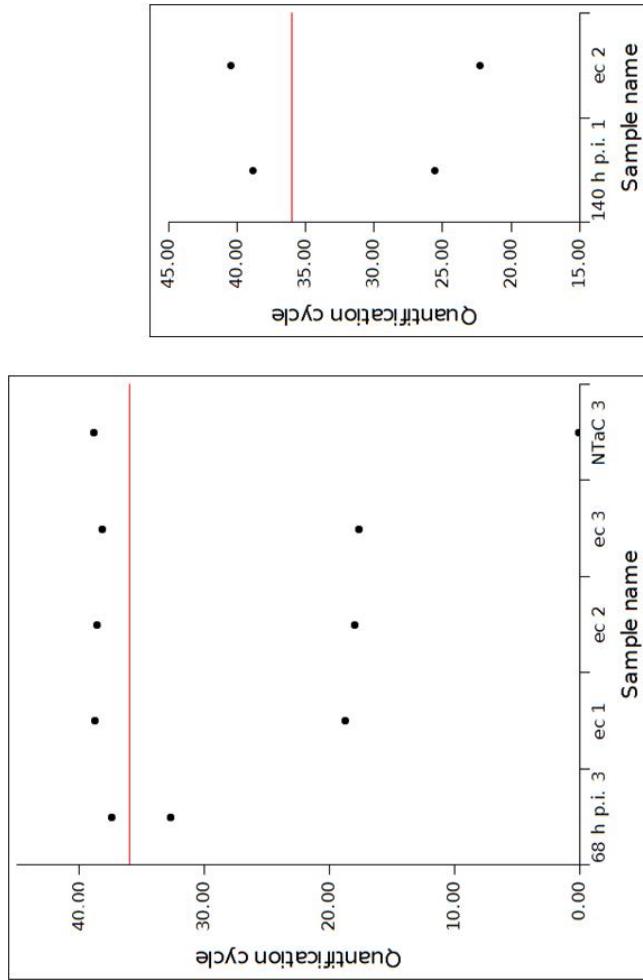


Figure S6. **Cq values obtained for cDNA samples and NRT controls.** Only samples where any amount of target is amplified in NRT controls are displayed. (a) Amplification with Aasi_1074 qPCR F1/R1, (b) with Aasi_1396 qPCR F2/R2. Red horizontal line shows limit of detection. In all NRT controls, Cq values are beyond limit of detection.

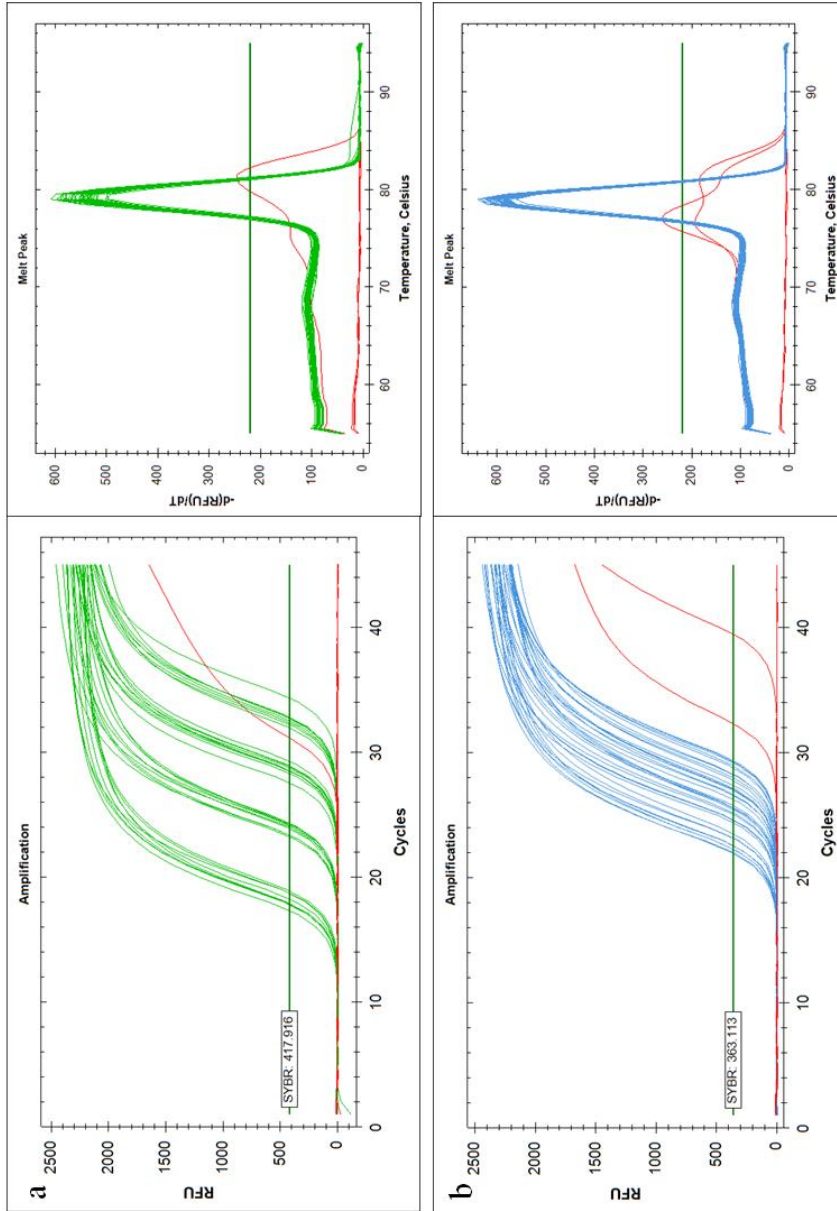


Figure S7. **C_q values obtained for cDNA samples and no template controls.** cDNA samples (in green and blue, respectively) and no-template controls (in red). (a) Amplification with primers Aasi_1074 qPCR F1/R1 (b) Amplification with Aasi_1396 qPCR F2/R2. Some no-template controls result in amplification curves above LOD, but Melt Peak indicates that amplified product is unspecific.

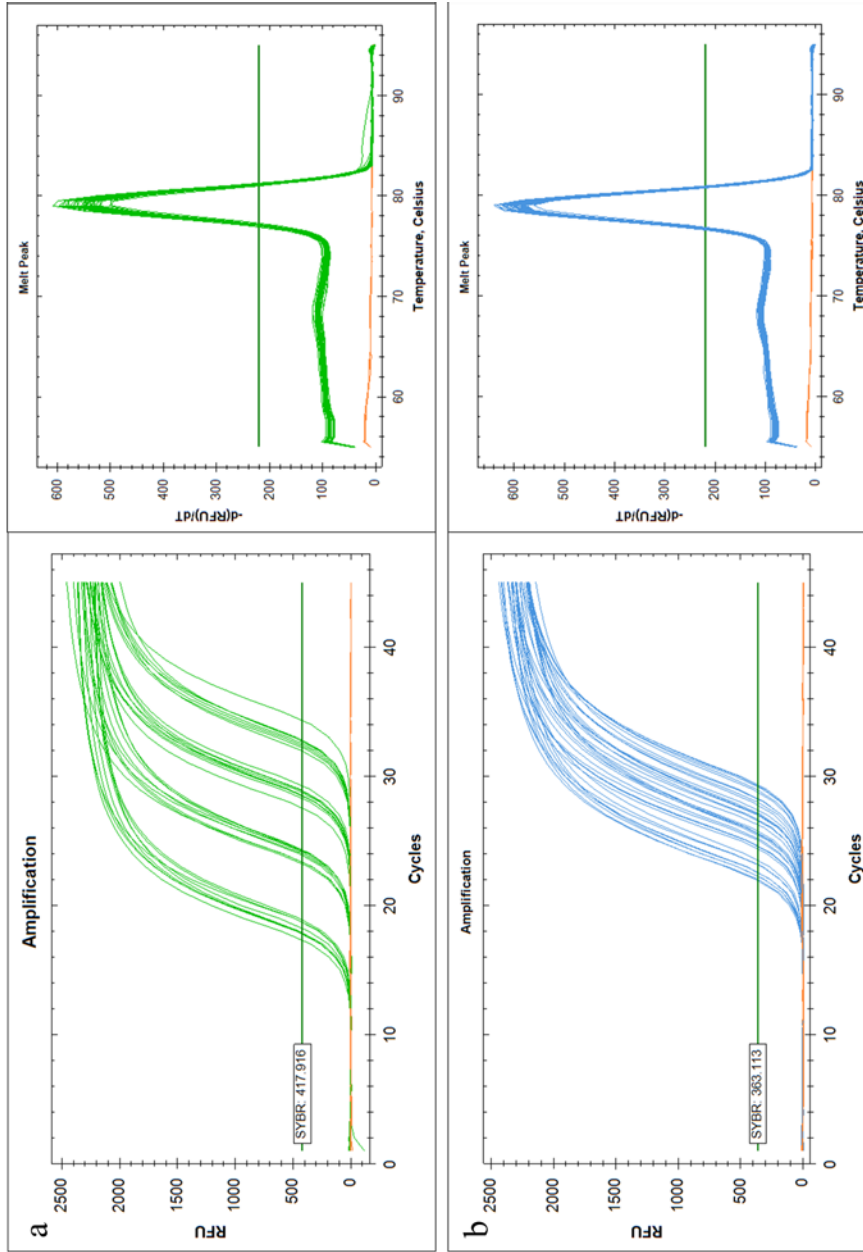


Figure S8. **Specificity of RT q-PCR assay.** cDNA samples (in green and blue, respectively) and non-target controls (in orange). (a) Amplification with primers Aasi_1074-qPCR F1/R1, (b) Amplification with Aasi_1396 qPCR F2/R2. Non-target controls are not amplified, indicating that the assay is specific.

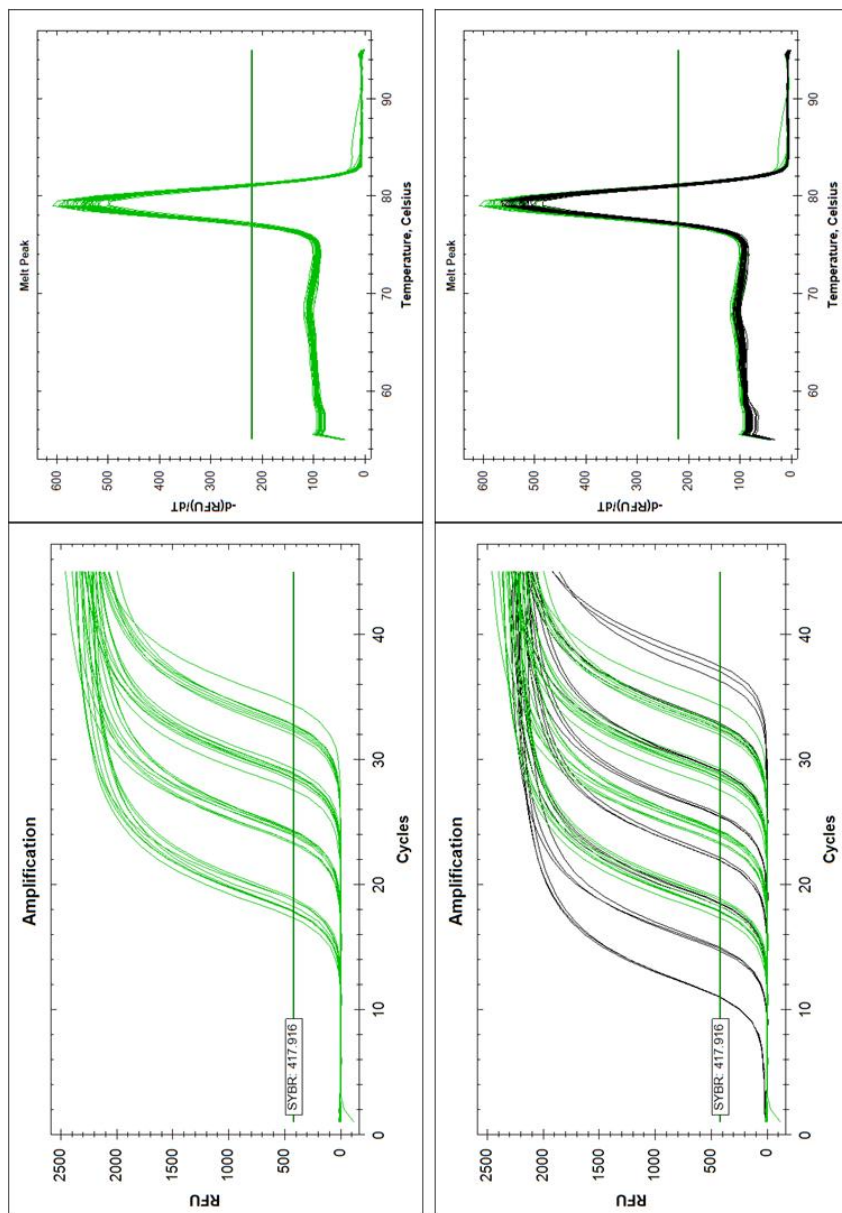


Figure S9. **Amplification and melt peak chart for phage tail sheath gene Aasi_1074.** (a) Without standards, (b) with standards. cDNA samples in green, standards in black. All cDNA samples are within the dynamic range.

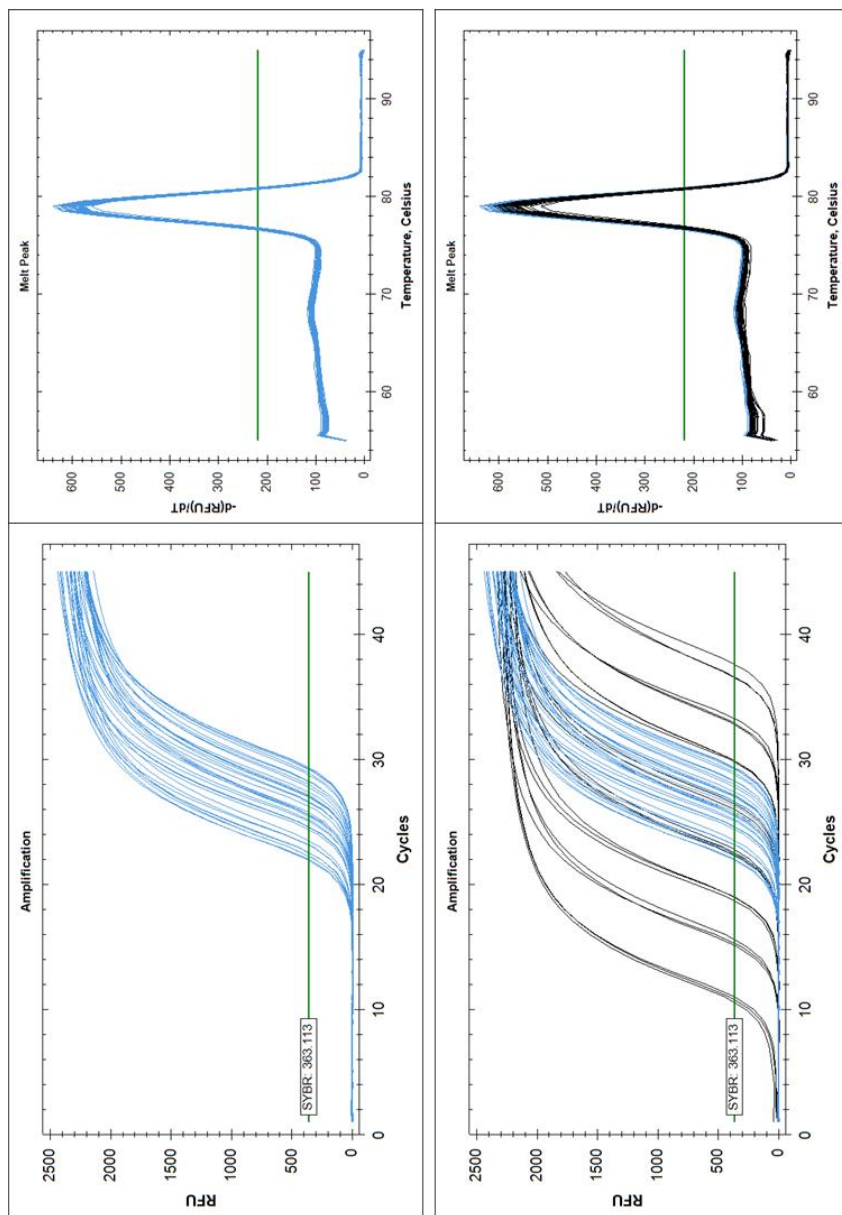


Figure S10. Amplification and melt peak chart for β subunit of RNA polymerase gene Aasi_1396. (a) Without standards, (b) with standards. cDNA samples in blue, standards in black. All cDNA samples are within the dynamic range.

Chapter VI

The endosymbiont *Amoebophilus asiaticus* encodes an S-adenosylmethionine carrier that compensates for its missing methylation cycle

published in J Bacteriol. 2013 Jul;195(14):3183-92. PMID: 23667233

The Endosymbiont *Amoebophilus asiaticus* Encodes an S-Adenosylmethionine Carrier That Compensates for Its Missing Methylation Cycle

Ilka Haferkamp, Thomas Penz, Melanie Geier, Michelle Ast, Tanja Mushak, Matthias Horn and Stephan Schmitz-Esser
J. Bacteriol. 2013, 195(14):3183. DOI: 10.1128/JB.00195-13.
Published Ahead of Print 10 May 2013.

Updated information and services can be found at:
<http://jb.asm.org/content/195/14/3183>

SUPPLEMENTAL MATERIAL	<i>These include:</i> Supplemental material
REFERENCES	This article cites 59 articles, 29 of which can be accessed free at: http://jb.asm.org/content/195/14/3183#ref-list-1
CONTENT ALERTS	Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), more»

Information about commercial reprint orders: <http://journals.asm.org/site/misc/reprints.xhtml>
To subscribe to to another ASM Journal go to: <http://journals.asm.org/site/subscriptions/>

Journals.ASM.org

The Endosymbiont *Amoebophilus asiaticus* Encodes an S-Adenosylmethionine Carrier That Compensates for Its Missing Methylation Cycle

Ilka Haferkamp,^a Thomas Penz,^b Melanie Geier,^a Michelle Ast,^a Tanja Mushak,^a Matthias Horn,^b Stephan Schmitz-Esser^c

Zelluläre Physiologie/Membrantransport, Technische Universität Kaiserslautern, Kaiserslautern, Germany^a; Department für Mikrobielle Ökologie, Universität Wien, Vienna, Austria^b; Institut für Milchhygiene, Veterinärmedizinische Universität Wien, Vienna, Austria^c

All organisms require S-adenosylmethionine (SAM) as a methyl group donor and cofactor for various biologically important processes. However, certain obligate intracellular parasitic bacteria and also the amoeba symbiont *Amoebophilus asiaticus* have lost the capacity to synthesize this cofactor and hence rely on its uptake from host cells. Genome analyses revealed that *A. asiaticus* encodes a putative SAM transporter. The corresponding protein was functionally characterized in *Escherichia coli*: import studies demonstrated that it is specific for SAM and S-adenosylhomocysteine (SAH), the end product of methylation. SAM transport activity was shown to be highly dependent on the presence of a membrane potential, and by targeted analyses, we obtained direct evidence for a proton-driven SAM/SAH antiport mechanism. Sequence analyses suggest that SAM carriers from *Rickettsiales* might operate in a similar way, in contrast to chlamydial SAM transporters. SAM/SAH antiport is of high physiological importance, as it allows for compensation for the missing methylation cycle. The identification of a SAM transporter in *A. asiaticus* belonging to the *Bacteroidetes* phylum demonstrates that SAM transport is more widely spread than previously assumed and occurs in bacteria belonging to three different phyla (*Proteobacteria*, *Chlamydiae*, and *Bacteroidetes*).

Methylation occurs in all organismic groups, from bacteria to eukaryotes, and is involved in general processes, such as RNA metabolism and the regulation of gene expression and protein function, as well as in more specific mechanisms, like modification of neurotransmitters and detoxification of heavy metals, etc. (1–3). In various synthetic and regulatory methylation reactions, S-adenosylmethionine (SAM) acts as a methyl group donor, and specific methyltransferases mediate the transfer of the reactive methyl group to the respective acceptor molecule (4–6). Methyl group transfer from SAM results in the formation of S-adenosylhomocysteine (SAH). SAH is an efficient competitive inhibitor of methyltransferases, and accordingly, its removal by catabolizing enzymes (such as specific hydrolases or nucleosidases) is required to guarantee maintenance of methylation processes (7, 8). In addition to its role in methylation, SAM is also an important reagent for posttranscriptional modification of tRNAs and is used as a source of ribosyl groups in the biosynthesis of queuosine, a hypermodified tRNA nucleoside occurring in tRNAs coding for asparagine, aspartic acid, histidine, and tyrosine (3). Moreover, SAM acts as a precursor (amino carboxylpropyl group donor) in polyamine generation, in bacterial N-acetylhomoserine lactone synthesis, as well as in ethylene and nicotinamine production in plants and also plays an important role as a radical source in various biological transformations during, e.g., DNA precursor, vitamin, or cofactor synthesis (3).

Most organisms are able to generate SAM from ATP and methionine via the enzyme SAM synthetase (MetK [EC 2.5.1.6]) (9–13). The *Escherichia coli* genome encodes only one single SAM synthetase isoform (*metK* gene), and the incapability to obtain *metK* deletion mutants demonstrated that SAM formation and, consequently, methylation are essential for cellular viability and growth (14). Interestingly, several obligate intracellular bacteria belonging to the *Rickettsiales* and *Chlamydiales* apparently have lost the capacity to synthesize this important cofactor because they

lack a functional *metK* gene (15, 16). In *Rickettsia prowazekii* and in related strains that cause spotted fever, the *metK* gene is defective due to internal stop codons or frameshifts, and SAH recycling also seems to be absent (16, 17). *R. prowazekii* harbors a drug metabolite transporter superfamily protein involved in SAM provision (RP076) (16). A possible H⁺/SAM symport was suggested to allow net uptake of SAM and compensation for the missing synthetic activity. Competition studies performed with the rickettsial carrier revealed that an excess of SAH caused significantly reduced SAM uptake, and therefore, SAH was discussed as a potential additional substrate of this transport protein (16). However, whether this carrier catalyzes SAM transport in exchange with SAH was not investigated in corresponding transport studies. A possible SAM/SAH antiport would supply SAM to the bacterium and synchronously facilitate the export of the demethylated backbone. Among the *Chlamydiales*, solely *Parachlamydia acanthamoebae* and *Waddlia chondrophila* harbor enzymes for SAM generation and SAH degradation and thus exhibit a complete methylation cycle (15, 18). Remarkably, SAM-dependent methylation (of 16S rRNA or class I release factors) is performed in these bacteria; however, methylation of DNA most likely seems to be of no relevance in chlamydiae due to the absence of DNA methyl-

Received 15 February 2013 Accepted 4 May 2013

Published ahead of print 10 May 2013

Address correspondence to Stephan Schmitz-Esser, stephan.schmitz-esser@vetmeduni.ac.at.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/JB.00195-13>.

Copyright © 2013, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.00195-13

The authors have paid a fee to allow immediate free access to this article.

transferase coding sequences in the corresponding genomes (15, 19, 20).

Recently, a carrier mediating SAM uptake (CTL0843) was also identified in *Chlamydia trachomatis* (15). Biochemical data led the authors of that study to the assumption that it exhibits diverse properties: the carrier is capable of proton-driven SAM net uptake as well as substrate counterexchange. It is noteworthy that significant counterexchange occurred in the presence or absence of a proton gradient. Moreover, slight SAM efflux was observed when both proton gradient and counterexchange substrates were missing (15). Although accepting identical/similar substrates and although belonging to the same transporter superfamily, the chlamydial and the rickettsial SAM carriers exhibit only low sequence similarities (approximately 20% amino acid sequence identity).

A reduced genome size accompanied by the loss of biosynthetic pathways and the recruitment of carriers for compensation for missing or truncated metabolic pathways is a characteristic common feature of *Chlamydiales* and *Rickettsiales* (21, 22): specific nucleotide transporters were shown to allow energy parasitism and complementation of missing purine and pyrimidine nucleotide or cofactor biosynthesis pathways (23–29). Deciphering of more and more genomes demonstrated that many intracellular bacteria are biosynthetically highly impaired. Moreover, several obligate endosymbionts and intracellular pathogens also contain a limited repertoire of transporters (30, 31). Analyses of the genome of the obligate intracellular *Acanthamoeba* symbiont *Amoebophilus asiaticus* strain 5a2, a member of the phylum *Bacteroidetes*, revealed an extraordinarily high degree of reduction in its biosynthetic capacities (32). The genome has a size of 1.89 Mbp, encodes 1,557 proteins, and is thus only moderately reduced in size compared to the sizes of many other obligate intracellular bacteria (33, 34). However, the biosynthetic capabilities of *A. asiaticus* are extremely limited; its genome does not encode pathways for *de novo* biosynthesis of cofactors, nucleotides, and almost all amino acids (32). Interestingly, its genome harbors one gene (Aasi_1859) with significant similarities to the rickettsial SAM carrier (45% amino acid similarity) (32). Characterization of the Aasi_1859 gene product in the heterologous host *E. coli* revealed that apart from rickettsial and chlamydial species, *A. asiaticus* also possesses a SAM transport protein. Its catalytic activity allows import of SAM from the host cell by the simultaneous removal of the end product of methyltransferase reactions.

MATERIALS AND METHODS

Sequence and phylogenetic analyses. The genome sequence of *A. asiaticus* 5a2 has recently been determined and analyzed (32) and is available at GenBank under accession no. CP001102. SAM transporter amino acid sequences were retrieved using BLASTP against GenBank by using Aasi_1859 as a query, and only those sequences having more than 30% amino acid identity to Aasi_1859 were used for phylogenetic analyses. Amino acid sequences (101 in total) were aligned with MAFFT (35), and phylogenetic trees were reconstructed with MEGA (36) by using the neighbor-joining method and the Poisson correction, the parsimony bootstrap method, and the maximum likelihood method (using the Jones-Taylor-Thornton [JTT] amino acid substitution model); all trees were calculated with 1,000× bootstrapping. All positions containing gaps and missing data were eliminated from the data sets.

Transcriptional analysis. *Acanthamoeba* sp. strain 5a2 (ATCC PRA-228) amoebae harboring *A. asiaticus* 5a2 cells were harvested by centrifugation (7,000 × g for 3 min at 27°C). The resulting cell pellet was resuspended in 750 μl TRIzol (Invitrogen Life Technologies), transferred into

a Lysing Matrix A tube (MP Biomedicals), and homogenized by using a BIO101/Savant FastPrep FP120 instrument (speed, 4.5 m/s; 30 s). RNA was extracted by phase separation, precipitation, washing, and redissolving according to the recommendations of the manufacturer (TRIzol; Invitrogen Life Technologies). The remaining DNA was removed by using the Turbo DNA-free kit (Ambion). After DNase treatment, RNA was resuspended in double-distilled water (ddH₂O) with diethyl pyrocarbonate (DEPC) and stored at –80°C until use. The absence of DNA contamination in the DNase-treated RNA was verified by performing a control PCR with 42 cycles by using primers targeting a 361-bp fragment of the Aasi_1859 gene (forward primer 5′-ATG GAG CCA GGG GAT TAA AG-3′ and reverse primer 5′-GTT GGT GGG AGT ACG CCA TA-3′) and an annealing temperature of 66.4°C. DNA-free total RNA (containing host and symbiont RNA) was used to synthesize cDNA by using the RevertAid first-strand cDNA synthesis kit (Fermentas) according to the recommendations of the manufacturer. cDNA was subsequently used as the template in standard PCRs (35 cycles and an annealing temperature of 66.4°C). Negative controls (no cDNA added) and positive controls (genomic DNA) were included in all PCRs. All experiments were performed in biologically independent triplicates.

Cloning of Aasi_1859 and heterologous protein synthesis in *E. coli*. *Acanthamoeba* sp. 5a2 (ATCC PRA-228) amoebae harboring *A. asiaticus* cells were used for DNA isolation with the DNeasy Blood and Tissue kit (Qiagen) according to the manufacturer's recommendations. The Aasi_1859 gene, coding for the putative SAM transport protein, was amplified by using High Fidelity PCR enzyme mix (MBI-Fermentas) according to the instructions of the manufacturer. A forward primer (5′-CCT GCG CAT ATG TTG AAA TAT TTT AAA GCA-3′), introducing an NdeI restriction site before the start codon, and a reverse primer (5′-CCT CGC CTC GAG TCA AGC TTT AGG TTG ATT-3′), containing an XhoI restriction site after the stop codon, were used. PCR conditions were as follows: a denaturation step at 94°C for 3 min, followed by 35 cycles of (i) denaturation at 94°C for 30 s, (ii) annealing at 56°C for 40 s, and (iii) elongation at 68°C for 90 s and a final elongation step at 68°C for 10 min. The resulting amplification products were gel purified and cloned into the cloning vector pCR-XL-TOPO by using the TOPO XL cloning kit (Invitrogen Life Technologies). The resulting plasmid was digested with the restriction endonucleases NdeI and XhoI, gel purified, and inserted in frame into the isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible expression vector pET16b containing a promoter site for the T7 RNA polymerase (Novagen). The newly constructed expression plasmid was transformed into and maintained in *E. coli* XL1-Blue cells (Stratagene). Integrity of the cloned gene was confirmed by sequencing on an ABI 3130 XL genetic analyzer using BigDye Terminator kit v3.1 (ABI). After the correctness of the insertion was proven, the construct was used for transformation of BLR(DE3) expression cells (Merck Biosciences). *E. coli* cells were cultured in standard yeast extract-tryptone (YT) medium at 37°C with vigorous shaking. Heterologous protein synthesis was induced by addition of 1 mM IPTG during exponential cell growth (at an optical density at 600 nm [OD₆₀₀] of 0.5). One hour after induction, cells were concentrated to an OD₆₀₀ of 5.0 by centrifugation (3,000 × g for 5 min at 8°C). Cells were either suspended in 50 mM potassium phosphate buffer (pH 7.0) (K_P) to an OD₆₀₀ of 5.0 and directly used for import studies or applied for protein fractionation and immune detection.

Protein fractionation and immune detection of the recombinant SAM carrier. Heterologous expression and insertion of the recombinant protein in the membrane fraction were analyzed by immune detection. First, cell wall integrity was reduced by freezing of the pellet in liquid nitrogen and subsequent thawing, and incubation for 5 to 10 min at 37°C resulted in release and activity of endogenous lysozyme of the BLR cells. Autolysis was conducted in the presence of the protease inhibitor phenylmethylsulfonyl fluoride (PMSF) (1 mM), and cell disruption was complemented by sonication (addition of RNase and DNase). In a first centrifugation step (20,000 × g for 15 min at 4°C), cell debris and incorrectly folded membrane protein aggregates, so-called inclusion bodies, were en-

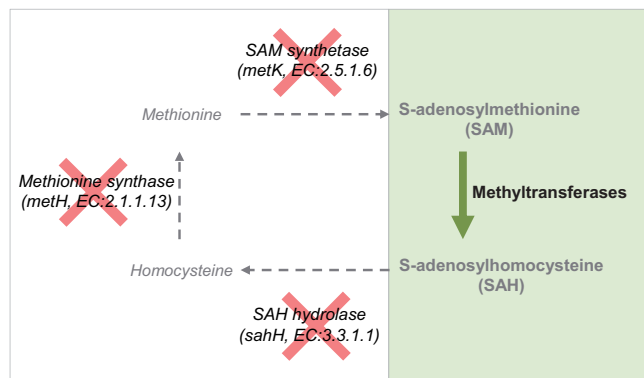


FIG 1 *A. asiaticus* harbors an incomplete methylation cycle. In total, 17 methyltransferases were identified in the *A. asiaticus* genome; the remaining enzymes of the methylation cycle required for SAM (re)generation are absent (crossed out in red). Methyl group transfer from SAM leads to the generation of SAH (shaded in green and in boldface type), whereas the capacities for SAH degradation and SAM regeneration are missing (italic type). Details on the methyltransferases present in *A. asiaticus* can be found in Table S1 in the supplemental material.

riched from the homogenate. Membrane proteins of the supernatant were afterwards separated from soluble proteins by ultracentrifugation ($100,000 \times g$ for 30 min at 4°C). Proteins of the membrane fraction were analyzed by SDS-PAGE (3% stacking and 15% separating gel) (37). Following electrophoresis, proteins were Coomassie stained or transferred onto a nitrocellulose membrane in a wet-blotting apparatus. Expression of the recombinant protein was verified by Western blotting and immune detection with anti-poly-His IgG combined with a secondary alkaline phosphatase-conjugated anti-mouse IgG (Sigma). Alkaline phosphatase activity was demonstrated by nitroblue tetrazolium chloride-5-bromo-4-chloro-3'-indolyl phosphate toluuidine staining. A prestained broad-range marker (7 to 175 kDa; New England BioLabs) was applied for estimation of the molecular protein masses.

Import studies with radioactively labeled SAM. Transport studies with intact cells are well suited to investigate SAM import because *E. coli* does not possess endogenous SAM uptake systems (15, 16). Import of radioactively labeled SAM was determined with induced and noninduced (control) *E. coli* cells harboring the corresponding plasmids. For this, *E. coli* cells were incubated at 30°C in 50 mM potassium phosphate buffer complemented with the indicated concentrations of labeled SAM (NEN). Optionally, the transport medium was supplemented with the indicated concentrations of nonlabeled substrates or molecules. Termination of transport was achieved by removal of the external substrate due to application of the cells onto pretreated filters (mixed cellulose ester, $0.45\text{-}\mu\text{m}$ pore size; Whatman), vacuum filtration, and washing (three times with 4 ml of KP₁ buffer). Radioactivity of the cell samples at the filters was quantified with a scintillation counter (Beckman LS6500; Beckman Coulter).

RESULTS

Metabolic requirement of SAM import and SAH export in *A. asiaticus*. The methylation pathway, including the SAM synthetase MetK, the SAH hydrolase SahH, and the methionine synthetase MetH, is completely absent in *A. asiaticus* (Fig. 1). However, the *A. asiaticus* genome encodes 17 putative methyltransferases as well as a homologue of the SAM-tRNA ribosyltransferase-isomerase (QueA [Aasi_0780]), which is responsible for the transfer of the ribose moiety of SAM into the modified tRNA (see Table S1 in the supplemental material). Therefore, there is clearly a need for SAM as a cofactor of methylation reactions as well as a donor of ribosyl groups in tRNA synthesis in *A. asiaticus*. Moreover, be-

cause *A. asiaticus* apparently lacks SAH-degrading enzymes, specific removal of SAH is mandatory to prevent inhibition of methyltransferases by accumulating SAH (7, 8). Consequently, a SAM import and SAH export system is predicted for *A. asiaticus*.

Comparative sequence analyses and phylogeny of SAM transport proteins. During analysis of the *A. asiaticus* genome, we identified a putative SAM transporter: Aasi_1859 is a 285-amino-acid protein with 10 predicted transmembrane helices and shows 45% amino acid sequence identity to the functionally characterized rickettsial SAM transporter encoded by the RP076 gene. Aasi_1859 and homologues contain a duplicated (functionally uncharacterized) EamA domain (Pfam accession no. PF00892) and belong to the drug-metabolite transporter (DMT) superfamily and the 10-transmembrane-segment (10-TMS) drug-metabolite exporter (DME) family (2.A.7.3) (38). Recently, a SAM transporter has also been identified in *Chlamydia trachomatis* (CTL0843) (15). Aasi_1859 and RP076 show only low amino acid sequence identity to CTL0843 (approximately 20%) (see Fig. S1 in the supplemental material). However, all carriers belong to the 10-TMS DME family. In addition, RP076 and Aasi_1859 homologues with more than 40% amino acid identity were also identified in other obligate intracellular bacteria belonging to the *Rickettsiales* and *Bacteroidetes*. The highest amino acid identity (47% amino acid identity) of Aasi_1859 is shared with a homologue found in “*Candidatus* *Odyssella thessalonicensis*,” an amoeba symbiont belonging to the *Rickettsiales*, as well as with “*Candidatus* *Cardinium hertigii*” cEper1, an obligate intracellular symbiont of parasitic wasps, representing the sister lineage of *A. asiaticus* (39). Interestingly, “*Ca. Odyssella thessalonicensis*” encodes at least three highly similar copies of Aasi_1859 homologues. Surprisingly, we also identified homologues in some members of the green algae (prasinophytes, order *Mamiellales*): *Ostreococcus* and *Micromonas* (40 to 42% amino acid sequence identity). We retrieved 101 homologues of Aasi_1859 and performed phylogenetic analyses of the (putative) SAM transport proteins. Among the Aasi_1859 and RP076 homologues used for phylogenetic analyses, no functionally characterized proteins were found; all homologues belong to the 10-TMS DME family. The application of maximum likelihood, neighbor-joining, and maximum parsimony treeing methods yielded stable phylogenetic relationships: all candidate SAM transporters and homologues of Aasi_1859 and RP076 clustered in a stable monophyletic lineage (Fig. 2). Due to the low sequence similarity, no calculation of phylogenetic relationships of Aasi_1859 and RP076 homologues with CTL0843 and homologues was possible.

Heterologous expression of Aasi_1859 stimulates [^{14}C]SAM uptake into *E. coli*. The high similarity to the rickettsial SAM carrier suggests that the homologue from *A. asiaticus* might act as a SAM transporter mediating the uptake of the essential cofactor into the endosymbiont. Reverse transcriptase PCR analysis with total RNA purified from amoebae harboring bacterial endosymbionts demonstrated transcription of Aasi_1859 during intracellular multiplication of *A. asiaticus* (see Fig. S2 in the supplemental material). Because functional analyses of carriers in *A. asiaticus* are hampered, if not impossible, due to its obligate intracellular lifestyle, we applied the heterologous *E. coli* expression system to investigate the biochemical properties of the Aasi_1859 gene product. Import measurements in intact *E. coli* cells synthesizing the recombinant carrier were previously successfully used to functionally characterize the rickettsial as well as the chlamydial

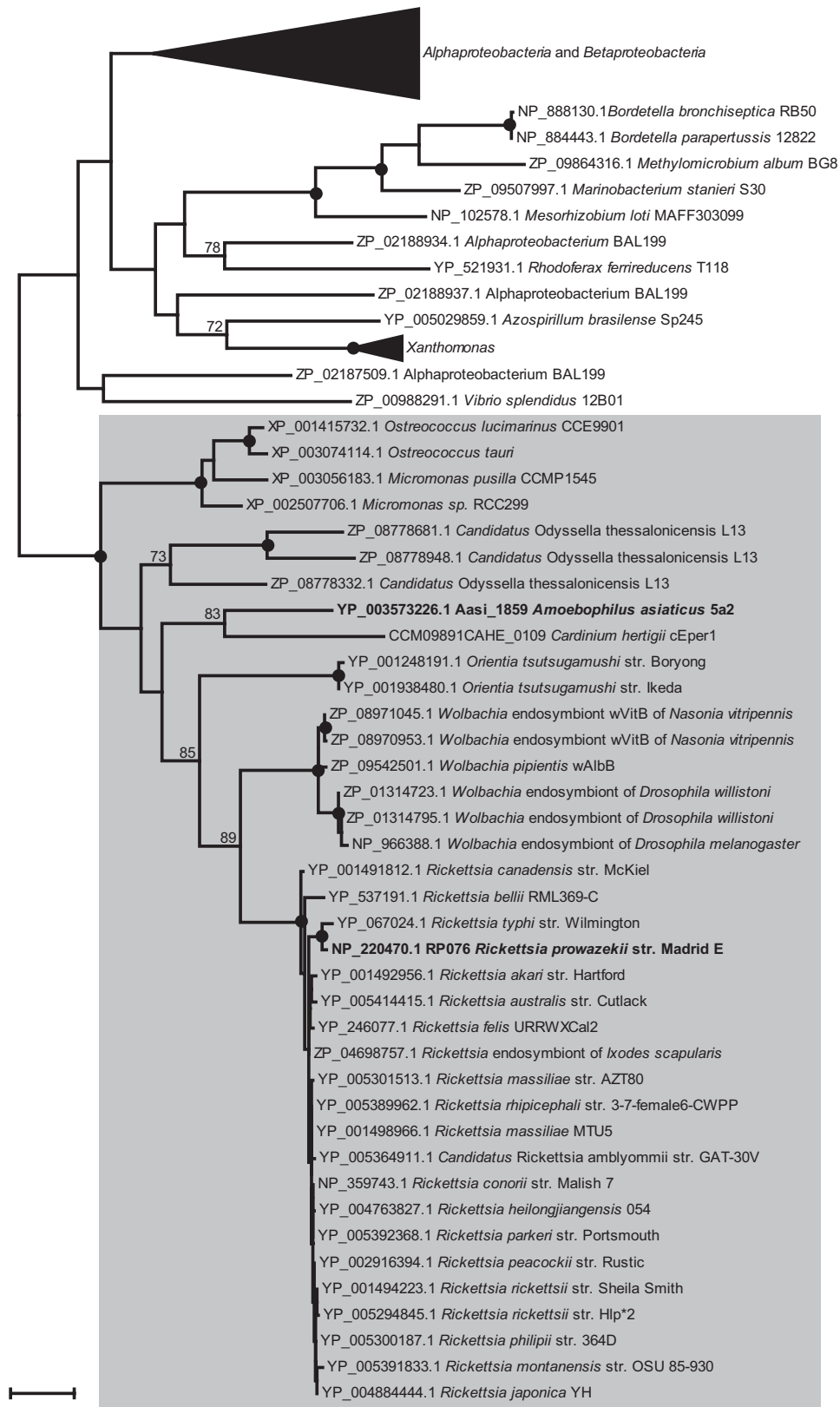


FIG 2 Phylogenetic relationships of Aasi_1859 and related characterized and putative SAM transport proteins. An amino-acid-based phylogenetic tree calculated with MEGA5 using the maximum likelihood algorithm with the JTT model is shown. Black dots indicate nodes which are supported by maximum likelihood, maximum parsimony, and neighbor-joining bootstrap values (1,000× resampling) greater than 90%. GenBank accession numbers are indicated. The bar represents 20% estimated evolutionary distance. Functionally characterized SAM transporters are shown in boldface type. The group comprising (putative and characterized) SAM transporters from *Rickettsiales*, prasinophytes, and *Bacteroidetes* is highlighted in gray.

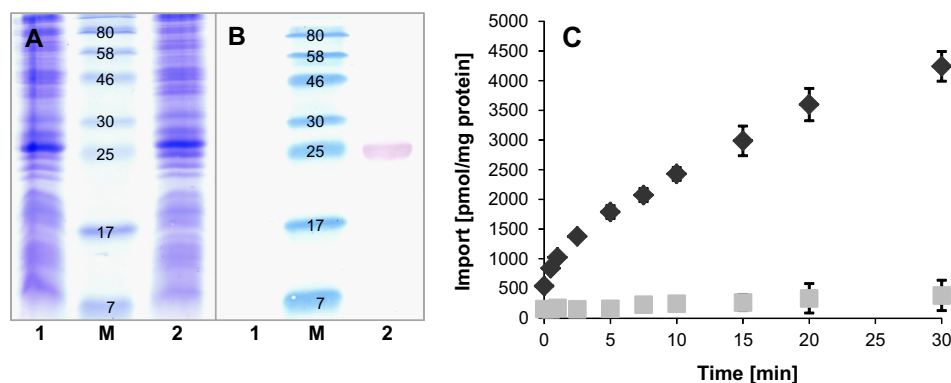


FIG 3 Heterologous expression and verification of functional membrane insertion of *AaSMT*. (A and B) Proteins of the *E. coli* membrane fraction (25 μg per lane) were separated by SDS-PAGE (A), and the presence of recombinant *AaSMT* was verified by immunodetection (B). Lanes: M, molecular mass marker (in kDa); 1, total membrane proteins of noninduced *E. coli* cells harboring the *AaSMT*-pET16b expression vector; 2, total membrane proteins from *E. coli* expressing *AaSMT*. (C) Time dependency of *AaSMT*-catalyzed uptake of radioactively labeled [*methyl*- ^{14}C]SAM. IPTG-induced (\blacklozenge) and noninduced (\square) *E. coli* cells were incubated with 10 μM ^{14}C -labeled SAM, and import was stopped at the indicated times by removal of external substrates via vacuum filtration and washing. Data are the means of data from three independent experiments, each with two technical replicates. Standard errors are displayed.

SAM carriers (15, 16). To allow comparison of our results with published data, we also applied the *E. coli* expression system and used *S*-adenosyl-L-[*methyl*- ^{14}C]methionine ([*methyl*- ^{14}C]SAM) for the majority of import studies.

Heterologous expression in *E. coli* and membrane insertion of the recombinant protein were analyzed by SDS-PAGE, Western blotting, and immunostaining (Fig. 3A and B). Induction of expression resulted in a significant accumulation of the recombinant carrier in the membrane fraction, whereas noninduced cells showed no comparable heterologous protein synthesis (Fig. 3B). Moreover, import studies with radioactively labeled [*methyl*- ^{14}C]SAM demonstrated that the carrier from *A. asiaticus* mediates a time-dependent uptake of radioactivity into induced cells (Fig. 3C). Import was linear for the first 2.5 min and slowly approached saturation of about 4,500 pmol mg protein $^{-1}$ at 30 min. Noninduced (control) cells showed no or comparatively low import of radioactivity, with maximal values of about 400 pmol mg protein $^{-1}$ (Fig. 3C). These data demonstrate that (i) *Aasi_1859* is heterologously expressed, (ii) the recombinant carrier is functional in the context of the *E. coli* membrane, and (iii) it accepts SAM as a substrate and thus might act as a SAM transporter in *A. asiaticus*. Here, we refer to this protein as *AaSMT*.

SAM transport depends on the presence of a proton gradient. SAM uptake via the rickettsial carrier was shown to be highly reduced by addition of the protonophore DNP (2,4-dinitrophenol) (1 mM) (16). Accordingly, rickettsial SAM translocation was suggested to be a proton gradient-dependent process. SAM import via the chlamydial carrier was also affected by protonophore addition (50 μM CCCP [cyanide *m*-chlorophenylhydrazone]) but to a lesser extent (50% residual activity) (15). To elucidate the transport mode of *AaSMT* and to identify whether *AaSMT* function is also influenced by the proton gradient, transport studies were performed in the presence of protonophores. Application of 10 μM CCCP already inhibited [*methyl*- ^{14}C]SAM uptake to a residual rate of about 13%, and the presence of 100 μM CCCP nearly completely abolished SAM accumulation (3.5% residual activity) compared to unaffected transport (set to 100%) (Fig. 4). In contrast to CCCP, higher concentrations of DNP (1 mM) were required to significantly reduce SAM import into *E. coli* cells ex-

pressing *AaSMT* (17% residual activity). This is because CCCP is known to be a more efficient protonophore than DNP, and comparably low concentrations of CCCP are sufficient to deplete the proton gradient across the *E. coli* membrane (40). The pronounced inhibitory effect of protonophores on SAM transport demonstrates that *AaSMT* operates not identically to the chlamydial SAM carrier but rather acts like the rickettsial SAM carrier.

Determination of the substrate specificity of *AaSMT*. The rickettsial as well as the chlamydial SAM transporters were shown to be specific for SAM uptake, and SAH was also suggested to be a potential substrate of these carriers (15, 16). To obtain a prelimi-

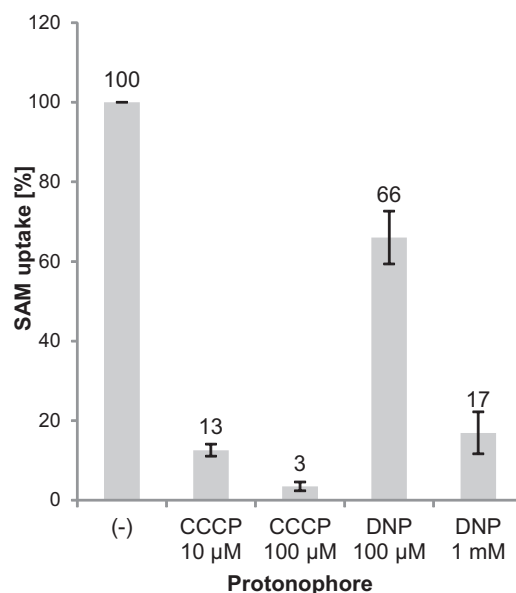


FIG 4 Analysis of the proton dependency of *AaSMT*. Shown are the effects of the protonophores CCCP and DNP on [*methyl*- ^{14}C]SAM uptake (5 min, 10 μM SAM). Transport rates are given as a percentage of nonaffected transport (-), which was set to 100%. Net values (minus control, noninduced cells) were used for calculation. Error bars are indicated. The values of the transport rates are displayed above the corresponding bars.

TABLE 1 Effects of various metabolites/inhibitors on [*methyl*-¹⁴C]SAM import by AaSAMT^a

Effector	SAM import (%)	SE (%)
None	100	±7
SAM	8	±2
SAH	6	±5
dc-SAH	34	±4
Sinefungin	69	±5
S-Adenosylcysteine	83	±3
ATP	97	±3
ADP	100	±5
AMP	102	±4
Methylthioadenosine	95	±3
Adenosine	130	±6
Adenine	95	±5
Cystathionine	101	±4
Methionine	98	±5
Homocysteine	99	±6
Cysteine	98	±4

^a Uptake of [*methyl*-¹⁴C]SAM by recombinant AaSAMT was measured at a substrate concentration of 10 μM, and nonlabeled effectors were present in a 10-fold excess. Structures of the tested molecules are shown in Fig. S3 in the supplemental material. Import was stopped after 2 min. Rates of SAM uptake are net values (minus control, noninduced *E. coli* cells) given as percentage of nonaffected transport (set to 100%). Data are the means of data from four independent experiments. Standard errors are given.

nary indication of the substrate spectrum of AaSAMT, we performed competition experiments with molecules highly as well as distantly related structurally to SAM (see Fig. S3 in the supplemental material). Import of [*methyl*-¹⁴C]SAM was measured in the presence of 14 different nonlabeled tested molecules applied in a 10-fold excess. The corresponding import was calculated in relation to nonaffected SAM uptake (set to 100%). A large reduction of the import rate might be indicative of transport inhibition or competition of the added compound with SAM during translocation. SAM uptake by recombinant AaSAMT was highly reduced by addition of SAM or SAH (<10% residual activity). A significant decrease of SAM import was also obtained by addition of the aminopropyl transferase inhibitor S-(5'-adenosyl)-3-thiopropylamine (dc-SAH) (~34% residual activity), whereas the methyltransferase inhibitor sinefungin caused only a slight reduction (~69% residual activity). All remaining tested molecules had no or comparably small effects (>80% residual activity) (Table 1). The observed effects suggest that the presence of the sulfur atom but not of the methyl and carboxyl group is required for substrate (or inhibitor) recognition. Apparently, AaSAMT exhibits quite high specificity for SAM uptake, but SAH might also represent another important substrate.

AaSAMT catalyzes counterexchange of SAM and SAH. We investigated the possible counterexchange capacity of AaSAMT to clarify its transport mode. Simultaneously, the effect of CCCP on the maintenance of the intracellular label was also analyzed. So-called chase or efflux experiments allow determination of whether external substrates/effectors can induce the release of labeled substrates previously loaded into *E. coli* cells. Because SAH might represent an additional substrate of AaSAMT, we also focused on its role during counterexchange transport.

It is very likely that viable recombinant *E. coli* cells exhibit efficient methylation activity and transfer the labeled methyl group of imported [*methyl*-¹⁴C]SAM to the diverse substrates of

methyltransferases. Therefore, label becomes at least partially “fixed” in the cell during transport measurements conducted with [*methyl*-¹⁴C]SAM. As a consequence, externally added substrates/effectors might be unable to induce complete efflux of radioactivity from the cells, and also, export of demethylated SAH cannot be monitored when cells are loaded with [*methyl*-¹⁴C]SAM. Therefore, we performed a chase experiment with S-adenosyl-L-[*carboxy*-¹⁴C]methionine ([*carboxy*-¹⁴C]SAM) carrying a labeled carboxyl group. Because the ¹⁴C label is not removed during endogenous methylation processes, the use of [*carboxy*-¹⁴C]SAM enables detection of SAM plus SAH export. We determined time-dependent uptake of [*carboxy*-¹⁴C]SAM (loading of radioactivity into the cell) and analyzed whether the addition of a 20-fold excess of nonlabeled SAM or SAH induces the efflux of label. Nonlabeled SAM as well as nonlabeled SAH caused fast and approximately complete efflux of radioactivity (Fig. 5). Therefore, it becomes evident that AaSAMT acts in an antiport manner and that both SAM and SAH represent efficient counterexchange substrates.

Addition of CCCP also led to a considerable depletion of cellular radioactivity. However, uncoupling of the proton gradient caused a slower decrease, and the plateau phase was reached at a slightly higher residual activity than the reduction of interior label by SAM or SAH addition (Fig. 5). This observation suggests that transport/maintenance of interior label is dependent on the proton gradient.

A closer examination of the time kinetics reveals that application of [*carboxy*-¹⁴C]SAM results in faster saturation and lower maximal transport rates than the time-dependent accumulation

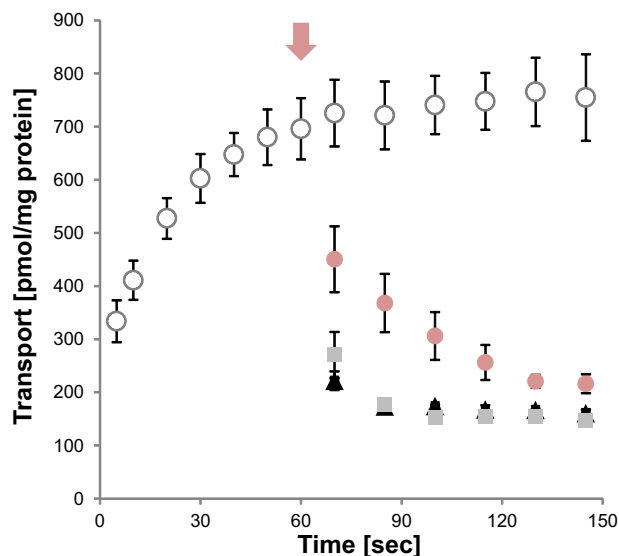


FIG 5 AaSAMT-driven import of [*carboxy*-¹⁴C]SAM and efflux analysis. Shown is the time-dependent import of 10 μM [*carboxy*-¹⁴C]SAM into IPTG-induced *E. coli* cells. Possible efflux of internal radioactivity was induced by addition of 200 μM nonlabeled SAM or 100 μM CCCP at the end of the linear phase of import (time point of addition is marked with a pink arrow). Shown are the time course of SAM import in the absence of effectors/nonlabeled substrates (white circles) and the time course of reduction of interior radioactivity after addition of nonlabeled SAM (black triangles), SAH (gray squares), or CCCP (pink circles). Data are the means of data from three independent experiments and represent net values of transport (calculated by subtraction of uptake rates of noninduced cells from import into cells expressing AaSAMT). Standard errors are displayed.

TABLE 2 Determination of kinetic parameters of AaSMT^a

Substrate	K_m (μM) (SE)	V_{max} (nmol mg protein ⁻¹ h ⁻¹) (SE)
[methyl- ¹⁴ C]SAM	13.89 (0.81)	104.80 (6.49)
[carboxy- ¹⁴ C]SAM	10.78 (1.21)	46.59 (3.20)

^a For determination of the respective K_m and V_{max} values, import was measured with increasing concentrations of [methyl-¹⁴C]SAM or [carboxy-¹⁴C]SAM. Determination of [methyl-¹⁴C]SAM uptake (5 μM) was performed in the presence of increasing concentrations of unlabeled SAH for determination of the K_i value ($8.10 \pm 1.20 \mu\text{M}$) and IC_{90} (concentration resulting in 90% inhibition of SAM uptake) ($66.67 \pm 7.45 \mu\text{M}$). Transport of [methyl-¹⁴C]SAM was allowed for 2 min and transport of [carboxy-¹⁴C]SAM was allowed for 1 min and was stopped by vacuum filtration and washing. Data represent net values (minus control, noninduced *E. coli* cells) and are means of data from at least three independent experiments; standard errors are also indicated.

of [methyl-¹⁴C]SAM (compare Fig. 5, white circles, and 3C, black diamonds). Knowledge about the counterexchange activity of AaSMT helps to interpret these differences. Reduced accumulation of [carboxy-¹⁴C]SAM might result from the simultaneous export of label, more precisely from the export of [carboxy-¹⁴C]SAM and [carboxy-¹⁴C]SAH. Whereas uptake of [methyl-¹⁴C]SAM is not accompanied by a comparably high loss of interior label, apparently small amounts of [methyl-¹⁴C]SAM but large amounts of nonlabeled SAH are exported. Therefore, we propose that *E. coli* efficiently converts SAM into SAH, the methyl group becomes trapped in the cell, and significant amounts of SAH are exported during counterexchange.

AaSMT exhibits high affinity for SAM and SAH import. Finally, we determined the biochemical parameters of AaSMT for SAM and SAH import. The apparent affinities and maximal velocities for [methyl-¹⁴C]SAM and for [carboxy-¹⁴C]SAM import were analyzed by application of increasing exterior substrate concentrations. Because the K_m values for import of [methyl-¹⁴C]SAM and [carboxy-¹⁴C]SAM are quite similar ($13.9 \pm 0.8 \mu\text{M}$ and $10.8 \mu\text{M} \pm 1.2 \mu\text{M}$, respectively), the affinity of the carrier for SAM apparently is not or is only marginally influenced by the position of the label (Table 2). Moreover, due to the low SAM K_m value, AaSMT is considered, like its chlamydial and rickettsial homologues, to be a high-affinity SAM transporter. AaSMT imported [methyl-¹⁴C]SAM with about a 2-fold-higher V_{max} ($104.8 \text{ nmol mg protein}^{-1} \text{ h}^{-1}$) than that of [carboxy-¹⁴C]SAM ($46.6 \pm 3.2 \text{ nmol mg protein}^{-1} \text{ h}^{-1}$) (Table 2), demonstrating that the observed difference in the time courses of [methyl-¹⁴C]SAM (Fig. 3C) and [carboxy-¹⁴C]SAM (Fig. 5) uptake resulted from different maximal import velocities of the corresponding transport processes. Generally, V_{max} values are influenced by the amount of functional recombinant transport protein in the *E. coli* membrane. However, import studies for V_{max} determination were performed in parallel with exactly the same *E. coli* cells, which guarantees that identical amounts of recombinant carriers were analyzed. A plausible explanation for the different V_{max} values is that AaSMT mediates measurable SAM counterexchange with demethylated products. As mentioned above, intact, metabolically active *E. coli* cells are capable of using SAM as a methyl group donor. Accordingly, endogenous methylation processes at least partially trap the methyl group of [methyl-¹⁴C]SAM in the cell, and export of SAH is not accompanied by a loss of label. However, after demethylation of [carboxy-¹⁴C]SAM, the ¹⁴C label still remains at SAH, and counterexchange of SAM and SAH causes a greater loss of internal radioactivity. Accordingly, import and ex-

port of labeled substrates ([carboxy-¹⁴C]SAM versus [carboxy-¹⁴C]SAH) result in a faster equilibrium, a faster saturation, and, thus, a lower apparent V_{max} value of the corresponding transport.

Two important observations indicate that SAH represents an additional substrate of AaSMT: first, it competes with SAM for import (Fig. 5), and second, SAH—just like SAM—induces the efflux of label from *E. coli* cells loaded with [carboxy-¹⁴C]SAM (Table 1). Because radioactively labeled SAH is not commercially available, we applied increasing concentrations of unlabeled SAH to [methyl-¹⁴C]SAM import to get an idea about the affinity of AaSMT for SAH. By this approach, we identified an apparent K_i value of about $8.1 \pm 1.0 \mu\text{M}$, and 90% SAM transport inhibition was obtained by addition of $66.7 \pm 7.5 \mu\text{M}$ SAH. Therefore, AaSMT exhibits a high apparent affinity for SAH import, quite similar to that of SAM uptake. The determined characteristics suggest that AaSMT can efficiently mediate SAM/SAH exchange.

DISCUSSION

Obligate intracellular bacteria are generally characterized by a highly reduced genome size and an impaired metabolic capacity (21, 22, 33, 34, 41, 42). In these organisms, essential metabolic pathways are often truncated or missing completely, and hence, import of intermediates or products is of high physiological importance. In the past years, several carrier proteins that mediate the provision of diverse metabolically relevant molecules and thus compensate for the reduced biosynthetic activity in intracellular bacteria have been identified (15, 16, 23–29, 43). Analysis of the genome of *A. asiaticus* revealed that its size is comparable to those of other intracellular bacteria; however, it encodes an unusually small number of proteins involved in metabolic processes (32). *A. asiaticus* lacks the oxidative pentose phosphate pathway and is impaired in ATP regeneration via the electron transport chain, the tricarboxylic acid cycle, and glycolysis. Moreover, in *A. asiaticus*, pathways for the *de novo* synthesis of purine and pyrimidine nucleotides, cofactors, and almost all amino acids are absent (32). This metabolic reduction necessitates the uptake of diverse metabolites from the amoeba host.

A. asiaticus encodes 17 putative methyltransferases (see Table S1 in the supplemental material), and thus, methylation apparently still takes place in this endosymbiont. This observation, combined with the fact that *A. asiaticus* does not possess SAM-generating and SAH-degrading enzymes, implies that corresponding reactions have to be performed by the host cell and that SAM and SAH have to be shuttled across the bacterial membrane.

Our analyses suggest that the protein AaSMT, encoded by the Aasi_1859 gene, possesses the biochemical prerequisites required to fulfill SAM and SAH exchange in *A. asiaticus*. The recombinant carrier mediates significant import of radioactively labeled SAM when heterologously expressed in *E. coli* (Fig. 3C). A proton gradient across the *E. coli* membrane was shown to be required for accumulation (Fig. 4) and maintenance (Fig. 5) of interior label. At first glance, the proton dependency suggested that SAMT from *A. asiaticus* catalyzes a secondary active H^+ /SAM symport and thus might be capable of net SAM supply. However, effector studies, application of differentially labeled SAM, and efflux studies demonstrated that AaSMT mediates counterexchange of SAM and SAH (Table 1 and Fig. 3C and 5). Both substrates are transported with comparably high affinities (K_m of $\sim 12 \mu\text{M}$ for SAM and K_i of $\sim 8 \mu\text{M}$ for SAH upon SAM import) (Table 2).

The absence of the methylation cycle (Fig. 1) necessitates SAM import and SAH export in *A. asiaticus*. Because SAM and SAH represent import and export substrates of AaSMT, it is important to check which physiological conditions allow SAM exploitation of the host and removal of bacterial SAH. Eukaryotic organisms generally exhibit higher SAM than SAH concentrations, and even under conditions of methyl deficiency, cellular SAM/SAH ratios higher than 1 were still identified in almost all investigated tissues (except from kidney [ratio of 0.61]) of mice (44). Accordingly, a balanced nutrient supply guarantees that more SAM than SAH is available in the host amoeba and that mainly SAM enters the bacterium. Moreover, methylation in *A. asiaticus* results in SAM consumption and fuels the carrier with SAH at the bacterial inner face. Interestingly, a decrease in SAM content was observed in *Physarum flavicomum* amoebae during the developmental transition from a growing state to dormant cysts (45, 46). Transferring this situation to the host of *A. asiaticus*, efficient SAM exploitation by the endosymbiont is rather restricted to the growing state of the amoeba. A decrease of the SAM/SAH ratio in the host (methyl deficiency and transition to dormancy) will cause SAH uptake into the endosymbiont. As a consequence, endosymbiotic methylation processes will slow down due to inhibition of methyltransferases by excess SAH (7, 8) and/or due to substrate deprivation.

Moreover, decreased metabolic activity of the host generally affects metabolite provision to the symbiotic bacterium, resulting in alteration of physiological processes and most likely in an insufficient membrane potential (47). The proposed reduction of the bacterial proton gradient inactivates H⁺ symport and influences H⁺-regulated carriers, including AaSMT.

Interestingly, Aasi_1859 is located within a cluster of three genes involved in tRNA modification: *mmmA* (Aasi_1200), *mmmE* (Aasi_1201), and *tifs* (Aasi_1198) (see Fig. S4 in the supplemental material). This might suggest a role of Aasi_1859—more precisely, of its substrate SAM—in tRNA modification (3). The corresponding ribosyl group transfer results in formation of methionine. The incapability of methionine to compete with SAM for import (Table 1) suggests that methionine is no substrate of AaSMT, and therefore, AaSMT apparently does not catalyze SAM uptake in exchange with methionine.

To establish a basic pool of SAM and to fuel nonmethylation processes with this cofactor, net uptake of SAM or at least SAM exchange with substrates different from SAH is required in *A. asiaticus*. Remarkably, addition of CCCP resulted in a significant loss of radioactively labeled SAM and SAH from *E. coli* cells expressing AaSMT (Fig. 5). The corresponding substrate flux is most likely driven by the concentration gradient across the bacterial membrane. Accordingly, net provision of SAM or SAH to the bacterium by AaSMT is imaginable, at least under conditions of a reduced membrane potential. SAM uptake in exchange with other, not-yet-identified substrates might provide SAM non-methylation processes. However, it is also possible that further SAM import systems exist in *A. asiaticus*.

In 2003, the first bacterial SAM transporter was identified in *R. prowazekii* (16), and recently, SAM transport was also clarified for *C. trachomatis* (15). These carriers showed high affinities for SAM import (K_m values of ~2.5 μM for *R. prowazekii* RP076 and ~6 μM for *C. trachomatis* CTL0843) and also a quite low SAH K_i of SAM transport (K_i values of ~14.3 μM for *R. prowazekii* RP076 and ~4.2 μM for *C. trachomatis* CTL0843), and thus, these parameters are comparable to those of AaSMT. The rickettsial

SAM transporter was proposed to act as an H⁺/SAM symporter because its activity was highly influenced by the proton gradient (16). The capacity of the rickettsial SAM carrier to perform counterexchange was not investigated. Not only does *R. prowazekii* lack functional SAM synthetase, SAH degradation and SAM recycling are also missing in all sequenced *Rickettsia* species based on analyses using the KEGG (Kyoto Encyclopedia of Genes and Genomes) database (<http://www.genome.jp/kegg/>) (48). Therefore, SAM import and SAH export are also required in *R. prowazekii*. In fact, SAH was shown to efficiently compete with SAM for uptake and therefore might represent an additional substrate of the rickettsial SAM carrier (16). Moreover, the close phylogenetic relationship of AaSMT and the rickettsial SAM carrier (Fig. 2) is indicative of their common evolutionary origin and horizontal gene transfer. The high amino acid sequence identity of these carriers (see Fig. S1 in the supplemental material) suggests that both proteins still possess similar biochemical properties. Therefore, it is conceivable that the rickettsial SAM carrier does not represent an H⁺/SAM symporter but facilitates SAM/SAH counterexchange that is regulated by the proton gradient.

The SAM transporter from *C. trachomatis*, although belonging to the same transporter class, exhibits comparably low sequence similarities to the SAM carriers from *R. prowazekii* and *A. asiaticus* (see Fig. S1 in the supplemental material). Destruction of the proton gradient across the membrane of *E. coli* expressing the chlamydial SAM carrier resulted in partial reduction of SAM uptake and induced slight SAM efflux in the absence of suitable counterexchange substrates but apparently did not affect counterexchange transport (15). Therefore, the chlamydial carrier was assumed to catalyze a proton gradient-independent SAM/SAH exchange in addition to a proton-driven net SAM import. In this context, it is difficult to understand why significant counterexchange also occurs in the presence of a proton gradient. The postulated transport mode would imply that the recombinant chlamydial SAM carrier proteins act partially as symporters and partially as SAM/SAH counterexchangers when a proton gradient exists and hence presupposes a heterogeneous and inconsistent regulation.

The chlamydial SAM transporter differs from the rickettsial and *A. asiaticus* SAM carriers, at least in the regulatory impact of the proton gradient on SAM counterexchange. Establishment from different ancestral carriers of the DMT group might explain the functional differences of the chlamydial SAM carrier and the SAM transporter from *R. prowazekii* and *A. asiaticus*. In this context, it is interesting to note that chlamydiae are located within vacuoles (the so-called inclusion) inside their host cells (49), whereas rickettsiae and *A. asiaticus* as well as its relative “*Ca. Cardinium hertigii*” are located directly inside the host cytoplasm (32, 50, 51). This fundamental difference in subcellular location might explain the presence of different SAM transporters and thus different transport modes in the rickettsial/*Bacteroidetes* group of SAM transporters and chlamydiae.

Obviously, SAM transporter genes were spread due to horizontal gene transfer. However, until now, it has been impossible to determine the direction of transfer unambiguously (Fig. 2). Most likely, Aasi_1859 and RP076-like SAM transporters were invented in an intracellular ancestor of the *Rickettsiales* and then transferred to *A. asiaticus*, “*Ca. Cardinium hertigii*” (*Bacteroidetes*), and some members of the prasinophytes. Horizontal gene transfer between intracellular bacteria, including rickettsiae and *A. asiaticus*, has

been suggested previously (32, 52–54). Due to the high level of divergence of SAM transport proteins from their nearest neighbors, these horizontal gene transfer events are most likely evolutionarily ancient.

In parasitic or endosymbiotic bacteria, the establishment of SAM transporters apparently is tightly associated with the intracellular lifestyle and particularly with the loss of the SAM biosynthesis capacity. However, until now, it has been completely unclear why *Prasinophyceae*, comparably primitive, mainly marine green algae, harbor carrier proteins highly related to the bacterial SAM transporters. The absence of homologues in other algae and higher plants points to a special function of the Aasi_1859 and RP076 homologues in *Prasinophyceae*. Gene transfer from intracellular bacteria (particularly from chlamydiae) to plants has been suggested by several studies (21, 55–57). In higher plants, members of the mitochondrial carrier family (MCF) (which are unrelated to Aasi_1859 and homologues) were shown to catalyze SAM provision to mitochondria and plastids (58, 59). Whether the DME-type carriers in addition to or instead of MCF-type carriers mediate mitochondrial or plastidial SAM transport and whether they might act in another compartment or accept substrates other than SAM and SAH in *Prasinophyceae* are open questions to be investigated in further studies. However, a role of the DME-type SAM transporters in uptake of extracellular SAM into corresponding algae can most likely be ruled out because SAM is not freely available in their habitat.

Conclusion. Our analyses indicate that the SAM transporter of *A. asiaticus* operates as a proton gradient-dependent SAM/SAH antiporter and thus perfectly complements the restricted metabolic capabilities of *A. asiaticus*. Our results expand previous studies characterizing SAM transporters in *R. prowazekii* and *C. trachomatis*. The presence of functionally different SAM transporters in *Bacteroidetes* and *Rickettsiales* on the one hand and in chlamydiae on the other hand might be the result of different functional constraints due to their different intracellular localizations. Interestingly, SAM transporter-like genes were horizontally transferred between rickettsiae and *Bacteroidetes* and some members of the prasinophytes.

The analysis of additional bacterial SAM transporters, including homologues of prasinophytes, might help us to gain insights into structure-function relationships of this carrier subgroup.

ACKNOWLEDGMENTS

This work was funded by a grant from the Austrian Science Fund (FWF) (project no. P22703-B17) to S.S.-E. and M.H. and the Deutsche Forschungsgemeinschaft (DFG) (project no. HA 5423/1-1 and SPP 1580) to the laboratory of I.H. M.H. acknowledges support from the European Research Council (ERC StG EvoChlamy).

REFERENCES

- Cantoni GL. 1975. Biological methylation: selected aspects. *Annu. Rev. Biochem.* 44:435–451.
- Chiang PK, Gordon RK, Tal J, Zeng GC, Doctor BP, Pardhasaradhi K, McCann PP. 1996. S-Adenosylmethionine and methylation. *FASEB J.* 10:471–480.
- Fontecave M, Atta M, Mulliez E. 2004. S-Adenosylmethionine: nothing goes to waste. *Trends Biochem. Sci.* 29:243–249.
- Cheng X, Roberts RJ. 2001. AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res.* 29:3784–3795.
- Martin JL, McMillan FM. 2002. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* 12:783–793.
- Schubert HL, Blumenthal RM, Cheng X. 2003. Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci.* 28:329–335.
- Borchardt RT, Huber JA, Wu YS. 1976. Potential inhibitors of S-adenosylmethionine-dependent methyltransferases. 4. Further modifications of the amino and base portions of S-adenosyl-L-homocysteine. *J. Med. Chem.* 19:1094–1099.
- Borchardt RT, Shiong Y, Huber JA, Wycpalek AF. 1976. Potential inhibitors of S-adenosylmethionine-dependent methyltransferases. 6. Structural modifications of S-adenosylmethionine. *J. Med. Chem.* 19:1104–1110.
- Markham GD, DeParasis J, Gatmaitan J. 1984. The sequence of metK, the structural gene for S-adenosylmethionine synthetase in *Escherichia coli*. *J. Biol. Chem.* 259:14505–14507.
- Thomas D, Surdin-Kerjan Y. 1987. SAM1, the structural gene for one of the S-adenosylmethionine synthetases in *Saccharomyces cerevisiae*. Sequence and expression. *J. Biol. Chem.* 262:16704–16709.
- Cherest H, Surdin-Kerjan Y. 1978. S-Adenosyl methionine requiring mutants in *Saccharomyces cerevisiae*: evidences for the existence of two methionine adenosyl transferases. *Mol. Gen. Genet.* 163:153–167.
- Peleman J, Boerjan W, Engler G, Seurinck J, Botterman J, Alliotte T, Van MM, Inze D. 1989. Strong cellular preference in the expression of a housekeeping gene of *Arabidopsis thaliana* encoding S-adenosylmethionine synthetase. *Plant Cell* 1:81–93.
- Mato JM, Corrales FJ, Lu SC, Avila MA. 2002. S-Adenosylmethionine: a control switch that regulates liver function. *FASEB J.* 16:15–26.
- Hafner EW, Tabor CW, Tabor H. 1977. Isolation of a metK mutant with a temperature-sensitive S-adenosylmethionine synthetase. *J. Bacteriol.* 132:832–840.
- Binet R, Fernandez RE, Fisher DJ, Maurelli AT. 2011. Identification and characterization of the *Chlamydia trachomatis* L2 S-adenosylmethionine transporter. *mBio* 2(3):e00051–11. doi:10.1128/mBio.00051-11.
- Tucker AM, Winkler HH, Driskell LO, Wood DO. 2003. S-Adenosylmethionine transport in *Rickettsia prowazekii*. *J. Bacteriol.* 185:3031–3035.
- Andersson JO, Andersson SG. 1999. Genome degradation is an ongoing process in *Rickettsia*. *Mol. Biol. Evol.* 16:1178–1191.
- Sanchez-Perez GF, Bautista JM, Pajares MA. 2004. Methionine adenosyltransferase as a useful molecular systematics tool revealed by phylogenetic and structural analyses. *J. Mol. Biol.* 335:693–706.
- Binet R, Maurelli AT. 2009. The chlamydial functional homolog of KsgA confers kasugamycin sensitivity to *Chlamydia trachomatis* and impacts bacterial fitness. *BMC Microbiol.* 9:279. doi:10.1186/1471-2180-9-279.
- Pannekoek Y, Heurgué-Hamard V, Langerak AA, Speijer D, Buckingham RH, van der Ende AJ. 2005. The N5-glutamine S-adenosyl-L-methionine-dependent methyltransferase PrmC/HemK in *Chlamydia trachomatis* methylates class 1 release factors. *J. Bacteriol.* 187:507–511.
- Collingro A, Tischler P, Weinmaier T, Penz T, Heinz E, Brunham RC, Read TD, Bavoil PM, Sachse K, Kahane S, Friedmann MG, Rattei T, Myers GS, Horn M. 2011. Unity in variety—the pan-genome of the *Chlamydiae*. *Mol. Biol. Evol.* 28:3253–3270.
- Fuxelius HH, Darby A, Min CK, Cho NH, Andersson SG. 2007. The genomic and metabolic diversity of *Rickettsia*. *Res. Microbiol.* 158:745–753.
- Tjaden J, Winkler HH, Schwöppe C, van der Laan M, Möhlmann T, Neuhaus HE. 1999. Two nucleotide transport proteins in *Chlamydia trachomatis*, one for net nucleoside triphosphate uptake and the other for the transport of energy. *J. Bacteriol.* 181:1196–1202.
- Haferkamp I, Schmitz-Esser S, Linka N, Urbany C, Collingro A, Wagner M, Horn M, Neuhaus HE. 2004. A candidate NAD⁺ transporter in an intracellular bacterial symbiont related to Chlamydiae. *Nature* 432:622–625.
- Audia JP, Winkler HH. 2006. Study of the five *Rickettsia prowazekii* proteins annotated as ATP/ADP translocases (Tlc): only Tlc1 transports ATP/ADP, while Tlc4 and Tlc5 transport other ribonucleotides. *J. Bacteriol.* 188:6261–6268.
- Schmitz-Esser S, Linka N, Collingro A, Beier CL, Neuhaus HE, Wagner M, Horn M. 2004. ATP/ADP translocases: a common feature of obligate intracellular amoebal symbionts related to chlamydiae and rickettsiae. *J. Bacteriol.* 186:683–691.
- Haferkamp I, Schmitz-Esser S, Wagner M, Neigel N, Horn M, Neuhaus HE. 2006. Tapping the nucleotide pool of the host: novel nucleotide carrier proteins of *Protochlamydia amoebophila*. *Mol. Microbiol.* 60:1534–1545.

28. Knab S, Mushak TM, Schmitz-Esser S, Horn M, Haferkamp I. 2011. Nucleotide parasitism by *Simkania negevensis* (Chlamydiae). *J. Bacteriol.* 193:225–235.
29. Krause DC, Winkler HH, Wood DO. 1985. Cloning and expression of the *Rickettsia prowazekii* ADP/ATP translocator in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 82:3015–3019.
30. Ren Q, Paulsen IT. 2007. Large-scale comparative genomic analyses of cytoplasmic membrane transport systems in prokaryotes. *J. Mol. Microbiol. Biotechnol.* 12:165–179.
31. Moya A, Peretó J, Gil R, Latorre A. 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat. Rev. Genet.* 9:218–229.
32. Schmitz-Esser S, Tischler P, Arnold R, Montanaro J, Wagner M, Rattei T, Horn M. 2010. The genome of the amoeba symbiont “*Candidatus Amoebophilus asiaticus*” reveals common mechanisms for host cell interaction among amoeba-associated bacteria. *J. Bacteriol.* 192:1045–1057.
33. Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. 2009. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biol. Direct* 4:13. doi:10.1186/1745-6150-4-13.
34. Wernegreen JJ. 2005. For better or worse: genomic consequences of intracellular mutualism and parasitism. *Curr. Opin. Genet. Dev.* 15:572–583.
35. Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* 9:286–298.
36. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28:2731–2739.
37. Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680–685.
38. Saier MH, Jr. 2000. A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol. Mol. Biol. Rev.* 64:354–411.
39. Penz T, Schmitz-Esser S, Kelly SE, Cass BN, Muller A, Woyke T, Malfatti SA, Hunter MS, Horn M. 2012. Comparative genomics suggests an independent origin of cytoplasmic incompatibility in *Cardinium heritigii*. *PLoS Genet.* 8:e1003012. doi:10.1371/journal.pgen.1003012.
40. Jana B, Panja S, Saha S, Basu T. 2009. Mechanism of protonophore-mediated induction of heat-shock response in *Escherichia coli*. *BMC Microbiol.* 9:20. doi:10.1186/1471-2180-9-20.
41. Casadevall A. 2008. Evolution of intracellular pathogens. *Annu. Rev. Microbiol.* 62:19–33.
42. Horn M, Collingro A, Schmitz-Esser S, Beier CL, Purkhold U, Fartmann B, Brandt P, Nyakatura GJ, Droege M, Frishman D, Rattei T, Mewes HW, Wagner M. 2004. Illuminating the evolutionary history of chlamydiae. *Science* 304:728–730.
43. Schwöppe C, Winkler HH, Neuhaus HE. 2002. Properties of the glucose 6-phosphate transporter from *Chlamydia pneumoniae* (HPTcp) and the glucose 6-phosphate sensor from *Escherichia coli* (UhpC). *J. Bacteriol.* 184:2108–2115.
44. Caudill MA, Wang JC, Melnyk S, Pogribny IP, Jernigan S, Collins MD, Santos-Guzman J, Swendseid ME, Cogger EA, James SJ. 2001. Intracellular S-adenosylhomocysteine concentrations predict global DNA hypomethylation in tissues of methyl-deficient cystathionine beta-synthase heterozygous mice. *J. Nutr.* 131:2811–2818.
45. Zhu CM, Cumaraswamy A, Henney HR, Jr. 1989. Comparison of polyamine and S-adenosylmethionine contents of growing and encysted *Acanthamoeba* isolates. *Mol. Cell. Biochem.* 90:145–153.
46. Cumaraswamy A, Henney H, Jr. 1990. S-Adenosylmethionine and S-adenosylhomocysteine transitions in encysting *Physarum flavicomum* amoebae. *Biochem. Cell Biol.* 68:769–777.
47. Lopez-Amoros R, Comas J, Vives-Rego J. 1995. Flow cytometric assessment of *Escherichia coli* and *Salmonella typhimurium* starvation-survival in seawater using rhodamine 123, propidium iodide, and oxonol. *Appl. Environ. Microbiol.* 61:2521–2526.
48. Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28:27–30.
49. Saka HA, Valdivia RH. 2010. Acquisition of nutrients by *Chlamydiae*: unique challenges of living in an intracellular compartment. *Curr. Opin. Microbiol.* 13:4–10.
50. Hackstadt T. 1998. The diverse habitats of obligate intracellular parasites. *Curr. Opin. Microbiol.* 1:82–87.
51. Zchori-Fein E, Perlman SJ, Kelly SE, Katzir N, Hunter MS. 2004. Characterization of a ‘*Bacteroidetes*’ symbiont in Encarsia wasps (Hymenoptera: Aphelinidae): proposal of ‘*Candidatus Cardinium heritigii*’. *Int. J. Syst. Evol. Microbiol.* 54:961–968.
52. Bertelli C, Greub G. 2012. Lateral gene exchanges shape the genomes of amoeba-resisting microorganisms. *Front. Cell. Infect. Microbiol.* 2:110. doi:10.3389/fcimb.2012.00110.
53. Ogata H, La Scola B, Audic S, Renesto P, Blanc G, Robert C, Fournier PE, Claverie JM, Raoult D. 2006. Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genet.* 2:e76. doi:10.1371/journal.pgen.0020076.
54. Gillespie JJ, Joardar V, Williams KP, Driscoll T, Hostetler JB, Nordberg E, Shukla M, Walenz B, Hill CA, Nene VM, Azad AF, Sobral BW, Caler E. 2012. A *Rickettsia* genome overrun by mobile genetic elements provides insight into the acquisition of genes characteristic of an obligate intracellular lifestyle. *J. Bacteriol.* 194:376–394.
55. Tyra HM, Linka M, Weber AP, Bhattacharya D. 2007. Host origin of plastid solute transporters in the first photosynthetic eukaryotes. *Genome Biol.* 8:R212. doi:10.1186/gb-2007-8-10-r212.
56. Huang J, Gogarten JP. 2007. Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biol.* 8:R99. doi:10.1186/gb-2007-8-6-r99.
57. Moustafa A, Reyes-Prieto A, Bhattacharya D. 2008. Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. *PLoS One* 3:e2205. doi:10.1371/journal.pone.0002205.
58. Bouvier F, Linka N, Isner JC, Mutterer J, Weber AP, Camara B. 2006. *Arabidopsis* SAMT1 defines a plastid transporter regulating plastid biogenesis and plant development. *Plant Cell* 18:3088–3105.
59. Palmieri L, Arrigoni R, Blanco E, Carrari F, Zanon MI, Studart-Guimaraes C, Fernie AR, Palmieri F. 2006. Molecular identification of an *Arabidopsis* S-adenosylmethionine transporter. Analysis of organ distribution, bacterial expression, reconstitution into liposomes, and functional characterization. *Plant Physiol.* 142:855–865.

Chapter VII

Conclusion

Conclusion

The goal of this thesis was the characterization of host-cell interaction factors of two bacterial symbionts belonging to the diverse group of *Bacteroidetes*. These two endosymbionts, *Amoebophilus asiaticus* and *Cardinium hertigii* have different host spectra. *Amoebophilus* is an endosymbiont of *Acanthamoeba* (Horn et al, 2001), while *Cardinium* is an endosymbiont of wasps (Hunter et al, 2003). Both endosymbionts are moderately related, but nevertheless form a phylogenetic sister lineage and have many features in common. One outstanding feature of both genomes that we analyzed is the high content of effector proteins (Penz et al, 2012; Schmitz-Esser et al, 2010). Effector proteins are comparable with words of a foreign language that the symbiont needs to learn to be able to interact with the host. But how could it be possible for a symbiont to learn this foreign language? The answer is by horizontal gene transfer. Horizontal gene transfer is mostly triggered by mobile genetic elements such as phages, plasmids and transposases (Frost et al, 2005). The genome of the *Amoebophilus* is overrun by mobile genetic elements (Schmitz-Esser et al, 2011), while the genome of *Cardinium* is not (Penz et al, 2012). Bacteria that recently became an endosymbiont usually have a high load of mobile genetic elements (Moran et al, 2008), which facilitate the exchange of genetic material. The genetic exchange can be intragenomic or intergenomic. Host-cell interaction-mediating proteins, such as effector proteins, are usually translocated into the host via secretion systems (Basler et al, 2012). In both symbionts, *Amoebophilus* and *Cardinium*, we describe a new, putative secretion apparatus derived from a defective prophage (Penz et al, 2011; Penz et al, 2012). The prophage-derived secretion apparatus shows structural similarities to the type-six secretion system. This secretion apparatus was acquired via horizontal gene transfer, as were many transport proteins. In one study we could functionally characterize an S-adenosylmethionine carrier encoded in the *Amoebophilus*

genome that is used to gain energy from the host to compensate the reduced biosynthetic capabilities in small symbiontal genomes (Haferkamp et al, 2013).

From an evolutionary point of view, there is evidence that the *Cardinium* genome is associated with eukaryotic hosts for much longer than *Amoebophilus*. Compared to *Amoebophilus*, the *Cardinium* genome is much smaller, contains less mobile genetic elements and is more evolved in terms of host specificity (Penz et al, 2012). All these facts support the hypothesis that simple protists are hosts for bacteria that potentially became endosymbionts and thus protists serve as training grounds for eukaryotic host cell adaption of bacteria (Molmeret et al, 2005). After a first adaption to an intracellular environment in amoebae, bacteria have the potential to infect higher eukaryotes such as insects and mammals.

The genome of *Cardinium* is not only interesting from an evolutionary point of view. Some bacterial symbionts of insects such as *Cardinium* (Hunter et al, 2003) and the *Alphaproteobacterium Wolbachia* (Werren, 1997) form a group of reproductive manipulators. These reproductive manipulators are able to influence their hosts' biology, ecology and evolution. Acquisition of such a reproductive manipulator may change population structure, ecological interactions, behavior, and cause rapid evolution of life history, reproduction and sex determination systems. One of the most common phenotypes of reproductive manipulators appears to be cytoplasmic incompatibility (CI) (Werren, 1997), a type of reproductive failure, in which bacteria in insect males modify sperm in a way that reduces the reproductive success of uninfected female mates. In spite of considerable interest, and three sequenced genomes of CI-inducing *Wolbachia*, the genetic basis for this phenotype still remains largely unknown.

Cardinium is the only bacterial lineage that causes at least three distinct reproductive manipulator phenotypes in insects other than the well-studied *Wolbachia*. The publication included in this thesis is the first to have described a genome of this group.

That this lineage is in the distant *Bacteroidetes*, an ecologically dominant, but much less-studied phylum makes it of particular interest. Although the ‘master-manipulator’ status of *Cardinium* has been apparent since the early 2000s, the genetic resources for this symbiont have been limited to the availability of only two gene sequences. This is because the technical challenges of this project are huge. The organisms in which *Cardinium* phenotypes have been described are tiny insects and mites. In this study, the parasitic wasp host is 1/1000 the mass of a *Drosophila*, and must be reared on other insects. Only the combination of an enormous insect-rearing, symbiont purification and DNA sequencing effort allowed us to reconstruct the *Cardinium* genome from the sequenced insect/symbiont metagenome. The particular strain of *Cardinium* analyzed in this study is the only symbiont other than *Wolbachia* to cause CI. Comparative genome analysis indicates that CI evolved independently in the *Wolbachia* and *Cardinium* lineages and shows functional overlap in the suite of conserved proteins in both lineages that are likely involved in mediating host cell interactions and CI. We could also show that *Cardinium* lacks all major biosynthetic pathways with the exception of biotin and lipolate synthesis, suggesting a potential role in host nutrition. These findings provide a novel comparative context for understanding the mechanistic basis of cytoplasmic incompatibility. Despite great interest in the mechanism of CI, sequences of only *Wolbachia* genomes have not yielded a clear understanding of the genetic basis for this effect. In this study, comparisons with *Wolbachia* have shown types of genes in common, as well as notable differences that help sort lineage-specific effects from genes of interest for reproductive manipulation. Our findings substantially increase our knowledge on reproductive manipulator symbionts that do not only severely affect the population genetic structure of arthropods, but also may serve as powerful tools in pest management for the suppression or transformation of pest or vector populations.

Basler M, Pilhofer M, Henderson GP, Jensen GJ, Mekalanos JJ (2012) Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* **483**: 182-186

Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* **3**: 722-732

Haferkamp I, Penz T, Geier M, Ast M, Mushak T, Horn M, Schmitz-Esser S (2013) The endosymbiont *Amoebophilus asiaticus* encodes an S-adenosylmethionine carrier that compensates for its missing methylation cycle. *J Bacteriol* **195**: 3183-3192

Horn M, Harzenetter MD, Linner T, Schmid EN, Muller KD, Michel R, Wagner M (2001) Members of the *Cytophaga-Flavobacterium-Bacteroides* phylum as intracellular bacteria of acanthamoebae: proposal of 'Candidatus Amoebophilus asiaticus'. *Environ Microbiol* **3**: 440-449

Hunter MS, Perlman SJ, Kelly SE (2003) A bacterial symbiont in the *Bacteroidetes* induces cytoplasmic incompatibility in the parasitoid wasp *Encarsia pergandiella*. *Proc Biol Sci* **270**: 2185-2190

Molmeret M, Horn M, Wagner M, Santic M, Abu Kwaik Y (2005) Amoebae as training grounds for intracellular bacterial pathogens. *Appl Environ Microbiol* **71**: 20-28

Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* **42**: 165-190

Penz T, Horn M, Schmitz-Esser S (2011) The genome of the amoeba symbiont "*Candidatus Amoebophilus asiaticus*" encodes an *afp*-like prophage possibly used for protein secretion. *Virulence* **1**: 541-545

Penz T, Schmitz-Esser S, Kelly SE, Cass BN, Muller A, Woyke T, Malfatti SA, Hunter MS, Horn M (2012) Comparative genomics suggests an independent origin of cytoplasmic incompatibility in *Cardinium hertigii*. *PLoS Genet* **8**: e1003012

Schmitz-Esser S, Penz T, Spang A, Horn M (2011) A bacterial genome in transition - an exceptional enrichment of IS elements but lack of evidence for recent transposition in the symbiont *Amoebophilus asiaticus*. *BMC Evol Biol* **11**: 270

Schmitz-Esser S, Tischler P, Arnold R, Montanaro J, Wagner M, Rattei T, Horn M (2010) The genome of the amoeba symbiont "*Candidatus Amoebophilus asiaticus*" reveals common mechanisms for host cell interaction among amoeba-associated bacteria. *J Bacteriol* **192**: 1045-1057

Werren JH (1997) Biology of *Wolbachia*. *Annu Rev Entomol* **42**: 587-609

Chapter VIII

Summary

Zusammenfassung

Curriculum Vitae

Acknowledgments

Summary

Stable associations of bacteria with protozoa and insects are widespread and frequently found in nature. Bacteria of different phyla can be symbionts of these eukaryotes. The *Acanthamoeba* endosymbiont *Amoebophilus asiaticus* and the reproductive manipulator of parasitic wasps *Cardinium hertigii* belong to the diverse phylum of the *Bacteroidetes*. Both endosymbionts are distantly related to other symbionts within this phylum and phylogenetic analyses place these two symbionts together as sister lineages. Interestingly, the genomes of *Amoebophilus* and *Cardinium* have many features in common.

In this thesis I could show via comparative genome analysis that *Amoebophilus* and *Cardinium* have reduced genomes and encode for a putative, phage-derived secretion system. This phage-derived secretion system is similar to a defective prophage described in the genome of the entomopathogen *Serratia entomophila*. With qPCR and protein mass spectrometry data we could show that this putative secretion system is highly expressed in the infective extracellular stage of the complex life cycle of the *Acanthamoeba* endosymbiont *Amoebophilus*, which was described with fluorescence *in situ* hybridization, cryo and electron microscopy. The putative secretion apparatus might play an important role in the translocation of effector proteins to modulate host-cell interaction and might thus be important in host manipulation. The outcome of host manipulation of the wasp endosymbiont *Cardinium* is severe. Crosses between symbiont-infected males and uninfected females result in reproductive failure, which is increasing the relative fitness of infected females and thus leads to the spreading of the strictly maternally transmitted symbiont in the host population. The phenotype causing reproductive failure is called cytoplasmic incompatibility and is also found in the distantly related insect endosymbiont *Wolbachia*. With comparative genome analysis based on next generation sequencing platforms, we could show that in both cytoplasmic

incompatibility-inducing bacteria, cytoplasmic incompatibility is of independent evolutionary origin.

Profound effects in genome evolution of symbiotic bacteria are caused by mobile genetic elements such as phages, plasmids and transposable elements. A substantial part of this thesis is devoted to the analysis of the genome of *Amoebophilus*, which is overrun by transposable elements.

In conclusion, this thesis does not only increase the knowledge and understanding of evolution and host-cell interaction of two *Bacteroidetes* symbionts, it also provides a novel comparative context for understanding the mechanistic basis of cytoplasmic incompatibility and substantially increases our knowledge on reproductive manipulator symbionts that may serve as powerful tools in insect pest management.

Zusammenfassung

In der Natur gibt es vielfältige Assoziationen von Bakterien mit Insekten und einfachen Eukaryoten wie Protozoen. Der *Acanthamoeben* Endosymbiont *Amoebophilus asiaticus* sowie *Cardinium hertigii*, ein Bakterium, das die Reproduktion von Schlüpfwespen beeinflussen kann gehören zum relativ diversen Phylum der *Bacteroidetes*. Beide dieser Endosymbionten sind mit anderen gut charakterisierten Endosymbionten aus diesem Phylum nur entfernt verwandt und bilden gemeinsam eine Schwesterlinie. Zusätzlich haben die Genome von *Amoebophilus* und *Cardinium* viele Gemeinsamkeiten.

In dieser Arbeit konnte mittels vergleichender Genomanalyse gezeigt werden, dass *Amoebophilus* und *Cardinium* sehr reduzierte Genome haben und einen von einem Prophagen abstammenden Sekretionsapparat besitzen. Dieser Sekretionsapparat ist einem defekten Prophagen von dem Insektenpathogen *Serratia entomophila* von der genetischen Organisation her sehr ähnlich. Mittels qPCR und Massenspektrometrie konnte gezeigt werden, dass dieser Sekretionsapparat sehr stark während des infektiösen extrazellulären Stadiums im komplexen Lebenszyklus des *Acanthamoeben* Endosymbionts *Amoebophilus* exprimiert wird. Der Lebenszyklus von *Amoebophilus* wurde mittels Fluoreszenz in situ Hybridisierung, Cryo und Transmissions Elektronenmikroskopie charakterisiert. Der Sekretionsapparat spielt eine sehr entscheidende Rolle in der Translokation von Effektorproteinen, die die Wirts Symbionten Interaktion vermitteln und bei der Wirtsmanipulation wichtig sind. Die Auswirkung der Wirtsmanipulation des Wespensymbionts *Cardinium* ist schwerwiegend. Bei einer Kreuzung von Symbionten infizierten Wespenmännchen mit nicht infizierten Wespenweibchen gibt es keine Nachkommen. Dies erhöht die relative Fitness von infizierten Weibchen und führt zur Verbreitung dieser strikt mütterlich vererbten Symbionten in der Wirtspopulation. Der Phänotyp, der zu dieser Inkompatibilität in der Vermehrung des Wirtes führt wird cytoplasmatische Inkompatibilität genannt. Cytoplasmatische Inkompatibilität wird auch von

den entfernt verwandten *Wolbachien* ausgelöst. Mit vergleichender Genomanalyse unter der Verwendung von „next generation“ Sequenzierungs Plattformen konnte gezeigt werden, dass cytoplasmatische Inkompatibilität in beiden Bakterien unterschiedlichen evolutionären Ursprungs ist.

Starke Effekte auf die Genomevolution in Symbionten können mobile genetische Elemente wie Phagen, Plasmide oder Transposasen haben. Ein Teil dieser Arbeit beschäftigt sich mit der evolutionären Beschreibung des Genoms von *Amoebophilus*, das von transposablen genetischen Elementen übersät ist.

Zusammenfassend trägt diese Arbeit nicht nur zum besseren Verständnis der Evolution von Wirts Symbionten Interaktion zweier *Bacteroidetes* Symbionten bei, sondern liefert auch einen neuen Kontext für das bessere Verständnis von cytoplasmatischer Inkompatibilität und stellt neues Wissen über Bakterien, die die Reproduktion von Insektenschädlingen beeinflussen können, bereit.

Curriculum Vitae



Personal details

Name: Mag. rer. nat. Thomas Penz
 Date of birth: January 18, 1983
 Place of birth: Wolfsberg, Carinthia
 Nationality: Austrian

Education, compulsory service

May 2009 until February 2013	<u>PhD thesis in Genetics and Microbiology</u> at the Department of Microbial Ecology, University of Vienna, supervisor: Prof. Dr. Matthias Horn
October 2003 until March 2009	<u>Study of Genetics and Microbiology</u> , University of Vienna Graduation: Master of natural sciences (Mag. rer. nat.)
April 2008 until November 2008	<u>Diploma thesis</u> at the Department of Microbial Ecology, University of Vienna, supervisor Prof. Dr. Matthias Horn
January 2003 until August 2003	<u>Compulsory service</u> in the Goiginger casern in Bleiburg, Carinthia
June 2002	<u>High school diploma</u> at the secondary school (Bundesgymnasium und Bundesrealgymnasium) Völkermarkt, Carinthia

Career

Since March 2013	<u>Next Generation Sequencing Technologist</u> at the Biomedical Sequencing Facility of the Research Center for Molecular Medicine of the Austrian Academy of Sciences
May 2009 until February 2013	<u>Research assistant</u> at the Department of Microbial Ecology, University of Vienna

Scientific publications

Haferkamp I, **Penz T**, Geier M, Ast M, Mushak T, Horn M, Schmitz-Esser S. The endosymbiont *Amoebophilus asiaticus* encodes an S-adenosylmethionine carrier that compensates for its missing methylation cycle. *J Bacteriol.* 2013 Jul;195(14):3183-92. PMID: 23667233

Penz T, Schmitz-Esser S, Kelly SE, Cass BN, Müller A, Woyke T, Malfatti SA, Hunter MS, Horn M. Comparative genomics suggests an independent origin of cytoplasmic incompatibility in *Cardinium hertigii*. *PLoS Genet* 8(10): e1003012. PMID: 23133394

Toenshoff ER, **Penz T**, Narzt T, Collingro A, Schmitz-Esser S, Pfeiffer S, Klepal W, Wagner M, Weinmaier T, Rattei T, Horn M. Bacteriocyte-associated gammaproteobacterial symbionts of the *Adelges nordmannianae/piceae* complex (Hemiptera: Adelgidae). *ISME J.* 2012 Feb;6(2):384-96. PMID: 21833037

Collingro A, Tischler P, Weinmaier T, **Penz T**, Heinz E, Brunham RC, Read TD, Bavoil PM, Sachse K, Kahane S, Friedman MG, Rattei T, Myers GS, Horn M. Unity in variety-the pan-genome of the *Chlamydiae*. *Mol Biol Evol.* 2011 Dec;28(12):3253-70. PMID: 21690563

Schmitz-Esser S, **Penz T**, Spang A, Horn M. A bacterial genome in transition-an exceptional enrichment of IS elements but lack of evidence for recent transposition in the symbiont *Amoebophilus asiaticus*. *BMC Evol Biol.* 2011 Sep 26;11:270. PMID: 21943072

Penz T, Horn M, Schmitz-Esser S. The genome of the amoeba symbiont "*Candidatus Amoebophilus asiaticus*" encodes an *afp*-like prophage possibly used for protein secretion. *Virulence.* 2010 Nov-Dec;1(6):541-5. PMID: 21178499

Schmitz-Esser S, Haferkamp I, Knab S, **Penz T**, Ast M, Kohl C, Wagner M, Horn M. *Lawsonia intracellularis* contains a gene encoding a functional rickettsia-like ATP/ADP translocase for host exploitation. *J Bacteriol.* 2008 Sep;190(17):5746-52. PMID: 18606736

Presentations

Talks

Penz T, Schmitz-Esser S, Kelly SE, Cass BN, Müller A, Woyke T, Malfatti SA, Hunter MS, Horn M. *Cardinium hertigii* – a reproductive manipulator with a role in host nutrition. 14th International Symposium on Microbial Ecology (ISME) August 2012; Copenhagen, Denmark

Penz T, Spang A, Wagner M, Horn M, Schmitz-Esser S. Massive accumulation of IS elements in the evolutionary stable genome of the intracellular amoeba symbiont '*Candidatus Amoebophilus asiaticus*'. 4th Congress of European Microbiologists, Federation of European Microbiological Societies (FEMS) June 2011; Geneva, Switzerland

Penz T, Schmitz-Esser S, Tischler P, Arnold R, Rattei T, Montanaro J, Wagner M, Horn M. The genome of the amoeba symbiont '*Candidatus Amoebophilus asiaticus*' shows a unique enrichment of proteins harboring eukaryotic domains and a putative prophage possibly used for host cell interaction. 2nd Workshop on Symbiotic Interactions of the Association for General and Applied Microbiology (VAAM) October 2011; Würzburg, Germany

Poster

Penz T, Horn M, Schmitz-Esser S. The genome of the amoeba symbiont “*Candidatus Amoebophilus asiaticus*” encodes an afp-like prophage possibly used for protein secretion. 4th Congress of European Microbiologists, Federation of European Microbiological Societies (FEMS) June 2011; Geneva, Switzerland

Penz T, Spang A, Tischler P, Arnold R, Wagner M, Rattei T, Schmitz-Esser S, Horn M. Massive accumulation of IS elements in the genome of the obligate intracellular amoeba symbiont ‘*Candidatus Amoebophilus asiaticus*’. Annual Conference of the Association for General and Applied Microbiology (VAAM) March 2010; Hannover, Germany

Teaching

Supervisor of a student course: Fluorescence-in-situ-hybridization (FISH) – Identification of non-cultivable Microorganisms, Department of Microbial Ecology, University of Vienna, September 2012, February 2011 and July 2009

Supervisor of a diploma student: Agnes Harreither, Department of Microbial Ecology, University of Vienna, August 2011 until August 2012

Supervisor of a course: 1st Annual Symbiomics Meeting, Fluorescence-in-situ-hybridization (FISH) – Identification of non-cultivable Microorganisms, Department of Microbial Ecology, University of Vienna, February 2011

Supervisor of a diploma student: Han Fei-Tsao, Department of Microbial Ecology, University of Vienna, August 2009 until March 2010

Acknowledgments

I would like to thank all persons that made contributions to my PhD thesis.

Special thanks to my supervisor Matthias Horn for teaching me science and everything else besides science. Thanks for being such a great supervisor.

I'm grateful to my second supervisor Stephan Schmitz-Esser. Thanks for the funding, helpful discussions, advices and friendship during my PhD.

Thanks to my Diploma students Han-Fei Allen Tsao and Agnes Harreither for work that contributed a lot to this thesis.

Thanks to all my collaborators especially to Molly Hunter and Rok Kostanjsek for their interest on my projects.

Many thanks to my working colleagues Assi, Karin, Alex and Faris. Thanks for all the fun we had during the last years, for all the support and of course for becoming good friends. Eheheh frreindin...

Thanks to all other DOME members for creating an extraordinary nice and exciting working atmosphere.

Special thanks to my friends from CeMM for their support, especially to Michael Schuster for his helpful comments.

Many thanks to all my friends for having fun together.

Last but not least I want to thank my family Josefine and Harald Penz, Gabriele, Franz and Sabrina Lauer. Thanks for supporting me at any time as much as you could.