DISSERTATION

Titel der Dissertation

Novel intracellular bacteria
with unusual lifestyles and unique features

Verfasser

Frederik Schulz MSc

angestrebter akademischer Grad

Doctor of Philosophy (PhD)

Wien, 2015

Table of Contents

Chapter I

Introduction

Introduction

Symbiotic interactions are believed to have been the driving force in the emergence of complex life. For example, the origin of eukaryotes, which includes all animals, plants and fungi, most likely involved ancient symbiotic interaction between an archaeon and one or more bacteria (Koonin, 2010; Martijn and Ettema, 2013). Today, microbial symbioses are ubiquitous and of central importance for ecology and evolution of nearly all life forms on earth (McFall-Ngai et al., 2013). Bacteria can be found in close association with eukaryotes and appear either as beneficial microbes or as pathogens.

A good model system to study host-symbiont interactions are amoebae. These predatory protists control microbial communities in many different environments, such has soil, sea- and freshwater and anthropogenic habitats. It was suggested that amoebae have played a key role in the adaptation of bacteria to intracellular life in higher eukaryotes (Molmeret, 2005). Basic bacterial strategies for a successful infection of animals and humans, such as survival of phagocytosis, establishment of a niche for intracellular replication and host cell lysis, have supposedly been acquired wile thriving in amoebae (Molmeret, 2005).

Figure 1. Diversity of bacterial symbionts of amoebae. Shown is a maximum-likelihood phylogenetic tree based on the 16S rRNA gene. Bacteria in stable association with amoebae are highlighted in red, bacteria which transiently exploit amoebae in turquoise.
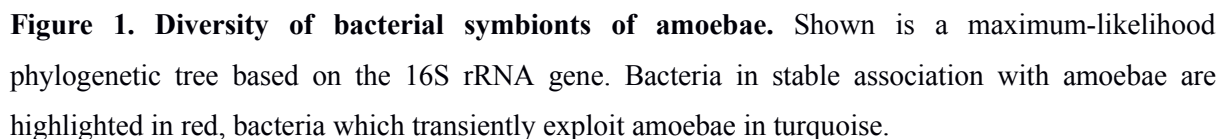
Obligate intracellular amoeba symbionts are ubiquitous, e.g. it has been suggested that about 25% of all acanthamoebae live in stable association with intracellular bacteria (Fritsche et al., 1993). Phylogenetically, most of these symbionts can be assigned to the *Alpha-* and *Betaproteobacteria*, the *Bacteroidetes* or the *Chlamydiae* (Schmitz-Esser et al., 2008) (Figure 1). However, the real diversity is much higher, e.g. just recently a novel bacterium belonging to the TM6 candidate phylum has been found in close association with its amoeba host (Pagnier et al., 2015). In addition, several pathogens of higher eukaryotes are able to thrive in amoebae, such as *Legionella*, *Franciscella*, *Coxiella*, *Burkholderia*, *Mycobacterium, Vibrio* and *Rickettsia* (Greub & Raoult, 2004, Ogata et al., 2006) (Figure 1). For these bacteria, amoebae may serve as a vehicle to reach higher eukaryotic hosts. One example are amoebae getting transmitted via aerosols, e.g. from biofilms in air conditioning systems or shower heads. Each amoeba potentially harbors hundreds of bacteria, which then can successfully establish infection in higher eukaryotic hosts. Furthermore, it has been shown that some pathogens, such as *Legionella*, express an increased virulence in macrophages after they have undergone replication in amoebae (Cirillo et al., 1995).

Figure 2. Niches for bacterial replication in a eukaryotic cell. Bacteria have been found thriving in vacuoles (blue), directly in their host's cytoplasm (yellow), in mitochondria (violet), in chloroplasts (green) and in the nucleus (red).

While the majority of intracellular bacteria thrive either in the cytoplasm or in cytoplasmic vacuoles, some have been found to occupy more unusual intracellular niches, such as chloroplasts, mitochondria and the eukaryotic nucleus (Wilcox, 1986, Schmidt, 1987, Sassera et al., 2006, Fujishima & Kodoma, 2012) (Figure 2). These bacteria likely have evolved unique genetic patterns facilitating colonization of these special compartments. However, our current knowledge about these enigmatic bacteria is very scarce.

In this thesis, I used amoebae as a model system to obtain novel insights into diversity, evolution and underlying mechanisms of host-symbiont interactions; with a particular focus on bacteria occupying unusual intracellular niches, such as the eukaryotic nucleus.

The review entitled "Intranuclear bacteria: inside the cellular control center of eukaryotes" (Chapter III) summarizes our current knowledge on bacteria exploiting this special intracellular niche and thus extends the introduction of this thesis. In "Life in an unusual intracellular niche: a bacterial symbiont infecting the nucleus of amoebae" (Chapter IV), I then present a concrete example of isolation and characterization of an intranuclear symbiont of amoebae, which we named "*Nucleicultrix amoebiphila*". In "What it takes to live inside the nucleus - An '-omics' perspective on the intranuclear lifestyle" (Chapter V) I further characterize molecular mechanisms underlying this unique host-symbiont relationship with the help of genomics and proteomics. Besides intranuclear bacteria (Chapter III, IV, V), I focus on a close relative of *Nucleicultrix*, which exclusively thrives in the cytoplasm of amoebae. A detailed analysis of the symbiont's life cycle, its remarkable genomic repertoire and its phylogenetic position in the *Rickettsiales* is comprised in "A Rickettsiales symbiont of amoebae with ancient features" (Chapter VI). To further deepen our current understanding of the diversity of intracellular bacteria, I describe the attempt to uncover new lineages of amoeba associated bacteria in "Marine amoebae with cytoplasmic and perinuclear symbionts deeply branching in the *Gammaproteobacteria*" (Chapter VII).

References

Cirillo, J.D., Falkow, S., and Tompkins, L.S. (1994). Growth of *Legionella pneumophila* in *Acanthamoeba castellanii* enhances invasion. Infect. Immun. *62*, 3254–3261.

Fritsche, T.R., Gautom, R.K., Seyedirashti, S., Bergeron, D.L., and Lindquist, T.D. (1993). Occurrence of bacterial endosymbionts in *Acanthamoeba* spp. isolated from corneal and environmental specimens and contact lenses. Journal of Clinical Microbiology *31*, 1122–1126.

Fujishima, M., and Kodama, Y. (2012). Endosymbionts in *Paramecium*. European Journal of Protistology *48*, 124–137.

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

Horn, M., Wagner, M., Santic, M., and Molmeret, M. (2005). Amoebae as training grounds for intracellular bacterial pathogens. Applied and Environmental Microbiology *71*, 20–28.

Koonin, E.V. (2010). The origin and early evolution of eukaryotes in the light of phylogenomics. Genome Biology *11*, 209–209.

Martijn, J., and Ettema, T.J.G. (2013). From archaeon to eukaryote: the evolutionary dark ages of the eukaryotic cell. Biochemical Society Transactions *41*, 451–457.

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

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

Ogata, H., La Scola, B., Audic, S., Renesto, P., Blanc, G., Robert, C., Fournier, P.-E., Claverie, J.-M., and Raoult, D. (2006). Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. PLoS Genet. *2*, e76.

Pagnier, I., Yutin, N., Croce, O., Makarova, K.S., Wolf, Y.I., Benamar, S., Raoult, D., Koonin, E.V., and Scola, B.L. (2015). *Babela massiliensis*, a representative of a widespread bacterial phylum with unusual adaptations to parasitism in amoebae. Biol Direct *10*, 1–17.

Sassera, D., Beninati, T., Bandi, C., Bouman, E.A.P., Sacchi, L., Fabbi, M., and Lo, N. (2006). "*Candidatus* Midichloria mitochondrii", an endosymbiont of the tick *Ixodes ricinus* with a unique intramitochondrial lifestyle. Int J Syst Evol Microbiol *56*, 2535–2540.

Schmidt, H., Goertz, HD, Quackenbush RL (1987). *Caedibacter caryophila* sp. nov., a killer symbiont inhabiting the macronucleus of *Paramecium caudatum*. International Journal of Systematic and Evolutionary Microbiology *37*, 459–462.

Schmitz-Esser, S., Toenshoff, E.E.R., Haider, S., Heinz, E., Hoenninger, V.M., Wagner, M., Horn, M., and Isolates, A. (2008). Diversity of bacterial endosymbionts of environmental acanthamoeba isolates. Applied and Environmental Microbiology *74*, 5822–5831.

Wilcox, L. (1986). Prokaryotic endosymbionts in the chloroplast stroma of the dinoflagellate *Woloszynskia pascheri*. Protoplasma *135*, 71–79.

Chapter II

Synopsis of the publications

Synopsis of the publications

Chapter III summarizes our current knowledge on intranuclear bacteria. By revisiting available literature originating from over 150 years of research, we could show that the intranuclear lifestyle is rather widespread phenomenon in the environment. We provided a comprehensive overview of the known diversity of intranuclear bacteria, which are affiliated with the *Alpha-* and *Gammaproteobacteria*, the *Chlamydiae* and the *Verrucomicrobia*, and their respective hosts, ranging from unicellular eukaryotes over arthropods and marine invertebrates to mammals. We summarized and analyzed evidences and hypotheses about infection processes and different strategies applied for the invasion of the nuclear compartment. We discussed possible advantages of exploiting this particular intracellular niche, and described evolutionary implications. We concluded the review by suggesting approaches that potentially allow to answer open questions and enable deeper insights into this unusual lifestyle.

Authors names: **Frederik Schulz**, Matthias Horn

Manuscript title: **Intranuclear bacteria: inside the cellular control center of eukaryotes**

Reference: Trends in Cell Biology 2015, 25 (6): 339-346

Contributions: Phylogenetic analysis was performed by FS. Microscopic images were taken by FS. All authors discussed and commented the results. The figures were prepared by FS. The draft manuscript was composed by FS and then edited by FS and MH.

The manuscript in Chapter IV reports on the discovery of a bacterial symbiont that specifically targets the nucleus of amoebae. We identified this bacterium as novel member of the *Alphaproteobacteria* that is only distantly related to known microbes. With the help of live cell imaging, fluorescence *in situ* hybridization and transmission electron microscopy we described the infection process of the symbiont, its impact on the amoeba host, and the mode of transmission. Furthermore, we provided evidence that the symbiont and its close relatives occurs world-wide in a variety of environments, yet mainly in freshwater and soil habitats. Our findings provide invaluable clues on the diversity of microbe-host interactions among protists. This unique symbiosis and the ability to maintain this association in the lab open up new avenues to explore the unusual intranuclear life, its evolution and underlying mechanisms.

Authors names: **Frederik Schulz**, Ilias Lagkouvardos, Florian Wascher, Karin Aistleitner, Matthias Horn

Manuscript title: **Life in an unusual intracellular niche: a bacterial symbiont infecting the nucleus of amoebae**

Contributions: Isolation and cultivation of the intranuclear symbiont together with its host, infection experiments, life-cell imaging and phylogenetic analyses were performed by FS. Growth experiments were conducted by FW. Transmission electron microscopy was performed by RK. Environmental distribution of the symbiont was analyzed by FS and IL. All authors discussed and commented the results. The figures were prepared by FS. The draft manuscript was composed by FS and then edited by FS and MH. MH initiated and supervised the project.

The study described in Chapter V provides detailed insights into the unique lifestyle of the obligate intranuclear amoeba symbiont *Nucleicultrix amoebiphila*. We analyzed its genome with a special emphasize on its unique genetic repertoire and genes involved in virulence and host-symbiont nutritional interplay. *In silico* protein structure prediction helped to uncover candidate proteins likely hijacking host intranuclear processes. With the help of phylogenomics we revisited the phylogenetic position of *Nucleicultrix* in the *Alphaproteobacteria*. Finally, we analysis of the proteomes of infected and uninfected amoeba nuclei to confirm the presence of predicted virulence factors and to assess the host response to an infection by intranuclear bacteria.

Authors names:	**Frederik Schulz**, Florian Wascher, Joran Martijn, Thijs Ettema, Dörte Becher, Stephanie Markert, Thomas Schweder, Thomas Weinmaier, Thomas Rattei, Matthias Horn

Manuscript title: **What it takes to live inside the nucleus - An '-omics' perspective on the intranuclear lifestyle**

Reference: *Draft manuscript*

Contributions: Isolation and cultivation of the intranuclear symbiont together with its host, genome sequencing, genome annotation and comparative analysis, analysis of the genomic island, the type VI secretion system, putative effectors and proteomics datasets were conducted by FS. Genome assembly was performed by FS and TW. Phylogenetic analysis of the *Alphaproteobacteria* was performed by FS and JM. Samples for proteomics were prepared by FW and FS, proteomics measurements were undertaken by DB, SM and TS. All figures were prepared by FS. The draft manuscript was composed by FS. MH supervised the project.

Chapter VI contains a manuscript in which we describe in detail the infection process of a rickettsial symbiont of amoebae, which is only distantly related to known members of the *Rickettsiales*. We provide evidence on diversity and environmental distribution of the symbiont. Comparative analysis of the symbiont's genome with other members of the *Rickettsiales* and free-living *Alphaproteobacteria* revealed several unexpected features, including an atypical genome organization, the presence of multiple type IV secretion systems alongside with a uniquely large number of potential effector proteins, and a nearly complete set of flagellar genes. These findings and our experimental observations allowed us to propose a model of the interaction of the symbiont with its eukaryotic host. The study represents the first report comprehensively describing infection process and genetic repertoire of an amoeba-associated, deeply branching member of the *Rickettsiales*.

Authors names: **Frederik Schulz**, Joran Martijn, Florian Wascher, Rok Kostanjsek, Thijs Ettema, Matthias Horn

Manuscript title: **A *Rickettsiales* symbiont of amoebae with ancient features**

Contributions: Cultivation of the intranuclear symbiont together with its host and infection experiments were performed by FS and FW. Genome sequencing, assembly and phylogenetic analysis of the *Alphaproteobacteria* were undertaken by JM and TE. Genome annotation, comparative genome analysis and phylogenetic analyses of flagellar genes and type IV secretion systems were performed by FS. Transmission electron microscopy was performed by RK. Environmental distribution of the symbiont was analyzed by FS and IL. The figures were prepared by FS. The draft manuscript was composed by FS and then edited by FS and MH. MH initiated and supervised the project.

The manuscript in Chapter VII describes the discovery of two novel marine amoeba isolates harboring bacterial endosymbionts. Based on morphological and phylogenetic criteria we showed that while one of the isolated amoeba strains is a representative of a clade of marine amoebae, the other represents a completely novel lineage in the Amoebozoa. We identified these bacterial endosymbionts as *Gammaproteobacteria* that are distantly related to the *Coxiellaceae*. We characterized both symbiotic associations and showed that one symbiont thrives in the amoeba nucleus, more specifically between inner and outer nuclear membrane. Taken together, this is the first report on bacterial symbionts in marine amoebae and on the perinuclear space as niche for obligate intranuclear bacteria.

Authors names: **Frederik Schulz***, Tomáš Tyml*, Ilaria Pizzetti, Iva Dyková, Stefano Fazi, Martin Kostka, Matthias Horn
*contributed equally

Manuscript title: **Marine amoebae with cytoplasmic and perinuclear symbionts deeply branching in the *Gammaproteobacteria***

Contributions: IP, SF, TT performed the sampling. FS, TT, IP and MK conducted the experiments; FS, TT and MK analyzed the results. All authors discussed and commented the results. FS and TT prepared figures. The draft manuscript was composed by FS and TT and then edited by FS and MH. All authors reviewed the manuscript.

Chapter III

Intranuclear bacteria: inside the cellular control center of eukaryotes

CellPress

Intranuclear bacteria: inside the cellular control center of eukaryotes

Frederik Schulz and Matthias Horn

Department of Microbiology and Ecosystem Science, University of Vienna, Althanstraße 14, 1090 Vienna, Austria

Intracellular bacteria including major pathogens live in the cytoplasm or in cytoplasmic vacuoles within their host cell. However, some can invade more unusual intracellular niches such as the eukaryotic nucleus. Phylogenetically diverse intranuclear bacteria have been discovered in various protist, arthropod, marine invertebrate, and mammalian hosts. Although targeting the same cellular compartment, they have apparently developed fundamentally-different infection strategies. The nucleus provides a rich pool of nutrients and protection against host cytoplasmic defense mechanisms; intranuclear bacteria can directly manipulate the host by interfering with nuclear processes. The impact on their host cells ranges from stable associations with a neutral or beneficial effect on host fitness to rapid host lysis. The analysis of the intranuclear lifestyle will extend our current framework for understanding host–pathogen interactions.

Bacteria in eukaryotic cells

Bacteria living within eukaryotic cells are ubiquitous. As beneficial symbionts or major pathogens they are integral to the biology, ecology, and evolution of their animal and human hosts. Intracellular bacteria generally reside directly in the host cytoplasm or in host-derived vacuoles [1]. To successfully thrive in vacuolar compartments, bacteria frequently hijack the host endocytic and secretory pathway, recruit host proteins and ribosomes to the vacuolar membrane, and ultimately structure a specialized organelle for their replication [2]. Only rarely have bacteria been observed directly within other intracellular compartments such as the endoplasmic reticulum, Golgi-related vesicles, plastids, mitochondria, or the nucleus [3–8]. Bacteria in the latter compartment are referred to as intranuclear (or endonuclear) bacteria, and currently represent the largest group of microbes targeting unusual subcellular niches.

The first evidence for intranuclear bacteria reaches back to the 19th century when enigmatic particles were observed in the nucleus of paramecia (Box 1). Today, intranuclear bacteria are known to occur in various protists, insects, marine invertebrates, and mammals. While most remain elusive, some have been described in more detail by light and electron microscopy. Information on their phylogeny and distribution, the infection process, and their impact on host cells is available for very few intranuclear bacteria. The advent of cultivation-independent methods for the identification and functional analysis of microbes in conjunction with current 'omics' technologies provides a new momentum, facilitating the characterization of intranuclear bacteria.

This review, for the first time, summarizes major findings on intranuclear bacteria obtained during the past 150 years. We first recapitulate current knowledge about the diversity of intranuclear bacteria and their respective hosts. We then provide an overview about the evidence and hypotheses regarding the infection process and invasion of the nuclear compartment. We then discuss possible advantages of exploiting this particular intracellular niche, and we end with brief evolutionary considerations.

Intranuclear bacteria in protozoa

The majority of known intranuclear bacteria are associated with protists (Figure 1). One of the most thoroughly studied and best-understood intranuclear bacteria are *Holospora* spp., which infect paramecia and are affiliated with rickettsiae (Alphaproteobacteria). In the late 19th century, three *Holospora* spp. were described and distinguished based on their characteristic morphology and their intracellular location, either in the micro- or the macronucleus [9]. The infection process of *Holospora* has been studied extensively and reviewed recently [10]. *Holospora* spp. show two distinct developmental stages: an infectious form up to 20 μm in length and a short reproductive morphotype which forms in the nucleus. Most *Holospora* spp. thrive inside the macronucleus of paramecia where up to 100% of a host population can be infected, despite a negative impact on host growth rate (reviewed in [11]). *Holospora* spp. exploiting the micronucleus are less prevalent (about 10% in the host population) because they impair ciliate sexual reproduction. The association of *Holospora* with paramecia can be stable over long time-periods, but starvation may trigger differentiation back into the infective form, which then escapes from the nuclear compartment and eventually causes host cell lysis (reviewed in [11]). Interestingly, infection with a *Holospora* sp. does not necessarily represent a burden for the *Paramecium* host but may impose a selective advantage. The presence of the symbiont enhances expression of host heat-shock proteins, resulting in increased host survival rates in the face of adverse environmental conditions such as altered salinity or temperature shifts [12–14].

CrossMark

Box 1. The discovery of intranuclear bacteria

In the middle of the 19th century, rod-shaped particles were observed for the first time in the nuclei of ciliates [79]. At that time it was believed that the nucleus played an important role in 'embryo formation', and the intranuclear structures were proposed to represent spermatozoa produced in this compartment [79–81]. In the following years, similar findings were reported several times; it was noted that the particles were immobile, and the hypothesis arose that these intranuclear rods are in fact bacterial parasites [82–84]. In the early 20th century, coccoid intranuclear parasites in the flagellate *Euglena* [31] were reported that caused hypertrophy of the nucleus, discoloration of plastids, and inhibition of host cell division. Morphologically-similar bacteria were subsequently found in *Trichonympha* flagellates isolated from termites and were described as *Caryococcus* spp. [32]. Bacteria in the nuclei of multicellular eukaryotes were first noted during the analysis of rickettsiae as the causative agent of Rocky Mountain spotted fever. In addition to rod-shaped bacteria being found in the cytoplasm of tick and guinea pig cells, these bacteria were also occasionally found inside the nuclear compartment [85,86].

A second lineage of Alphaproteobacteria, represented by *Caedibacter caryophilus* and *Caedibacter macronucleorum*, is able to thrive in the nuclei of paramecia [15–17]. These bacteria were observed for the first time in the early 20th century. They are mainly transmitted vertically upon cell division of the *Paramecium* host and were originally described based on an outstanding feature, the presence of characteristic phage-like structures, so-called reticulate bodies (R-bodies; reviewed in [18]). These R-bodies mediate the 'killer trait' of *Paramecium* hosts, which serves as a defense mechanism to outcompete related symbiont-free paramecia ([19], reviewed in [18]). Two different *Caedibacter* phenotypes can be distinguished by the infective form that expresses R-bodies and the reproductive form that lacks R-bodies. Infective forms are released from symbiont-carrying hosts and can be taken up by symbiont-free paramecia through phagocytosis. Upon acidification of the phagosome, the R-body unrolls, elongates, penetrates
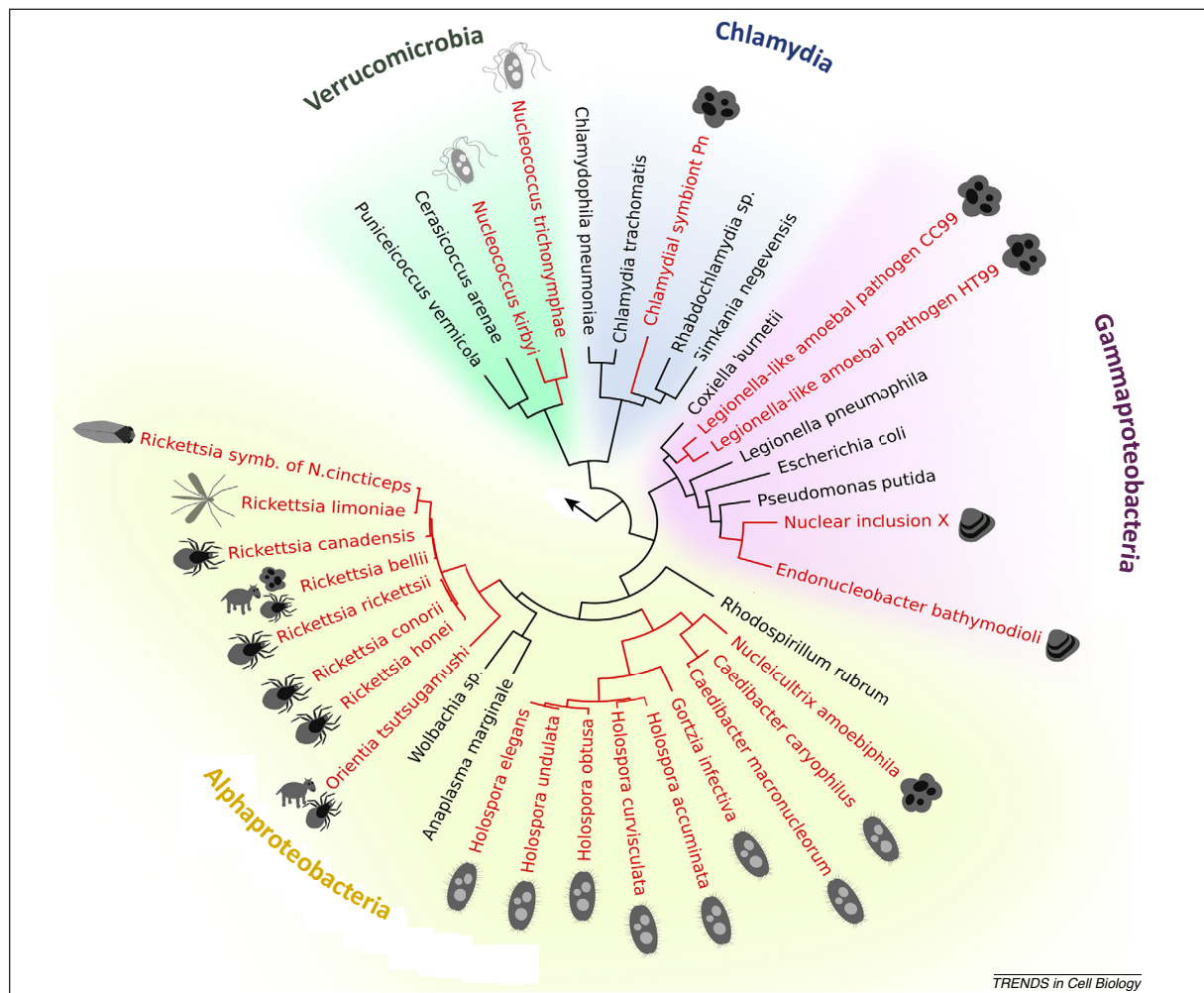
Figure 1. Phylogenetic relationships of known intranuclear bacteria. Microbes exploiting the nucleus of their hosts (shown in red) can be found in four major clades; the Alpha- and Gammaproteobacteria, the Chlamydiae, and the Verrucomicrobia. The respective host organisms are schematically indicated. A maximum-likelihood tree calculated with FastTree2 ([87] GTR+CAT approximation) based on a 16S rRNA alignment performed in Arb is shown [88].

the vacuolar membrane, and potentially delivers an as-yet unidentified toxin into the host cytoplasm, resulting in death of the *Paramecium* host cell [20].

Dinoflagellates may also serve as hosts for intranuclear bacteria [21,22]. To date only the betaproteobacterial symbionts of *Gyrodinium instriatum* have been studied to some extent [22]. Ultrastructural analysis revealed that the symbionts are distributed throughout the nucleoplasm, but mainly concentrate in dense clusters adjacent to the nuclear envelope, while bacteria are only rarely observed in direct contact with chromosomes. No negative effect on the dinoflagellate host was observed. Single bacteria are released from the nuclear compartment into the cytoplasm, and might serve as a food source for starving host cells [21,22].

Bacteria inside the nucleus have also been found in euglenoids [23–25]. A more detailed analysis is available for the yet unclassified intranuclear bacteria in *Euglena hemichromata* [26]. The bacteria are dispersed in the entire nucleoplasm but preferentially locate in close proximity to the nucleolus. At later stages of infection, host chromatin is reduced, and host cells are ultimately lysed. Although only a low percentage of *Euglena* populations is generally infected, these intranuclear parasites might have a strong impact on population dynamics and algal blooms [26].

Bacteria colonizing the nucleus have been observed multiple times in flagellates isolated from termite guts [27,28]. A recent study described the parasitic behavior of as-yet unidentified intranuclear bacteria in *Staurojoenina assimilis* [29]. The symbionts were seen in direct contact with host chromatin, and a pronounced structural change of the host nuclear envelope was observed. The infection led to highly enlarged nuclei (nuclear hypertrophy), eventually resulting in its destruction and bacterial release. The intranuclear bacteria of the termite gut flagellate *Trichonympha agilis* were recently identified as novel Verrucomicrobia, a group of predominantly free-living bacteria found in freshwater and soil environments. '*Candidatus* Nucleococcus trichonymphae' and '*Candidatus* Nucleococcus kirbyi' (Figure 1) did not cause nuclear hypertrophy or rupture of the host nucleus [30]. They likely represent the genus *Caryococcus* that had been described a century ago (Box 1) [31,32]. Verrucomicrobial symbionts in the nucleus seem to be more widespread among termite-associated flagellates as indicated by the detection of intranuclear bacteria in different hosts with a Verrucomicrobia-specific fluorescence *in situ* hybridization assay [30]. Of note, coinfection of the nucleus of flagellates with a verrucomicrobial and another bacterial symbiont has been observed [30,33].

Intranuclear bacteria have also been found in free-living amoebae, for the first time in *Naegleria clarki* [34,35]. The bacteria has a coccoid morphology and mainly surrounds the nucleolus (Figure 2). These bacteria, tentatively referred to as strain 'Pn', are members of Chlamydiae, a group of obligate intracellular bacteria comprising well-known pathogens as well as ubiquitous symbionts [36]. Another study reported on the identification of two bacteria, strains HT99 and CC99, which can infect and thrive in *Acanthamoeba polyphaga* [37]. These bacteria are Gammaproteobacteria, which are moderately related to
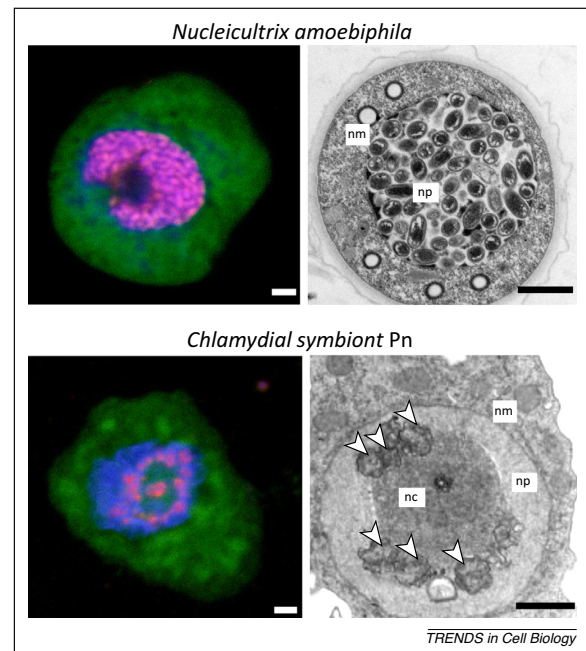

Figure 2. Intranuclear bacteria in Amoebae. Fluorescence *in situ* hybridization images (left) and transmission electron microscopy (TEM, right) of *Nucleicultrix amoebiphila* and the chlamydial symbiont Pn (arrowheads) reveal different locations inside the nuclei of free-living amoebae (bacteria in pink, nucleoplasm in blue, amoeba cytoplasm in grey or green). Whereas *Nucleicultrix amoebiphila* takes up large parts of the nucleoplasm and is closely associated with host heterochromatin (dark, granular material), Pn is arranged around the nucleolus. The TEM image of Pn was modified from [44]. Scale bars indicate 1 μm. Abbreviations: nc, nucleolus; nm, nuclear membrane; np, nucleoplasm.

Legionella and *Coxiella* spp. Recently a symbiont in the nucleus of yet another amoeba strain, *Hartmannella* sp., was discovered [38]. The rod-shaped bacterium, named '*Candidatus* Nucleicultrix amoebiphila', is related to *Holospora* and *Caedibacter*, and a member of a clade of protist symbionts in the Alphaproteobacteria (Figures 1 and 2). Infection with this symbiont does not appear to negatively influence its native *Hartmannella* sp. host, whereas host cell lysis is observed upon infection of *Acanthamoeba* spp. [38].

Intranuclear bacteria in arthropods
Several members of the Rickettsiaceae are able to enter the nuclei of their arthropod host cells, namely the tick-associated bacterial pathogens *Rickettsia rickettsii*, *Rickettsia monacensis*, *Rickettsia honei*, and *Rickettsia canada* [39–42]. However, invasion of the nucleus seems to be a rare event because these bacteria generally locate in the host cytoplasm. Rickettsiae have also been found in the nuclei of Malpighian tubuli or gut epithelial cells of the booklouse *Liposcelis bostrychophila*, the moss bug *Xenophyes cascus*, and the leafhopper *Nephotettix cincticeps* [43–45]. Remarkably, the intranuclear bacteria of *N. cincticeps* are transmitted paternally (i.e., from one generation to the next) via the nuclei of sperm cells [45]. These symbionts are also able to infect the nuclear compartment of a mosquito and a silkworm cell line, indicating its potential to exploit various other insect hosts [45].

There are two additional reports about intranuclear bacteria in arthropods; however, the identity of the symbionts remains elusive. One study described bacteria in ovarian cells of two termite spp. [46]. Unlike most intranuclear bacteria, they were surrounded by a vacuolar membrane. Even though their presence caused a reduction in chromatin density, and led to the formation of granular material in the nucleus, no negative impact on meiotic processes or host cell integrity was seen [46]. Intranuclear bacteria enclosed in vacuoles inside the host nucleus were also found in the planthopper *Pentastiridius leporinus* [47].

Intranuclear bacteria in marine invertebrates

Intranuclear symbionts have been found in several clams and bathymodiolin mussels [48–51]. Originally described as rickettsia-like 'nuclear inclusion X', the intranuclear bacteria found in the pacific razor clam *Siliqua patula* were identified as Gammaproteobacteria most closely related to *Pseudomonas* spp. [48,52,53]. Infection with these bacteria causes nuclear hypertrophy; they are considered to be serious pathogens and have been associated with extensive declines in host populations [48]. 'Nuclear inclusion X' belongs to a clade of symbionts of marine invertebrates, which also includes '*Candidatus* Endonucleobacter bathymodioli' described recently as an intranuclear parasite of bathymodiolin mussels [51]. These bacteria show a distinct developmental cycle with a short rod-shaped stage infecting the nuclear compartment, enlarging to long filaments, and dividing by transverse binary fission (Figure 3). Infection is accompanied by a reduction of chromatin and enlargement of the nuclei, and eventually leads to host cell lysis [51]. Interestingly, gill tissue cells harboring intranuclear bacteria are free of the thiotrophic and methanotrophic cytoplasmic symbionts that are typically present in these tissues. Intranuclear bacteria have also been observed in sponges but have not yet been further investigated [54].

Potential to infect mammalian cells

Several lines of evidence suggest that some intranuclear bacteria can also infect mammalian cells. *Rickettsia bellii*, *Rickettsia conorii*, and *Rickettsia rickettsia* have been found to occasionally invade the nucleus *in vitro* in cell culture experiments [39,55]. The rickettsia-related organism *Orientia tsutsugamushi* was detected within the nuclei of human liver cells and was observed *in vitro* in corresponding cell lines [56,57]. However, an intranuclear location has only been observed in a fraction of host cells. It has thus been suggested that rickettsiae might accidentally enter the nucleus. These bacteria induce host actin nucleation and polymerization, and propel through the host cytoplasm by the force generated in this process [58]. When they accidentally penetrate the nuclear envelope, the different nature of actin in this compartment may lead to loss of bacterial motility and restrict its exit [59]. The bacteria then start replicating, which ultimately leads to the enlargement and bursting of the nucleus [60]. Additional evidence for the ability of intranuclear bacteria to infect mammalian cells is available for the amoeba symbiont C99, which is capable of invading nuclei of human macrophages and HeLa cells, resulting in host cell lysis [37].

Infection process and life cycle of intranuclear bacteria

Before the nucleus can be exploited for replication, bacteria must enter the host cell, withstand cytoplasmic defense mechanisms, migrate to the nuclear compartment and cross the nuclear envelope, which is fundamentally different from the cytoplasmic membrane and represents a major boundary. It is composed of a double-membrane: the outer nuclear membrane, which is continuous with the endoplasmic reticulum, and the inner nuclear membrane, which plays an important role in maintaining the structural integrity of the nucleoplasm. Both membranes are fused at insertion sites of nuclear pore complexes; these large protein channels structure the nuclear envelope as a
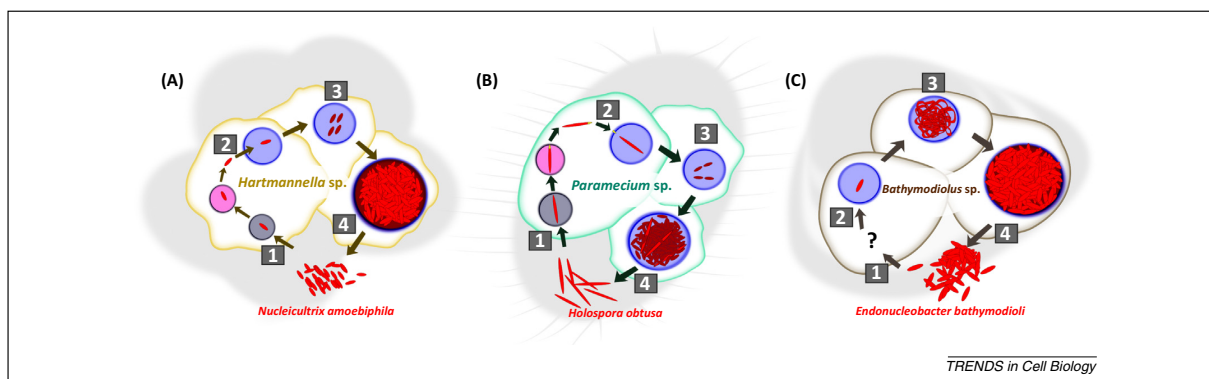

Figure 3. Infection processes of intranuclear bacteria. **(A)** The amoeba symbiont *Nucleicultrix amoebiphila* is (1) taken up by phagocytosis, followed by (2) acidification of the phagosome, escape into the cytoplasm, and ultimately invasion of the nuclear compartment, possibly during host cell division. The symbiont starts replicating in the nucleoplasm (3) which may lead to hypertrophy and bursting of the nucleus (4), host lysis, and bacterial release. **(B)** The long infective form of the *Paramecium* symbiont *Holospora* sp. (1) is taken up by phagocytosis, escapes the phagolysosome, traffics to the nucleus, and penetrates the nuclear envelope with the help of a macromolecular structure known as the invasion tip (2). Within the nucleus, the symbionts differentiate from an elongated morphotype to a short replicative form (3). Adverse environmental conditions can trigger the re-differentiation into the infective form, and ultimately to enlargement of the nuclear compartment and eventual release of bacteria after host cell lysis (4). **(C)** The infection process of the mussel symbiont *Endonucleobacter bathymodioli*. How these bacteria enter the host cell and invade the nuclear compartment remains elusive (1,2). However, *E. bathymodioli* undergoes a developmental cycle. Once inside the nucleus (2) the short infective form elongates to up to 20 μm in length and divides by longitudinal fission (3). This is followed by repeated binary fissions at later stages, which lead to the presence of a large number of short infective forms in a hypertrophic nucleus (4) and the destruction of the host cell together with the release of bacterial aggregates that eventually disintegrate.

diffusion barrier and control the bidirectional exchange of substrates between the nucleus and the cytoplasm [61,62]. To overcome this barrier, intranuclear bacteria have to apply sophisticated strategies.

To date, the infection processes and life cycles of intranuclear bacteria have been described in some detail for *E. bathimodioli*, *N. amoebiphila*, *Holospora* spp., and for bacteria occupying the nucleus of *Euglena* (Figure 3) [10,26,38,51]. *E. bathimodioli* shows a pronounced developmental cycle involving morphologically very distinct stages (Figure 3), but information on its uptake into the host cell and invasion of the nucleus is lacking [51]. *N. amoebiphila* seems to exit the phagolysosome and presumably invades the nuclear compartment upon the breakdown of the nuclear envelope during open mitosis of its amoeba host (Figures 3 and 4) [38]. *Holospora* spp. escape the acidified phagosome and uses their so-called 'invasion tip' to further infect (Figure 3) [63,64]. This macromolecular structure has a length of about 1 µm and is coated with a 89 kDa protein containing two actin-binding motifs, one coiled-coil and a cadherin domain [10]. It was hypothesized that this protein induces polymerization of host actin, allowing *Holospora* to move through the host cytoplasm [64]. Furthermore, the 89 kDa protein facilitates binding to the nuclear membrane and entry of the bacterium, but is left behind in the course of this process (Figure 4) [10].

An alternative strategy to invade the nucleus has been described for intranuclear bacteria in *Euglena*. The symbiont is taken up by phagocytosis and survives

phagosome–lysosome fusion. The vacuole is then translocated to the nucleus, where its membrane fuses with the outer nuclear membrane. The bacterium is released into the perinuclear space and presumably enters the nucleoplasm by invagination of the inner nuclear membrane (Figure 4) [26]. Once inside the nucleus, all four symbionts initiate replication, which may ultimately lead to nuclear hypertrophy and eventually host cell lysis.

The nucleus as a niche
Life inside the nucleus has many potential advantages. Its invasion might represent a strategy to escape host cellular defense mechanisms, providing a long-term and safe niche where bacteria are protected from cytoplasmic processes such as autophagy [65].

It seems plausible that the close proximity to host chromosomes facilitates delivery of bacterial effector molecules that can manipulate the host cell in favor of bacterial replication by directly targeting nuclear processes. Such effectors, also called nucleomodulins, are increasingly recognized in intracellular bacterial pathogens such as *Listeria*, *Mycobacterium*, *Shigella*, and *Helicobacter* [8]. These bacteria manipulate chromatin and epigenetic machineries of their host cells [66] either indirectly through cytoplasmic signaling cascades [67,68] or directly in the nucleus by nucleomodulins mediating histone modifications [69,70]. Intranuclear bacteria are likely to use the 'direct' strategy. Similarly to known nucleomodulins, their effectors could act as host chromatin regulators or eukaryotic transcription factors [8], thereby controlling host gene expression, DNA replication and repair, cell cycle regulation, and substrate exchange with the cytoplasm. However, so far only the manipulation of host cell transcription has been described for intranuclear bacteria. The presence of *Holospora obtusa* in the nucleus alters the expression of host genes involved in stress resistance, intracellular signaling, transcription, and aerobic metabolism [12,71].

Replication in the nucleus might also provide a nutritional advantage. A rich pool of proteins (up to 90%), small ribonucleoproteins, nucleic acids (up to 30%), and nucleoside triphosphates are readily available to be utilized by the bacterial symbiont. To facilitate the uptake of nucleotides, intranuclear rickettsiae, *C. caryophilus*, and *Holospora* spp. encode one or more nucleotide transporters, which are well known from other intracellular bacteria thriving as energy parasites in vacuoles or the cytoplasm [72,73].

A reduction of the host heterochromatin has been observed for several intranuclear bacteria, which might indicate that host chromosomes can serve as a substrate [74]. Such a scenario would rely on the secretion of exoenzymes able to digest high molecular weight DNA and the presence of specific transport systems encoded by competence genes facilitating the import of host DNA fragments into the bacterial cell [75,76]. It might be assumed that degradation of host genetic material would be severely detrimental for the host cell, but observations of several intranuclear bacteria suggest they do not confer an apparent fitness decrease on their hosts [22,38,46]. A strategy to avoid damaging host chromosomes could be the uptake of
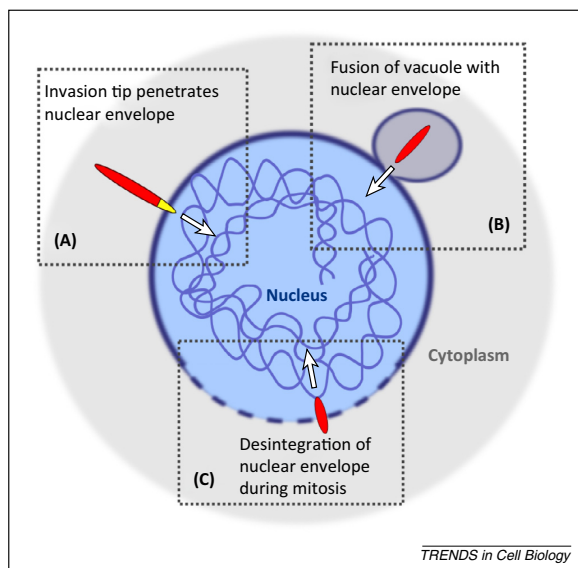
Figure 4. Invading the nucleus. Different strategies of how intranuclear bacteria enter the nucleus have been proposed. (A) *Holospora* spp. employ a specialized macromolecular structure known as the invasion tip. This structure includes an 89 kDa protein with actin-binding activity, which is secreted upon contact with the vacuolar membrane and facilitates escape into the cytoplasm and penetration of the nuclear envelope. (B) Intranuclear symbionts of *Euglena* spp. possibly enter the nucleus by fusion of bacteria-containing host-derived vacuoles with the outer nuclear membrane followed by invagination of the inner nuclear membrane and release into the nucleoplasm. (C) *Nucleicultrix amoebiphila* may use the disintegration of the nuclear envelope taking place during open mitosis to permit invasion of the nucleus.

other nucleic acids such as mRNAs or intronic RNAs after splicing, but this would have major consequences on host gene expression and regulation.

An alternative explanation for the observed reduction of heterochromatin in the nucleus could be its transition to the electron-lucent euchromatin because intranuclear bacteria potentially affect host chromatin structure by interfering with DNA methylation or by altering chromatin marks. Several histone modifications, such as acetylation, methylation, and phosphorylation, result in a change in the compaction of chromatin fibers [66]; however, the manipulation of these processes would severely affect host gene expression and genome stability.

The localization of bacteria in the nuclear compartment has great potential for horizontal gene transfer and recombination between both partners. Indeed, the genomes of amoebae, such as *Acanthamoeba castellanii*, have been shaped to a large extent by the acquisition of bacterial genes [77], and a role for intranuclear bacteria in this process is conceivable. Although systematic analyses are currently missing, such a scenario is supported by the detection of 16S rRNA gene fragments of 'Candidatus Nucleococcus' symbionts in the genome of their flagellate host [30].

Independent evolutionary origin of an intranuclear life style

Intranuclear bacteria are manifold and diverse. (i) They infect a wide range of eukaryotic hosts. (ii) They use different strategies for infection and invasion of the nucleus (Figure 4). (iii) They show different life cycles that sometimes involve unusual developmental stages (Figure 3). (iv) The effects on their host cells differ, ranging from parasitism to mutualism, and (v) they are affiliated with five different bacterial taxonomic clades (Figure 1). Each of these phylogenetic groups comprises other symbionts of protists or higher eukaryotes. However, the majority of those bacteria are located in their host cell cytoplasm, and an intranuclear lifestyle seems to be a comparably rare phenomenon. Taken together, these findings are strongly indicative of an independent, convergent evolution of the ability to use the nuclear compartment as a niche. Of note, most of the intranuclear bacteria identified so far form deep branches in phylogenetic trees, suggesting that adaptation to an intranuclear lifestyle occurred a long time ago. It seems most parsimonious that the ability to invade the nucleus is a secondary adaptation of bacteria that had already developed an intracellular lifestyle. This perception is currently supported by the phylogeny of chlamydial and alphaproteobacterial symbionts (Figure 1).

Concluding remarks

Despite recent advances, many intranuclear bacteria have not yet been clearly identified and have not yet been the subject of targeted research efforts. We are still far from understanding the basis of this remarkable lifestyle (Box 2). For example, we cannot explain how intranuclear bacteria thrive inside the nucleus while maintaining nuclear function and cellular integrity. It also remains elusive how the intranuclear bacteria acquire nutrients; targeted labeling of host nuclear compounds (e.g., using stable isotopes) could

---

Box 2. Outstanding questions

- Successful invasion of the nucleus requires a complex infection process; in particular, the nuclear envelope represents a major barrier. What are the molecular mechanisms underlying trafficking of intranuclear bacteria through the host cytoplasm, and recognition and invasion of the nuclear compartment?
- What are the key effector proteins used by intranuclear bacteria to manipulate host nuclear processes? Are they similar to known nucleomodulins found in major intracellular pathogens?
- Different patterns of colonization of the nucleus have been observed. Further, bacteria are distributed throughout the nucleoplasm or in association with either the nuclear envelope or the nucleolus. Therefore, what are the implications of these findings for the host and symbiont?
- Although some intranuclear bacteria promote host cell lysis, others can stably thrive in the host nucleus for prolonged time-periods. How do the latter intranuclear bacteria establish their niche without severely damaging host cell integrity or strongly affecting host cell physiology? How does the host cell cope with an intranuclear infection?
- The nucleus can be considered a pantry, rich in proteins, small ribonucleoproteins, nucleic acids, and nucleoside triphosphates; therefore, what molecules are utilized by intranuclear bacteria?
- Evolution of the intranuclear lifestyle has occurred multiple times independently. Are there distinct mechanisms and genetic patterns that are shared among all intranuclear bacteria?

---

help to elucidate which substrates are utilized by intranuclear bacteria. One of the major points for future research will be to reveal factors and molecular mechanisms that allow invasion of the nuclear compartment and facilitate manipulation of host nuclear processes directly in the cellular control center. This will be challenging because all known intranuclear bacteria are obligate intracellular pathogens and – in contrast to most pathogenic model organisms – cannot be maintained in host-free media. In addition, there are no available genetic tools for these organisms. Nevertheless, an important first step towards deciphering the molecular basis of the intranuclear life style was the recent sequencing of the genomes of three *Holospora* spp. [78]. The availability of genome sequences of intranuclear bacteria of different evolutionary lineages and their hosts, combined with functional studies involving transcriptomic, proteomic, and metabolomic approaches will contribute significantly to our understanding of this unique lifestyle.

References

1 Ray, K. and Marteyn, B. (2009) Life on the inside: the intracellular lifestyle of cytosolic bacteria. *Nat. Rev. Microbiol.* 7, 333–340
2 Canton, J. and Kima, P.E. (2012) Interactions of pathogen-containing compartments with the secretory pathway. *Cell. Microbiol.* 14, 1676–1686
3 Wilcox, L. (1986) Prokaryotic endosymbionts in the chloroplast stroma of the dinoflagellate *Woloszynskia pascheri*. *Protoplasma* 135, 71–79
4 Vogt, G. (1992) Enclosure of bacteria by the rough endoplasmic reticulum of shrimp hepatopancreas cells. *Protoplasma* 171, 89–96
5 Schmid, A.M.M. (2003) Endobacteria in the diatom *Pinnularia* (Bacillariophyceae). I. 'Scattered ct-nucleoids' explained: DAPI–DNA

complexes stem from exoplastidial bacteria boring into the chloroplasts. *J. Phycol.* 39, 122–138

6 Sassera, D. *et al.* (2006) '*Candidatus Midichloria mitochondrii*', an endosymbiont of the tick *Ixodes ricinus* with a unique intramitochondrial lifestyle. *Int. J. Syst. Evol. Microbiol.* 56, 2535–2540

7 Cho, K-O. *et al.* (2011) *Wolbachia* bacteria reside in host Golgi-related vesicles whose position is regulated by polarity proteins. *PLoS ONE* 6, e22703

8 Bierne, H. *et al.* (2012) When bacteria target the nucleus: the emerging family of nucleomodulins. *Cell. Microbiol.* 14, 622–633

9 Hafkine, M.W. (1890) Maladies infectieuses des paramécies. *Ann Inst Pasteur* 4, 148–162

10 Fujishima, M. and Kodama, Y. (2012) Endosymbionts in *Paramecium*. *Eur. J. Protistol.* 48, 124–137

11 Görtz, H.D. (2001) Intracellular bacteria in ciliates. *Int. Microbiol.* 4, 143–150

12 Hori, M. and Fujishima, M. (2003) The endosymbiotic bacterium *Holospora obtusa* enhances heat-shock gene expression of the host *Paramecium caudatum*. *J. Eukaryot. Microbiol.* 50, 293–298

13 Hori, M. *et al.* (2008) Micronucleus-specific bacterium *Holospora elegans* irreversibly enhances stress gene expression of the host *Paramecium caudatum*. *J. Eukaryot. Microbiol.* 55, 515–521

14 Fujishima, M. *et al.* (2005) *Paramecium caudatum* acquires heat-shock resistance in ciliary movement by infection with the endonuclear symbiotic bacterium *Holospora obtusa*. *FEMS Microbiol. Lett.* 243, 101–105

15 Schmidt, H. *et al.* (1987) *Caedibacter* caryophila sp. nov., a killer symbiont inhabiting the macronucleus of *Paramecium caudatum*. *Int. J. Syst. Evol. Microbiol.* 37, 459–462

16 Fokin, S. and Görtz, H-D. (1993) *Caedibacter macronucleorum* sp. nov., a bacterium inhabiting the macronucleus of *Paramecium duboscqui*. *Arch. Für Protistenkd.* 143, 319–324

17 Schrallhammer, M. *et al.* (2006) Molecular characterization of the obligate endosymbiont 'Caedibacter macronucleorum' Fokin and Görtz, 1993 and of its host Paramecium duboscqui strain Ku4-8. *J. Eukaryot. Microbiol.* 53, 499–506

18 Schrallhammer, M. and Schweikert, M. (2009) The killer effect of *Paramecium* and its causative agents. In *Endosymbionts in Paramecium* (Fujishima, M., ed.), pp. 227–246, Springer

19 Preer, J.R. and Stark, P. (1953) Cytological observations on the cytoplasmic factor 'kappa' in *Paramecium aurelia*. *Exp. Cell Res.* 5, 478–491

20 Schrallhammer, M. *et al.* (2012) Tracing the role of R-bodies in the killer trait: absence of toxicity of R-body producing recombinant *E. coli* on paramecia. *Eur. J. Protistol.* 48, 290–296

21 Silva, E.S. and Franca, S. (1985) The association dinoflagellate-bacteria: their ultrastructural relationship in two species of dinoflagellates. *Protistologica* 21, 429–446

22 Alverca, E. *et al.* (2002) In situ identification and localization of bacteria associated with *Gyrodinium instriatum* (Gymnodiniales, Dinophyceae) by electron and confocal microscopy. *Eur. J. Phycol.* 37, 523–530

23 Roth, L.E. (1957) An electron microscope study of the cytology of the protozoan Euplotes patella. *J. Biophys. Biochem. Cytol.* 3, 985–1000

24 Leedale, G.F. (1969) Observations on endonuclear bacteria in euglenoid flagellates. *Österr. Bot. Z.* 116, 279–294

25 Dragos, N. *et al.* (1990) Intracellular bacteria in euglenoid flagellates. *Rev. Roum. Biol. Ser. Biol. Veg.* 35, 51–54

26 Shin, W. *et al.* (2003) Endonuclear bacteria in *Euglena hemichromata* (Euglenophyceae): a proposed pathway to endonucleobiosis. *Phycologia* 42, 198–203

27 Hollande, A. and Carruette-Valentin, J. (1971) Les atractophores, l'induction du fuseau et la division cellulaire chez les Hypermastigines, étude infrastructurale et révision systématique des Trichonymphines et des Spirotrichonymphines. *Protistologica* 7, 5–100

28 D'Ambrosio, U. *et al.* (1999) Devescovinid trichomonad with axostyle-based rotary motor ('Rubberneckia'): taxonomic assignment as *Caduceia versatilis* sp. nov. *Eur. J. Protistol.* 35, 327–337

29 Dolan, M.F. *et al.* (2004) Cysts and symbionts of *Staurojoenina assimilis* Kirby from *Neotermes*. *Eur. J. Protistol.* 40, 257–264

30 Sato, T. *et al.* (2014) Intranuclear verrucomicrobial symbionts and evidence of lateral gene transfer to the host protist in the termite gut. *ISME J.* 8, 1008–1019

31 Dangeard, P.A. (1902) Sur le caryophyseme des Eugleniens. *CR Acad Sci Paris* 134, 1365–1366

32 Kirby, H. (1944) *The Structural Characteristics and Nuclear Parasites of Some Species of Trichonympha in Termites*, University of California Press

33 Radek, R. *et al.* (1992) Ectobiotic and endocytobiotic bacteria associated with the termite flagellate *Joenia annectens*. *ACTA Protozool.* 31, 93–107

34 Michel, R. *et al.* (1999) An environmental *Naegleria*-strain, unable to form cysts – turned out to harbour two different species of endocytobionts. *Endocytobiosis Cell Res.* 118, 115–118

35 Walochnik, J. *et al.* (2005) An endocytobiont harbouring Naegleria strain identified as N. clarki De Jonckheere, 1994. *Acta Protozool.* 44, 301–310

36 Horn, M. (2008) Chlamydiae as symbionts in eukaryotes. *Annu. Rev. Microbiol.* 62, 113–131

37 Farone, A.L. *et al.* (2010) *The Isolation and Characterization of Naturally-Occurring Amoeba-Resistant Bacteria from Water Samples*, Final Report R833102, US Environmental Protection Agency

38 Schulz, F. *et al.* (2014) Life in an unusual intracellular niche: a bacterial symbiont infecting the nucleus of amoebae. *ISME J.* 8, 1634–1644

39 Burgdorfer, W. *et al.* (1968) Intranuclear growth of *Rickettsia rickettsii*. *J. Bacteriol.* 96, 1415–1418

40 Burgdorfer, W. and Brinton, L. (1970) Intranuclear growth of *Rickettsia canada*, a member of the typhus group. *Infect. Immun.* 2, 112–114

41 Simser, J. and Palmer, A. (2002) *Rickettsia monacensis* sp. nov., a spotted fever group *Rickettsia*, from ticks (*Ixodes ricinus*) collected in a European city park. *Appl. Environ. Microbiol.* 68, 4559–4566

42 Whitworth, T. *et al.* (2003) Ultrastructural and genetic evidence of a reptilian tick, *Aponomma hydrosauri*, as a host of *Rickettsia honei* in Australia: possible transovarial transmission. *Ann. N. Y. Acad. Sci.* 990, 67–74

43 Perotti, M.A. *et al.* (2006) *Rickettsia* as obligate and mycetomic bacteria. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* 20, 2372–2374

44 Kuechler, S.M. *et al.* (2013) Diversity of bacterial endosymbionts and bacteria–host co-evolution in Gondwanan relict moss bugs (Hemiptera: Coleorrhyncha: Peloridiidae). *Environ. Microbiol.* 15, 2031–2042

45 Watanabe, K. *et al.* (2014) Intrasperm vertical symbiont transmission. *Proc. Natl. Acad. Sci. U.S.A.* 111, 7433–7437

46 Grandi, G. *et al.* (1997) Endonuclear bacterial symbionts in two termite species: an ultrastructural study. *J. Submicrosc. Cytol. Pathol.* 29, 281–292

47 Arneodo, J.D. *et al.* (2008) Ultrastructural detection of an unusual intranuclear bacterium in *Pentastiridius leporinus* (Hemiptera: Cixiidae). *J. Invertebr. Pathol.* 97, 310–313

48 Elston, R.A. (1986) An intranuclear pathogen [nuclear inclusion X (NIX)] associated with massive mortalities of the Pacific razor clam, *Siliqua patula*. *J. Invertebr. Pathol.* 47, 93–104

49 Azevedo, C. (1989) Fine structure of endonucleobiotic bacteria in the gill epithelium of *Ruditapes decussatus*. *Mar. Biol.* 100, 339–341

50 Jensen, S. *et al.* (2010) Intracellular *Oceanospirillales* bacteria inhabit gills of *Acesta* bivalves. *FEMS Microbiol. Ecol.* 74, 523–533

51 Zielinski, F.U. *et al.* (2009) Widespread occurrence of an intranuclear bacterial parasite in vent and seep bathymodiolin mussels. *Environ. Microbiol.* 11, 1150–1167

52 Olson, R.E. and Pierce, J. (1988) Occurrence of the nuclear inclusion parasite in Oregon razor clams (*Siliqua patula*). *J. Invertebr. Pathol.* 52, 198–199

53 Kerk, D. *et al.* (1992) Phylogenetic placement of 'Nuclear Inclusion X (NIX)' into the gamma subclass of Proteobacteria on the basis of 16S ribosomal RNA sequence comparisons. *Syst. Appl. Microbiol.* 15, 191–196

54 Friedrich, A. *et al.* (1999) Microbial diversity in the marine sponge *Aplysina cavernicola* (formerly *Verongia cavernicola*) analyzed by fluorescence in situ hybridization (FISH). *Mar. Biol.* 134, 461–470

55 Ogata, H. *et al.* (2006) Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genet.* 2, e76

56 Urakami, H. *et al.* (1982) Intranuclear *Rickettsia tsutsugamushi* in cultured mouse fibroblasts (L cells). *Microbiol. Immunol.* 26, 445–447

57 Pongponratn, E. *et al.* (1998) Electron-microscopic examination of *Rickettsia tsutsugamushi*-infected human liver. *Trop. Med. Int. Health TM IH* 3, 242–248

58 Gouin, E. *et al.* (2005) Actin-based motility of intracellular pathogens. *Curr. Opin. Microbiol.* 8, 35–45

59 Heinzen, R. *et al.* (1993) Directional actin polymerization associated with spotted fever group *Rickettsia* infection of Vero cells. *Infect. Immun.* 61, 1926–1935

60 Wisseman, C.L. *et al.* (1976) Infection cycle of *Rickettsia rickettsii* in chicken embryo and L-929 cells in culture. *Infect. Immun.* 14, 1052–1064

61 Capelson, M. *et al.* (2011) Nuclear pore complexes guardians of the nuclear genome. *Cold Spring Harb. Symp. Quant. Biol.* 75, 585–597

62 Hetzer, M.W. (2010) The nuclear envelope. *Cold Spring Harb. Perspect. Biol.* 2, a000539

63 Iwatani, K. *et al.* (2005) Translocation of an 89-kDa periplasmic protein is associated with *Holospora* infection. *Biochem. Biophys. Res. Commun.* 337, 1198–1205

64 Sabaneyeva, E.V. *et al.* (2009) Actin-based mechanism of *Holospora obtusa* trafficking in *Paramecium caudatum*. *Protist* 160, 205–219

65 Huang, J. and Brumell, J.H. (2014) Bacteria–autophagy interplay: a battle for survival. *Nat. Rev. Microbiol.* 12, 101–114

66 Bierne, H. *et al.* (2012) Epigenetics and bacterial infections. *Cold Spring Harb. Perspect. Med.* 2, a010272

67 Hamon, M.A. and Cossart, P. (2008) Histone modifications and chromatin remodeling during bacterial infections. *Cell Host Microbe* 4, 100–109

68 Eskandarian, H.A. *et al.* (2013) A role for SIRT2-dependent histone H3K18 deacetylation in bacterial infection. *Science* 341, 1238858

69 Rolando, M. *et al.* (2013) *Legionella pneumophila* effector RomA uniquely modifies host chromatin to repress gene expression and promote intracellular bacterial replication. *Cell Host Microbe* 13, 395–405

70 Lebreton, A. *et al.* (2011) A bacterial protein targets the BAHD1 chromatin complex to stimulate type III interferon response. *Science* 331, 1319–1321

71 Nakamura, Y. *et al.* (2004) Differences in gene expression of the ciliate *Paramecium caudatum* caused by endonuclear symbiosis with *Holospora obtusa*, revealed using differential display reverse transcribed PCR. *FEMS Microbiol. Lett.* 240, 209–213

72 Schmitz-Esser, S. *et al.* (2004) ATP/ADP translocases: a common feature of obligate intracellular amoebal symbionts related to Chlamydiae and Rickettsiae. *J. Bacteriol.* 186, 683–691

73 Haferkamp, I. *et al.* (2006) Tapping the nucleotide pool of the host: novel nucleotide carrier proteins of *Protochlamydia amoebophila*. *Mol. Microbiol.* 60, 1534–1545

74 Görtz, H.D. (1986) Endonucleobiosis in ciliates. *Int. Rev. Cytol.* 102, 169–213

75 Palchevskiy, V. and Finkel, S.E. (2009) A role for single-stranded exonucleases in the use of DNA as a nutrient. *J. Bacteriol.* 191, 3712–3716

76 Finkel, S. and Kolter, R. (2001) DNA as a nutrient: novel role for bacterial competence gene homologs. *J. Bacteriol.* 183, 6288–6293

77 Clarke, M. *et al.* (2013) Genome of *Acanthamoeba castellanii* highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. *Genome Biol.* 14, R11

78 Dohra, H. *et al.* (2014) Draft genome sequences of three *Holospora* species (*Holospora obtusa*, *Holospora* undulata, and *Holospora* elegans), endonuclear symbiotic bacteria of the ciliate *Paramecium caudatum*. *FEMS Microbiol. Lett.* 359, 16–18

79 Müller, J. (1856) Einige Beobachtungen an Infusorien. *Monatsbericht Königl Preuß Akad Wiss.* Juli, 389–393

80 Claperede, E. and Lachmann, J. (1857) Note sur la reproduction des Infusoires. *Ann. Sci. Nat. Zool.* 4, 221–244

81 Stein, F. (1859) Allgemeiner Theil und Naturgeschichte der hypotrichen Infusionsthiere. In *Der Organismus der Infusionsthiere* (1st edn), pp. 96-100

82 Balbiani, E.G. (1861) Recherches sur les phenomenes sexueles des Infusoires. *Arch. Physiol. Norm. Path.* 4, 102–520

83 Bütschli, O. (1876) Studien über die ersten Entwicklungsvorgänge der Eizelle, die Zelltheilung und die Conjugation der Infusorien. *Abh Senckenberg Naturf Ges Frankf.* 10, 213–464

84 Engelmann, T.W. (1875) Über Entwicklung und Fortpflanzung der Infusorien. *Morphol. Jahrb.* 1, 573–635

85 Wolbach, S.B. (1919) Studies on the Rocky Mountain spotted fever. *J. Med. Res.* 41, 1–198

86 Pinkerton, H. and Hass, G. (1932) Spotted fever I. Intranuclear Rickettsiae in spotted fever studied in tissue culture. *J. Exp. Med.* 56, 151–156

87 Price, M.N. *et al.* (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5, e9490

88 Ludwig, W. *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res.* 32, 1363–1371

Chapter IV

Life in an unusual intracellular niche: a bacterial symbiont infecting the nucleus of amoebae

npg

ORIGINAL ARTICLE

Life in an unusual intracellular niche: a bacterial symbiont infecting the nucleus of amoebae

Frederik Schulz[1], Ilias Lagkouvardos[1], Florian Wascher[1], Karin Aistleitner[1], Rok Kostanjšek[2] and Matthias Horn[1]

[1]*Division of Microbial Ecology, Department of Microbiology and Ecosystem Science, University of Vienna, Vienna, Austria and* [2]*Department of Biology, University of Ljubljana, Ljubljana, Slovenia*

Amoebae serve as hosts for various intracellular bacteria, including human pathogens. These microbes are able to overcome amoebal defense mechanisms and successfully establish a niche for replication, which is usually the cytoplasm. Here, we report on the discovery of a bacterial symbiont that is located inside the nucleus of its *Hartmannella* sp. host. This symbiont, tentatively named '*Candidatus* Nucleicultrix amoebiphila', is only moderately related to known bacteria (∼90% 16S and 23S rRNA sequence similarity) and member of a novel clade of protist symbionts affiliated with the *Rickettsiales* and *Rhodospirillales*. Screening of 16S rRNA amplicon data sets revealed a broad distribution of these bacteria in freshwater and soil habitats. '*Candidatus* Nucleicultrix amoebiphila' traffics within 6 h post infection to the host nucleus. Maximum infection levels are reached after 96–120 h, at which time point the nucleus is pronouncedly enlarged and filled with bacteria. Transmission of the symbionts occurs vertically upon host cell division but may also occur horizontally through host cell lysis. Although we observed no impact on the fitness of the original *Hartmannella* sp. host, the bacteria are rather lytic for *Acanthamoeba castellanii*. Intranuclear symbiosis is an exceptional phenomenon, and amoebae represent an ideal model system to further investigate evolution and underlying molecular mechanisms of these unique microbial associations.

Introduction

Free-living amoebae are important members of microbial communities in terrestrial and aquatic habitats, where they interact with other microorganisms in various ways. They primarily feed on bacteria, fungi and other protists, and thus represent main predators controlling the microbial populations (Rodríguez-Zaragoza, 1994). However, continuous grazing pressure has facilitated the evolution of mechanisms allowing some bacteria to evade degradation during phagocytosis and to thrive within amoebae (Barker and Brown, 1994; Casadevall, 2008). After uptake, these bacteria either transiently exploit their eukaryotic host eventually leading to its lysis, or they establish long-term relationships with these protists (Molmeret *et al.*, 2005).

Many bacteria that are able to infect and propagate within amoebae are well-known human pathogens, such as *Legionella pneumophila* (Swanson and Hammer, 2000), *Francissella tularensis* (Abd and Johansson, 2003) and *Mycobacterium tuberculosis* (Mba Medie *et al.*, 2011). In addition, amoebae contain frequently obligate intracellular symbionts that appear to be cosmopolitan in distribution (Fritsche *et al.*, 1993; Horn and Wagner, 2004; Schmitz-Esser *et al.*, 2008). These include members of diverse evolutionary lineages, most notably the *Chlamydiae* (Greub *et al.*, 2003; Horn, 2008). The study of these amoeba-associated microbes provided novel insights in the evolution of pathogenicity and the molecular mechanisms underlying the interaction with eukaryotic host cells (Cazalet *et al.*, 2004; Horn *et al.*, 2004; Collingro *et al.*, 2005).

Intracellular bacteria generally replicate either within host-derived vacuoles or directly in the cytoplasm (Schmitz-Esser *et al.*, 2008; Friedrich *et al.*, 2012). For example, *L. pneumophila* blocks phagosome–lysosome fusion, modulates the endosome by manipulating host processes and finally uses it for replication (Isberg *et al.*, 2009). In contrast, *F. tularensis* is able to escape the phagosome and replicates inside the host cytoplasm (Santic *et al.*, 2009). There are only rare cases where other intracellular niches are exploited. *Midichloria mitochondrii* targets mitochondria (Sassera *et al.*, 2006), *Rickettsia bellii* and *Endonucleobacter*

bathymodioli infect the nucleus of ticks or deep sea mussels (Ogata *et al.*, 2006; Zielinski *et al.*, 2009), and *Holospora* and *Caedibacter* species thrive in the nuclei of *Paramecium* hosts (Goertz, 1986; Fokin, 2004).

Here, we report the first isolation and in-depth characterization of bacteria thriving in the nucleus of free-living amoebae. These symbionts, for which we propose the name 'Candidatus Nucleicultrix amoebiphila', are members of a novel clade within the *Alphaproteobacteria*. We describe the infection cycle of these bacteria, analyze their host range and their impact on host fitness, and we show that this novel symbiont clade represents a diverse group of bacteria with global distribution.

Materials and methods

Isolation and cultivation of amoebae
An activated sludge sample taken from a nitrifying bioreactor (maintained in our laboratory) was applied to a non-nutrient agar plate based on Page's amoebic saline (PAS; $0.12\,g\,l^{-1}$ NaCl, $0.004\,g\,l^{-1}$ MgSO$_4 \times 7$ H$_2$O, $0.004\,g\,l^{-1}$ CaCl$_2 \times 2$H$_2$O, $0.142\,g\,l^{-1}$ Na$_2$HPO$_4$, $0.136\,g\,l^{-1}$ KH$_2$PO$_4$) containing a layer of heat-inactivated *Escherichia coli* (Page, 1988). Amoebae propagating on these plates were repeatedly transferred to new non-nutrient agar plates seeded with heat-inactivated *E. coli* and incubated at room temperature. Finally, amoebae were transferred to culture flasks (Nunclon delta-surface, Thermoscientific, St Leon Rot, Germany) containing PAS and live *E. coli tolC⁻* as well as ampicillin ($200\,ng\,ml^{-1}$). Amoeba cultures were screened using fluorescence *in situ* hybridization (FISH) for the presence of bacterial symbionts.

To obtain clonal cultures with a high percentage of intracellular bacteria, single amoeba from cultures containing bacterial symbionts was picked with a micromanipulator (Eppendorf, Hamburg, Germany) and placed in PAS containing *E. coli tolC⁻* and ampicillin ($200\,ng\,ml^{-1}$) in 96-well culture dishes (Corning, Corning, NY, USA). Cultures were screened using FISH; uninfected cultures were pooled and used as symbiont-free host cells.

Hartmannella vermiformis A1Hsp, *Acanthamoeba castellanii* Neff, *Acanthamoeba polyphaga* DOME, *Acanthamoeba* sp. C1, *Acanthamoeba* sp. 5a2, *Acanthamoeba* sp. UWC8 and *Naegleria gruberi* NEG-M used for infection experiments were incubated at 24 °C in PAS supplemented with *E. coli tolC⁻* and ampicillin ($200\,ng\,ml^{-1}$).

Transmission electron microscopy
Amoebae were detached from culture flasks by shaking, and concentrated by centrifugation ($3000\,g$, 8 min). The supernatant was discarded and samples were fixed for transmission electron microscopy (TEM) 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% OsO$_4$ for 1 h and dehydrated in an increasing ethanol series. Dehydrated samples were embedded in agar 100 resin and cut. Ultrathin sections were stained with uranyl acetate and Reynold's lead citrate before examination with a Philips CM 100 transmission electron microscope (FEI, Hillsboro, OR, USA) operating at 80 kV.

Fluorescence in situ hybridization
Amoeba cells were harvested using centrifugation ($3000\,g$, 8 min), washed with PAS and applied on microscope slides for 30 min to allow attachment of cells before fixation with 4% formaldehyde (15 min at room temperature). A symbiont-specific probe (CBR125, 5′-ACCATTCGGCATGTTCCC-3′) was designed using the probedesign/probematch tools of the ARB software package and deposited at probeBase (Ludwig *et al.*, 2004; Loy *et al.*, 2007). Optimal hybridization conditions for symbiont-specific probes were determined in a series of hybridization experiments with increasing formamide concentrations in the hybridization buffer. In addition, probe EUK516 (5′-ACCAGACTTGCCC TCC-3′, Amann and Binder, 1990) targeting most eukaryotes, the probe mix EUB338 I-III (5′-GCTGCC TCCCGTAGGAGT-3′, 5′-GCAGCCACCCGTAGGTGT-3′, 5′-GCTGCCACCCGTAGGTGT-3; Amann and Binder, 1990; Daims *et al.*, 1999) targeting most bacteria and the negative control probe NONEUB (5′-ACTCCTAC GGGAGGCAGC-3′) were used. All probes were purchased from Thermo Fisher Scientific (St Leon Rot, Germany). Hybridization was performed overnight at 46 °C at a formamide concentration of 20% using standard hybridization and washing buffers (Daims *et al.*, 2005). Cells were subsequently stained with 4′, 6-diamidino-2-phenylindole ($0.5\,\mu g\,ml^{-1}$ in PAS, 3 min), washed once with PAS and embedded in Citifluor (Agar-Scientific, Stansted, UK). Slides were examined using a confocal laser scanning microscope (LSM 510 Meta, Zeiss, Oberkochen, Germany) or an epifluorescence microscope (Axioplan 2 imaging, Zeiss) equipped with a CCD camera (AxioCam HRc, Zeiss).

DNA extraction, polymerase chain reaction (PCR), cloning and sequencing
DNA was extracted from infected amoeba cultures using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany). Amoebal 18S rRNA genes were amplified by PCR using primers 18SF (5′-GTAGTCA TATGCTTGTCTC-3′) and 18SR (5′-CGRARACCTTG TTACGAC-3′; Schmitz-Esser *et al.*, 2008) at an annealing temperature of 50 °C. Bacterial 16S rRNA genes were amplified using primers 616V (5′-AGAG TTTGATYMTGGCTCAG-3′) and 1492R (5′-GGYTAC CTTGTTACGACTT-3′) at an annealing temperature of 52 °C (Juretschko *et al.*, 1998; Loy *et al.*, 2002).

Bacterial 23S rRNA genes were amplified using primers 129F (5'-CYGAATGGGRVAACC-3') and 2490R (5'-CGACATCGAGGTGCCAAAC-3') at an annealing temperature of 56 °C (Hunt and Klepac-Ceraj, 2006). PCRs typically contained 100 ng template DNA, 50 pmol of each primer, 1 unit of Taq DNA polymerase (Fermentas, St Leon Rot, Germany), 10 × Taq buffer with KCl and 2 μM MgCl$_2$ and 0.2 μM of each deoxynucleotide in a total volume of 50 μl. PCR products were purified using the PCR Purification Kit (Qiagen) and were either sequenced directly or cloned using the TOPO TA Cloning Kit (Invitrogen, Darmstadt, Germany) following the manufacturers' instructions. Nucleotide sequences were determined using the BigDye Terminator kit v3.1 (Applied Biosystems, Vienna, Austria) and an ABI 3130 XL genetic analyzer (Applied Biosystems). Newly obtained rRNA gene sequences were deposited at Genbank/EMBL/DDBJ under accession numbers KF697195, KF697196 and KF697197.

Infection experiments
Supernatant of infected amoebae was collected and passed through a 1.2-μm filter (Sartorius, Goettingen, Germany). The filtrate was centrifuged at 10 500 g for 10 min, and the resulting bacterial pellet was washed once with PAS, resuspended in PAS and directly used for infection experiments. Bacterial cell numbers were estimated by counting DAPI-stained bacteria in a Neubauer counting chamber.

All experiments were performed at room temperature, if not stated otherwise. Amoebae were seeded into 12-well plates (Thermoscientific, Hvidovre, Denmark; 3 × 10^5 cells per well), and bacteria at an estimated multiplicity of infection of 500 were added. After 6 h, amoeba cells were washed twice with PAS to remove extracellular bacteria. At different time points post infection (p.i.), amoebae were detached and amoeba cell numbers were determined with a Neubauer counting chamber. Half of the cell suspension was then used for FISH, and the other half for viability staining.

The percentage of amoebae-containing bacteria in the nucleus was determined by analyzing FISH-stained samples. At least 100 amoeba cells were counted per sample. To assess the viability of amoeba cells during infection, amoebae were incubated with propidium iodide (PI; 1.5 μM; Molecular Probes, Eugene, OR, USA) in PAS for 20 min at room temperature in the dark. Samples were washed once with PAS and transferred to a black plate reader dish (Greiner Bio-One, Frickenhausen, Germany), and fluorescence intensities were determined with a Tecan Infinite M200 plate reader (Tecan, Groeding, Austria). Negative controls included PAS and suspensions of uninfected amoebae.

For each time point and condition, at least three replicate infection experiments were analyzed. To test for statistically significant differences ($P < 0.05$)

between different conditions and time points, a two-way analysis of variance including the Bonferroni post test was performed using GraphPad Prism (version 6; La Jolla, CA, USA).

Phylogenetic analysis
Phylogenetic analyses were conducted using MEGA version 5 (Tamura et al., 2011) and ARB (Ludwig et al., 2004). The alignment editor integrated in ARB was used to build alignments based on the current Silva ARB 16S and 23S rRNA databases (Quast et al., 2013), which were updated with sequences from GenBank/EMBL/DDBJ obtained by sequence homology searches using BLASTn available at the NCBI web site (National Center for Biotechnology Information; Altschul et al., 1997; Wheeler et al., 2008). The alignment was trimmed to the length of the shortest sequence, manually curated and exported from ARB using a 50% conservation filter. The resulting alignments comprised 103 sequences and 1252 positions for the 16S rRNA and 75 sequences including 2235 positions for the 23S rRNA; the concatenated 16S–23S rRNA data set included 3487 positions. MEGA was used to find the optimal maximum likelihood model, and the GTR + I + G model of evolution was selected. Maximum likelihood trees were calculated in MEGA. To evaluate the robustness of the tree topology, 1000 bootstrap resamplings were performed. In addition, MrBayes 3.2 (Ronquist et al., 2012) was used for tree reconstruction. The Bayesian analysis included four Markov chains, consisting of 5 000 000 generations and a sampling every 100 generations. Two simultaneous runs with different random start trees were performed, and the first 25% of samples were discarded. The resulting phylogenetic trees were visualized in iTOL (Letunic and Bork, 2007).

Screening of 16S rRNA gene amplicon data sets
The screening of amplicon data sets is based on a recently published approach (Lagkouvardos et al., 2013). Briefly, all raw 454-based 16S rRNA gene amplicon sequence data from environmental samples in the databases SRA (Kodama et al., 2012) and VAMPS (Huse et al., 2010; http://vamps.mbl.edu/) were extracted and organized by sample in independent data sets. The databases were searched using BLAST (Altschul et al., 1997) and the full-length 16S rRNA gene sequence of 'Candidatus Nucleicultrix amoebiphila' as query. The detected amplicon sequences were filtered with respect to size (>200 nucleotides) and alignment length (>80% of their size). Extracted sequences that were at least 95% similar to the 16S rRNA of 'Candidatus Nucleicultrix amoebiphila' and that had a length of at least 300 nucleotides were assigned to either of the three data sets covering different regions of a reference alignment including the full-length sequence of 'Candidatus Nucleicultrix amoebiphila'

and the most similar sequences from GenBank/ EMBL/DDBJ (Wheeler *et al.*, 2008). To decrease the complexity of the data sets and to account for artifacts caused by sequencing errors, each set of sequences was clustered with UCLUST at a 97% threshold (Edgar, 2010). Singletons were omitted, and in the case where operational taxonomic units contained multiple sequences from the same sample, only one sequence representative for this sample was kept. Sequences were aligned using SSU-ALIGN for each of the three data sets (Nawrocki and Adviser-Eddy, 2009), manually curated and trimmed to the length of the shortest sequence in Jalview (Waterhouse *et al.*, 2009). Phylogenetic analysis was performed with FastTree2 using the GTR model and the gamma distribution with 20 categories option (Price *et al.*, 2010). The resulting phylogenetic tree was visualized in iTOL (Letunic and Bork, 2007).

Results

Isolation of a Hartmannella sp. with bacterial symbionts in the nucleus

During a routine screening of a nitrifying bioreactor for free-living amoebae, we obtained an amoeba isolate morphologically resembling *Hartmannella* species. This was confirmed by 18S rRNA gene sequencing, demonstrating that this amoeba strain, termed *Hartmannella* sp. FS5, is closely related to known *Hartmannella* species (Supplementary Figure S1). Subsequent analysis of *Hartmannella* sp. FS5 using FISH demonstrated the presence of bacterial endosymbionts. Surprisingly, we observed that the bacterial FISH signal overlapped with DAPI (4′, 6-diamidino-2-phenylindole)-stained amoeba nuclei, although no bacteria were detected in the cytoplasm (Supplementary Figure S2). This suggested that the bacterial symbionts in *Hartmannella* sp. FS5 are located in the host nucleus. TEM confirmed the intranuclear location (Figures 1a–c). Coccoid rods with a length of 0.5–1 µm, a diameter of 0.3–0.4 µm and a Gram-negative-type cell wall were observed directly in the nucleoplasm of infected amoebae.

A novel symbiont related to the Rickettsiales and the Rhodospirillales

Sequencing of the 16S and 23S rRNA genes revealed that the *Hartmannella* sp. FS5 symbionts are novel and only distantly related to known bacteria (~90% sequence similarity). Phylogenetic analysis showed that together with other symbionts of amoebae and
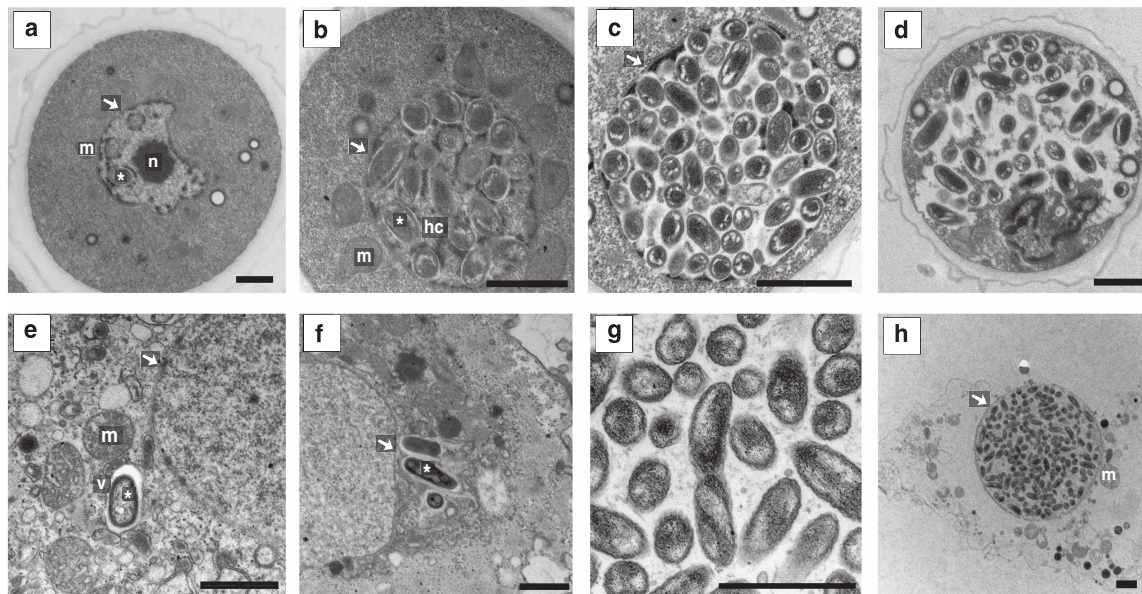
Figure 1 The intracellular niche of 'Candidatus Nucleicultrix amoebiphila'. Transmission electron micrographs showing different infection stages of *Hartmannella* sp. (**a–d**) and *A. castellanii* (**e–h**). (**a**) Early infection stage in *Hartmannella* with three *Nucleicultrix* particles present in the nucleus, bacteria are surrounded by host electron-dense heterochromatin. (**b**) Intermediate infection stage; the nucleus is filled with bacteria. (**c**) Late infection stage with an enlarged nucleus. (**d**) Intact cyst containing symbionts; the nuclear membrane is ruptured. (**e**) Early infection stage in *A. castellanii*; after uptake into the cytoplasm, the bacteria remain enclosed by a vacuole and are surrounded by an electron-translucent space indicating the presence of extracellular polymeric substances. (**f**) Symbionts in close proximity to the host nuclear membrane; no clear phagosomal membrane can be recognized. (**g**) Ultrastructure of the symbionts inside an *Acanthamoeba* sp. nucleus; coccoid rods with a Gram-negative-type cell wall can be seen, one dividing cell is shown. (**h**) Lysis of an *A. castellanii* host cell; the nucleus is densely populated by bacteria, while the nuclear membrane is still intact. Asterisks indicate 'Candidatus Nucleicultrix amoebiphila'; arrows indicate the nuclear membrane; m, mitochondria; n, nucleolus; hc, heterochromatin; v, vacuole. Bars, 1 µm.

Figure 2 Phylogenetic relationship of 'Candidatus Nucleicultrix amoebiphila' with the Alphaproteobacteria. (**a**) Maximum likelihood tree based on a concatenated 16S–23S rRNA alignment. 'Candidatus Nucleicultrix amoebiphila' (indicated by an asterisk) together with other protist symbionts forms a deeply branching lineage within the Alphaproteobacteria. (**b**) 16S rRNA-based maximum likelihood tree showing members of the protist symbiont clade containing 'Candidatus Nucleicultrix amoebiphila'. Circles displayed at the nodes represent bootstrap values (upper half) and Bayesian posterior probability values (lower half), respectively. Black indicates support values > 80% or probability values > 0.8; gray indicates support values of 50–80% or probability values of 0.5–0.8, white indicates no support. Fully resolved trees are shown in Supplementary Figures S3–S5; accession numbers of sequences used for phylogenetic analyses are listed in Supplementary Table S1.

ciliates they form a deep-branching lineage in the Alphaproteobacteria (Figure 2). However, the relationship of this symbiont clade with other Alphaproteobacteria could not be resolved unambiguously. Different data sets (16S rRNA, 23S rRNA or both genes concatenated) and different treeing methods resulted in conflicting, yet generally well-supported, tree topologies. The symbiont clade branched off within the Rickettsiales in 16S rRNA-based trees but grouped together with the Rhodospirillales in the 23S rRNA-based trees and in the trees inferred with the concatenated genes (Figure 2, Supplementary Figures S3–S5). The name 'Candidatus Nucleicultrix amoebiphila' is proposed for tentative classification of these novel intranuclear bacteria (a formal description is provided as Supplementary Text S1).

'Candidatus Nucleicultrix amoebiphila' also infects Acanthamoeba spp.
We next tested whether 'Candidatus Nucleicultrix amoebiphila' (hereafter referred to as Nucleicultrix) is able to infect amoebae other than its original Hartmannella sp. FS5 host. The symbionts were purified and added to various Acanthamoeba,

Hartmannella and Naegleria strains; amoeba cultures were screened with FISH 7 days p.i. Nucleicultrix infected Hartmannella vermiformis and four Acanthamoeba strains with different efficiencies, with A. castellanii Neff being the most susceptible Acanthamoeba strain. In these amoebae, Nucleicultrix was always located within the host nucleus (Supplementary Figure S7). It should be noted, however, that successful infection of Hartmannella spp. was dependent on the simultaneous presence of E. coli as food source (Supplementary Figure S8), and that no infection was observed in nutrient-rich media. Infection with Nucleicultrix was not detected for Acanthamoeba sp. UWC8 and Naegleria gruberi. Growth of purified Nucleicultrix was also not observed under host-free conditions using a variety of standard laboratory media.

Close association of Nucleicultrix with host cell heterochromatin
To obtain more detailed insights into the lifestyle of Nucleicultrix, we examined Hartmannella sp. and A. castellanii at different infection stages using TEM. In both amoebae, the symbiont was only rarely seen in the host cytoplasm, suggesting that

after uptake *Nucleicultrix* migrates to the nucleus and does not replicate in the cytoplasm. Bacteria in the cytoplasm were mostly enclosed in vacuoles (Figures 1e and f), which contained single or few bacterial cells often surrounded by an electron-translucent space, which might indicate the presence of extracellular polymeric substances. Within the host nucleus, the symbionts colocalized with electron-dense heterochromatin (Figures 1a and b). These structures more and more disappeared in strongly infected nuclei, and areas of white, unstructured electron-translucent space increased (Figure 1c). The infection leads to highly enlarged nuclei, densely packed with bacteria. Infected nuclei may burst (Figure 1d) or stay intact after host cell lysis (Figure 1h).

Infection cycle and impact on amoeba hosts
Monitoring of the infection of *Hartmannella* and *A. castellanii* using FISH and staining of endocytic compartments showed that *Nucleicultrix* is taken up by phagocytosis, resides within the phagosome and survives acidification by lysosome fusion. The symbionts escape from the phagolysosome and are detected as individual cells in the cytoplasm as early as 2 h p.i (Supplementary Figure S6). *Nucleicultrix* was first seen within the nuclear compartment 4–6 h (Figure 3, Supplementary Figure S7) p.i. Following invasion of the nucleus, the symbionts start replicating, and after 24 h, two to four bacteria were present per nucleus. Up to 32 bacteria could be found 48 h p.i., suggesting a doubling time of about 8 h at this stage of infection. Nuclei were densely

packed with *Nucleicultrix* after 72–96 h, and this compartment strongly increased in size at 120 h.

Host lysis was the predominant result of infection of *A. castellanii*, occurring from 72 h p.i. onwards and suggesting horizontal transmission of the bacterial symbionts (Supplementary Figure S7). In contrast, host lysis occurred later (at 96 h) and only rarely in *Hartmannella*. Instead, we observed dividing *Hartmannella* trophozoites with infected nuclei throughout the course of the experiment, with both daughter cells carrying the symbiont (Figure 4). This suggests that *Nucleicultrix* in *Hartmannella* sp. is transmitted both horizontally and vertically (Figure 4).

Host cell lysis in infected *Acanthamoeba* cultures was further analyzed and quantified using a PI assay, which detects dead cells with disintegrated membranes (Figure 5a). A significant increase in PI fluorescence intensity was observed at 72 h p.i. in *A. castellanii* cultures ($P < 0.05$). This assay also demonstrated the general lack of host cell lysis in the original *Hartmannella* sp. host (Figure 5b), in which no significant increase in PI fluorescence intensities was detected. This is particularly noteworthy as the proportion of infected cells was generally higher in *Hartmannella* than in *A. castellanii* cultures (Figures 5a and b). At 24 h p.i., *Nucleicultrix* was present in 30% of all *Hartmannella* sp. cells but in only 10% of *A. castellanii* trophozoites when the same multiplicity of infection was used. Although the percentage of infected cells remained at a similar level during the course of the experiment in case of *A. castellanii*, nearly 80% of all *Hartmannella* sp. trophozoites contained
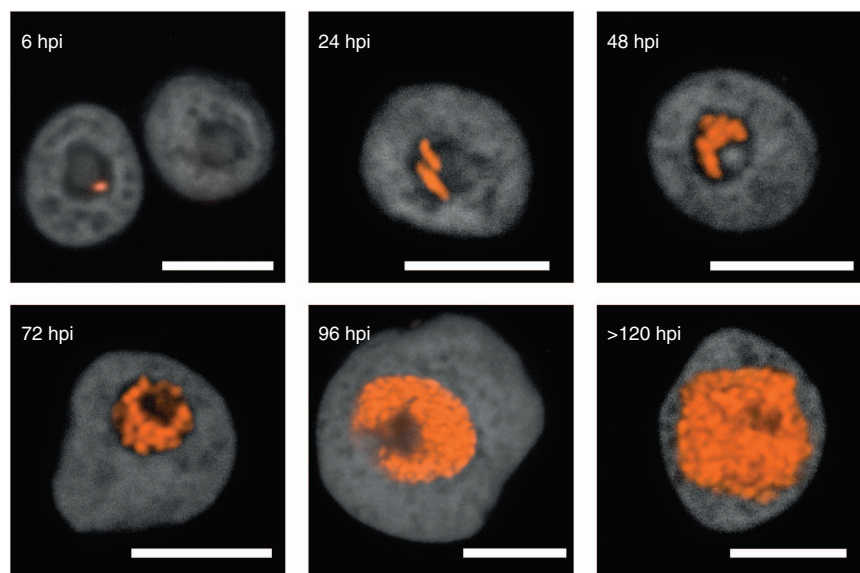

Figure 3 Infection cycle of 'Candidatus Nucleicultrix amoebiphila' in *Hartmannella* sp. FS5. Infection was monitored over a course of 144 h and visualized with FISH using probes CBR125 (orange) and EUK516-Cy5 (gray). Single bacterium starts colonizing the nuclear compartment within the first 6 h post infection; the nucleus is completely filled with bacteria after 96 h. Fully infected and highly enlarged nuclei can be observed after 120 h. The course of infection of *Acanthamoeba* sp. cells is similar (Supplementary Figure S7). Bar, 5 μm.
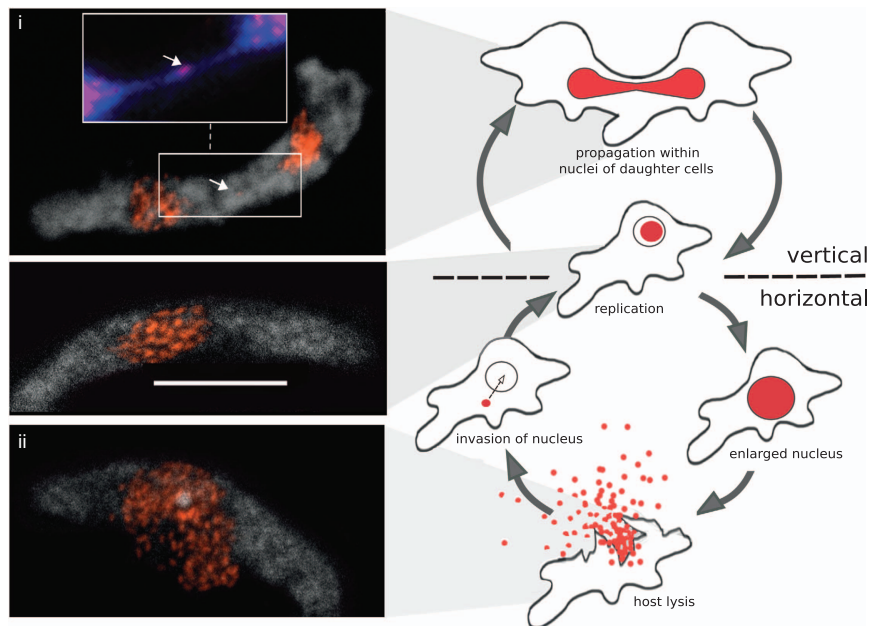
Figure 4 Modes of transmission of 'Candidatus Nucleicultrix amoebiphila'. Key events during transmission of *Nucleicultrix* in *Hartmannella* sp. visualized by FISH are shown in the left panel. Bacteria and amoebae were labeled with probes CBR125 (orange) and EUK516 (gray), respectively. A dividing *Hartmannella* trophozoite is shown (i) in which the distribution of *Nucleicultrix* among both daughter cells can be seen. Note the single bacterial cell (arrows) within the amoeba nucleus undergoing segregation. Host cell lysis (ii) occurs only rarely in *Hartmannella* sp.; bar, 10 μm. A model of the life cycle and transmission of *Nucleicultrix* in amoebae is shown in the right panel. Although vertical transmission is predominant in *Hartmannella* sp., host lysis and horizontal transmission is the major route of *Nucleicultrix* in acanthamoebae.

Nucleicultrix at 144 h p.i. (Figures 5a and b). However, the presence of *Nucleicultrix* did not have a negative effect on the net growth of both *Acanthamoeba* and *Hartmannella* cultures (Figures 5c and d). The presence of *Nucleicultrix* also did not affect *Hartmannella* growth in continuous cultures, even when different incubation temperatures were used (Supplementary Figure S9).

We were able to maintain *Nucleicultrix* in continuous *Hartmannella* sp. cultures for more than 2 years so far, with 30–80% of all amoebae being infected. Moreover, infected *Acanthamoeba* cultures could be propagated for several months, yet with lower infection rates of 5–20%. *Nucleicultrix* is able to persist in *Hartmannella* sp. cysts over longer periods of starvation (Figure 1d), whereas we never observed infected *Acanthamoeba* cysts.

Environmental distribution of Nucleicultrix
To investigate the environmental distribution and diversity of *Nucleicultrix*, we searched 16S rRNA amplicon data sets deposited at SRA and the VAMPS (Huse *et al.*, 2010; Kodama *et al.*, 2012; http://vamps.mbl.edu; Lagkouvardos *et al.*, 2013). We retrieved a total of 1356 sequences with a similarity >95% to the 16S rRNA of *Nucleicultrix*, with the majority of sequences being nearly identical to that of the symbiont (⩾99%; Supplementary Figure S10; Supplementary Table S2).

Grouping these sequences into operational taxonomic units, dereplication and subsequent phylogenetic analysis showed that *Nucleicultrix*-related sequences form two clearly distinct clades (Figure 6). The vast majority of sequences originated from freshwater samples (Supplementary Figure S10, Supplementary Table S2); however, both clades also contained sequences from a variety of environments including soil, anthropogenic habitats, humans and animals (Figure 6). Nearly identical sequences were found in different environmental samples originating from geographically distinct regions.

Discussion

This study represents the first identification and characterization of bacterial symbionts colonizing the nucleus of amoebae. So far, bacteria sharing this unusual lifestyle have been mainly found in protozoan hosts: in ciliates (Goertz, 1986; Fokin, 2004) and also in *Euglenida* (Shin *et al.*, 2003), *Trichonymphida* (Dolan *et al.*, 2004), *Cristamonadida* (d'Ambrosio *et al.*, 1999), *Spirotrichonymphida* (Hollande and Carruette-Valentin, 1971) and dinoflagellates (Leedale, 1969; Alverca *et al.*, 2002). Most of these bacteria were described only at the phenotypic level, using light- and electron microscopy to confirm the intranuclear location. The phenomenon of intranuclear symbiosis was investigated in more detail only for the *Paramecium*
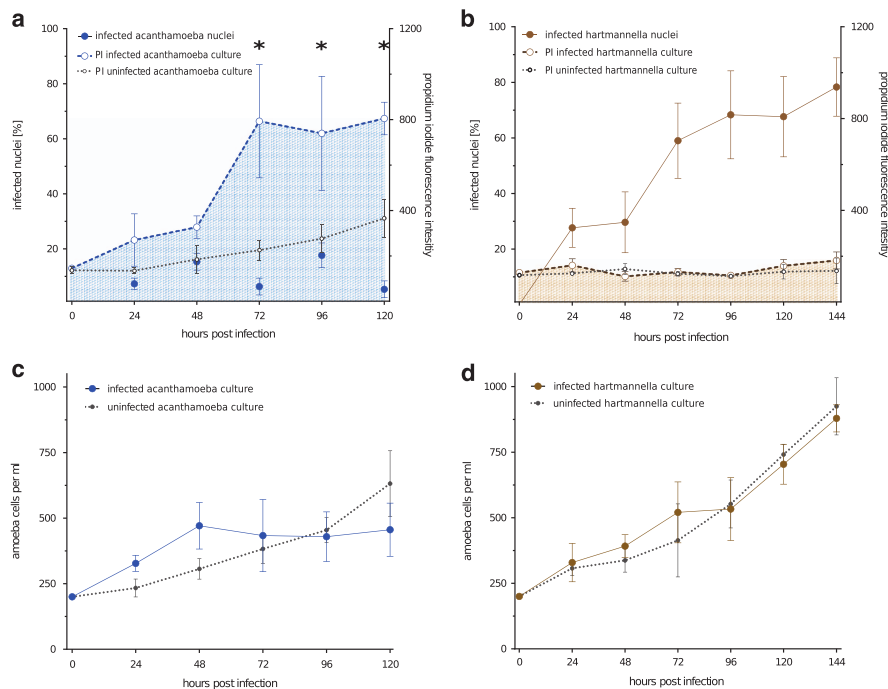
Figure 5 Fate of amoeba hosts during infection with 'Candidatus Nucleicultrix amoebiphila'. (**a**, **b**) Viability of amoeba cells during infection with Nucleicultrix. PI fluorescence intensities indicating dead amoeba cells were determined during the course of infection. A. castellanii but hardly Hartmannella sp. cells are lysed in the presence of Nucleicultrix. The percentage of infected amoebae during the course of infection is shown by filled circles. (**c**, **d**) Amoeba cell numbers during the course of infection. Asterisks indicate significant differences to the control ($P < 0.05$), error bars show s.d. based on three replicate infection experiments.
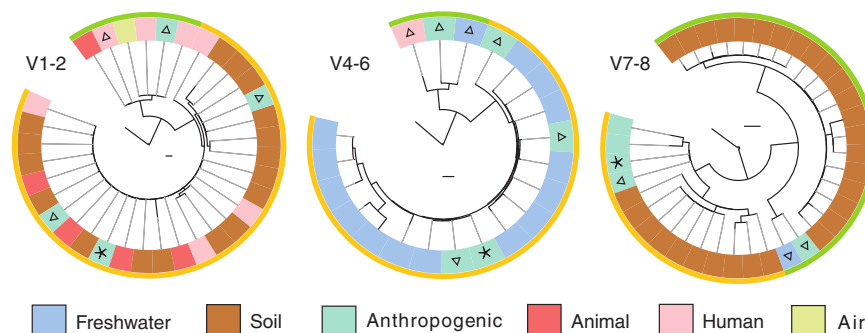

Figure 6 Diversity and environmental distribution of 'Candidatus Nucleicultrix amoebiphila'. Phylogenetic trees based on amplicon sequences (> 300 nucleotides length, $> 95\%$ sequence similarity to the Nucleicultrix 16S rRNA) are shown. As 16S rRNA sequences from amplicon-based studies generally cover only short regions of the full-length gene, separate trees comprising sequences spanning variable regions V1–V2, V4–V6 and V7–V8, respectively, were calculated. Each tree depicts a subset of representative sequences retrieved from public databases. With all three data sets, Nucleicultrix-related sequences group into two distinct clades (marked yellow and green in the outer circle). Colors in the inner circle represent the environmental origin of the sequences; triangles indicate full-length sequences; the position of Nucleicultrix is labeled with an asterisk. Bars indicate an estimated evolutionary distance of 0.01.

symbionts *Caedibacter* spp. (Schrallhammer and Schweikert, 2009), *Holospora* spp. (Fujishima and Fujita, 1985; Fokin and Görtz, 2009), 'Candidatus Paraholospora nucleivisitans' (Eschbach et al., 2009) and for the recently described 'Candidatus Endonucleobacter bathymodioli' in deep sea mussels (Zielinski et al., 2009).

The affiliation of most intranuclear bacteria is unknown; however, E. bathymodioli has been identified as a member of the *Gammaproteobacteria*, and the *Paramecium* symbionts *Holospora* and *Caedibacter* were described as *Alphaproteobacteria*, forming a monophyletic group in the order *Rickettsiales* (Zielinski et al., 2009; Boscaro et al., 2013). *Nucleicultrix* is only distantly related to *Holospora* and *Caedibacter* species, but it groups with the latter, as well as with the cytoplasmic *Acanthamoeba* symbiont S40 (Figure 2b) (Matsuo

et al., 2009). We noted that—in contrast to previous analysis and 16S rRNA-based phylogenetic trees—the 23S rRNA-based tree and a tree inferred from concatenated 16S and 23S rRNA genes placed this symbiont clade outside of the *Rickettsiales* and together with the *Rhodospirillales*, which comprise mainly free-living bacteria (Figure 2a). It has previously been shown that the relationships among *Alphaproteobacteria* are difficult to resolve based on rRNA gene analysis alone (Williams *et al.*, 2007). Although the exact affiliation of *Nucleicultrix* with other *Alphaproteobacteria* remains elusive at this point, future phylogenomic analysis might help to clarify this. Despite this uncertainty, the occurrence of the intranuclear lifestyle in only distantly related evolutionary lineages (*Alpha-* and *Gammaproteobacteria*, respectively) suggests that the ability to use the nucleus of eukaryotic cells as a niche has evolved multiple times independently in different bacterial groups.

Querying sequence databases containing environmental 16S rRNA gene sequences derived from amplicon-based studies provided a first glimpse at diversity and environmental distribution of *Nucleicultrix*-related bacteria. In addition to only 17 near-full-length sequences in the main public databases Genbank/EMBL/DDBJ, this analysis recovered more than 1300 partial 16S rRNA gene sequences with similarity to the 16S rRNA of *Nucleicultrix*. On the basis of this data set, *Nucleicultrix*-related bacteria belong to either of the two phylogenetically distinct clades and occur in a variety of habitats (Figure 6, Supplementary Figure S10). The presence of nearly identical sequences in samples from geographically distinct locations and in environments where protists are typically found (such as freshwater and soil) suggests that *Nucleicultrix*-related bacteria occur world-wide and are likely dispersed by protozoa.

Known bacterial symbionts targeting the nucleus of their eukaryotic host cells display a number of remarkable morphologic features different from those of *Nucleicultrix*. Both the mussel symbiont *E. bathymodioli* and the *Paramecium* symbionts *Holospora* spp. show a developmental cycle with morphologically distinct stages (Zielinski *et al.*, 2009; Fujishima and Kodama, 2012). The shape of these symbionts alters between short coccoid rods and elongated cells with a length of up to 15–20 μm. These elongated forms represent the reproductive stage of *E. bathymodioli* or the infectious form of *Holospora* species. The latter is further characterized by a so-called invasion tip, a subcellular structure involved in the invasion of the host nucleus (Kawai and Fujishima, 2000). In contrast, we found no evidence for different cell shapes or developmental forms for *Nucleicultrix* during the infection of its amoeba hosts. *Nucleicultrix* is also different from the *Paramecium* symbionts *Caedibacter caryophila* and *C. macronucleorum*, which contain characteristic 'refractile bodies' known as

R-bodies that mediate the killer trait of their hosts (Schmidt and Goertz, 1987; Schrallhammer and Schweikert, 2009). Taken together, this suggests that the strategies for host interaction are fundamentally different between known endonuclear bacteria.

After uptake into the host cell by phagocytosis, *Nucleicultrix* escapes the phagolysosome and is temporarily located in the cytoplasm (Supplementary Figure S6). The exact mode of entry into the nucleus remains unknown. However, during cell division, *Hartmannella* and *Acanthamoeba* undergo open mitosis (Levandowsky and Hutner, 1979; Khan, 2009), in which the nuclear envelope disintegrates. A conceivable scenario is that *Nucleicultrix* interacts with chromatin during host cell division and is enclosed in the nucleus after reassembly of the nuclear envelope. Once inside the nucleus, *Nucleicultrix* colocalizes with host chromatin and initiates replication. During the course of infection, the electron-dense heterochromatin gradually disappears and a white, unstructured electron-lucent space around the symbionts occurs in TEM images (Figure 1). This is reminiscent of nuclei of paramecia infected with *Holospora* spp. and suggests that endonuclear bacteria might feed directly on host chromatin (Goertz, 1986). However, the overall neutral effect of *Nucleicultrix* on host fitness seems to rule out an extensive degradation of host DNA. In its original *Hartmannella* host, *Nucleicultrix* does not negatively affect growth (Supplementary Figure S9); the symbionts are transmitted vertically during binary fission of their host cells and amoeba cultures can be maintained for prolonged periods at high infection rates. This situation is different in *Acanthamoeba* host cells where *Nucleicultrix* is more lytic and the percentage of infected amoeba is generally lower (Figure 5).

Although rarely used as intracellular niche, the nucleus is a potentially attractive compartment for intracellular bacteria, as it is rich in proteins, nucleic acids and nucleoside triphosphates required for DNA replication and transcription. It also represents a shelter protecting bacteria from cellular defense mechanism in the cytoplasm. A number of bacterial pathogens that reside in the cytoplasm have evolved sophisticated strategies to target bacterial effector proteins (nucleomodulins) to the nucleus (Bierne *et al.*, 2012). Residing within this compartment should facilitate manipulations of the host cell by targeting basic processes in the nucleus such as gene expression or DNA replication. The identification of bacteria living in the nucleus of amoebae, which can easily be kept in culture, represents a great opportunity to further investigate molecular mechanisms underlying intranuclear symbiosis.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

References

Abd H, Johansson T. (2003). Survival and growth of *Francisella tularensis* in *Acanthamoeba castellanii*. *Appl Environ Microbiol* **69**: 600–606.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Alverca E, Biegala I, Kennaway G. (2002). In situ identification and localization of bacteria associated with *Gyrodinium instriatum* (*Gymnodiniales, Dinophyceae*) by electron and confocal microscopy. *Eur J Phycol* **37**: 37–41.

Amann R, Binder B. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* **56**: 1919–1925.

Barker J, Brown M. (1994). Trojan horses of the microbial world: protozoa and the survival of bacterial pathogens in the environment. *Microbiology* **140**: 1253–1259.

Bierne H, Cossart P, Bact I. (2012). When bacteria target the nucleus: the emerging family of nucleomodulins. *Cell Microbiol* **14**: 622–633.

Boscaro V, Fokin S, Schrallhammer M. (2013). Revised systematics of *Holospora*-like bacteria and characterization of ''*Candidatus* Gortzia infectiva'' a novel macronuclear symbiont of *Paramecium jenningsi*. *Microb Ecol* **65**: 255–267.

Cazalet C, Rusniok C, Brüggemann H, Zidane N, Magnier A, Ma L *et al.* (2004). Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. *Nat Genet* **36**: 1165–1173.

Casadevall A. (2008). Evolution of intracellular pathogens. *Annu Rev Microbiol* **62**: 19–33.

Collingro A, Toenshoff ER, Taylor MW, Fritsche TR, Wagner M, Horn M. (2005). ''*Candidatus* Protochlamydia amoebophila'' an endosymbiont of *Acanthamoeba* spp. *Int J Syst Evol Microbiol* **55**: 1863–1866.

d'Ambrosio U, Dolan M, Wier AM, Margulis L. (1999). *Devescovinid trichomonad* with axostyle-based rotary motor ('*Rubberneckia*'): taxonomic assignment as *Caduceia versatilis* sp. nov. *Eur J Protistol* **35**: 327–337.

Daims H, Brühl A, Amann R, Schleifer KH, Wagner M. (1999). The domain-specific probe EUB338 is insufficient for the detection of all bacteria: development and evaluation of a more comprehensive probe set. *Syst Appl Microbiol* **22**: 434–444.

Daims H, Stoecker K, Wagner M. (2005). Fluorescence in situ hybridization for the detection of prokaryotes. *Mol Microb Ecol* **213**: 239.

Dolan MF, Wier AM, Melnitsky H, Whiteside JH, Margulis L. (2004). Cysts and symbionts of *Staurojoenina assimilis* Kirby from *Neotermes*. *Eur J Protistol* **40**: 257–264.

Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Eschbach E, Pfannkuchen M, Schweikert M, Drutschmann D, Brümmer F, Fokin S *et al.* (2009). '*Candidatus* Paraholospora nucleivisitans' an intracellular bacterium in *Paramecium sexaurelia* shuttles between the cytoplasm and the nucleus of its host. *Syst Appl Microbiol* **32**: 490–500.

Fokin SI. (2004). Bacterial endocytobionts of *Ciliophora* and their interactions with the host cell. *Int Rev Cytol* **236**: 181–249.

Fokin SI, Görtz HD. (2009). Diversity of *Holospora* bacteria in *Paramecium* and their characterization. In: Fujishima M (ed) *Endosymbionts in Paramecium*. Springer-Verlag: Heidelberg, Germany, pp 161–199.

Friedrich N, Hagedorn M, Soldati-Favre D, Soldati T. (2012). Prison break: pathogens' strategies to egress from host cells. *Microbiol Mol Biol Rev* **76**: 707–720.

Fritsche TR, Gautom RK, Seyedirashti S, Bergeron DL, Lindquist TD. (1993). Occurrence of bacterial endosymbionts in *Acanthamoeba* spp isolated from corneal and environmental specimens and contact lenses. *J Clin Microbiol* **31**: 1122–1126.

Fujishima M, Fujita M. (1985). Infection and maintenance of *Holospora obtusa* a macronucleus-specific bacterium of the ciliate *Paramecium caudatum*. *J Cell Sci* **76**: 179–187.

Fujishima M, Kodama Y. (2012). Endosymbionts in paramecium. *Eur J Protistol* **48**: 124–137.

Gortz H. (1986). Endonucleobiosis in ciliates. *Int Rev Cytol* **102**: 169–213.

Greub G, Mege J, Raoult D. (2003). *Parachlamydia acanthamoeba* enters and multiplies within human macrophages and induces their apoptosis. *Infect Immun* **59**: 5979–5985.

Hollande A, Carruette-Valentin J. (1971). Les atractophores l'induction du fuseau et la division cellulaire chez les *Hypermastigines*. *Protistologica* **7**: 5–100.

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

Horn M, Collingro A, Schmitz-Esser S, Beier CL, Purkhold U, Fartmann B *et al.* (2004). Illuminating the evolutionary history of *Chlamydiae*. *Science* **304**: 728–730.

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

Hunt D, Klepac-Ceraj V. (2006). Evaluation of 23S rRNA PCR primers for use in phylogenetic studies of bacterial diversity. *Appl Environ Microbiol* **72**: 2221–2225.

Huse SM, Welch DM, Morrison HG, Sogin ML. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ Microbiol* **12**: 1889–1898.

Isberg RR, O'Connor TJ, Heidtman M. (2009). The *Legionella pneumophila* replication vacuole: making a cosy niche inside host cells. *Nat Rev Microbiol* **7**: 13–24.

Juretschko S, Timmermann G, Schleifer KH, Schmid M, Pommerening-Röser A, Koops HP *et al.* (1998). Molecular and conventional analyses of nitrifying bacterium diversity in activated sludge: *Nitrosococcus mobilis* and *Nitrospira*-like bacteria as dominant populations. *Appl Environ Microbiol* **64**: 3042–3051.

Kawai M, Fujishima M. (2000). Invasion of the macronucleus of *Paramecium caudatum* by the bacterium *Holospora obtusa*: fates of the bacteria

and timings of invasion steps. *Eur J Protistol* **36**: 46–52.

Khan NA. (2009). *Acanthamoeba: biology and pathogenesis*. Causter Academic Press: Norfolk, UK.

Kodama Y, Shumway M, Leinonen R. (2012). The Sequence read archive: explosive growth of sequencing data. *Nucleic Acids Res* **40**: D54–D56.

Lagkouvardos I, Weinmaier T, Lauro FM, Cavicchioli R, Rattei T, Horn M. (2013). Integrating metagenomic and amplicon databases to resolve the phylogenetic and ecological diversity of the *Chlamydiae*. *ISME J* **8**: 115–125.

Leedale G F. (1969). Observations on endonuclear bacteria in euglenoid flagellates. *Österreichische Bot Zeitschrift* **116**: 279–294.

Letunic I, Bork P. (2007). Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**: 127–128.

Levandowsky M, Hutner SH. (1979). *Biochemistry and physiology of protozoa* 1. Academic Press: New York, USA.

Loy A, Lehner A, Lee N. (2002). Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Appl Environ Microbiol* **68**: 5064–5081.

Loy A, Maixner F, Wagner M, Horn M. (2007). probeBase—an online resource for rRNA-targeted oligonucleotide probes: new features 2007. *Nucleic Acids Res* **35**: 800–804.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Buchner A et al. (2004). ARB:a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.

Matsuo J, Kawaguchi K, Nakamura S, Hayashi Y, Yoshida M, Takahashi K et al. (2009). Survival and transfer ability of phylogenetically diverse bacterial endosymbionts in environmental *Acanthamoeba* isolates. *Environ Microbiol Rep* **2**: 524–533.

Mba Medie F, Ben Salah I, Henrissat B, Raoult D, Drancourt M. (2011). *Mycobacterium tuberculosis* complex mycobacteria as amoeba-resistant organisms. *PLoS One* **6**: e20499.

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

Nawrocki E, Eddy S. (2009). Structural RNA homology search and alignment using covariance models. *Electronic Theses and Dissertations.* Paper 256. Washington University.

Ogata H, La Scola B, Audic S, Renesto P, Blanc G, Robert C et al. (2006). Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genet* **2**: e76.

Page FC. (1988). *A new key to freshwater and soil Gymnamoebae: with instructions for culture*. Freshwater Biological Association: Cumbria, UK, pp 95–96.

Price MN, Dehal PS, Arkin AP. (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590–D596.

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

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**: 539–542.

Santic M, Akimana C, Asare R, Kouokam JC, Atay S, Kwaik YA. (2009). Intracellular fate of *Francisella tularensis* within arthropod-derived cell. *Environ Microbiol* **11**: 1473–1481.

Sassera D, Beninati T, Bandi C, Bouman EP, Sacchi L, Fabbi M et al. (2006). ''*Candidatus* Midichloria mitochondrii'' an endosymbiont of the tick *Ixodes ricinus* with a unique intramitochondrial lifestyle. *Int J Syst Evol Microbiol* **56**: 2535–2540.

Schmidt HJ, Goertz HD. (1987). *Caedibacter caryophila* sp. nov., a killer symbiont inhabiting the macronucleus of *Paramecium caudatum*. *Int J Syst Evol Microbiol* **37**: 6–9.

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

Schrallhammer M, Schweikert M. (2009). The killer effect of *Paramecium* and its causative agents. In: Fujishima M (ed) *Endosymbionts in Paramecium*. Springer-Verlag: Heidelberg, Germany, pp 227–245.

Shin W, Boo S, Fritz L. (2003). Endonuclear bacteria in *Euglena hemichromata* (*Euglenophyceae*): a proposed pathway to endonucleobiosis. *Phycologia* **42**: 198–203.

Swanson M S, Hammer BK. (2000). Legionella pneumophila pathogenesis: a fateful journey from amoebae to macrophages. *Annu Rev Microbiol* **54**: 567–613.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood evolutionary distance, maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.

Waterhouse AM, Procter J B, Martin DM, Clamp M, Barton GJ. (2009). Jalview version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189–1191.

Wheeler DL, Barrett T, Benson D, Bryant SH, Canese K, Chetvernin V et al. (2008). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **36**: D13–D21.

Williams KP, Sobral BW, Dickerman AW. (2007). A robust species tree for the alphaproteobacteria. *J Bacteriol* **189**: 4578–4586.

Zielinski FU, Pernthaler A, Duperron S, Raggi L, Giere O, Borowski C et al. (2009). Widespread occurrence of an intranuclear bacterial parasite in vent and seep bathymodiolin mussels. *Environ Microbiol* **11**: 1150–1167.

Life in an unusual intracellular niche – a bacterial symbiont infecting the nucleus of amoebae

Frederik Schulz, Ilias Lagkouvardos, Florian Wascher, Karin Aistleitner, Rok Kostanjšek, Matthias Horn

Supplementary Information

Supplementary Text S1

Description of "*Candidatus* <u>Nucleicultrix</u> <u>amoebiphila</u>" (*Nu.cle.i.cul'trix. L. n. nucleus, a nut, and in biology a nucleus; L. fem. n. cultrix, inhabitant; a.mo.e.bi'phi.la. N.L. n. amoeba; N.L. fem. adj. phila, friend, loving; N.L. fem. adj. amoebiphila, amoeba-loving*). Bacteria infecting and replicating in the nucleus of *Hartmannella* and *Acanthamoeba* species; the organism was originally discovered in a *Hartmannella* sp. isolate obtained from a denitrifying bioreactor; it has not been cultured in host-free media. The bacteria appear as coccoid rods of 0.5 to 1 μm in length and 0.3 to 0.4 μm in diameter and show a Gram-negative type cell wall. The organism is a member of the *Alphaproteobacteria*; its classification is based on 16S rRNA and 23S rRNA gene sequences (Genbank acc. numbers KF697195, KF697196) and fluorescence in situ hybridization with the 16S rRNA-targeted oligonucleotide probe CBR125 (5'-TTCACTCTCAAGTCGCCC-3').
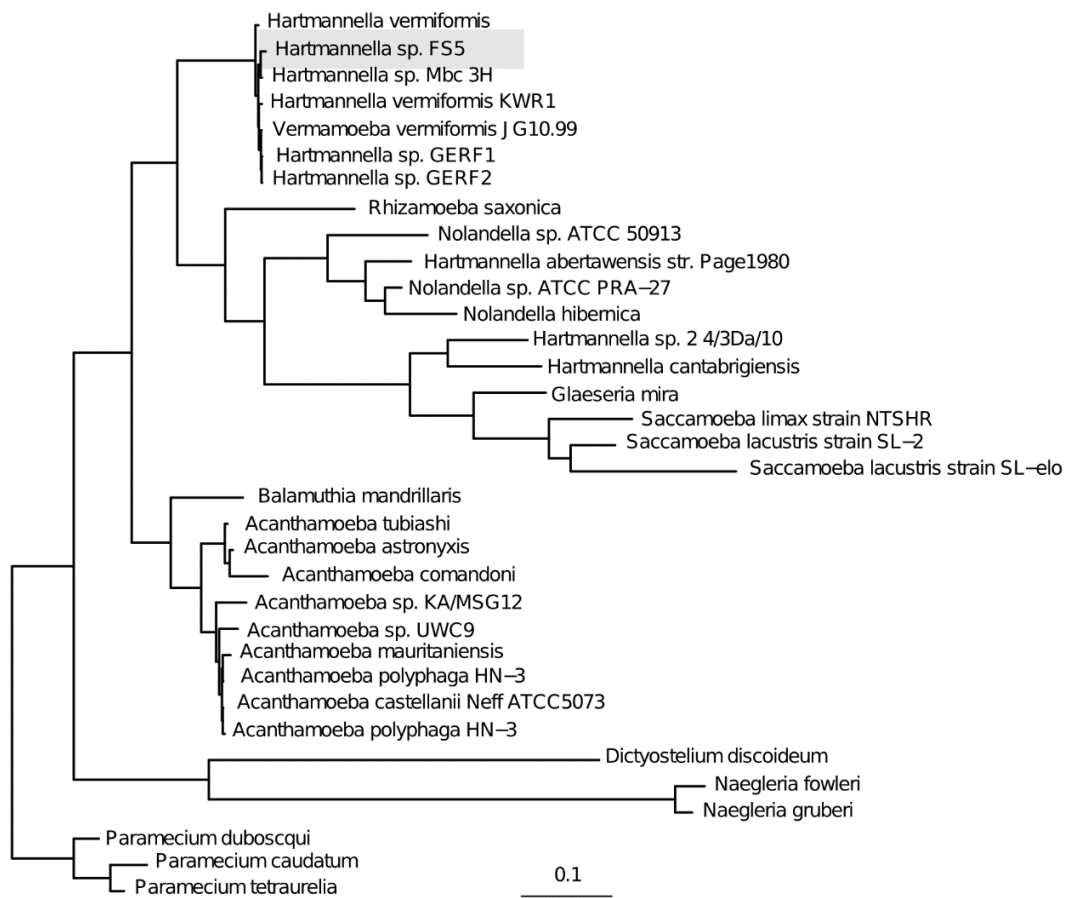
Figure S1. Phylogenetic relationship of *Hartmannella* sp. FS5. The tree is based on a 18S rRNA alignment (Schmitz-Esser, 2008); the isolate *Hartmannella* sp. FS5 and selected relatives in Genbank/EMBL/DDBJ were added, and the tree was calculated with FastTree2 (Price et al., 2010). *Hartmannella* sp. FS5 is highlighted in grey. Accession numbers are provided in Table S1.
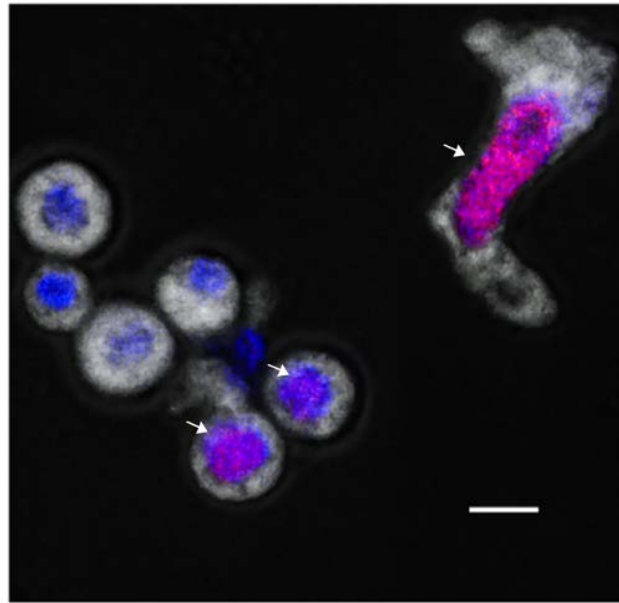
Figure S2. Identification and intranuclear location of "*Candidatus* Nucleicultrix amoebiphila" in the original amoeba isolate *Hartmannella* sp. FS5. *Hartmannella* sp. cells infected by *Nucleicultrix* were stained by DAPI (blue) and FISH using the general bacterial probe mix EUB338 I-III (labeled in red resulting in a pink color due to the overlay with the blue DAPI signal) and the eukaryotic probe EUK516 (shown in grey). Arrows indicate amoeba trophozoites with bacteria in the nuclear compartment. Bar, 5 μm.

Figure S3. 16S rRNA-based maximum likelihood tree showing the relationship of "*Candidatus* Nucleicultrix amoebiphila" with other *Alphaproteobacteria. Nucleicultrix* and its closest relatives are highlighted in grey. Maximum likelihood bootstrap values and Bayesian posterior probabilities are indicated at the nodes.

Figure S4. 23S rRNA-based maximum likelihood tree showing the relationship of "*Candidatus* Nucleicultrix amoebiphila" with other *Alphaproteobacteria*. *Nucleicultrix* and its closest relatives are highlighted in grey. Maximum likelihood bootstrap values and Bayesian posterior probabilities are indicated at the nodes.

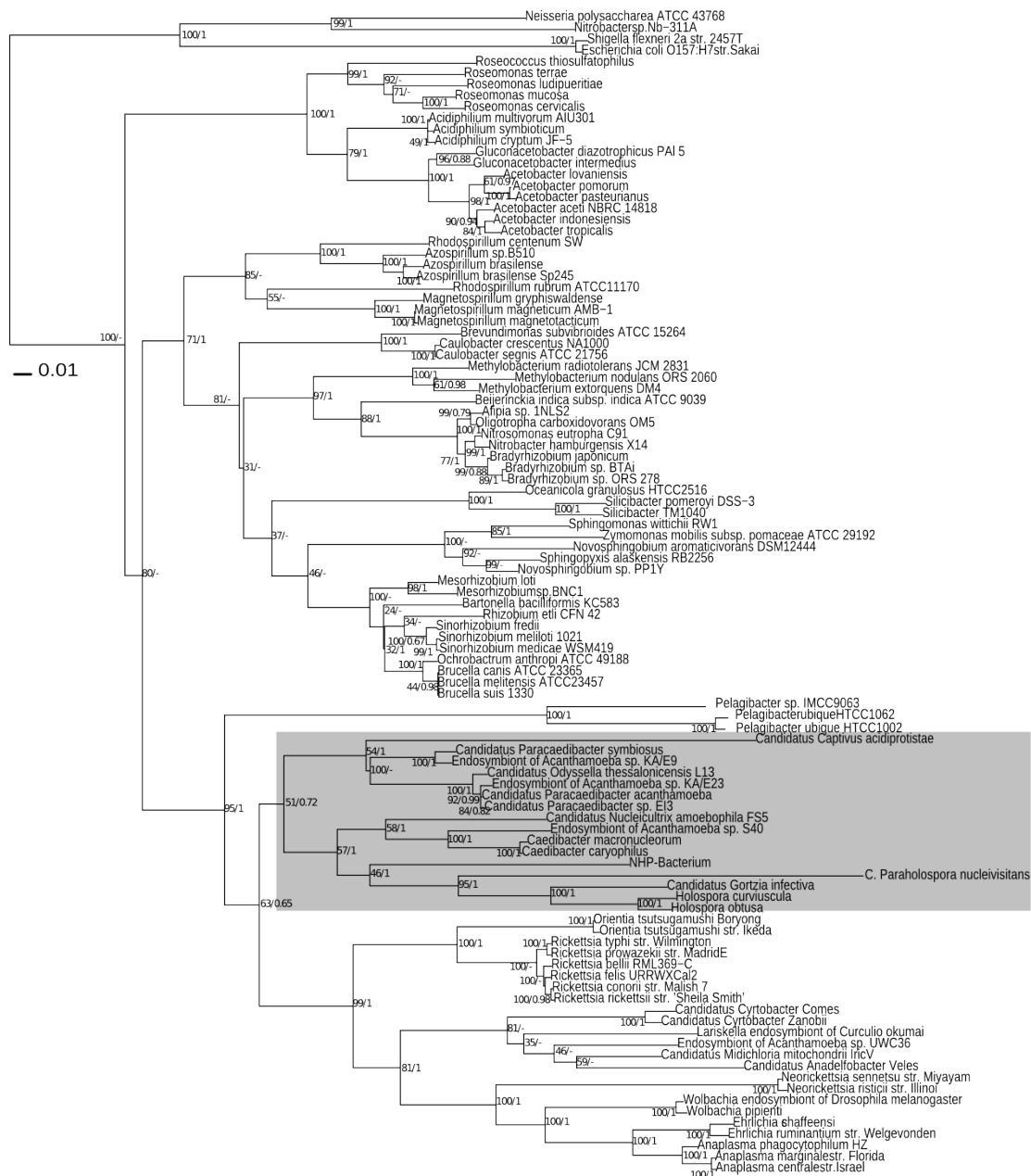**Figure S5. Maximum likelihood tree showing the relationship of "*Candidatus* Nucleicultrix amoebiphila" with other *Alphaproteobacteria* based on concatenated 16S and 23S rRNA.** *Nucleicultrix* and its closest relatives are highlighted in grey. Maximum likelihood bootstrap values and Bayesian posterior probabilities are indicated at the nodes.
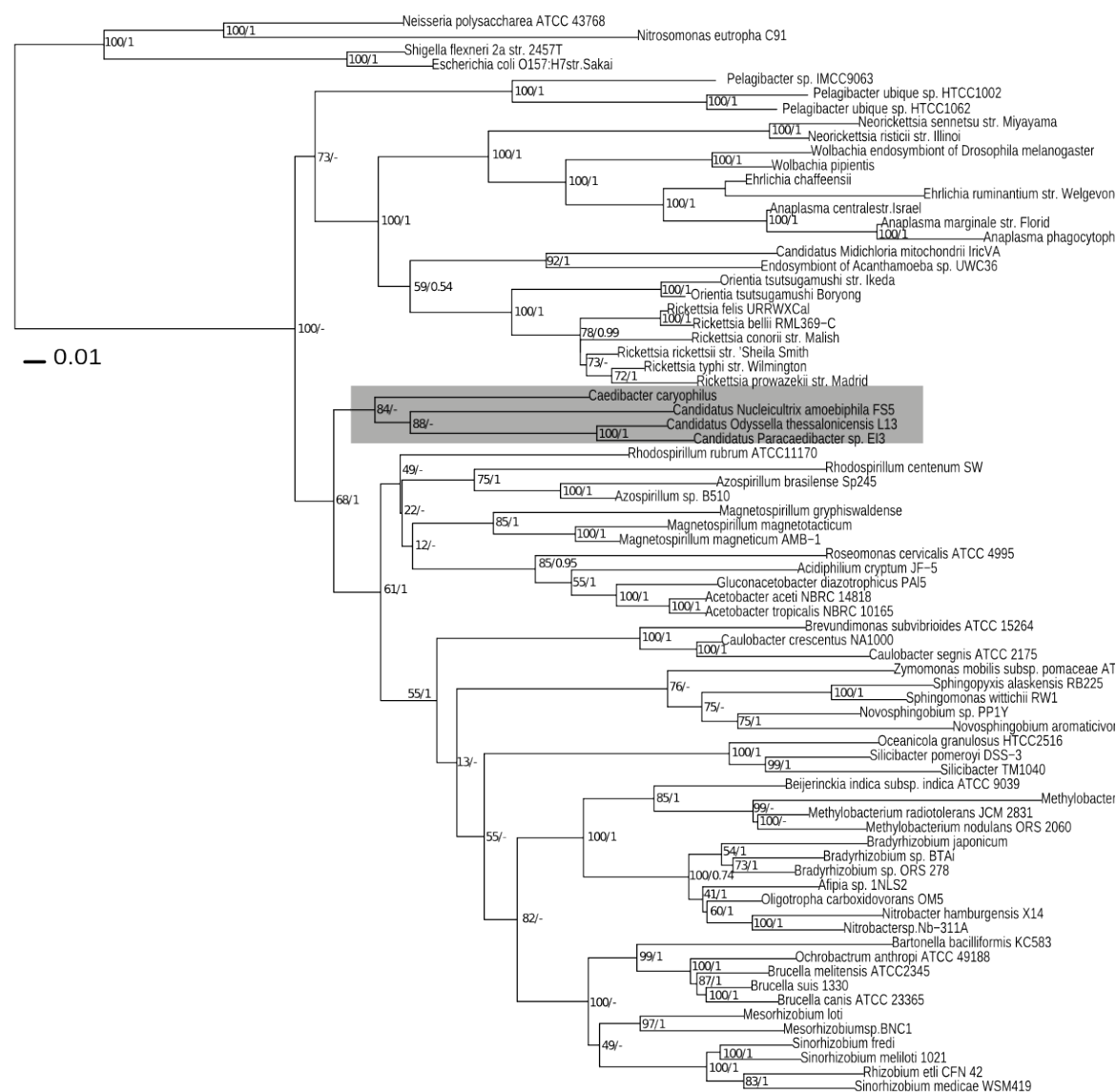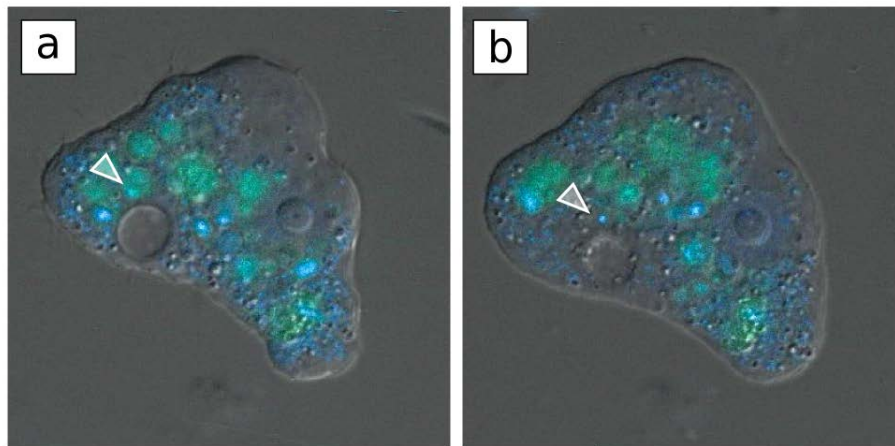
Figure S6. Escape of 'Candidatus Nucleicultrix amoebiphila' from phagolysosomes. *A. castellanii* cells were incubated with LysoTracker-Yellow HCK-123 (Lifetechnologies; 2 μM) in 1xPAS for 1 hour and infected with *Nucleicultrix* as described. The cell suspension was then transferred to a chambered cover glass system (NalgeNunc Int.), and DAPI (0.2 μg/ml) was added. The infection process was monitored over a period of 6 hours with a confocal microscope (Leica SP8). The images shown here display the same amoeba trophozoites approximately 2 h post infection. (a) *Nucleicultrix* (blue, arrowhead) was initially located in phagolysosomes (shown in green). (b) The bacteria remained intact and 5 minutes later escaped into the cytoplasm.
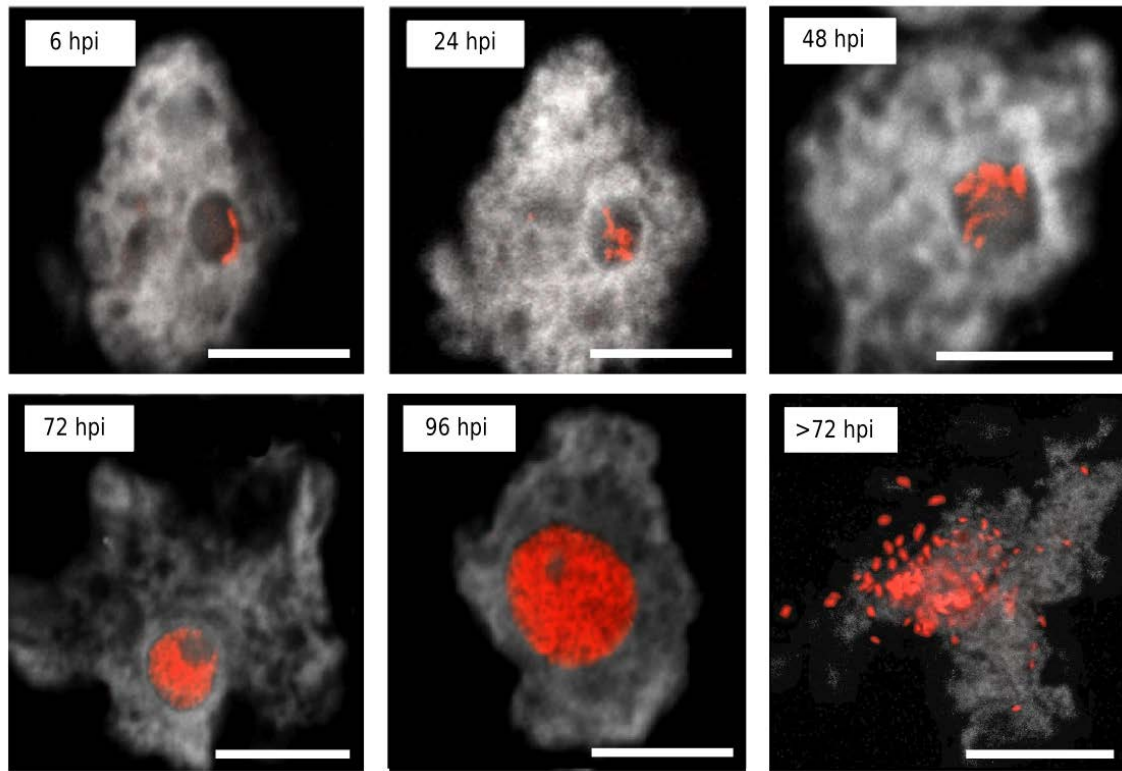
**Figure S7. Infection cycle of '*Candidatus* Nucleicultrix amoebiphila' in *Acanthamoeba castellanii.* The infection process was monitored over a course of 120 hours and visualized by FISH using the *Nucleicultrix*-specific probe CBR125 (shown in orange) and probe EUK516 targeting the amoeba host (shown in grey). Single bacteria colonize the nucleus within 6 hours p.i., the nucleoplasm is completely filled with bacteria after 72 - 96 hours, which is accompanied by significantly enlarged nuclei. Host cell lysis occurs from 72 hours p.i. on. Bar, 10 μm.
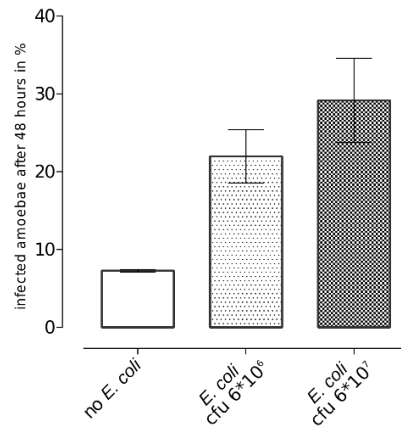
Figure S8. The presence of *E. coli* increases infection rate of '*Candidatus* Nucleicultrix amoebiphila' in *Hartmannella* sp. FS5. The percentage of infected amoebae 48 hours p.i. is shown in the absence or presence of *E. coli* (at two different cell numbers given in colony forming units, CFU). Error bars indicate standard deviation based on three replicate infection experiments.
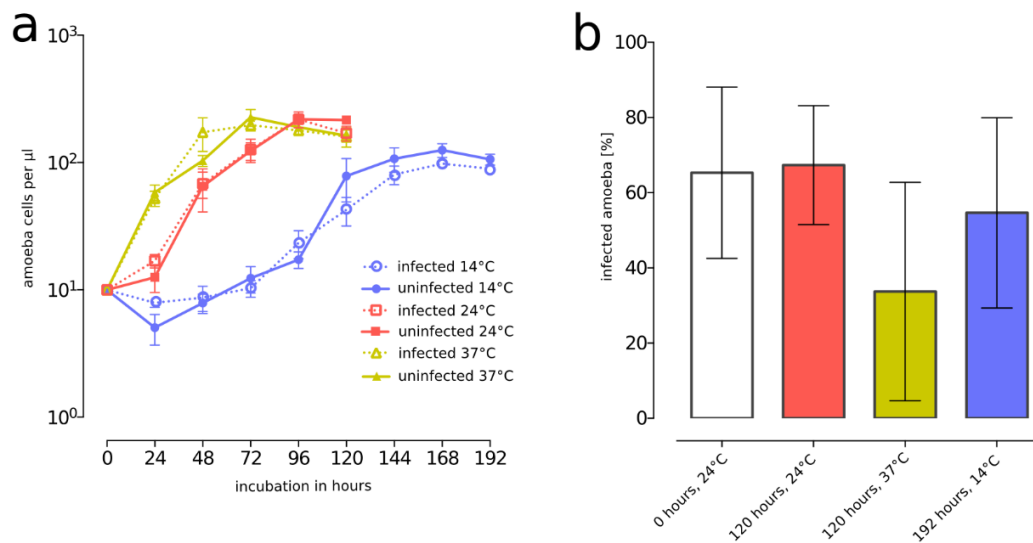
Figure S9. Influence of '*Candidatus* Nucleicultrix amoebiphila' on *Hartmannella* sp. growth at different temperatures. 10^4 amoebae from either infected or uninfected cultures were seeded in 12-well plates, containing 1 ml PAS supplemented with *E. coli tolC-* and ampicillin (200 ng/ml). Cultures were incubated at three distinct temperatures (14°C, 24°C, and 37°C), and cells were detached and counted every 24 hours. As soon as stationary growth phase was reached, amoebae were harvested and fixed with 4% PFA. The percentage of infected amoebae was determined using FISH combined with DAPI staining. (a) Number of amoeba trophozoites with or without symbionts during incubation at 14°C, 24°C, and 37°C, respectively. (b) Percentage of infected amoebae at the beginning and end of the experiment. Error bars indicate standard deviation based on at least three replicate experiments. Differences in amoeba cell numbers between infected and uninfected cultures were not significant at all incubation temperatures ($p > 0.05$; two-way ANOVA).
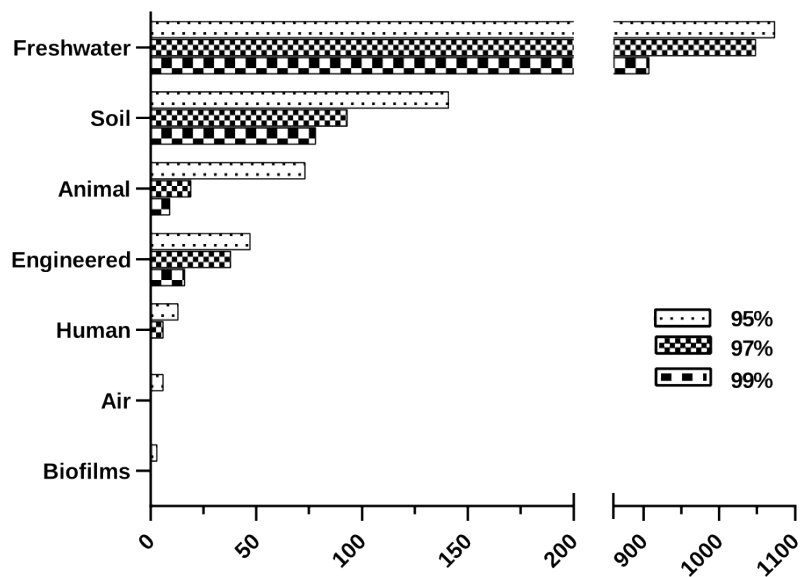
Figure S10. Environmental sequences similar to the 16S rRNA of '*Candidatus Nucleicultrix amoebiphila*'. The number of amplicon sequences (> 200 nt) found in SRA and VAMPs using different similarity thresholds are shown and classified based on their environmental origin. Additional details are provided in Table S2.

Table S1. Genbank/EMBL/DDBJ accession numbers of 16S, 18S and 23S rRNA sequences used for phylogenetic analysis. Provided as Excel file.

Table S2. List of samples in SRA and VAMPS that include sequences similar to the 16S rRNA gene of '*Candidatus* Nucleicultrix amoebiphila'. Provided as Excel file.

References

Price MN, Dehal PS, Arkin AP. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.

Schmitz-Esser S, Toenshoff E, 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.

Chapter V

What it takes to live inside the nucleus - An '-omics' perspective on the intranuclear lifestyle

Draft manuscript

What it takes to live inside the nucleus: an 'omics' perspective on the intranuclear lifestyle

Frederik Schulz[1], Florian Wascher[1], Joran Martijn[2], Thijs Ettema[2], Dörte Becher[3], Stephanie Markert[4], Thomas Schweder[4], Thomas Weinmaier[1], Thomas Rattei[1], Matthias Horn[1]

[1]*Department of Microbiology and Ecosystem Science, University of Vienna, Althanstraße 14, Vienna, Austria*

[2]*Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Husargatan 3, Uppsala, Sweden*

[3]*Institute of Microbiology, Ernst-Moritz-Arndt-University Greifswald, Friedrich-Ludwig-Jahn-Str. 17, 17487 Greifswald, Germany*

[3]*Institute of Marine Biotechnology, Ernst-Moritz-Arndt-University Greifswald, Friedrich-Ludwig-Jahn-Str. 17, 17487 Greifswald, Germany*

Running title: *What it takes to live inside the nucleus*

Abstract

Intracellular bacteria usually thrive in host-derived vacuoles or directly in the cytoplasm, only a few invade unusual niches, such as the eukaryotic nucleus. We now provide a detailed analysis of the genetic repertoire of the obligate intranuclear amoeba symbiont *Nucleicultrix amoebiphila*. We deliver phylogenomic evidence on its position in the *Holosporales*, a distinct clade of intracellular alphaproteobacteria. The symbiont compensates for truncated nucleotide de-novo biosynthesis pathways by encoding for three nucleotide transporters, among them a UTP transporter with a so-far unparalleled substrate specificity. In addition, type IV pili and a competence complex potentially govern the uptake of host DNA. *Nucleicultrix* heavily modifies its cell envelope by glycosylation and by a polysialic acid capsule and encodes type II and type VI secretion systems likely facilitating the translocation of bacterial effector proteins into the host nucleus. *In silico* structure analysis revealed a comprehensive complement of candidate nucleomodulins, such as histone methyl- and acetyltransferases and eukaryotic-like proteins involved in transcriptional regulation, nuclear import/export, mRNA degradation and mitosis. The analysis of the proteomes of infected and uninfected amoeba nuclei could confirm predicted effectors and proteins essential for the assembly of flagellum, pilus and type VI secretion system. The intranuclear infection had a broad impact on host cellular processes, proteins involved in DNA transcription, signal transduction and translation were differentially expressed. This is the first study illustrating metabolic capabilities and nucleomodulins encoded in the genome of an intranuclear bacterium and the impact of an infection on host cellular and nuclear processes. The findings enhance our knowledge of host-symbiont interactions and the intranuclear lifestyle in particular.

Introduction

Symbiotic associations are manifold and affect most aspects of life on earth, in particular symbiosis between bacteria and higher eukaryotes are of central importance; nearly all eukaryotes are associated with bacteria which either beneficially affect their hosts or appear as pathogens (McFall-Ngai et al. 2013). The more tightly linked a symbiont gets to its host, the more it becomes dependent. A concrete example are obligate intracellular bacteria. The availability of nutrients inside the host cell reduces selective constrains and at the same time mutation rates increase which ultimately cause the loss of bacterial genes necessary to perform major metabolic functions (Toft and Andersson 2010). However, despite their often strongly reduced and streamlined genomes, these bacteria encode for a variegated set of genes necessary to maintain the symbiotic relationship, such as transport systems, secretion systems and attendant effector molecules (McCutcheon and Moran 2012; Hentschel et al. 2000). These secreted proteins manipulate the host cell in favor of bacterial replication.

Intracellular bacteria typically replicate directly inside their host's cytoplasm or in host-derived vacuoles. Comparably few cases were described in which bacteria invade other intracellular compartments such as the endoplasmic reticulum, plastids, mitochondria, or the nucleus (Wilcox 1986; Vogt 1992; Sassera et al. 2006; Schulz et al. 2014). Among them, the nucleus seems to be the most widely used niche for bacterial replication. However, the intranuclear life-style is not well understood; more detailed studies exist for *Holospora,* which mainly infects the macronuclei of paramecia (Görtz et al. 1987; Fujishima and Kodama 2012). These symbiont use a macro-molecular structure, called invasion tip, to evade phagocytosis in the host cytoplasm, migrate to the nucleus and invade the nucleoplasm (Fujishima and Kodama 2012). At the moment, little is known about how intranuclear bacteria interact with their hosts on a metabolic level and how the symbiont manipulate host nuclear processes, such as gene transcription and the eukaryotic cell cycle.

We recently identified *Nucleicultrix amoebiphila* (hereafter: *Nucleicultrix*), a bacterium which infects nuclei of amoebae and which can be stably maintained together with its host in laboratory

culture (Schulz et al. 2014). Here we now give a first detailed analysis of the unique genetic repertoire of this obligate intranuclear bacterium. We illuminate how this potential may facilitate virulence, host interaction and securing the symbiont's need for nutrients in the nuclear compartment and show that an intranuclear infection strongly impacts host cellular processes.

Results & Discussion

The genome of Nucleicultrix is reduced in size and has a high coding density

The complete 1.85 Mb genome of *Nucleicultrix* comprises two complete rRNA operons, 15 non-coding RNAs, 44 tRNAs specifying all 20 amino acids, 1850 predicted protein-encoding genes and has an overall coding density of 91.4 % (Figures 1, S1). Although the genome's GC content with in average 39.7 % is in a typical range of *Holosporales* and *Rickettsiales* members capable to infect amoebae, its size is at the lower end (Figure S1). Most amoeba associated alphaproteobacteria tend to have larger genomes than their arthropod associated relatives, with sizes of up to 3 Mb (Georgiades et al. 2011; Wang and Wu 2015; Schulz et al. 2015, Figure S1). One possible explanation for this phenomenon is that by thriving in the cytoplasm of amoeba, a host mainly feeding on microbes present in the environment, the symbionts are frequently exposed to other intracellular bacteria, giant viruses and food bacteria and are thus subject of a frequent inflow of novel genetic material by horizontal gene transfer and recombination (Moliner et al. 2010). Compared to *Nucleicultrix*, the three intranuclear ciliate symbionts *Holospora* spp. have even smaller genomes ((Dohra et al. 2014), Figure S1). It is tempting to speculate that the colonization of a sheltered compartment such as the eukaryotic nucleus limits the exposure to intracytoplasmic gene flow and due to lower selective pressure eventually enhances genome reductive processes.

The intranuclear lifestyle evolved two times independently in the Holosporales

In a previous study the phylogenetic position of *Nucleicultrix* could not be clearly resolved; although the symbiont grouped with the *Rickettsiales* in a phylogenetic tree inferred solely on the 16S rRNA gene and it clustered together with the *Rhodospirillales* if a concatenated alignment of

16S and 23S rRNA was used (Schulz et al. 2014). Here we assessed the phylogenetic position of *Nucleicultrix* based on a concatenated alignment of 38 carefully evaluated single copy marker proteins. The symbiont represented a sister lineage to the intracytoplasmic amoeba symbiont *Caedibacter acanthamoeba* (Bayesian posterior probability 1.0) in the *Holosporales* (Figure 1), which grouped with the *Rhodospirillales* (Bayesian posterior probability 0.91). In general, resolving the phylogeny of the *Alphaproteobacteria* is a challenging task, mainly due to the presence of free-living bacteria and obligate intracellular symbionts which are subject of highly accelerated rates of evolution and feature extremely reduced genomes with high AT-content (Williams et al. 2007; Viklund et al. 2013). To alleviate artifacts caused by compositional bias and long branch attraction, we applied a two-step filtering process, consisting of 1) a discordance filter which removes genes which eventually have been affected by horizontal gene transfer and 2) a chi2 filter (Viklund et al. 2012), which reduces compositional bias by removing the highly biased sites in the alignment. To allow the CAT model to optimally account for the base heterogeneity in the alignment we included a similar number of representatives from each of the major alphaproteobacterial taxonomic groups and did not include mitochondrial genomes, which due to their extremely divergent sequences likely inflate treeing artifacts (Wang and Wu 2015). Moreover, we constructed phylogenetic trees without the deeply branching *Holospora* lineages, which however had no remarkable influence on the branching order. Compared to previous studies, our analysis included the most comprehensive genomic data set of the *Holosporales* (n=9) to date (Figure 1). Nevertheless, it represents only snapshot of the actual diversity of the *Holosporales* (Boscaro et al. 2013a; Schulz et al. 2014), and we thus expect this symbiont clade to become further elaborated in the near future. Of note, the inferred alphaproteobacterial phylogeny contradicts a common evolutionary origin of the *Holosporales* with the *Rickettsiales*, which was suggested in recent phylogenomic and 16S rRNA gene-based studies (Boscaro et al. 2013a; Wang and Wu 2015). The observed topology rather suggests a scenario of two independent origin of the obligate intracellular lifestyle out of the *Rhodospirillales*.

Nucleicultrix shares only few genes with other intranuclear alphaproteobacteria

Clustering of gene families of 80 mostly alphaproteobacterial genomes revealed orthologs for 82 % of the symbiont's genes (n=1508)(Figure 1). *Nucleicultrix* shared most genes and had the

highest overall proteome sequence similarity with cytoplasmic amoeba symbionts in the *Paracaedibacteriaceae* (up to 50 % and 32 %, respectively) and its free-living relatives in the *Rhodospirillales* (49 % and 30 %, respectively) (Figure 1). The symbiont had a much lower number of orthologs in common with *Holospora* species (up to 28 % and 16 %, respectively) and arthropod associated rickettsiae (up to 35 %) (Figures 1, S2). In total 372 genes were shared among all members of the *Holosporales* (Figures 2A, S2). To better understand the unique genetic repertoire of *Nucleicultrix* we next focused on genes not found in other *Holosporales* members; most of these "accessory" genes were either unique to *Nucleicultrix* (n=294) or at least not present in any other alphaproteobacterium in our data set (n=178), followed by genes exclusively shared with free-living alphaproteobacteria (n=183) or with the *Rickettsiales* (n=114) (Figure 2). The majority of genes in the *Nucleicultrix* accessory genome could functionally not be annotated (n=648) and several completely lacked homology to any sequence deposited in the NCBI nr database (n=109) (Figure 2A). To better understand the unique coding potential of *Nucleicultrix*, we assigned functional categories to genes in the non-core genomes of *Holosporales* members. Surprisingly, the following comparative analysis revealed the overall pattern found in *Nucleicultrix* to be most similar to the one in the cytoplasmic shrimp pathogen *Hepatobacter penaei* (Nunan et al. 2013) (Figure 2B). In both organisms genes likely involved in cell wall/membrane/envelope biogenesis were in particular enriched (n=108), while genes playing a role in replication, recombination and repair (n=87) were underrepresented. Compared to other *Holosporales* members, *Nucleicultrix* possesses many genes implicated in carbohydrate transport and metabolism (n=103), inorganic ion transport (n=91), lipid transport and metabolism (n=76) and posttranslational modifications and protein turnover (n=54). Interestingly, despite occupying the same niche but in different hosts, *Nucleicultrix* and *Holospora* spp. exclusively shared only 20 genes (Figure S2), such as methyl- and acetyltransferases, transposases, putative phosphate and peptide transporters and several enzymes involved in glycosylation (Supplementary table S1).

Nucleicultrix possesses enzymes needed for energy generation and can store carbon

The *Nucleicultrix* genetic repertoire comprises all enzymes necessary for energy generation (Figure 3); it can perform glycolysis, citric acid cycle and oxidative phosphorylation. Similar to

most members of the *Holosporales*, but contrary to many other obligate intracellular bacteria, the symbiont encodes for two aldehyde dehydrogenases (FS5_A_06640, FS5_A_10820) and a acetyl-CoA synthetase (FS5_A_00690), allowing to convert aldehyde to acetate and to utilize acetate to acetyl-CoA. With its acetyl-CoA carboxylase the symbiont can provide malonyl-CoA as substrate for fatty acid biosynthesis. Interconversions of sugars can be achieved by the non-oxidative branch of the pentose phosphate pathway and phosphoribosyl pyrophosphate can be synthesized. In general, the genome of *Nucleicultrix* contains an exceptionally comprehensive arsenal of enzymes involved in carbohydrate and aminosugar metabolism. It can potentially convert fructose via mannose to fucose and also to mannan and D-Glucose-1P to UDP-glucose and then either with UDP-glucose 6-dehydrogenase to UDP-D-glucuronate or with a UDP-glucose 4-epimerase to UDP-galactose. UDP-galactose is an important building block of carbohydrate polymers, such as the bacterial virulence determinants LPS and exopolysaccharide (Misra et al. 2014). Moreover, with its two poly-beta-hydroxybutyrate synthases (FS5_A_00900, FS5_A_08910) the symbiont is able to store carbon in form of the polymer poly-hydroxybutyrate (PHB). The accumulation of this polymer as an energy reserve for the survival under nutrient-limiting conditions is a feature mostly found in free-living bacteria, but also facultative intracellular symbionts such as *Legionella* spp. (James et al. 1999).

Nucleicultrix has capacities to take up amino acids and nucleic acids from its host

One common characteristic of the obligate intracellular lifestyle is the absence of essential metabolic pathways (Toft and Andersson 2010). This also applies for *Nucleicultrix*; despite the presence of enzymes necessary for the biosynthesis of several amino acids, none of the essential ones can be build (Figure 3). To compensate, the symbiont potentially can make use of 8 putative amino acid transport systems. Besides its detrimentally reduced potential to synthesize amino acids, *Nucleicultrix* is also devoid of enzymes involved in purine and pyrimidine de-novo synthesis. However, nucleotide interconversions are possible, a CTP synthase (FS5_A_09960) likely allows the symbiont to convert UTP to CTP, a uracil phosphoribosyltransferase (FS5_A_09220) the creation of UMP from uracil and with its nucleoside triphosphate pyrophosphohydrolase (FS5_A_09050) the symbiont can potentially hydrolyze all eight ribo- and deoxynucleoside triphosphates to their respective monophosphates. Considering the

intranuclear lifestyle, an attractive substrate for intranuclear bacteria are nucleic acids, either as high molecular DNA and RNA or in the form of single nucleotides. In theory *Nucleicultrix* can take up both, 1) it harbors genes involved in the assembly of two type IV pili and a competence complex, which previously have been shown to function together in the uptake of DNA and RNA (Chen et al. 2005; Muschiol et al. 2015) and 2) the symbiont encodes for three putative nucleotide transporters, which are a common feature of obligate intracellular bacteria enabling the uptake of host nucleotides as an energy source (Linka et al. 2003). Considering the uptake of high molecular DNA only fragments in a certain length can be imported (Finkel and Kolter 2001). Free-living bacteria achieve this with the help of extracellular endonucleases which pre-digest high molecular DNA (Gödeke et al. 2011). However, a scenario in which the symbiont directly is gnawing on the host chromosomes would likely have a devastating impact on the host and thus contradict previous experimental findings on the remarkably stable association of *Nucleicultrix* with its native *Hartmannella* host (Schulz et al. 2014). Nevertheless, the uptake of high molecular (ribo)nucleic acids might still be a strategy applied by *Nucleicultrix*; one possible scenario could be the utilization of RNA intermediates, such as intronic RNAs after splicing. However, the most convincing strategy would simply be the uptake of single nucleotides. We thus recently characterized the three putative nucleotide transporters encoded in the symbiont's genome (Schulz et al., in prep). Besides one typical ATP/ADP translocase and one carrier which takes up different nucleotides at very low rates, *Nucleicultrix* possesses a transporter capable to import UTP with exceptionally high affinity and specificity (Schulz et al., in prep). The establishment of such an effective uptake system definitely brings an evolutionary advantage, considering the ample availability of UTP at the place of DNA transcription.

Nucleicultrix uses a versatile set of transporters to upgrade its metabolic capacities

In general, the eukaryotic nucleus is a niche rich in nutrients, such as amino acids, phosphate, peptides, fatty acids, lipids, free nucleic acids and high molecular RNA and DNA. It is thus not surprising that the *Nucleicultrix* recruits a plethora of transporters allowing to tap from this land of plenty (Figure 3). With a sugar importer (FS5_A_07340), a homolog of the UhpC transporter which was shown to facilitate uptake of glucose-6-phosphate in chlamydiae (Schwöppe et al. 2002), and a PTS system (FS5_A_10870, FS5_A_10880), the symbiont can potentially take up

alternative carbon sources. Moreover, it possesses two putative C4-dicarboxylate carrier, likely mediating the uptake, exchange or efflux of aspartate, malate, fumarate and succinate (FS5_A_16450, FS5_A_12730). Additional carbon sources for bacterial growth can potentially be made available with the help of a homolog of FadL, a transporter facilitating the import of long-chain fatty acids (van den Berg 2005), and a homolog to UgpB, allowing the uptake of Glycerol-3-phosphate, a degradation product of phospholipids (Wuttge et al. 2012). The symbiont's genome also harbors a suite of genes necessary to assemble PstSCAB, a high-affinity transport system for phosphate (Rao and Torriani 1990). Similar to other obligate intracellular bacteria, *Nucleicultrix* can import the co-factor and methyl group donor S-adenosylmethionine (SAM) via a SAM transporter (FS5_A_07260) (Haferkamp et al. 2013). Finally, *Nucleicultrix* also encodes for a putrescine transporter system potFGHI, which might directly impact polyamine-mediated chromatin remodeling in the nucleus by tapping putrescine (Pasini et al. 2014). With its homospermidine synthase (FS5_A_00310) the symbiont is potentially able to convert putrescine to nitrate and *sym*-homospermidine. Both polyamines, *sym*-homospermidine and putrescine, have been shown to positively influence transcriptional and translational processes in the bacterial cell and enhance virulence (Shah and Swiatlo 2008; Nasrallah et al. 2011) .

Nucleicultrix can persist and withstand stress and adverse environmental conditions

The life cycle of *Nucleicultrix* can be divided into 1) persistence in host-free environments, 2) uptake into the host cell, 3) escape from the phagosome, 4) intracellular trafficking through the host cytoplasm, 5) invasion of the nuclear compartment 6) bacterial replication inside the nucleus and 7) exit from the intracellular niche upon eventual host cell lysis (Schulz et al. 2014). The symbiont encodes a vast number of genes to accomplish these major steps. In the course of horizontal transmission, *Nucleicultrix* supposedly has to survive extended periods outside its host. To ensure bacterial cell integrity and the availability of energy needed to initiate infection upon contact with a potential host cell, the symbiont likely utilizes stored PHB granules with its intracellular PHB depolymerase (FS5_A_11490). In addition, the genome of *Nucleicultrix* is surprisingly enriched in cation efflux transport systems (Figure 3), contributing to ion homeostasis and thus being essential for persistence and stress resistance under adverse

environmental conditions and for survival in the host cytoplasm (Agranoff and Krishna 1998). The symbiont encodes for several antioxidants; a co-enzyme F420-dependent NADP oxidoreductase, the super-oxide dismutase SodB (FS5_A_15050), two alkyl hydro-peroxide reductases (FS5_A_01020,FS5_A_04010) and a homolog to the peroxiredoxin OsmC (FS5_A_16760). By securing bivalent cations, *Nucleicultrix* can equip redox enzymes with necessary co-factors and thus further reduce oxygen stress in the intracellular milieu (Agranoff and Krishna 1998). The following escape from the phagosome is likely triggered by the expression of genes encoding for phosphoplipases, such as six patatins (FS5_A_05160, FS5_A_05380, FS5_A_05530, FS5_A_07360, FS5_A_08590, FS5_A_13590) and three enzymes with homology to phospholipase D (FS5_A_05210, FS5_A_07910, FS5_A_17180). It has been shown that in particular rickettsial patatins may function as phospholipases, share a common phylogenetic origin with patatins found in other obligate intracellular bacteria, and are possibly involved in escape from the phagosome and host cell lysis (Rahman et al. 2013).

A flagellum and multiple pili as possible key determinants of virulence

The symbiont's genome harbors an elaborated suite of genes necessary to assemble a flagellum (Figure 4). Although flagellar genes could not be detected in *Holospora* spp., a flagellum is encoded in the genomes of all other *Holosporales* members and furthermore, *Odyssella thessalonicensis* and *H. penaei* have been described as being motile (Birtles et al. 2000; Nunan et al. 2013; Wang and Wu 2014). Flagella have also been observed in protist-associated *Rickettsiales* members (Boscaro et al. 2013b; Schulz et al. 2015; Martijn et al. 2015) and flagellar-based motility potential plays an important role in symbiont transmission in aquatic habitats, where in particular *Nucleicultrix* and close relatives have been described to occur (Schulz et al. 2014). However, the exact role of the flagellum, such as establishing contact with the host cell, facilitating phagosomal escape or triggering host exit, remains currently elusive. Another key feature for host interaction are the pili encoded in the *Nucleicultrix* genome (Figure 4). They are organized in two operons, one consisting of the genes cpaABCDEF, tadBC and two pilins and the other comprises two copies of tadE, tadG and one pilin. Type IV pili potentially contribute to a wide range of cellular processes, such as protein secretion, DNA uptake and competence, adherence, motility and pathogenesis (Craig et al. 2004; Giltner et al. 2012).

Nucleicultrix extensively upgrades its cell envelope

A defining feature of *Nucleicultrix* is its highly modified cell envelope (Figure 4). The symbiont encodes for the enzymes lptABCDEF (FS5_A_03700, FS5_A_03710, FS5_A_03690, FS5_A_01720, FS5_A_00050, FS5_A_01740) and capDIL (FS5_A_05020, FS5_A_09810, FS5_A_09800), which are necessary to synthesize a capsule and lipopolysaccharide (LPS). Moreover, a surprisingly large number the symbiont's genes are potentially involved in the glycosylation of the symbiont's cell envelope, its pilus and its flagellum. Many of these genes are found on a 65 kb large genomic region with a reverse GC skew, an atypical nucleotide composition and is associated with IS elements and tRNAs at its ends. We analyzed synteny and phylogeny of genes on this putative genomic island and could show that several of them are found in diverse bacterial lineages, such as *Legionella pneumophila*, *Chromobacter violaceum*, *Pelodictyon phaeclathratiforme* and *Nitrobacter winogradsky*, but not in other members of the *Holosporales* or *Rickettsiales* (Figure S3). Many of the encoded genes are likely involved in the generation of a polysialic acid capsule, such as diacetyllegionaminic acid synthases (FS5_A_16360, FS5_A_16290) and N,N'-diacetylbacillosamine 2-epimerase (FS5_A_16320). The addition of sialic acids to surface structures is a strategy frequently applied by intracellular pathogens; it promotes the infection of a eukaryotic host by mimicking surface patterns of the host cell and thus significantly contributes to host invasion, protection from digestive processes in the cytoplasm and evasion of the host immune system (Severi et al. 2007). As an additional modification, *Nucleicultrix* encodes for a homolog of PE-PGRS (FS5_A_00070), a surface antigen reportedly involved in blocking phagosome maturation and inhibiting iNOS expression in the host cytoplasm and eventually also in the nucleus (Thi et al. 2013, Figure 4).

A transmembrane DNA binding protein potentially involved in the invasion of the nucleus

After escape into the host cytoplasm, *Nucleicultrix* has to reach the nuclear compartment. It has been suggested, that the symbiont migrates through the cytoplasm and approaches the interior of the nucleus during open mitosis of the host (Schulz et al. 2014). *Nucleicultrix* possesses two proteins with eukaryotic-like structure potentially impacting the host cytoskeleton; one actin interacting protein (FS5_A_10200) and one vinculin (FS5_A_10200). Induction of actin polymerization and cytoskeletal rearrangements are known triggers of uptake and intracellular

motility of bacterial pathogens, among them *Holospora* species and rickettsiae, and in particular molecular mimicry to subvert vinculin has been shown before (Izard et al. 2006; Park et al. 2011; Sabaneyeva et al. 2009). Interestingly, *Nucleicultrix* encodes for a homolog of the histone H1-like nucleoprotein HC2 found in chlamydiae (FS5_A_01890, Brickman et al. 1993). It was hypothesized that this protein has similarity to the eukaryotic histone H1 and it was shown to have DNA binding activity and later that it is involved in DNA condensation during the chlamydial developmental cycle (Grieshaber et al. 2004). In contrast to its chlamydial counterpart, the protein encoded by *Nucleicultrix* is larger (258 amino acids) and expresses three transmembrane regions at its 5' end, while the HC2 domain is located at the 3' end (Figure S4). A possible scenario could be that this protein resides in the symbiont's outer membrane and confers binding to the host chromosomes and thus allows the symbiont to establish intranuclear infection during open mitosis and to stably remain in the nucleus of dividing host cells.

Nucleicultrix encodes for a novel type VI secretion system

To ensure successful bacterial replication, *Nucleicultrix* has to interact with its host cell. To accomplish this task, the symbiont comprises all genes necessary for a type II secretion system (T2SS), which facilitates the translocation of proteins into the periplasm either by the *sec* or the *tat* system, and then with the help of pore-forming secretins (FS5_A_07160, FS5_A_10540) across the outer cell membrane. Furthermore, the symbiont encodes for a conjugative type IV secretion system (*tra*), a feature frequently found in intracellular bacteria, such as environmental chlamydiae and some rickettsiae (Collingro et al. 2011; Gillespie et al. 2015; Schulz et al. 2015). However, the functionality of this system could experimentally not be validated in neither one of these obligate intracellular bacteria. Lastly, *Nucleicultrix* encodes for all proteins needed to deploy a type VI secretion system (T6SS) (Figure 4, Figure S5). By injecting effector proteins T6SS are crucial for virulence of several bacterial pathogens, such as *Salmonella*, *Francisella* and *Pseudomonas*, and they also facilitate interactions among bacteria (Ho et al. 2014). In contrast to known members of the *Rickettsiales*, which deliver effector proteins with the help of a *vir* type IV secretion system (T4SS) (Gillespie et al. 2010), the T6SS appears as distinctive feature of the *Holosporales*, but not of *Holospora* species. The genes encoding for this macromolecular structure are split in multiple genetic loci (Figure S5). Strikingly, in one of these

operons, adjacent to the gene encoding for T6SS extracellular component VgrG, a putative effector molecule is encoded containing two MORN repeats domains,, which are eukaryotic like repeat domains potentially governing nuclear segregation (Gubbels et al. 2006). *In silico* structure prediction could show that this protein is most similar to a SET7/9 histone methyltransferase, however with only low confidence (c-score = -2.5). Furthermore, *Nucleicultrix* also possesses a putative histon deacetylase (FS5_A_18260) with a typical domain structure (Hist_deacetyl, PF00850.14), and a GCN5-related N-acetyl-transferase with highest structural similarity to a putative histone acetyltransferase (FS5_A_00390). None of these three putative effectors contains signal peptides, suggesting that these proteins may be injected into the host cell by the T6SS.

Nucleicultrix encodes for a so far unseen number of potential nucleomodulins

The secretome of *Nucleicultrix* comprises a compelling array of candidate effector proteins (n = 424). To not only limit functional annotation of these proteins to sequence homology and the presence of distinct PFAM domains, we conducted an *in silico* structure prediction with I-TASSER and COACH (Yang et al. 2015). This approach helped to identify several potential nucleomodulins (Table 1); one of them (FS5_A_07780) with a high structural similarity to a TIP120 domain containing protein, which is a TATA box binding transcriptional enhancer and functions as a global regulator of gene expression (Makino et al. 1999). Furthermore, a protein (FS5_A_14450) possibly mimics the function of the a Ski2-like helicase, a factor involved in the unwinding of double-stranded DNA (Fairman-Williams et al. 2010). Moreover, we found in total three proteins with similarity to nuclear import and export factors; two putative exportins (FS5_A_07770, FS5_A_17880), and a Snurportin-1-GTP-binding nuclear protein Ran-Exportin-1 complex (FS5_A_07940). Two additional candidate nucleomodulins showed structural similarity to a O-linked N-acetylglucosamine (GlcNAc) transferase (FS5_A_04260), which by glycosylation of the nuclear pore complexes potentially affects nucleocytoplasmic transport processes (Li and Kohler 2014) and a Ras GTPase-activating-like protein, which might get implemented into the contractile F-actin ring during cell division and a leucine-rich repeat containing protein (FS5_A_15450) potentially mimicking the GTP-binding nuclear protein Ran, which is regarded as key manipulator of various intranuclear processes (Silverman et al. 2012;

Güttler and Görlich 2011). Taken together, *Nucleicultrix* encodes for a so-far unseen wealth of intranuclear effectors; hijacking regulatory mechanisms of gene expression and RNA turnover, interfering with cytokinesis and taking over control of nuclear import and export of proteins and RNA.

The Nucleicultrix proteome is enriched in factors mediating host interaction

In order to identify which of the predicted features actually play a role during infection, we analyzed the proteome of *Nucleicultrix* inside *Acanthamoeba castellanii* nuclei with a LC-MS/MS-based approach. Infected nuclei were isolated and measured and a fraction of abundant symbiont proteins (n=118) were found expressed. Most of them (n=68) were widely distributed in the alphaproteobacterial genomes in our data set (>75 %) (Figure 5). These proteins are typically involved in the maintenance of basic cellular functions; either as ribosomal proteins (n=30) or as enzymes facilitating transcription, replication and energy generation. In total 14 of the detected proteins contained transmembrane helices and are thus likely membrane-bound. Four of them showed the by far highest detected normalized spectral abundance factor (NSAF%), in particular the outer membrane components Omp16 and OmpA, the cell envelope antigen PE-PGRS and the long-chain fatty acids transporter FADL (Figure 5). Moreover, the presence of several macromolecular structures potentially involved in host invasion and effector delivery could be confirmed; three components of the T6SS, namely Hcp and the ImpJ and ImC homologs, Flagellin, which is the extracellular compound of the flagellum, and the Flp pilus protein CpaD. Strikingly, 39 (76 %) of the 51 expressed but less commonly distributed proteins contained a signal peptide and thus represent a part of the symbiont's secretome. Among them were patatins (FS5_A_05160, FS5_A_07360), a thermolysin (FS5_A_12840), a hemolysin (FS5_A_03910) and a cytolysin (FS5_A_17560), all enzyme likely contributing to either phagosomal escape or host cell lysis (Schnepf et al. 2005; Vandenesch et al. 2012; Rahman et al. 2013). Intriguingly, several of the *in silico* predicted intranuclear effectors were found in the proteome (6 of 13). Considering the relatively low sensitivity of our approach, we expect additional potential nucleomodulins to be actually expressed but in abundances below the detection threshold. To note, the sequences of four of the five expressed intranuclear candidate effectors have absolutely no similarity to known proteins.

The impact of the intranuclear infection on host cellular processes

To better understand the host reaction to an intranuclear infection, we compared the proteomes of infected and uninfected *Acanthamoeba* nuclei. In total 435 proteins could be retrieved in the proteome of symbiont-free nuclei and 561 in infected nuclei. A tremendous number of proteins involved in translational processes and posttranslational modifications were up-regulated, whereas many proteins conducting transcription, RNA processing and signal transduction mechanisms were down-regulated (Figure 6). Furthermore, proteins important for cytoskeletal structure and dynamics, carbohydrate transport and metabolism and energy production and conversion were differentially expressed (Figure 6). The intranuclear infection had strongest impact on two EF-hand domain proteins (ACA1_391420, ACA1_368250), likely involved in calcium homeostasis and signal transduction, both were strongly down-regulated (11 and 16 fold, respectively) in infected *Acanthamoeba* nuclei whereas a calponin and calreticulin, which potentially scavenge Ca^{2+} were found up-regulated (3 and 5 fold, respectively) (Figure S6). Interestingly, one of the proteins highly expressed in the symbiont's proteome was an unusual bacterial EF-hand domain containing protein (FS5_A_11940), suggesting a modulation of the eukaryotic cell signaling by bacterial mimicry (Figure 5). The regulation of the Ca^{2+} flux affects most intracellular processes and also plays a crucial role in the induction of cytoskeletal rearrangements (Berridge et al. 2003; Clapham 2007). Along with this we found proteins involved in actin polymerization and depolymerization differential expressed in the amoeba proteome, such as subunits of the ARP2/3 complex (2 fold down-regulated) and Fascine (3 fold down-regulated), likely resulting in drastic changes in the regulation of host cell motility. The suppression of host gene expression was demonstrated by an 11-fold lower NSAF of the DNA directed RNA polymerase (ACA1_162130) and also two subunits of the splicing factor 3b (ACA1_064880, ACA1_039280) which were both about 5 fold lower in infected cells. Importantly, besides the reduced abundance of Histone H2A and Histone H4 (1.6-fold and 2-fold down-regulated, respectively) we did not detect a major impact of an intranuclear infection on the histone content in the nucleus. This finding contributes to the hypothesis that *Nucleicultrix* does not degrade host chromatin. Moreover, an up-regulation of several host stress-related proteins, such as a high molecular weight heat shock protein (4-fold), a chaperonine (5folg), Hsp20 and Hsp90 (2-fold) could be observed, likely being the natural consequence of the burden of an intranuclear infection.

Conclusion

This is the first detailed analysis of the genome and proteome of an obligate intranuclear bacterial symbiont. We expected to find common features in the genomes of *Nucleicultrix* and *Holospora* species, as both represent members of the *Holosporales* and share the protist nucleus as niche for replication. However, the genomes of these two symbiont lineages were rather divergent. Intriguingly, for the proteins expressed in the *Nucleicultrix* proteome which potentially facilitate symbiosis, no orthologs exist in the *Holospora* genomes, indicating a fundamentally different set of genes involved in host interaction. Recently it was shown for several intracellular pathogens, that the usage of effectors targeting the eukaryotic nucleus is a rather widespread phenomenon (Bierne et al. 2012; Rennoll-Bankert et al. 2015). Here we now for the first time illustrate that a bacterium which directly resides in this compartment takes advantage of the location and likely modulates the host directly out of the cellular control center. Considering the vast number of effectors found in the symbiont's proteome, it is not surprising that there is a dramatic impact on a wide-range of cellular processes. A further experimental evaluation of the determined candidate effectors will help to better understand how the life in this narrow eukaryotic niche shaped the evolution of bacterial effector molecules with a likely impact on the current understanding on nucleomodulins widely used by cytoplasmic pathogens.

Methods

Genome sequencing & Assembly

Cells and DNA were prepared as previously described (Schmitz-Esser et al., 2010). Paired-end Illumina HiSeq sequencing and Pacbio sequencing were performed at the Norwegian Sequencing Center (Oslo, Norway) using 1 μg template DNA each. Prior to assembly, the quality of the sequence data was assessed with FastQC 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). By combining Illumina reads (>50 bp read length) and Pacbio reads (>2 kb read length) genome was assembled de novo using SPAdes 3.0 (Nurk et al. 2013) with k-mer sizes (21, 33, 55, 63), and using the 'error correction', 'careful' and 'pacbio' flags. The resulting assembly contained three contigs (coverage > 500), which could be manually merged by inserting the rRNA operon twice in reverse complementary

direction. The quality of the assembly was evaluated by mapping back the reads on the genome sequence, followed by manual inspection of the coverage distribution in TABLET (Milne et al. 2013) and evaluation of the assembly with QUAST (Gurevich et al. 2013).

Genome annotation and comparative genomics

The genome of *Nucleicultrix* was annotated with an in-house annotation pipeline (Schulz et al., 2015). Genome data was visualized with Circos (Krzywinski et al. 2009). This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession [SUB552699 PRJNA251793]. Clustering of orthologous genes was performed using orthoMCL 2.0 on a dataset of 84 mainly alphaproteobacterial genomes (Li et al. 2003) with settings and downstream analysis as described previously (Schulz et al., 2015). Orthologs for *Nucleicultrix* genes found in selected reference genomes were visualized with Circos (Krzywinski et al. 2009). Genome statistics, such as GC content, coding density, proteome sequence similarity, shared and unique orthologs, were assessed with custom python scripts and visualized with the matplotlib graphics environment in (Hunter 2007). Syntenies of T6SS genes and genes encoded on the putative genomic island were visualized in genoplotR (Guy et al. 2010).

Taxa selection for phylogenomic analyses

At the time of the start of the project, genomic sequences were available for almost 600 alphaproteobacteria. To reduce the number of taxa to a computationally feasible amount while keeping the maximum phylogenetic diversity, 80 taxa were handpicked based on a concatenated 16S+23S phylogeny of all sequenced alphaproteobacteria. To build this phylogeny, we first extracted the 16S and 23S gene sequences from all GenBank records of sequenced alphaproteobacteria for which both genes were present. In case of multiple copies, the longest representative was picked. Incomplete genes were removed as much as possible by removing all 16S sequences shorter than 1200 bp and all 23S sequences shorter than 2200 bp. The genes from *Acetobacter pasteurianus* and *Candidatus Hodgkinia cicadicola* were omitted due to their extreme long branches. As outgroup we collected the genes from four gammaproteobacteria and four betaproteobacteria based on a recent study (Wang and Wu 2015). The final 16S+23S dataset

consisted of 1011 taxa. The 16S and 23S genes were then separately aligned with MAFFT-7.050b (Katoh and Standley 2013) using the 'genafpair' option (mafft-einsi) and '--adjustdirection' to correct for reverse complemented sequences and the resulting alignments trimmed with trimAl 1.4 (Capella-Gutiérrez et al. 2009), selecting for sites for which less than 30% of the taxa contain a gap. The trimmed alignments were concatenated and then used to infer a maximum likelihood phylogeny with RAxML 7.2.8 (Stamatakis 2006) under the GTRCAT model and 100 rapid bootstraps.

Phylogenomics

For each COG, sequences were aligned with MAFFT-7.050b (Katoh and Standley 2013) using the local pair option (mafft-linsi) and subsequently trimmed with trimAl 1.4 (Capella-Gutiérrez et al. 2009), selecting sites for which less than half of the taxa contain a gap. To reduce effects of horizontal gene transfer on phylogeny inference, 11 (20%) COGs that displayed the most divergent phylogenies were removed using a discordance filter (Williams et al. 2010; Guy et al. 2014). The single gene tree bootstraps required for the discordance filter were inferred using RAxML 7.2.8 (Stamatakis 2006) with 100 non-parametric bootstraps, the GAMMA model for rate heterogeneity among sites and the LG substitution matrix (Le and Gascuel 2008). For each the remaining 43 COGs, single gene phylogenies were inferred with RAxML 7.2.8 under the PROTCATLG model and 100 rapid bootstraps and subsequently visually checked for presence of non-orthologous sequences and horizontal gene transfers that might have been missed by the discordance filter. This left us with a final 38 COGs consisting of orthologs and vertically evolving sequences only. These were concatenated, resulting in an alignment of 10,512 sites. To account for compositional bias an alignment trimming strategy was applied in which a χ^2-filter (Viklund et al. 2012) was used to remove half of all the sites in the alignment that contributed most to compositional heterogeneity. Phylogenies were inferred with PhyloBayes-MPI 1.4f (Lartillot et al. 2013) using four independent Markov Chain Monte Carlo chains under the CAT-Poisson model until the chains converged (maxdiff < 0.3). The log likelihood, total tree length, α-parameter and number of categories of all trees per chain were traced and visually inspected to choose the burn-in cutoff and consensus trees were calculated such that a total of approximately 200 trees were sampled per chain.

In-silico prediction of candidate effector proteins

Candidate effectors were selected in three ways, 1) proteins which comprised a signal peptide but no transmembrane domain, 2) proteins which contained eukaryotic-like repeat domains (LRRs, TPRs, ANKs, F-box) or 3) proteins for which several of the best blast hits belonged to eukaryotic proteins. Selected protein sequences were submitted to the I-TASSER suite (Yang et al. 2015). Calculated models were evaluated by their c-score, which ranges from -5 to 2, but only models with c-score of > -1.5 were considered to be of sufficient confidence (Yang et al. 2015). The best model was compared to proteins in the PDB database and fitting was assessed by the TM-score. Models with a high confidence were functionally annotated with COACH (Yang et al. 2013).

Proteomics

Acanthamoeba castellanii Neff cultures were infected as described previously (Schulz et al. 2014), and infection progress was evaluated after 5 days with fluorescence in situ hybridization using the symbiont specific probe CBR125 (Schulz et al. 2014). *A. castellanii* cultures were harvested by centrifugation at 6600g for 8 min at 4°C, lysed by adding pellet in 0.5ml EZ Lysis buffer (Sigma, Germany) and vortexing for 30 seconds at maximum speed. Then 5 ml EZ ysis buffer was added and the samples were incubated for 6 min at 4 °C, followed by 15 seconds vortexing and 30 strokes with a tight-fitting Dounce homogenizer. Nuclei were collected by centrifugation at 1500g for 8 minutes at 4°C and washed twice in 200µl EZ storage buffer (Sigma, Germany). Finally, purified nuclei were collected by centrifugation at 1500g for 8 minutes at 4°C and stored in 100 µl EZ storage buffer at -80 °C.

Pellets were dissolved in 50 mM TEAM, 1% Sodium-desoxycholate and 4 % SDS, collected on Microcon Ultracel-30 membranes, washed, reduced alkylated and digested. The peptides were seperated by liquid chromatography and measured at an Orbitrap Elite. Proteomics data were normalized by calculating the normalized spectral abundance factor (NSAF); protein abundance was estimated by dividing the spectral count for each protein by protein length and dividing again by the sum of all length-normalized spectral counts. The resulting number was then multiplied with a factor of 100.

References

Agranoff DD, Krishna S. 1998. Metal ion homeostasis and intracellular parasitism. *Mol Microbiol* **28**: 403–412.

Berridge MJ, Bootman MD, Roderick HL. 2003. Calcium signalling: dynamics, homeostasis and remodelling. *Nat Rev Mol Cell Biol* **4**: 517–529.

Bierne H, Cossart P, Bact I. 2012. When bacteria target the nucleus: the emerging family of nucleomodulins. *Cell Microbiol* **14**: 622–33.

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**: 63–72.

Boscaro V, Fokin SI, Schrallhammer M, Schweikert M, Petroni G. 2013a. Revised systematics of Holospora-like bacteria and characterization of "Candidatus Gortzia infectiva", a novel macronuclear symbiont of Paramecium jenningsi. *Microb Ecol* **65**: 255–67.

Boscaro V, Schrallhammer M, Benken KA, Krenek S, Szokoli F, Berendonk TU, Schweikert M, Verni F, Sabaneyeva EV, Petroni G. 2013b. Rediscovering the genus Lyticum, multiflagellated symbionts of the order Rickettsiales. *Sci Rep* **3**: 3305–3305.

Brickman TJ, Barry CE, Hackstadt T. 1993. Molecular cloning and expression of hctB encoding a strain-variant chlamydial histone-like protein with DNA-binding activity. *J Bacteriol* **175**: 4274–4281.

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

Chen I, Christie PJ, Dubnau D. 2005. The ins and outs of DNA transfer in bacteria. *Science* **310**: 1456–1460.

Clapham DE. 2007. Calcium Signaling. *Cell* **131**: 1047–1058.

Collingro A, Tischler P, Weinmaier T, Penz T, Heinz E, Brunham RC, Read TD, Bavoil PM, Sachse K, Kahane S, et al. 2011. Unity in variety--the pan-genome of the Chlamydiae. *Mol Biol Evol* **28**: 3253–3270.

Craig L, Pique ME, Tainer JA. 2004. Type IV pilus structure and bacterial pathogenicity. *Nat Rev Microbiol* **2**: 363–378.

Dohra H, Tanaka K, Suzuki T, Fujishima M, Suzuki H. 2014. Draft genome sequences of three Holospora species (Holospora obtusa , Holospora undulata , and Holospora elegans), endonuclear symbiotic bacteria of the ciliate Paramecium caudatum. *FEMS Microbiol Lett* n/a–n/a.

Fairman-Williams ME, Guenther U-P, Jankowsky E. 2010. SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol* **20**: 313–324.

Finkel S, Kolter R. 2001. DNA as a nutrient: novel role for bacterial competence gene homologs. *J Bacteriol* **183**: 6288–6293.

Fujishima M, Kodama Y. 2012. Endosymbionts in Paramecium. *Eur J Protistol* **48**: 124–37.

Georgiades K, Madoui M-A, Le P, Robert C, Raoult D. 2011. Phylogenomic analysis of Odyssella thessalonicensis fortifies the common origin of Rickettsiales, Pelagibacter ubique and Reclimonas americana mitochondrion. *PLoS ONE* **6**: e24857.

Gillespie JJ, Brayton KA, Williams KP, Diaz MAQ, Brown WC, Azad AF, Sobral BW. 2010. Phylogenomics reveals a diverse Rickettsiales type IV secretion system. *Infect Immun* **78**: 1809–23.

Gillespie JJ, Kaur SJ, Rahman MS, Rennoll-Bankert K, Sears KT, Beier-Sexton M, Azad AF. 2015. Secretome of obligate intracellular Rickettsia. *FEMS Microbiol Rev* **39**: 47–80.

Giltner CL, Nguyen Y, Burrows LL. 2012. Type IV pilin proteins: versatile molecular modules. *Microbiol Mol Biol Rev* **76**: 740–772.

Gödeke J, Heun M, Bubendorfer S, Paul K, Thormann KM. 2011. Roles of Two Shewanella oneidensis MR-1 Extracellular Endonucleases. *Appl Environ Microbiol* **77**: 5342–5351.

Görtz H, Schmidt H, GÖRTZ. 1987. Killer particles in the macronucleus of Paramecium caudatum. *Ann N Y Acad ...* 580–583.

Grieshaber NA, Fischer ER, Mead DJ, Dooley CA, Hackstadt T. 2004. Chlamydial histone-DNA interactions are disrupted by a metabolite in the methylerythritol phosphate pathway of isoprenoid biosynthesis. *Proc Natl Acad Sci U S A* **101**: 7451–7456.

Gubbels M-J, Vaishnava S, Boot N, Dubremetz J-F, Striepen B. 2006. A MORN-repeat protein is a dynamic component of the Toxoplasma gondii cell division apparatus. *J Cell Sci* **119**: 2236–2245.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: Quality assessment tool for genome assemblies. *Bioinforma Oxf Engl* 1–4.

Güttler T, Görlich D. 2011. Ran-dependent nuclear export mediators: a structural perspective. *EMBO J* **30**: 3457–3474.

Guy L, Kultima JR, Andersson SGE. 2010. genoPlotR: comparative gene and genome visualization in R. *Bioinforma Oxf Engl* **26**: 2334–5.

Guy L, Saw JH, Ettema TJG. 2014. The archaeal legacy of Eukaryotes: A phylogenomic perspective. *Cold Spring Harb Perspect Biol* **6**: a016022.

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.

Hentschel U, Steinert M, Hacker J. 2000. Common molecular mechanisms of symbiosis and pathogenesis. *Trends Microbiol* **8**: 226–231.

Ho BT, Dong TG, Mekalanos JJ. 2014. A view to a kill: the bacterial type VI secretion system. *Cell Host Microbe* **15**: 9–21.

Hunter JD. 2007. Matplotlib: A 2D Graphics Environment. *Comput Sci Eng* **9**: 90–95.

Izard T, Tran Van Nhieu G, Bois PRJ. 2006. Shigella applies molecular mimicry to subvert vinculin and invade host cells. *J Cell Biol* **175**: 465–475.

James BW, Mauchline WS, Dennis PJ, Keevil CW, Wait R. 1999. Poly-3-Hydroxybutyrate in Legionella pneumophila, an Energy Source for Survival in Low-Nutrient Environments. *Appl Environ Microbiol* **65**: 822–827.

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

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra M a. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* **19**: 1639–45.

Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst Biol* **62**: 611–615.

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

Li B, Kohler JJ. 2014. Glycosylation of the Nuclear Pore. *Traffic* **15**: 347–361.

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

Linka N, Hurka H, Lang BF, Burger G, Winkler HH, Stamme C, Urbany C, Seil I, Kusch J, Neuhaus HE. 2003. Phylogenetic relationships of non-mitochondrial nucleotide transport proteins in bacteria and eukaryotes. *Gene* **306**: 27–35.

Makino Y, Yogosawa S, Kayukawa K, Coin F, Egly JM, Wang Z x, Roeder RG, Yamamoto K, Muramatsu M, Tamura T a. 1999. TATA-Binding protein-interacting protein 120, TIP120, stimulates three classes of eukaryotic transcription via a unique mechanism. *Mol Cell Biol* **19**: 7951–7960.

Martijn J, Schulz F, Zaremba-Niedzwiedzka K, Viklund J, Stepanauskas R, Andersson SGE, Horn M, Guy L, Ettema TJG. 2015. Single-cell genomics of a rare environmental alphaproteobacterium provides unique insights into Rickettsiaceae evolution. *ISME J*. http://www.nature.com/ismej/journal/vaop/ncurrent/full/ismej201546a.html (Accessed June 4, 2015).

McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* **10**: 13–26.

McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF, et al. 2013. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci U S A* **110**: 3229–36.

Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, Shaw PD, Marshall D. 2013. Using Tablet for visual exploration of second-generation sequencing data. *Brief Bioinform* **14**: 193–202.

Misra S, Sharma V, Srivastava AK. 2014. Bacterial Polysaccharides: An Overview. In *Polysaccharides* (eds. K.G. Ramawat and J.-M. Mérillon), pp. 1–24, Springer International Publishing http://link.springer.com/referenceworkentry/10.1007/978-3-319-03751-6_68-1 (Accessed May 31, 2015).

Moliner C, Fournier P-E, Raoult D. 2010. Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. *FEMS Microbiol Rev* **34**: 281–294.

Muschiol S, Balaban M, Normark S, Henriques-Normark B. 2015. Uptake of extracellular DNA: Competence induced pili in natural transformation of Streptococcus pneumoniae. *BioEssays* **37**: 426–435.

Nasrallah GK, Riveroll AL, Chong A, Murray LE, Lewis PJ, Garduno RA. 2011. Legionella pneumophilaRequires Polyamines for Optimal Intracellular Growth. *J Bacteriol* **193**: 4346–4360.

Nunan LM, Pantoja CR, Gomez-Jimenez S, Lightner DV. 2013. "Candidatus Hepatobacter penaei," an Intracellular Pathogenic Enteric Bacterium in the Hepatopancreas of the Marine Shrimp Penaeus vannamei (Crustacea: Decapoda). *Appl Environ Microbiol* **79**: 1407–1409.

Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, et al. 2013. Assembling single-cell genomes and mini-metagenomes from chimeric

MDA products. *J Comput Biol J Comput Mol Cell Biol* **20**: 714–737.

Park H, Lee JH, Gouin E, Cossart P, Izard T. 2011. The Rickettsia Surface Cell Antigen 4 Applies Mimicry to Bind to and Activate Vinculin. *J Biol Chem* **286**: 35096–35103.

Pasini A, Caldarera CM, Giordano E. 2014. Chromatin remodeling by polyamines and polyamine analogs. *Amino Acids* **46**: 595–603.

Rahman MS, Gillespie JJ, Kaur SJ, Sears KT, Ceraul SM, Beier-Sexton M, Azad AF. 2013. Rickettsia typhi possesses phospholipase A2 enzymes that are involved in infection of host cells. *PLoS Pathog* **9**: e1003399.

Rao NN, Torriani A. 1990. Molecular aspects of phosphate transport in Escherichia coli. *Mol Microbiol* **4**: 1083–1090.

Rennoll-Bankert KE, Garcia-Garcia JC, Sinclair SH, Dumler JS. 2015. Chromatin bound bacterial effector AnkA recruits HDAC1 and modifies host gene expression. *Cell Microbiol* n/a–n/a.

Sabaneyeva EV, Derkacheva ME, Benken KA, Fokin SI, Vainio S, Skovorodkin IN. 2009. Actin-based mechanism of holospora obtusa trafficking in Paramecium caudatum. *Protist* **160**: 205–19.

Sassera D, Beninati T, Bandi C, Bouman EAP, Sacchi L, Fabbi M, Lo N. 2006. "Candidatus Midichloria mitochondrii", an endosymbiont of the tick Ixodes ricinus with a unique intramitochondrial lifestyle. *Int J Syst Evol Microbiol* **56**: 2535–2540.

Schnepf HE, Lee S, Dojillo J, Burmeister P, Fencil K, Morera L, Nygaard L, Narva KE, Wolt JD. 2005. Characterization of Cry34/Cry35 Binary Insecticidal Proteins from Diverse Bacillus thuringiensis Strain Collections. *Appl Environ Microbiol* **71**: 1765–1774.

Schulz F, Lagkouvardos I, Wascher F, Aistleitner K, Kostanjšek R, Horn M. 2014. Life in an unusual intracellular niche: a bacterial symbiont infecting the nucleus of amoebae. *ISME J* **8**: 1634–1644.

Schulz F, Martijn J, Wascher F, Lagkouvardos I, Kostanjšek R, Ettema TJG, Horn M. 2015. A rickettsiales symbiont of amoebae with ancient features. *Environ Microbiol*.

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.

Severi E, Hood DW, Thomas GH. 2007. Sialic acid utilization by bacterial pathogens. *Microbiology* **153**: 2817–2822.

Shah P, Swiatlo E. 2008. A multifaceted role for polyamines in bacterial pathogens. *Mol Microbiol* **68**: 4–16.

Silverman JM, Brunet YR, Cascales E, Mougous JD. 2012. Structure and regulation of the type VI secretion system. *Annu Rev Microbiol* **66**: 453–472.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinforma Oxf Engl* **22**: 2688–90.

Thi EP, Hong CJH, Sanghera G, Reiner NE. 2013. Identification of the Mycobacterium tuberculosis protein PE-PGRS62 as a novel effector that functions to block phagosome maturation and inhibit iNOS expression. *Cell Microbiol* **15**: 795–808.

Toft C, Andersson SGE. 2010. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat Rev Genet* **11**: 465–475.

Van den Berg B. 2005. The FadL family: unusual transporters for unusual substrates. *Curr Opin Struct Biol* **15**: 401–407.

Vandenesch F, Lina G, Henry T. 2012. Staphylococcus aureus Hemolysins, bi-component Leukocidins, and Cytolytic Peptides: A Redundant Arsenal of Membrane-Damaging Virulence Factors? *Front Cell Infect Microbiol* **2**. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3417661/ (Accessed June 3, 2015).

Viklund J, Ettema TJG, Andersson SGE. 2012. Independent genome reduction and phylogenetic reclassification of the oceanic SAR11 clade. *Mol Biol Evol* **29**: 599–615.

Viklund J, Martijn J, Ettema TJG, Andersson SGE. 2013. Comparative and phylogenomic evidence that the alphaproteobacterium HIMB59 is not a member of the oceanic SAR11 clade. *PloS One* **8**: e78858–e78858.

Vogt G. 1992. Enclosure of bacteria by the rough endoplasmic reticulum of shrimp hepatopancreas cells. *Protoplasma* **171**: 89–96.

Wang Z, Wu M. 2015. An integrated phylogenomic approach toward pinpointing the origin of mitochondria. *Sci Rep* **5**: 7949.

Wang Z, Wu M. 2014. Phylogenomic reconstruction indicates mitochondrial ancestor was an energy parasite. *PLoS ONE* **9**: e110685.

Wilcox L. 1986. Prokaryotic endosymbionts in the chloroplast stroma of the dinoflagellate Woloszynskia pascheri. *Protoplasma* **135**: 71–79.

Williams KP, Gillespie JJ, Sobral BWS, Nordberg EK, Snyder EE, Shallom JM, Dickerman AW. 2010. Phylogeny of gammaproteobacteria. *J Bacteriol* **192**: 2305–14.

Williams KP, Sobral BW, Dickerman AW. 2007. A robust species tree for the alphaproteobacteria. *J Bacteriol* **189**: 4578–86.

Wuttge S, Bommer M, Jäger F, Martins BM, Jacob S, Licht A, Scheffel F, Dobbek H, Schneider E. 2012. Determinants of substrate specificity and biochemical properties of the sn-glycerol-3-phosphate ATP binding cassette transporter (UgpB–AEC2) of Escherichia coli. *Mol Microbiol* **86**: 908–920.

Yang J, Roy A, Zhang Y. 2013. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinforma Oxf Engl* **29**: 2588–2595.

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**: 7–8.
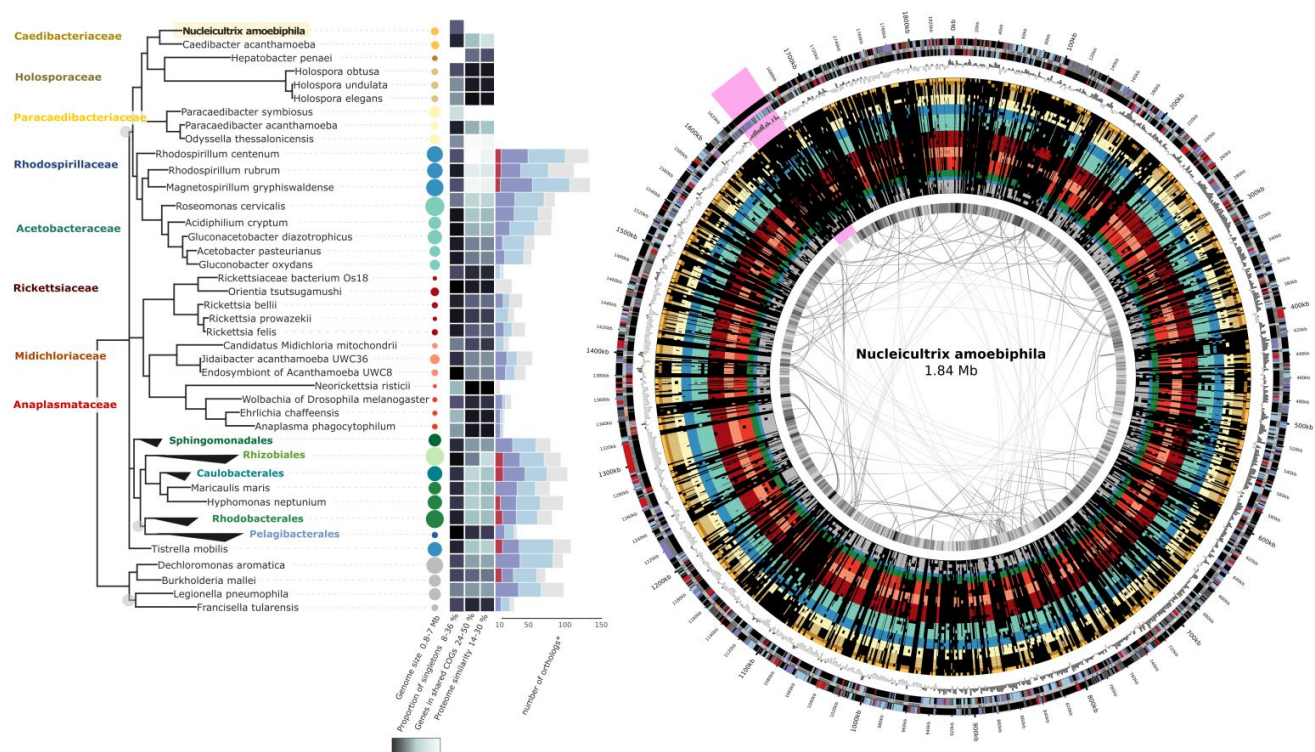
Figure 1. Genetic repertoire and phylogenetic position of *Nucleicultrix* in the *Alphaproteobacteria*. The left panel shows a Bayesian phylogenetic tree of the Alphaproteobacteria (Phylobayes CAT+GTR, 55000 generations, 4 chains, maxdiff = 0.18) based on a concatenated alignment of 38 universally distributed single copy proteins. Bayesian posterior probabilities below 0.99 are indicated as filled circles at the nodes, probabilities of 0.8 - 0.98 in light gray. Major monophyletic clades are collapsed and depicted as wedges. Colors of filled circles next to taxon names represent taxonomic classification and circle diameters correlate with respective genome sizes. Additional information is given as heatmaps; proportion of singletons in the respective genome, proportion of shared orthologs with the *Nucleicultrix* proteome, average amino acid similarities of all orthologs shared with *Nucleicultrix*, number of orthologs *Nucleicultrix* shares with the respective organism but not with any other member of the *Holosporales* assigned to major functional categories (red = information storage and processing, blue = metabolism, purple = cellular processes and signaling, gray = poorly characterized). The right panel displays a circular plot of the *Nucleicultrix* amoebiphila genome. Information included in the tracks from the outside inwards as follows: Scale of the *Nucleicultrix* genome in Kb, genes assigned to major functional categories on the either plus or minus strand (red = information storage and processing, blue = metabolism, purple = cellular processes and signaling, gray = poorly characterized); GC-skew (G-C/G+C ratio), presence/absence of orthologs to *Nucleicultrix* amoebiphila genes in selected reference genomes, reference organisms are in order based on their position in the phylogenetic tree shown in the left panel and colored accordingly, GC-content between 30 % (light grey) and 50 % (dark grey, grey links indicate the assignment of connected genes to the same gene family. A putative genomic island is highlighted in purple.
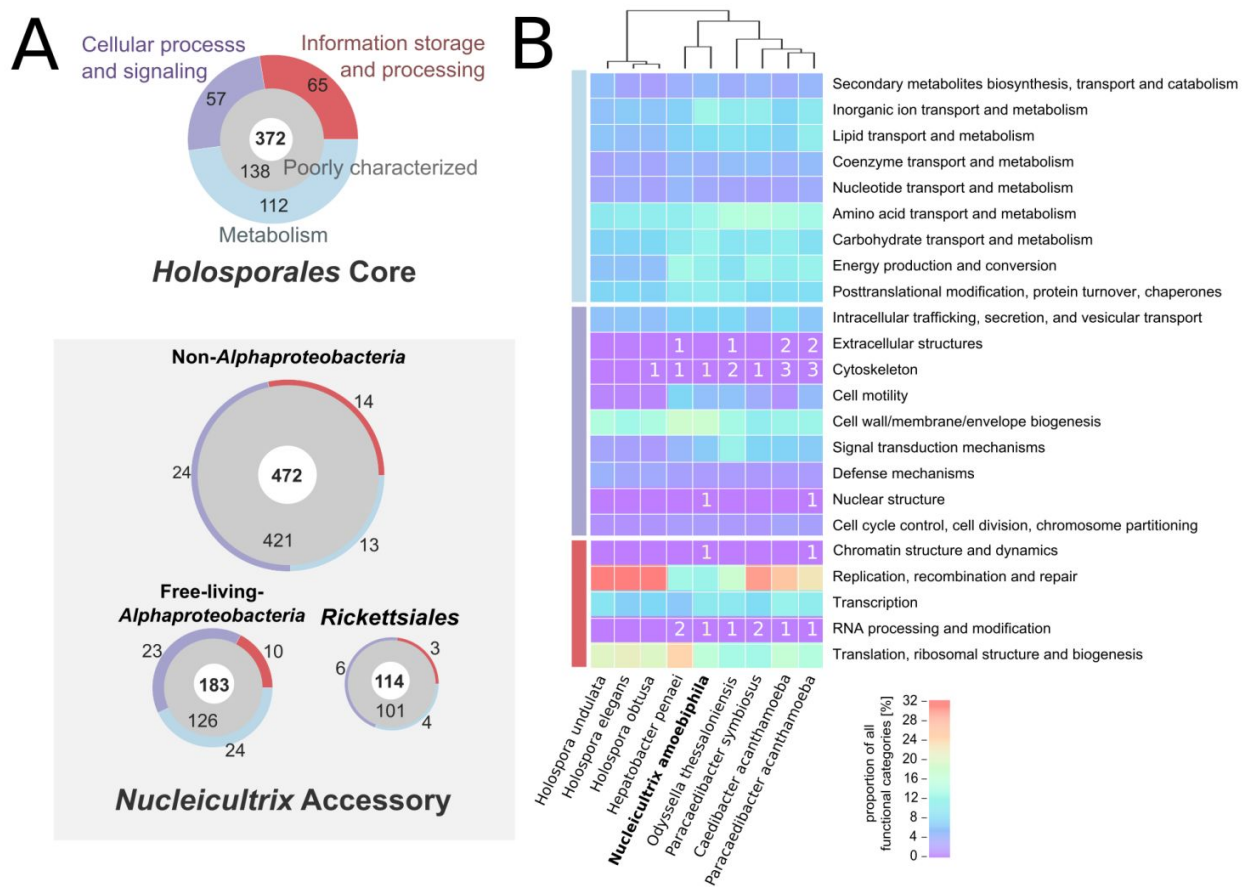
Figure 2. Core and accessory genome of *Nucleicultrix*. (A) Number of genes of the core genome of *Nucleicultrix* amoebiphila and other members of the *Holosporales* in major functional categories. The gray box shows the accessory genome of *Nucleicultrix* shared exclusively with non-alphaproteobacteria, free-living alphaproteobacteria and the *Rickettsiales*. The outer ring of each circle is divided into the number orthologs belonging to major functional categories, the gray fill of the circle indicates the number of orthologs for which no distinct function could be assigned. (B) A heatmap showing the proportion of genes assigned to distinct functional categories of each of the *Holosporales* members, genes present in all *Holosporales* members are excluded. The categories "general function prediction only" and "hypothetical" are not shown.
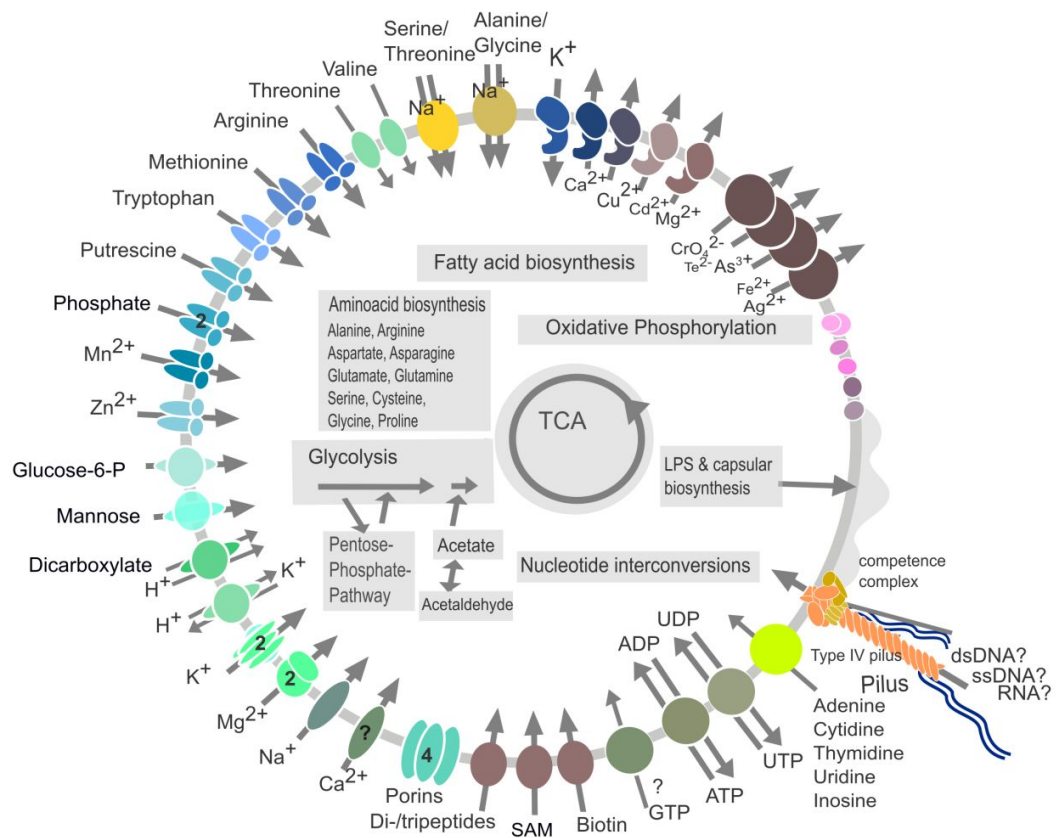
Figure 3. Metabolic capabilities of *Nucleicultrix*. Depicted are all major metabolic processes *Nucleicultrix* can perform. In addition, transporters encoded in the *Nucleicultrix* genome which could functionally be annotated are shown.
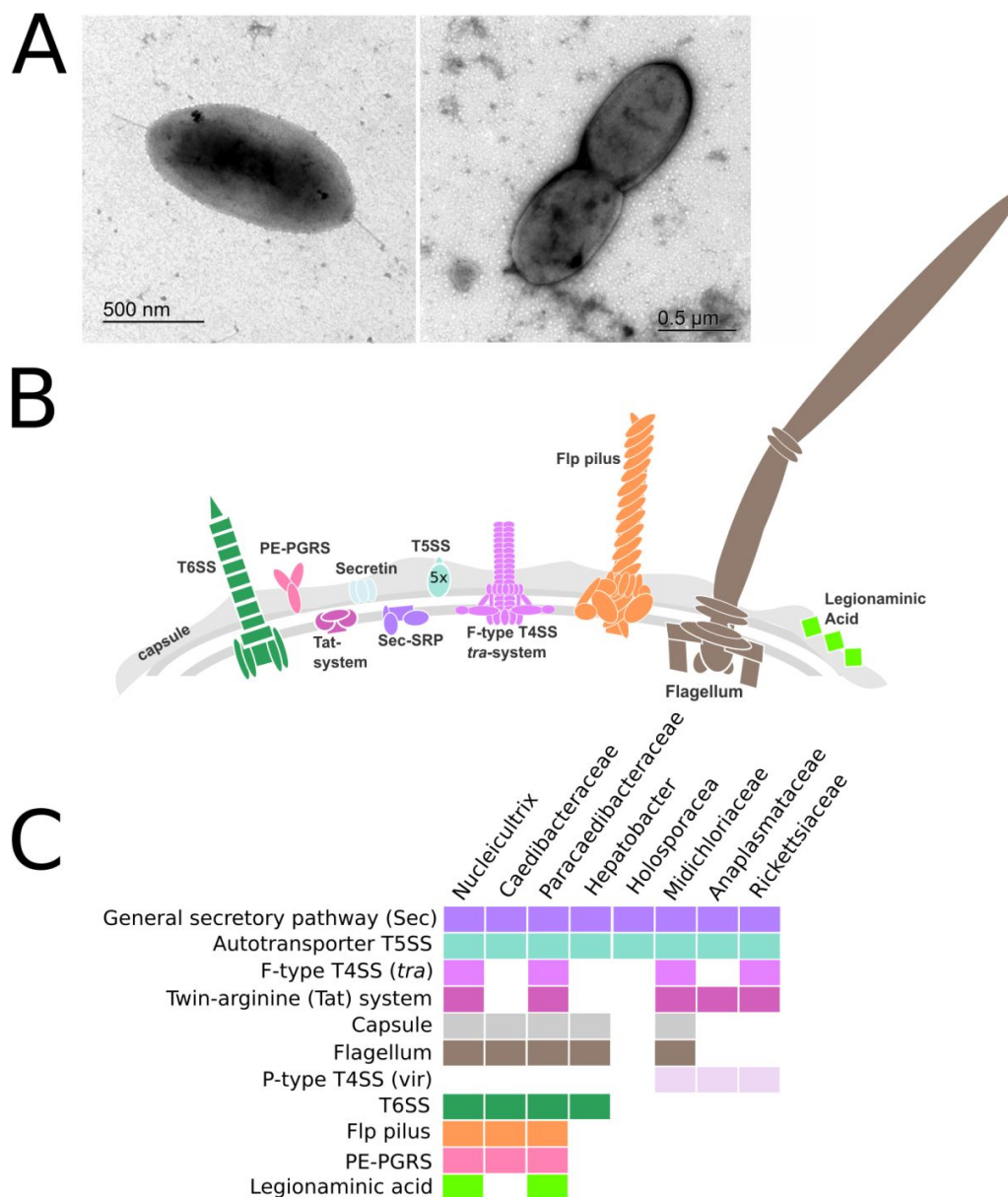
Figure 4. Virulence factors of *Nucleicultrix*. (A) Negative stain of *Nucleicultrix* cells shows features of the cell envelope, putative bipolar flagella and fimbriae. (B) The upper panel illustrates *Nucleicultrix* cell envelope modifications and macromolecular structures of the *Nucleicultrix* cell envelope likely involved in host interaction. (C) Presence or absence of the respective structure in defined taxonomic groups in the *Holosporales* and the *Rickettsiales*.
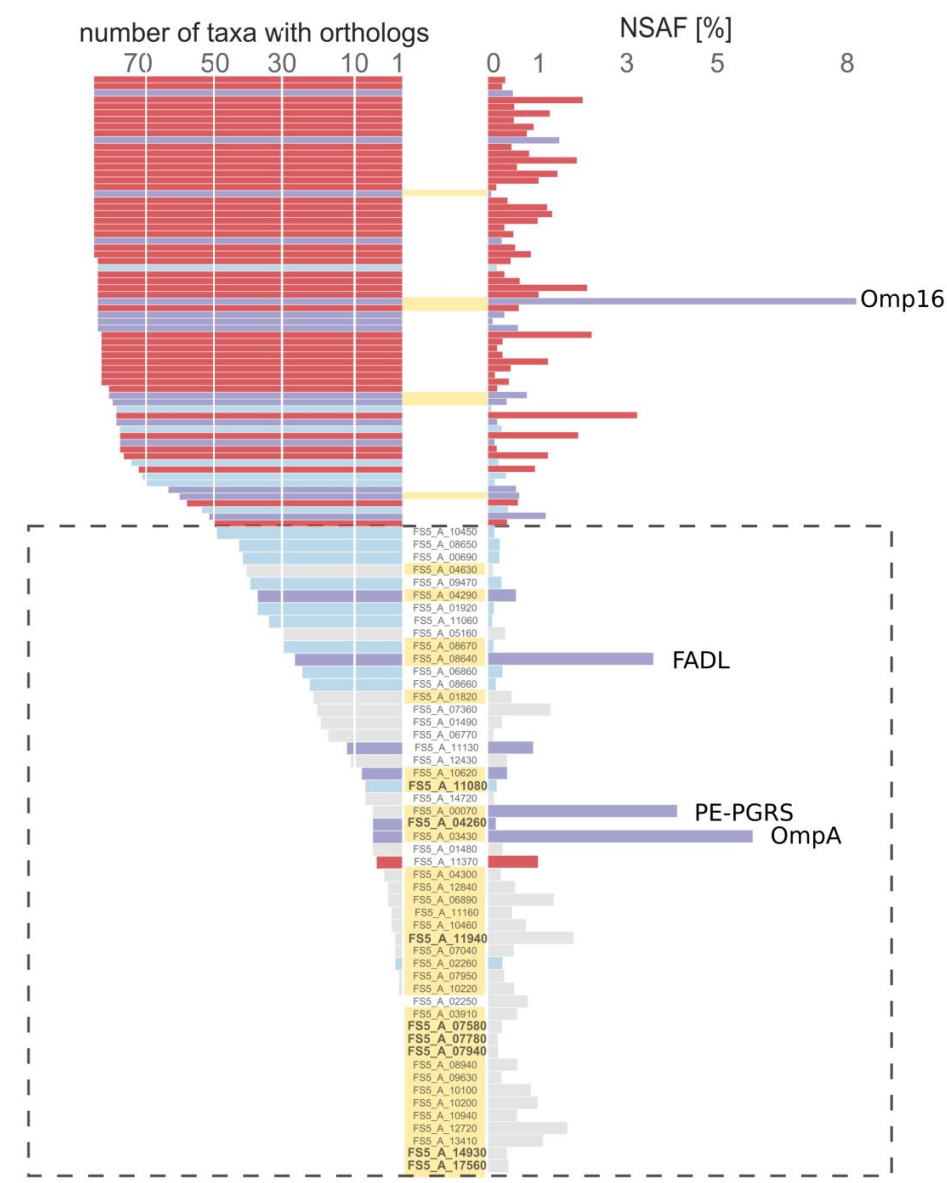
Figure 5. The proteome of *Nucleicultrix* in the course of an *Acanthamoeba castellanii* infection. The left side shows the distribution of orthologs of proteins expressed by *Nucleicultrix*, the right side depicts the normalized spectral abundance factor for each expressed protein. Expressed proteins have been assigned to major functional categories (red = information storage and processing, blue = metabolism, purple = cellular processes and signaling, gray = poorly characterized). Accessions number of proteins encoded in genomes of more than 75% of taxa in our data set are not shown and bars are compressed. Yellow highlighting indicates the presence of a signal peptide.
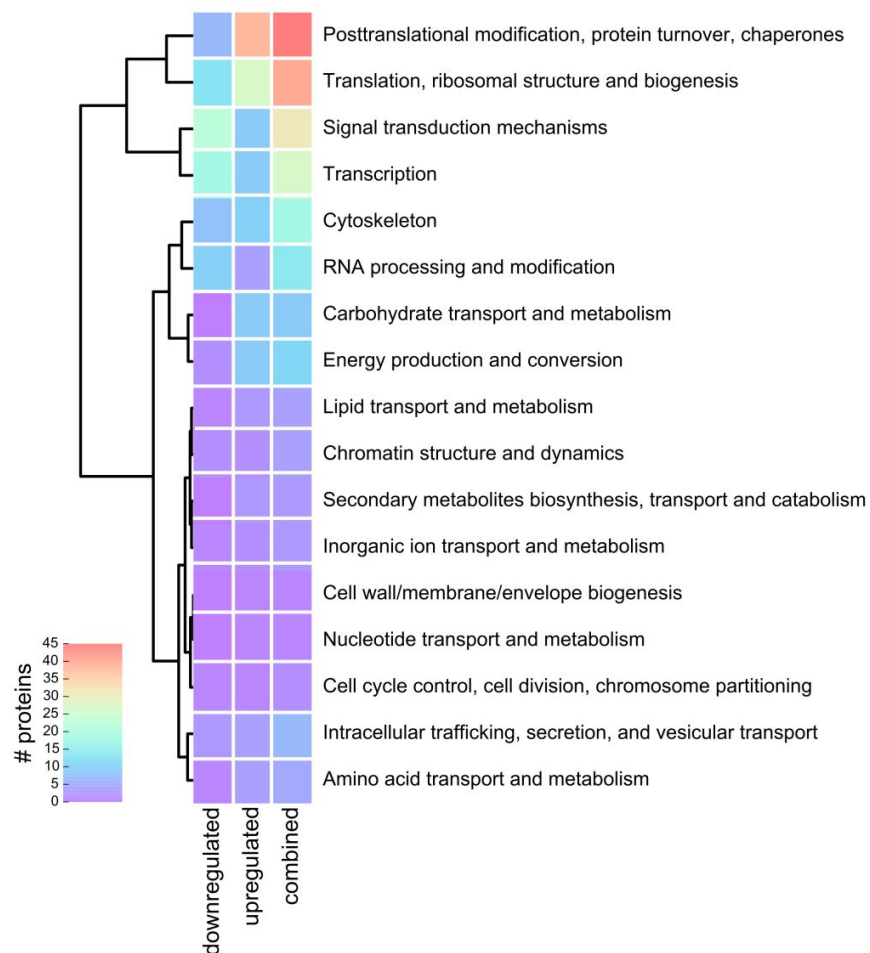
Figure 6. Impact of an intranuclear infection on *Acanthamoeba castellanii*. The number of differentially expressed (> 2-fold up- or down-regulated) proteins assigned to distinct functional categories depicted as heatmap.

Accession	GC%	aa	TM-score	C-score	Best fitting structure	PDB	PFAM	SP
FS5_A_00340	40.42	761	0.57	-1.19	GYP1, Rab-33B	2G77	SUFU (PF05076)	+
FS5_A_00390	37.95	101	0.75	0.29	Protein At1g77540	2EVN	Acetyltransf_CG (PF14542.1)	-
FS5_A_04260*	36.77	1106	0.59	-1	O-linked GlcNAc transferase	1W3B	TPR_19 (PF14559.1),TPR_2 (PF07719.12)	+
FS5_A_07440	37.73	258	0.36	-3.2	Histone H3 methyltransferase	1H3I	2x MORN_2 (PF07661.8)	-
FS5_A_07580*	41.89	1485	0.42	-2.52	Ski2-3-8 complex	4BUJ	-	+
FS5_A_07780*	40.23	1058	0.66	-0.45	TIP120 protein	1U6G	-	+
FS5_A_07940*	35.08	1106	0.53	-1.52	Snurportin-1,Ran,Exportin-1	3GJX	-	+
FS5_A_11940*	40.29	254	0.34	-3.34	Calcium uptake protein 1	4NSC	EF-hand_5 (PF13202.1)	+
FS5_A_14450	36.08	401	0.53	-1.5	Putative ski2-type helicase	2ZJ8	Mito_fiss_Elm1 (PF06258.6)	+
FS5_A_14930*	38.9	461	0.54	-1.45	DDB1-Cul4A-Rbx1-SV5V	2HYE	-	+
FS5_A_15450	38.91	644	0.76	0.34	GTP-binding nuclear protein Ran	1K5D	LRR_6 (PF13516.1)	+
FS5_A_17880	36.31	548	0.53	-1.49	Chromosome region maintenance 1	1N11		+
FS5_A_18260	40.23	319	0.95	1.67	Histone deacetylase 8	3F07	Hist_deacetyl (PF00850.14)	-

Table 1. Candidate nucleomodulins encoded in the *Nucleicultrix* genome. Information on GC content and on the protein length in amino acids is given. Asterisk indicate expression of the respective protein in the symbiont's proteome.

Accession	Description	PFAM
FS5_A_00510	Transposase	HTH_28 (PF13518.1)
FS5_A_02220	Antibiotic (virginiamycin and lincomycin) resistance protein	ABC_tran (PF00005.22), AAA_21 (PF13304.1), AAA_21 (PF13304.1), ABC_tran (PF00005.22), AAA_21 (PF133
FS5_A_03040	Acetyltransferase, GNAT family	Acetyltransf_10 (PF13673.1), Acetyltransf_7 (PF13508.1), Acetyltransf_1 (PF00583.19)
FS5_A_03870	Mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase	Glucosaminidase (PF01832.15)
FS5_A_04100	Putative carboxypeptidase YocD	Peptidase_S66 (PF02016.10)
FS5_A_05810	Aminoglycoside N(6')-acetyltransferase type 1	Acetyltransf_3 (PF13302.1), Acetyltransf_8 (PF13523.1), Acetyltransf_1 (PF00583.19)
FS5_A_08580	Insertion element iso-IS1n protein InsB	DDE_Tnp_IS1 (PF03400.8)
FS5_A_11870	Spectinomycin phosphotransferase	APH (PF01636.18), Choline_kinase (PF01633.15), EcKinase (PF02958.15)
FS5_A_12300	Uncharacterized ABC transporter	MacB_PCD (PF12704.2), FtsX (PF02687.16)
FS5_A_12310	Uncharacterized ABC transporter ATP-binding protein MJ1508	Rad17 (PF03215.10), ABC_tran (PF00005.22)
FS5_A_12320	Multidrug resistance efflux pump	HlyD_2 (PF12700.2), Biotin_lipoyl_2 (PF13533.1), HlyD (PF00529.15), HlyD_3 (PF13437.1)
FS5_A_13390	Uncharacterized transmembrane protein	
FS5_A_16400	Imidazole glycerol phosphate synthase subunit HisH	GATase (PF00117.23), SNO (PF01174.14), GATase_3 (PF07685.9), Peptidase_C26 (PF07722.8)
FS5_A_16420	Ubiquinone biosynthesis O-methyltransferase	Methyltransf_23 (PF13489.1), Methyltransf_31 (PF13847.1), Methyltransf_12 (PF08242.7), Methyltransf_11 (PF08
FS5_A_16560	UDP-glucose 4-epimerase, SP	RmlD_sub_bind (PF04321.12), Polysacc_synt_2 (PF02719.10), Epimerase (PF01370.16), 3Beta_HSD (PF01073.14
FS5_A_16570	UDP-glucuronic acid decarboxylase 3	RmlD_sub_bind (PF04321.12), Epimerase (PF01370.16), Polysacc_synt_2 (PF02719.10), 3Beta_HSD (PF01073.14
FS5_A_16700	GCN5-related N-acetyltransferase	Acetyltransf_3 (PF13302.1)
FS5_A_17020	Aminoglycoside phosphotransferase	APH (PF01636.18)
FS5_A_17200	Methyltransferase type 11	Methyltransf_23 (PF13489.1), Methyltransf_31 (PF13847.1), Ubie_methyltran (PF01209.13)

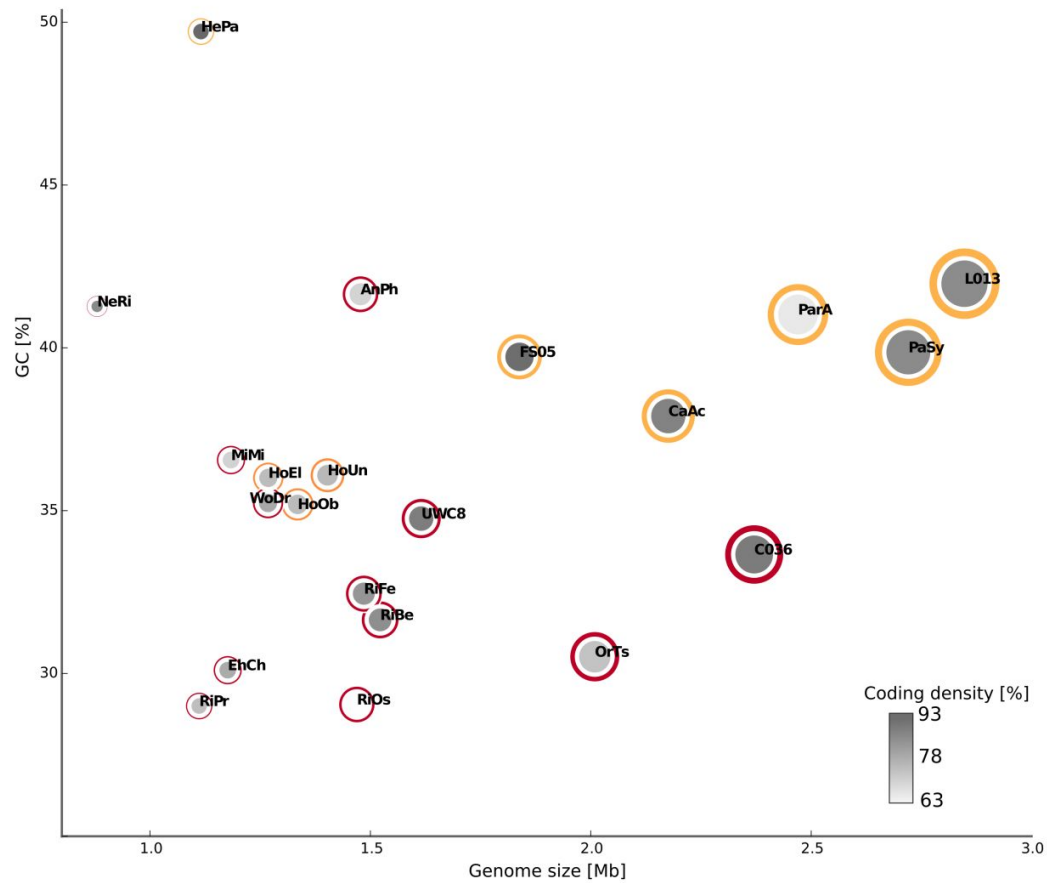**Table S1. Genes exclusively shared between *Nucleicultrix* and *Holospora* species.**

Figure S1. Distribution of genome size versus GC content in the *Holosporales* and the *Rickettsiales*. Colored rings indicate taxonomic affiliation, yellow for members of the *Holosporales* and red or members of the *Rickettsiales*. Filled gray circles indicate coding density.
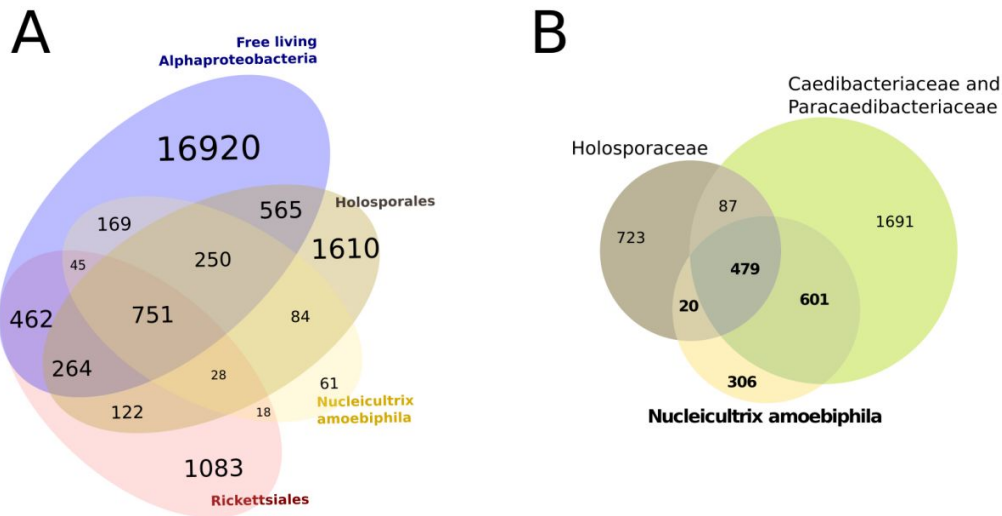
Figure S2. Shared and unique orthologs between *Nucleicultrix* and selected taxonomic groups. (A) Four-set Venn diagram illustrating genes shared between *Nucleicultrix* and different bacterial groups, intersections indicate the number of shared orthologs. Unique genes for each group include orthologs shared between members of the respective group or with bacteria not included in the analysis, singletons are not shown. (B) Three-set Venn diagram showing the overlap in gene content between *Nucleicultrix*, *Holospora* species and cytoplasmic amoeba symbionts in the *Holosporales* (*Caedibacteriaceae* and *Paracaedibacteriaceae*). *Hepatobacter penaei* was excluded in the analysis.
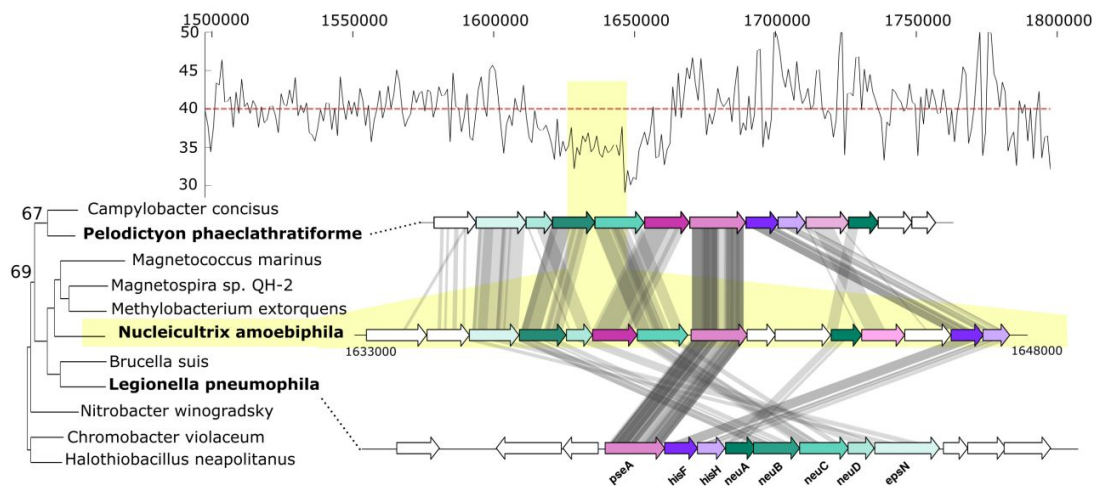
Figure S3. Putative genomic island encoded in the *Nucleicultrix* genome. Shown is the GC % distribution and conservation of encoded genes in *Pelodictyon phaeoclathratiforme, Nucleicultrix amoebiphila* and *Legionella pneumophila.* Orthologs are labeled using identical colors; highly conserved regions are connected by vertical gray shading. Phylogenetic tree is based on a concatenated alignment of the four proteins NeuABC and EpsN and was calculated with RaxML GAMMA+GTR and 100 rapid bootstraps. Support values of below 95 are indicated.
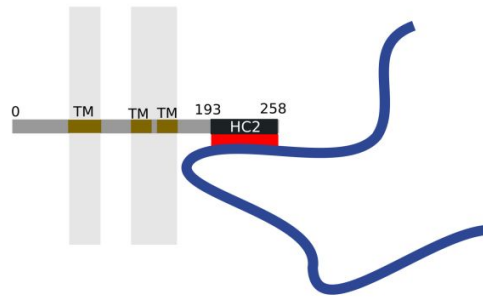
Figure S4. Model for membrane inserted DNA binding protein FS5_A_01890. Transmembrane domains (TM, brown) and DNA-binding region (HC2, black) and bacterial cell membrane (light grey) are shown. Red interconnection indicates binding to host DNA (blue).
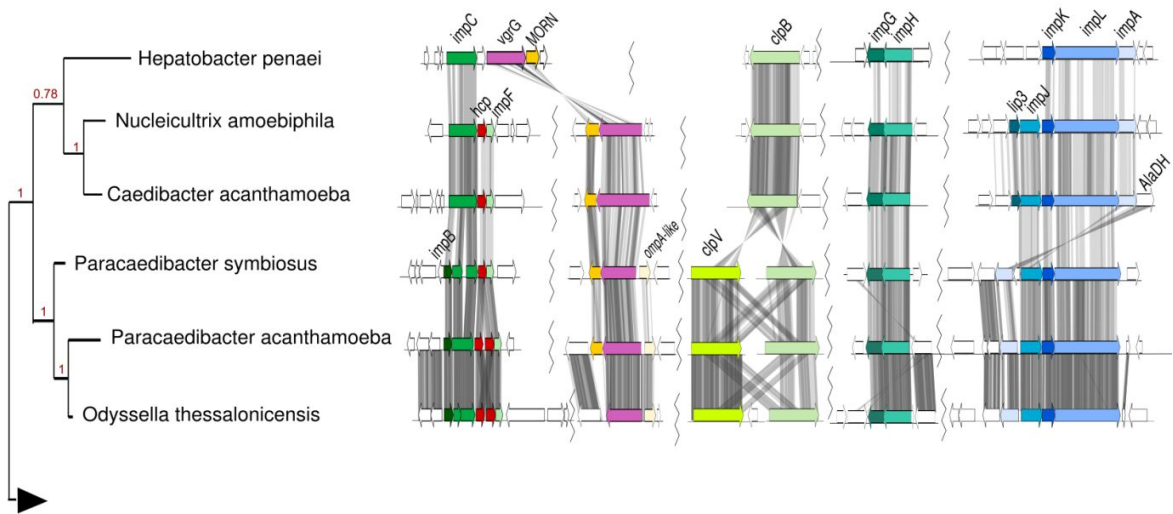
Figure S5. The type VI secretion system encoded in the *Holosporales*. Genomic origin of type VI secretion systems is shown for members of the *Holosporales*. The phylogeny was inferred with Phylobayes CAT+GTR based on concatenated alignment of VgrG, ImpL, ImpH, ImpK and ImpG. Orthologs are labeled using identical colors; highly conserved regions are connected by vertical gray shading.

Figure S6. Differentially expressed *Acanthamoeba castellanii* proteins during an intranuclear infection. Shown are protein either at least 3-fold up-(red) or down-regulated Expression values are shown as the log2 fold-change. Differential coloring indicates assignment of proteins to functional categories.

Chapter VI

A *Rickettsiales* symbiont of amoebae with ancient features

A *Rickettsiales* symbiont of amoebae with ancient features

Frederik Schulz,[1] Joran Martijn,[2] Florian Wascher,[1] Ilias Lagkouvardos,[1] Rok Kostanjšek,[3] Thijs J. G. Ettema[2] and Matthias Horn[1]*

[1]*Department of Microbiology and Ecosystem Science, University of Vienna, Althanstraße 14, Vienna, Austria.*
[2]*Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Husargatan 3, Uppsala, Sweden.*
[3]*Department of Biology, University of Ljubljana, Večna pot 111, Ljubljana, Slovenia.*

Summary

The *Rickettsiae* comprise intracellular bacterial symbionts and pathogens infecting diverse eukaryotes. Here, we provide a detailed characterization of 'Candidatus Jidaibacter acanthamoeba', a rickettsial symbiont of *Acanthamoeba*. The bacterium establishes the infection in its amoeba host within 2 h where it replicates within vacuoles. Higher bacterial loads and accelerated spread of infection at elevated temperatures were observed. The infection had a negative impact on host growth rate, although no increased levels of host cell lysis were seen. Phylogenomic analysis identified this bacterium as member of the *Midichloriaceae*. Its 2.4 Mb genome represents the largest among *Rickettsiales* and is characterized by a moderate degree of pseudogenization and a high coding density. We found an unusually large number of genes encoding proteins with eukaryotic-like domains such as ankyrins, leucine-rich repeats and tetratricopeptide repeats, which likely function in host interaction. There are a total of three divergent, independently acquired type IV secretion systems, and 35 flagellar genes representing the most complete set found in an obligate intracellular *Alphaproteobacterium*. The deeply branching phylogenetic position of 'Candidatus Jidaibacter acanthamoeba' together with its ancient features place it closely to the rickettsial ancestor and helps to better understand the transition from a free-living to an intracellular lifestyle.

Introduction

Microbial symbioses are ubiquitous and of central importance for the ecology of most life forms on earth (McFall-Ngai *et al.*, 2013). From an evolutionary point of view, the *Alphaproteobacteria*, and in particular the order *Rickettsiales*, represents an interesting group of bacteria to study the emergence of a symbiotic lifestyle (Ettema and Andersson, 2009; Georgiades *et al.*, 2011a,b). The *Rickettsiales* comprise obligate intracellular bacteria infecting diverse eukaryotic hosts, ranging from protists to mammals. Prominent members of this group include *Wolbachia* species, the most prevalent bacterial symbionts known to date, infecting arthropods and nematodes (Werren *et al.*, 2008; Zug and Hammerstein, 2012). Many rickettsiae, such as *Anaplasma*, *Ehrlichia*, *Neorickettsia*, *Orientia* and *Rickettsia* species are able to perform a host switch; they can infect both invertebrate and vertebrate hosts and are recognized as important pathogens (Weinert *et al.*, 2009). Furthermore, they are considered to represent the closest extant relatives of mitochondria (Andersson *et al.*, 1998; Gray, 1998; Williams *et al.*, 2007; Driscoll *et al.*, 2013; Wang and Wu, 2015).

Consistent with their intracellular lifestyle, the genomes of *Rickettsiales* members strongly reflect host adaptation; they generally have small sizes of 1–1.5 Mb, a low coding density, truncated metabolic capabilities and a high degree of pseudogenization (Sällström and Andersson, 2005; Renvoisé *et al.*, 2011). Furthermore, most rickettsiae are considered energy parasites, i.e. they compensate for a restricted metabolism by importing and using host ATP (Andersson, 1998; Schmitz-Esser *et al.*, 2004; Gillespie *et al.*, 2012a,b). Despite their small genomes, they encode many genes contributing to antigenic variability, which help to overcome host defence mechanisms and ensure infectivity (Darby *et al.*, 2007). Moreover, *Rickettsiales* genomes encode a variety of putative effector proteins, which help to establish the intracellular niche and to manipulate host cellular processes in a sophisticated manner (Lockwood *et al.*, 2011; Renvoisé *et al.*, 2011; Liu *et al.*, 2012; Beyer *et al.*, 2014).

Some *Rickettsiales* exhibit remarkable features. Perhaps the best example is *Midichloria mitochondrii* (hereafter *Midichloria*), which thrives in the mitochondria of tick cells (Sassera *et al.*, 2006). Its 1.2 Mb genome

encodes for features unusual for rickettsial genomes, such as flagellar genes and a cbb3 cytochrome oxidase, but overall shows a high similarity to the genomes of other *Rickettsiales* (Sassera *et al.*, 2011). Besides *Midichloria*, several other deeply branching rickettsiae have been identified in the past decades, mainly thriving in protist hosts (Horn *et al.*, 1999; Vannini *et al.*, 2005; 2010; Schmitz-Esser *et al.*, 2008b; Ferrantini *et al.*, 2009; Kawafune *et al.*, 2012; Boscaro *et al.*, 2013a,b; Schulz *et al.*, 2014). Several of these bacteria, including the amoeba symbiont *Odyssella thessalonicensis* (hereafter: *Odyssella*), have been assigned to the family *Holosporaceae*, which was initially described as belonging to the *Rickettsiales*; however, its correct phylogenetic placement is still a matter of debate (Ferla *et al.*, 2013; Santos and Massard, 2014). Overall, the knowledge about these rickettsial organisms is rather scarce, especially from a genomic point of view.

Here, we report on a *Rickettsiales* symbiont of free-living amoebae, tentatively named 'Candidatus Jidaibacter acanthamoeba' (alluding to character names in the Star Wars saga; see Text S1 for etymological considerations and formal description; hereafter *Jidaibacter*), which was isolated together with its *Acanthamoeba* sp. UWC36 host from the cornea of a keratitis patient and first described more than 15 years ago (Fritsche *et al.*, 1999). We characterized the bacterium's infection process and provide a detailed analysis of its unique genomic repertoire: three type IV secretion systems, an arsenal of putative effector proteins and the most complete set of flagellar genes found so far in an obligate intracellular alphaproteobacterium.

Results and discussion

Diversity and environmental distribution

To assess the diversity and environmental distribution of *Jidaibacter*, we used a recently described approach (Lagkouvardos *et al.*, 2013) to search large amplicon sequencing databases. Our analysis revealed evidence for a number of closely related *Jidaibacter* strains ($> 99\%$ 16S rRNA similarity to *Jidaibacter*; $n = 33$) in diverse environments. At the genus level ($> 95\%$ 16S rRNA), *Jidaibacter*-related bacteria were predominantly found in soil ($n = 250$), followed by anthropogenic/freshwater ($n = 17$) and marine ($n = 6$) habitats. These numbers correspond to those observed for other bacterial symbionts of protists (Schulz *et al.*, 2014; Schulz *et al.*, unpubl. data) and suggest that soil-associated protists are a major reservoir of *Jidaibacter*.

Infection process and impact on the host

The infection process of *Jidaibacter* was studied in *Acanthamoeba castellanii* Neff, serving as a substitute for the symbiont's native host *Acanthamoeba* sp. UWC36, which could not be rendered symbiont-free even after intensive antibiotic treatments (Table S2). *Acanthamoeba castellanii* Neff has been used before to study the bacterial symbionts of amoebae (Aistleitner *et al.*, 2013; Pilhofer *et al.*, 2014; Schulz *et al.*, 2014), and both amoeba strains belong to the *Acanthamoeba* 18S rRNA sequence type (T4). Although differences to the native host might exist, we consider the type strain and model host *A. castellan*ii Neff a suitable system to investigate basic features of the infection process of *Jidaibacter*.

The progress of the infection with *Jidaibacter* was monitored over a period of 144 h at two different temperatures, 20°C and 30°C (with 20°C representing the standard cultivation condition for *Acanthamoebae* in the lab, and 30°C representing a temperature at which other bacterial symbionts of amoebae show a tendency towards host lysis). Ultrastructural analysis with transmission electron microscopy (TEM) showed that in the host cytoplasm, the bacteria remain enclosed in vacuoles of various sizes throughout the infection cycle (Fig. 1A). At both temperatures, single coccoid rods were located inside the amoeba host cell within 24 h post-infection (p.i.), and *Jidaibacter* started replicating (Fig. 1B). After 24 h, the bacterial morphology changed to elongated rods (Fig. 1B), and we observed dividing bacteria and an accumulation of ribosomes around the vacuolar membrane (Fig. 1Ai). Highly infected amoebae (> 30 bacteria in the cytoplasm) were detected for the first time after 48 h at 30°C and after 72 h at 20°C (Fig. 2). From 72 h p.i. onward, we observed vacuoles filled with bacteria in the supernatant at both temperatures. The majority of amoebae ($> 80\%$) was infected by *Jidaibacter* 96 h p.i. at 30°C and 144 h p.i. at 20°C (Fig. 2). After 144 h, less than 20% of host cells were highly infected at the lower temperature, whereas more than 90% of amoebae contained more than 30 bacterial particles at the higher temperature. At both temperatures, no significant difference in host cell viability compared with the uninfected controls was observed during the first 72 h p.i., as indicated by total amoeba cell numbers and propidium iodide fluorescence intensities (Fig. 2). Only from 96 h p.i. onwards, slightly but significantly ($P < 0.05$) lower numbers of amoebae were detected in infected cultures as compared with uninfected cultures. We could not measure a difference in propidium iodide fluorescence intensities or host cell lysis at any time point.

Our results showed that elevated temperature leads to a faster progression of the *Jidaibacter* infection and to higher bacterial loads in infected host cells. Despite no apparent increase in host cell lysis, we observed lower amoeba cell numbers at later time points compared with the uninfected control, suggesting a reduced amoeba growth rate due to the presence of the symbiont. This is consistent with previous reports on other intracellular
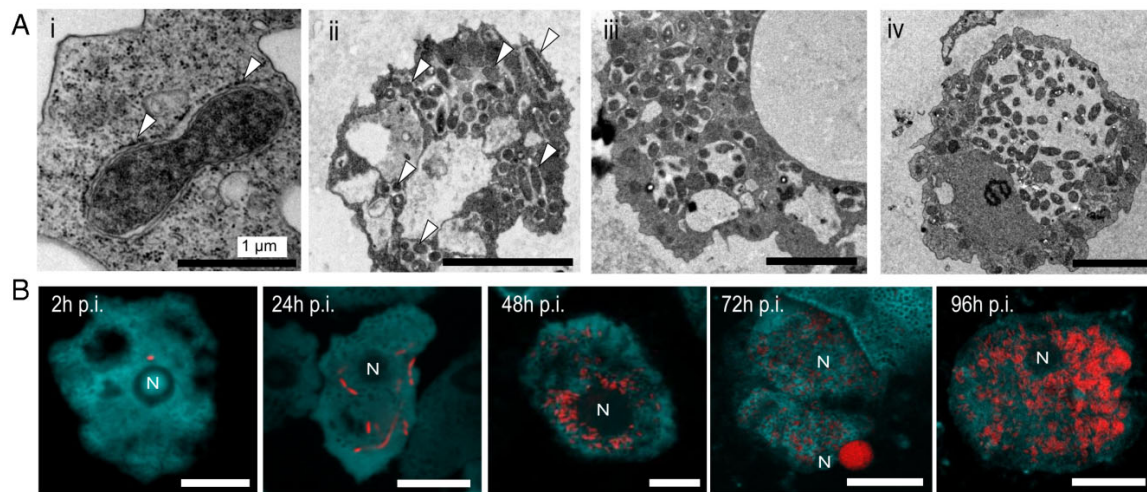
Fig. 1. Course of *Jidaibacter acanthamoeba* infection in *Acanthamoeba castellanii*.
A. Transmission electron microscopic images illustrating key steps during infection of *A. castellanii* Neff by *Jidaibacter*. (i) Cell division of *Jidaibacter* enclosed in a host-derived vacuole surrounded by ribosomes (arrow heads), (ii) *A. castellanii* heavily infected by rod-shaped *Jidaibacter* (arrow heads), (iii) the symbiont replicating in multiple vacuoles in the host cytoplasm and (iv) one large bacteria-filled vacuole in the cytoplasm of *A. castellanii* Neff.
B. Time course of *Jidaibacter* (red) infection at 30°C in *A. castellanii* Neff (turquoise) monitored by fluorescence in situ hybridization. A single bacterial cell is visible in the cytoplasm after 2 h; the morphology of the *Jidaibacter* changes from coccoid rods to rods and filamentous rods after 24 h, and bacterial cells are actively dividing; amoebae are heavily infected (> 30 bacteria/cell) from 48 h p.i. onward. Scale bars indicate 10 μm in the FISH images and 5 μm in the TEM images, if not stated otherwise. N, amoeba nucleus.

symbionts of amoeba, in particular environmental chlamydiae (Collingro *et al.*, 2004), and reflects the adverse impact on host cell fitness, likely due to the symbiont's consumption of host cell metabolites. Enhanced replication at elevated temperature was also observed for other rickettsiae, and it was suggested that the capability to grow at higher temperatures facilitated the switch from arthropod to mammalian hosts (Azad and Traub, 1985; Dreher-Lesnick *et al.*, 2008).

General features and composition of the Jidaibacter *genome*

To obtain insights into the genetic repertoire of *Jidaibacter* and the molecular basis of its association with amoebae, we generated a draft genome sequence comprising in a total of 249 contigs (144 contigs > 1 kb; largest contig = 261 kb; N50 = 77.4 kb) (Fig. S1). The presence of 36 tRNAs for all 20 proteinogenic amino acids, a split rRNA operon (16S RNA gene + 5S/23S rRNA genes), as well as all 31 universal bacterial marker proteins integrated in AMPHORA2 (Wu and Scott, 2012), indicates that our data set represents the symbiont's nearly complete genome. With about 2.4 Mb, the genome of *Jidaibacter* is more than two times larger than those found in the majority of obligate intracellular bacteria and represents the largest genome in the order *Rickettsiales* to date (Darby

et al., 2007; McCutcheon and Moran, 2012). It has an average G + C content of 34%, 13 pseudogenes and with 2267 predicted genes a coding density of around 87%, the latter of which is rather high for members of the order (Mavromatis *et al.*, 2006; Sassera *et al.*, 2011). A feature of many *Rickettsiales* genomes is a relatively large number of repetitive sequences (Darby *et al.*, 2007; Gillespie *et al.*, 2012b), which is also seen in the *Jidaibacter* genome where we observed 376 repetitive sequences that are > 100 bp and show a similarity of > 95% (representing about 7% of the genome; Fig. S1). In conclusion, the *Jidaibacter* genome shows typical features of other members of the *Rickettsiales*, but the relatively large genome size and only moderate pseudogenization suggest that it has not undergone strong reduction. This is similar to chlamydial symbionts of amoebae, which also show larger genomes in contrast to their medically relevant counterparts (Horn, 2008). Less reduced genomes might be associated with the environmental lifestyle in amoebae compared with animal or human hosts.

Phylogenomics pinpoints the position of Jidaibacter *within the* Rickettsiales

To assess the phylogenetic relationship of *Jidaibacter* in the *Rickettsiales* and to complement previous studies,
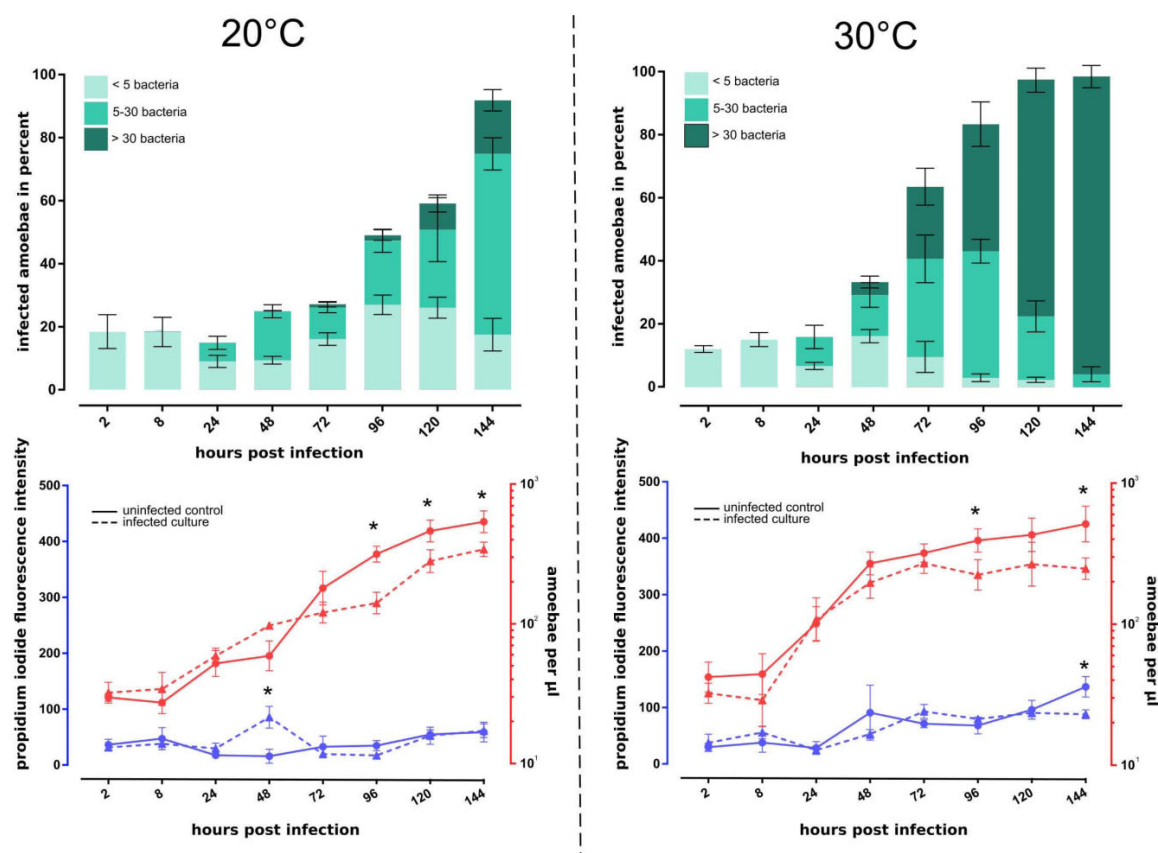
Fig. 2. Effect of *Jidaibacter acanthamoeba* on its amoeba host. *Jidaibacter* infection of *A. castellanii* Neff was monitored at an incubation temperature of 20°C or 30°C. The top panels show the percentage of infected amoebae until 144 h p.i., with the number of bacteria/amoeba indicated. The bottom panels show amoeba cell numbers during the course of infection (red graphs) and amoeba viability determined by propidium iodide staining (blue graphs). Asterisks indicate significant differences to the control ($P < 0.05$); error bars show standard deviation based on three replicate infection experiments.

which are based on the 16S and 23S rRNA genes, placed the symbiont in the *Midichloriacea* (Gillespie *et al.*, 2012a; Driscoll *et al.*, 2013; Montagna *et al.*, 2013; Schulz *et al.*, 2014), we searched for universally distributed single copy genes. In the maximum likelihood tree inferred from 173 single copy genes found in our data set, *Jidaibacter* grouped with the *Midichloriaceae*, which formed the basal branch in the *Rickettsiales* (Fig. S2A). To mitigate the effects of horizontal gene transfer on phylogeny inference, those genes that displayed the most divergent phylogenetic signal were removed using a discordance filter (Williams *et al.*, 2010; Guy *et al.*, 2014), reducing our data set to 139 proteins (Fig. S2B). It is known that the reduced genomes of intracellular bacteria often are AT-rich, leading to a compositional bias (Foster and Hickey, 1999; Rodríguez-Ezpeleta and Embley, 2012; Viklund *et al.*, 2012). To alleviate our data set for such compositional bias, a χ^2 filter was applied (Viklund *et al.*,

2012), removing the most biased sites in the remaining data set. Our final topology groups *Jidaibacter* again together with the *Midichloriaceae*, representing a sister clade to the *Anaplasmataceae* (Fig. 3). This branching is well supported in the Bayesian tree (maxdiff = 0.15, posterior probability = 0.99) but rather unresolved in the maximum likelihood tree (bootstrap support = 53%). Of note, this topology is in agreement with previous trees based on the 16S rRNA gene (Gillespie *et al.*, 2012a; Driscoll *et al.*, 2013; Montagna *et al.*, 2013; Schulz *et al.*, 2014), but disagrees with recent phylogenomic studies in which the *Midichloriaceae* represented a sister clade of the *Rickettsiaceae* (Driscoll *et al.*, 2013; Ferla *et al.*, 2013). The filtering steps used here to account for HGT and compositional bias might explain these differences. A better coverage of the *Midichloriaceae* in terms of genome sequences should help to resolve the phylogenetic position of this bacterial group.
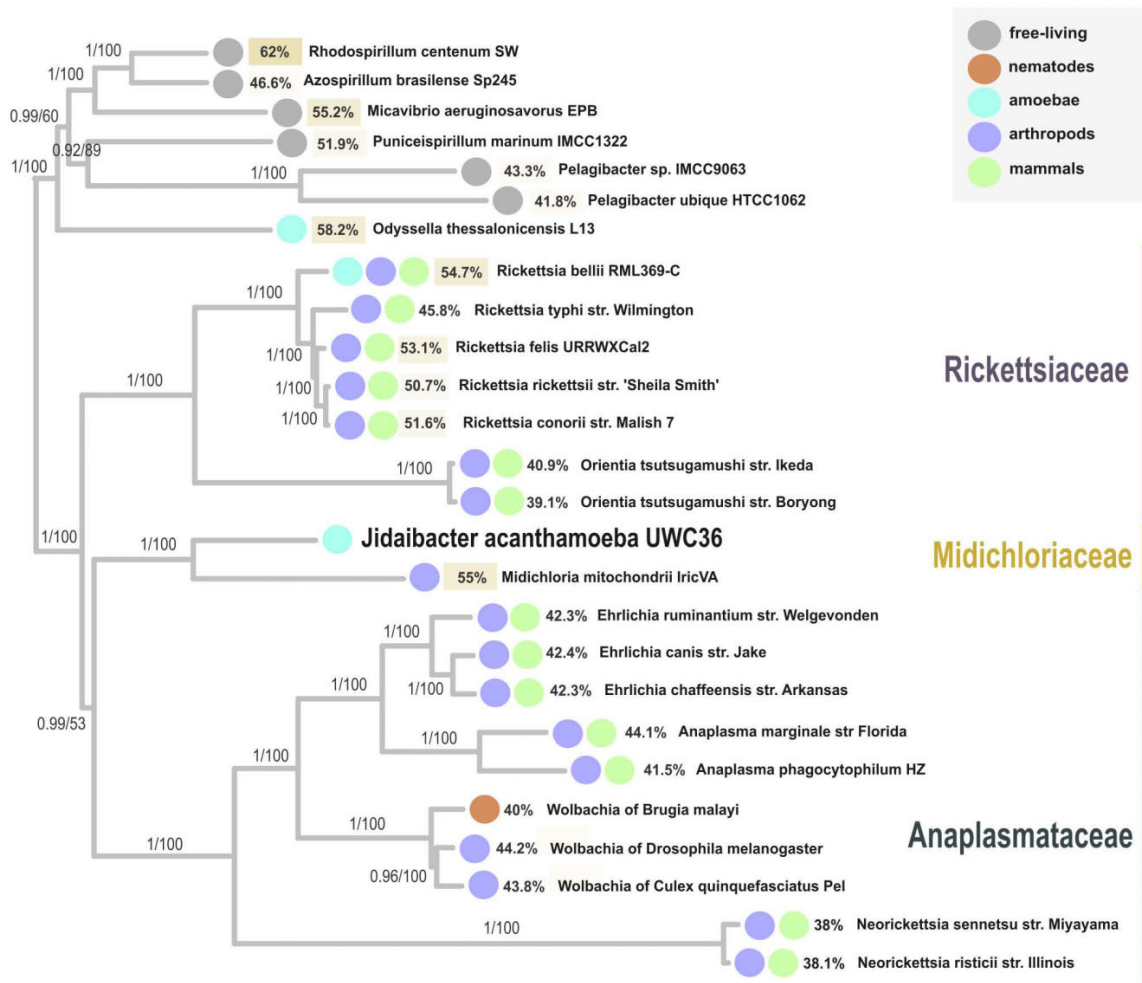
Fig. 3. Phylogenomic relationship of *Jidaibacter acanthamoeba* with members of the *Rickettsiales* and other alphaproteobacteria. The phylogenetic tree is based on a concatenated alignment of 139 single copy genes; discordant genes and sites with a high compositional bias were removed. Numbers at the nodes represent posterior probabilities (Phylobayes; CAT + Poisson) and maximum likelihood bootstrap values (RAxML; LG + GAMMA) respectively. Filled circles at the leaves indicate lifestyle and host specificity; percentages indicate the genome-wide similarity with *Jidaibacter* calculated as numbers of genes in shared COGs normalized to the total number of genes in the *Jidaibacter* genome.

Genetic repertoire

We compared the gene content of *Jidaibacter* with other *Alphaproteobacteria* (Table S1); in total, 330 clusters of orthologous groups (COGs) are shared between all members of the order *Rickettsiales*, which is similar to what was found previously (Fig. 4; Le *et al.*, 2012). However, the *Jidaibacter* genome is markedly distinct from other *Rickettsiales* because (i) it contains the highest number of COGs shared with bacteria outside the order or unique to *Jidaibacter* ($n = 341$), (ii) it encodes the highest number of singletons ($n = 451$), i.e. genes that are not clustered in COGs (Fig. S3A) and (iii) it possesses the highest number of COGs containing duplicated genes ($n = 209$) in our data set (Fig. S3B).

Within the *Rickettsiales*, *Jidaibacter* shared the highest number of orthologues with *Midichloria*, which is not surprising as the two organisms were the closest relatives to each other in our phylogenomic analysis (Fig. 3). But it is striking that *Jidaibacter* actually shares an even higher number of orthologues with free-living alphaproteobacteria outside the *Rickettsiales*, in particular *Rhodospririllum centenum* and *Micavibrio aeruginosavorus* (Fig. 3). This finding suggests that the
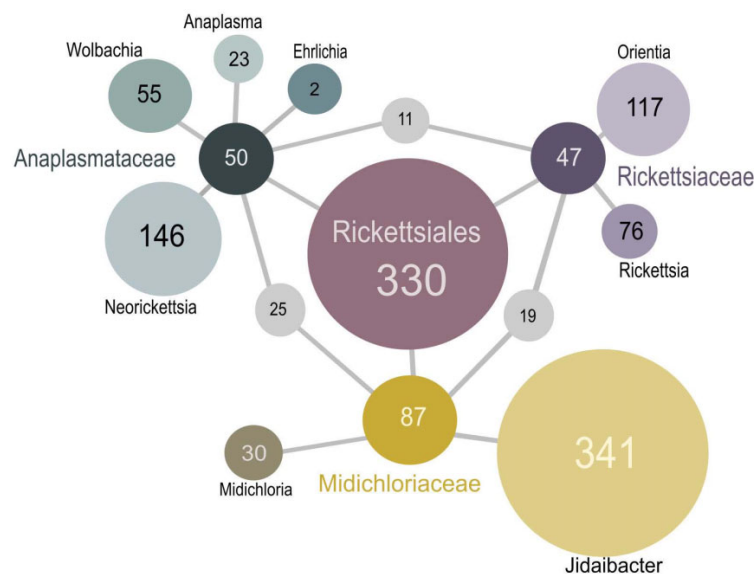
Fig. 4. The pan-genome of *Rickettsiales* and the distinctive position of *Jidaibacter acanthamoeba*. The network graph shows the total numbers of COGs shared among members of rickettsiae at the order level (*Rickettsiales*), family level (*Anaplasmataceae*, *Rickettsiaceae*, *Midichloriaceae*) and genus level (*Orientia*, *Rickettsia*, *Neorickettsia*, *Wolbachia*, *Anaplasma*, *Ehrlichia*, *Midichloria*, *Jidaibacter*). Intersections indicate the numbers of COGs exclusively shared between connected taxa.

Jidaibacter genome encodes many features characteristic for free-living alphaproteobacteria, but absent in the obligate intracellular *Rickettsiales* (Table S3).

Interestingly, *Jidaibacter* comprises a remarkably high number of genes involved in information storage and processing: almost 300 genes involved in replication, recombination and repair and 100 in transcription (Fig. S4). A nearly complete complement of proteins involved in recombination and repair might help to sustain genome integrity and could partially explain the large genome size compared with other members of the *Rickettsiales* (Dale *et al.*, 2003). Another outstanding feature of the *Jidaibacter* genome is the presence of 42 genes involved in cell motility of which 25 are shared with *Midichloria*; none of those genes was found in any other complete *Rickettsiales* genome in our data set.

Flagellar genes

Used mainly for chemotaxis, flagella are common in many free-living bacteria and represent an ancient trait of the *Alphaproteobacteria* (Boussau *et al.*, 2004). For decades, flagella have been considered absent among all *Rickettsiales*, but this view was challenged recently when flagellar genes were identified in the genomes of *Midichloria*, *Odyssella* and the rickettsial endosymbiont of *Trichoplax adhaerens* (RETA) (Georgiades *et al.*, 2011a; Sassera *et al.*, 2011; Mariconti *et al.*, 2012; Driscoll *et al.*, 2013). *Jidaibacter* encodes for 35 flagellar genes, the most complete set found so far in genomes of obligate intracellular alphaproteobacteria (Fig. 5; Table S3).

In our phylogenetic analysis, the *Jidaibacter* flagellar proteins form a clade with those found in *Midichloria*, representing together the most basal lineage in the *Alphaproteobacteria* (Fig. 5). This finding strengthens recent observations, which showed *Midichloria* and RETA at this position (Sassera *et al.*, 2011; Driscoll *et al.*, 2013). Lost in most *Rickettsiales* lineages, the flagellar genes in *Jidaibacter* can thus be considered an ancient feature inherited from free-living bacteria. Although we did not observe flagellar structures or motility for *Jidaibacter*, flagellar motility was just recently detected in two novel *Rickettsiaceae* symbionts of ciliates, and flagellar structures were found in *Lyticum* species, symbionts of paramecia and new members of the *Midichloriaceae* (Boscaro *et al.*, 2013a,b; Vannini *et al.*, 2014).

The maintenance of such a complex molecular machinery suggests a possible role in symbiont transmission and infection process, possibly for initial infection or to exit the host cell; similar to what has been reported from other facultative intracellular bacteria like *Legionella* sp. (Dietrich *et al.*, 2001). Not only both *Jidaibacter* and *Odyssella*, but also other protist symbionts, are predominantly transferred horizontally in aquatic environments or at boundaries between aquatic and soil habitats; a functional flagellum might help to facilitate host contact. Consequently, the paucity of flagellar genes in members of the *Rickettsiaceae* and *Anaplasmataceae*, likely represents the outcome of the adaption to a direct transmission between animal hosts in non-aquatic niches. In such a scenario, *Midichloria* represents an intermediate; after a presumably recent adaptation to an arthropod host, a functional flagellum became obsolete. The flagellar genes
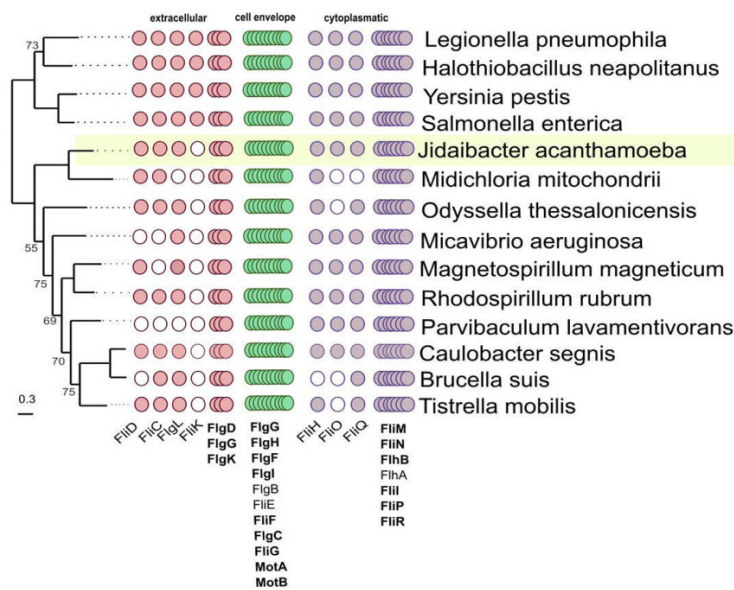
Fig. 5. A complete set of flagellar genes in *Jidaibacter acanthamoeba*. The presence (filled circles) or absence (empty circles) of genes involved in flagellar assembly of selected alpha- and gammaproteobacteria is indicated. The phylogeny of the flagellar apparatus was inferred with RAxML (GAMMA + I + LG) based on a concatenated alignment of 18 flagellar proteins (genes shown in bold); numbers at the nodes represent maximum likelihood bootstrap values; only values below 99% are indicated.

present in the *Midichloria* genome could on the one hand be subject of ongoing gene loss, whereas on the other hand, a subset of flagellar genes could also be of advantage in mediating host interaction, e.g. as a secretion system (Abby and Rocha, 2012; Mariconti *et al.*, 2012).
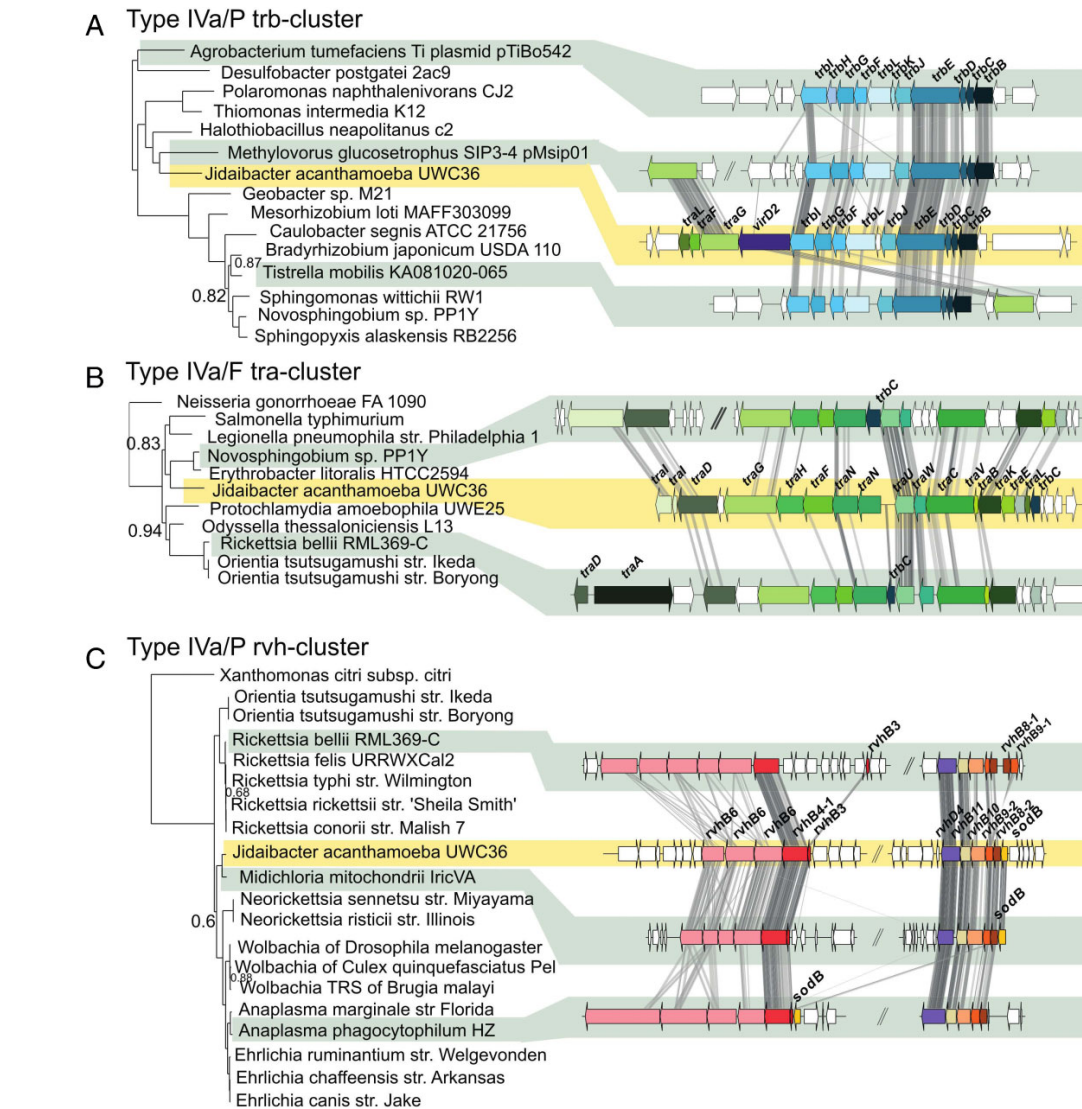
Three type IV secretion systems

Type IV secretion systems (T4SS) are macromolecular transporters that facilitate conjugation and/or move effectors, which can be both DNA and protein, across the bacterial cell envelope into eukaryotic host cells (Backert and Meyer, 2006). The *Jidaibacter* genome encodes for three T4SS clusters, from which two correspond to conjugative systems, *tra* (type IVa/F-type) and *trb* (type IVa/P-type), and one to the *vir* system (type IVa/P-type). We analysed the synteny of genes involved in each T4SS and inferred their phylogeny based on concatenated alignments of key proteins (Fig. 6).

The presence of a *trb* gene cluster is unusual for a *Rickettsiales* genome; in our Bayesian analysis, it groups together with those of the free-living gamma-proteobacterium *Halothiobacillus neapolitanus* and the betaproteobacterium *Methylovorus glucosotrophus* (Fig. 6A). Presence of members of three different proteobacterial classes in one monophyletic clade suggests that the *trb* cluster of *Jidaibacter* has been acquired horizontally. This is further supported by the presence of flanking transposases and tRNAs, which are frequently found near genomic islands. We did not find orthologues of these *trb* genes in any other *Rickettsiales*, but detected proteins with moderate similarity in free-living

Alphaproteobacteria, such as *Tistrella mobilis* or *Agrobacterium tumefaciens*. In the latter, the *trb* gene cluster is involved in the conjugal transfer of the Ti plasmid between *Agrobacterium* cells (Li *et al.*, 1998).

In addition, *Jidaibacter* encodes a conjugative *tra* system, which is also found in other members of the *Rickettsiales* and was recently termed *Rickettsiales* amplified genetic element (RAGE) (Gillespie *et al.*, 2012b, 2014). However, in our phylogenetic tree, the *Jidaibacter tra* genes group together with those found in free-living alphaproteobacteria, in particular *Novosphingobium* sp. and *Erythrobacter litoralis*; whereas the *tra* system of other *Rickettsiales* members and amoeba symbionts, such as *Odyssella* and the only distantly related *Protochlamydia amoebophila*, formed a sister clade (Fig. 6B). The *tra* systems in these intracellular microbes have been studied extensively; however, no experimental evidence for their function is available to date (Ogata *et al.*, 2006; Blanc *et al.*, 2007a; Collingro *et al.*, 2011).

One keystone feature of members of the *Rickettsiales* is the *Rickettsiales vir* homologue (*rvh*), which is considered a stable chronometer for the evolution of this bacterial group (Gillespie *et al.*, 2010, 2014). The *rvh* has likely been acquired horizontally from a free-living gammaproteobacterium and plays a crucial role in host interaction (Frank *et al.*, 2005; Gillespie *et al.*, 2010). Its function in some members of the *Rickettsiales* has recently been experimentally demonstrated to translocate effector proteins (Lockwood *et al.*, 2011; Liu *et al.*, 2012). The *Jidaibacter vir* system can be classified as an *rvh*; its phylogeny follows the species tree of the *Rickettsiales* (Figs 3 and 6C), and arrangement and order of the *rvh*

Fig. 6. The three type IV secretion systems of *Jidaibacter acanthamoeba*.
A. The genomic origin of the *trb* gene cluster found in *Jidaibacter* compared with the *Methylovorus glucosetrophus* SIP3-2 pMsip01 plasmid, with *Agrobacterium tumefaciens* Ti plasmid and with *Tistrella mobilis*.
B. The *tra* gene cluster of *Jidaibacter* compared with *Novosphingobium* sp. PP1Y and with *Rickettsia bellii*.
C. The *vir* gene clusters compared with *Rickettsia bellii*, *Midichloria mitochondrii* and *Anaplasma phagocytophilum*. The phylogeny of each type IV secretion system was inferred with Phylobayes CAT + GTR based on concatenated alignment of (a) TrbB, TrbE, TrbF, TrbG, TrbI; (b) TraC, TraD, TraF, TraG, TraH, TraN, TraU, TraW; (c) RvhB4-1, RvhB8-2, RvhB9-1, RvhB9-2, RvhB10, RvhB11. Orthologous genes are labelled using identical colours; highly conserved regions are connected by vertical grey shading. Bayesian posterior probabilities of 0.95 or lower are indicated.

genes are similar between *Jidaibacter*, *Midichloria* and other *Rickettsiales*.

An armada of putative effector proteins

Once inside the host cell, *Jidaibacter* is likely to interact with its host using a large number of putative effectors, in particular proteins with eukaryotic-like repeat domains such as ankyrin (ANK), leucine-rich repeats (LRR) and tetratricopeptide repeats (TPR), representing more than 10% of the total predicted proteome (Fig. S5). This high percentage is very unusual; it is the highest among *Alphaproteobacteria*, and a similarly expansion of genes encoding potential effector proteins with eukaryotic-like
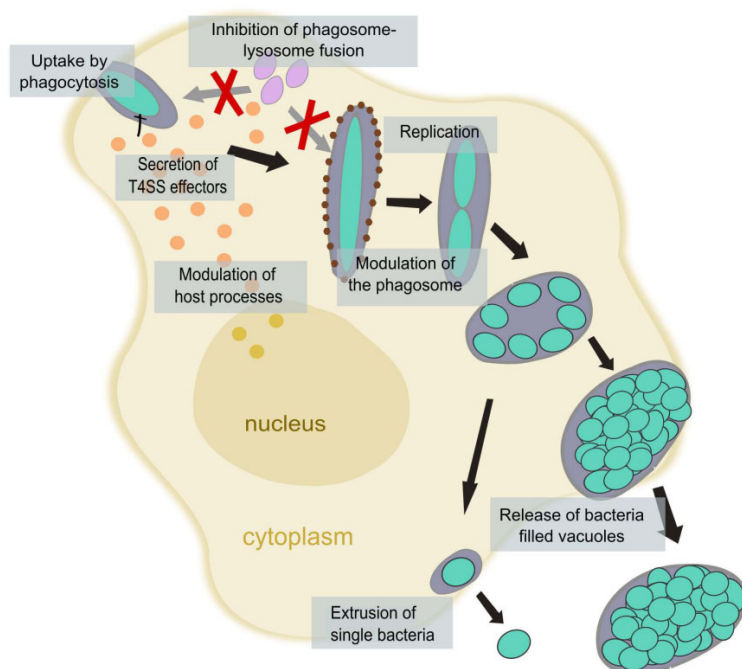
Fig. 7. Model of the *Jidaibacter acanthamoeba* infection process. A model based on experimental observations and genome data is depicted. The symbiont is taken up by phagocytosis and releases effector proteins containing eukaryotic-like domains via its type IV secretion systems. The effectors manipulate processes in the host cell, such as phagosome–lysosome fusion, gene expression and signal transduction. Ribosomes are recruited to the symbiont-containing endosome, and *Jidaibacter* replicates inside this compartment. The bacteria can exit the host cell by extrusion of single cells or vacuoles, or after host cell lysis. See text for further details.

domains has been observed previously only for few other bacteria such as the amoeba symbionts *Amoebophilus asiaticus*, *Protochlamydia amoebophila* or *Neochlamydia* species (Schmitz-Esser *et al.*, 2010; Collingro *et al.*, 2011; Domman *et al.*, 2014; Ishida *et al.*, 2014). Interestingly, the vast majority of putative effector proteins ($n = 132$) contains ANK domains (CL0465), which like LRR (CL0022) and TPR domains (CL0020) are generally involved in protein–protein interactions.

To get a better idea about the role of these putative effector proteins, we searched for the presence of additional functional domains. Several repeat containing proteins contained F-Box (PF00646), U-Box (PF04564), Apc3 (PF12895) and patatin (PF01734) domains, among others (Table S4). Proteins with these eukaryotic-like domains are also found in genomes of other intracellular bacteria, especially in members of the *Rickettsiales* and other amoeba-associated bacteria, where they may modulate the host ubiquitination system, proteasomal degradation, interfere with cell cycle regulation or mediate cytotoxicity (Cho *et al.*, 2007; Pan *et al.*, 2008; Price *et al.*, 2011; Domman *et al.*, 2014; Rolando and Buchrieser, 2014).

Remarkably, putative effectors of *Jidaibacter* with eukaryotic domains are among the largest gene families found in the genome (Table S5). Maintenance of a large number of copies with apparently similar functions in host interactions has been observed recently in chlamydial

symbionts (Domman *et al.*, 2014) and it might represent an adaptation to a larger (yet unknown) host range.

Infection process, host interaction and metabolic potential

By combining our experimental observations and genome-based findings, we propose a model for the *Jidaibacter* infection process (Fig. 7). The genome of *Jidaibacter* comprises a surprisingly large number of genes involved in the biosynthesis of cell envelope (Fig. S6); a full set for peptidoglycan and lipopolysaccharide biosynthesis is present, and in contrast to other *Rickettsiales*, *Jidaibacter* can synthesize a capsule (Fig. 1, Table S3; Fritsche *et al.*, 1999), likely increasing resistance against adverse environmental conditions and host defence mechanisms. After uptake by phagocytosis, the symbiont releases an array of T4SS effector molecules into the host cytoplasm, facilitating the establishment of the intracellular niche (Fig. 6, Fig. S5, Table S4). Our experimental observations show that the bacteria recruit endoplasmic reticulum studded with ribosomes (Fig. 1) to escape phagolysosomal degradation and to create a favourable environment for replication possibly in a process similar to *Legionella*. An effector with homology to *Legionella* VipD present in the *Jidaibacter* genome might contribute to inhibition of lysosomal protein trafficking and phagosome–lysosome fusion (Ku *et al.*,

2012; Gaspar and Machner, 2014). The symbiont may be able to manipulate its host's cell cycle as it encodes multiple proteins harbouring an Apc3 domain (PF12895), which facilitates interactions with the mitotic cyclosome. Furthermore, effector proteins containing ANK, U- and F-box domains have the potential to interfere with host ubiquitination, hijacking diverse cell-signalling processes (Angot *et al.*, 2007; Al-Quadan *et al.*, 2012; Rolando and Buchrieser, 2014).

Reconstruction of the *Jidaibacter* metabolism showed that its genome encodes for similar metabolic features as other members of the *Rickettsiales*, but possess some additional genes in its accessory genome involved in amino acid and carbohydrate metabolism (Fig. S6, Table S3). *Jidaibacter* encodes a full set of genes necessary for gluconeogenesis, but it lacks 6-phosphofructokinase and pyruvate kinase, which are fundamental for glycolysis. The pentose phosphate pathway is present, and the genome encodes all genes of the tricarboxylic acid cycle. Several amino acid biosynthesis pathways are incomplete; however, employing nine putative amino acid transporters, *Jidaibacter* should be able to import missing amino acids from its host cell. Furthermore, we found two putative nucleotide transporters, one of which likely represents an ATP/ADP translocase facilitating energy parasitism (Linka *et al.*, 2003; Daugherty *et al.*, 2004), a widespread feature among obligate intracellular bacteria. The other transporter might help to compensate for missing purine and pyrimidine de novo biosynthesis pathways by importing host nucleotides (Haferkamp *et al.*, 2006; Schmitz-Esser *et al.*, 2008a).

Conclusion

Our current knowledge on genomic patterns underlying symbiont–host interactions in the *Rickettsiales* is biased toward medically important members of the order, which often display highly reduced genomes and features evolved during the adaption to animal hosts. However, the *Rickettsiales* comprise many bacteria affiliated with protists. The amoeba symbiont *Jidaibacter* shows a number of ancient features, such as a large genome size, a complete set of flagellar genes and a high overlap in gene content with free-living bacteria. As suggested previously (Blanc *et al.*, 2007b), it thus represents a missing link between a free-living and a symbiotic lifestyle in the *Alphaproteobacteria*. The *Rickettsiales* have been proposed as the closest relatives to mitochondria, but the origin of these organelles is still controversial (Gray, 2012). The deeply branching phylogenetic position of *Jidaibacter* places it closely to a putative last common ancestor of mitochondria; its analysis will thus contribute to a better understanding of the origin of this organelle and the evolution of eukaryotes.

Experimental procedures

Cultivation of amoebae and infection experiments

Acanthamoeba sp. UWC36 containing *Jidaibacter* (ATCC PRA-6) and symbiont-free *Acanthamoeba castellanii* Neff were maintained in TSY medium (30 g l^{-1} trypticase soy broth, 10 g l^{-1} yeast extract) at 20°C.

For infection experiments, the supernatant of infected amoebae was collected and passed through a 5 μm filter (Sartorius, Germany). The filtrate was centrifuged at 10 500*g* for 10 min, and the resulting bacterial pellet was washed once with Page's amoebic saline (PAS; 0.12 g l^{-1} NaCl, 0.004 g l^{-1} MgSO$_4 \times$ 7H$_2$O, 0.004 g l^{-1} CaCl$_2 \times$ 2H$_2$O, 0.142 g l^{-1} Na$_2$HPO$_4$, 0.136 g l^{-1} KH$_2$PO$_4$), re-suspended in PAS and directly used for infection experiments. Bacterial cell numbers were determined by counting of bacteria, which were filtered onto a 0.2 μm filter (Sartorius, Germany) and stained with 4,6-diamidino-2-phenylindole (DAPI), at an epifluorescence microscope (Axioplan 2 imaging; Zeiss, Germany). *Acanthamoeba castellanii* Neff was seeded in 12-well plates (Thermoscientific, Denmark; 3 × 10^5 cells per well) containing TSY and bacteria at an estimated multiplicity of infection of 100 were added, and infection was enhanced by centrifugation at 370*g* for 15 min. After 2 h, amoeba cells were washed twice with PAS to remove extracellular bacteria. At different time points p.i. amoebae were detached and amoeba cell numbers were determined with a Neubauer counting chamber. Half of the cell suspension was then used for fluorescence in situ hybridization (FISH), the other half for viability staining.

The percentage of infected amoebae was determined by analysing FISH-stained samples. At least 100 amoeba cells were counted per sample. To assess the viability of amoeba cells during infection, amoebae were incubated with propidium iodide (1.5 μM; Molecular Probes, USA) in PAS for 20 min at room temperature in the dark. Samples were washed once with PAS and transferred to a black plate reader dish (Greiner Bio-One, Germany), and fluorescence intensities were determined with a Tecan Infinite M200 plate reader (Tecan, Austria). Negative controls included PAS and suspensions of uninfected amoebae.

For each time point and condition, at least three replicate infection experiments were analysed. To test for statistically significant differences ($P < 0.05$) between different conditions and time points, a two-way analysis of variance (ANOVA) including the Bonferroni post-test was performed using GraphPad Prism (version 6, La Jolla, CA, USA).

Fluorescence in situ hybridization

Amoeba cells were harvested by centrifugation (3000 × *g*, 8 min), washed with PAS and applied on microscope slides for 30 min to allow attachment of cells prior to fixation with 4% formaldehyde (15 min at room temperature). The samples were hybridized for 2 h at 46°C at a formamide concentration of 20% using standard hybridization and washing buffers (Daims *et al.*, 2005) and a combination of the following

probes: a symbiont specific probe (AcRic90, 5′-TGCCA CTAGCAGAACTCC-3′; Fritsche *et al.*, 1999) together with EUK-516 (5′-ACCAGACTTGCCCTCC-3′; Amann and Binder, 1990) targeting most eukaryotes and the EUB338 I-III (5′-GCTGCCTCCCGTAGGAGT-3′, 5′-GCAGCCACCCGTA GGTGT-3′, 5′-GCTGCCACCCGTAGGTGT-3′; Amann and Binder, 1990; Daims *et al.*, 1999) probe mix targeting most bacteria. All probes were purchased from Thermo Fisher Scientific (Germany). Cells were subsequently stained with DAPI (0.5 μg ml⁻¹ in PAS, 3 min), washed once with PAS and embedded in Citifluor (Agar-Scientific, UK). Slides were examined using a confocal laser scanning microscope (SP8, Leica, Germany).

Transmission electron microscopy

Amoebae were detached from culture flasks by shaking and concentrated by centrifugation ($3000 \times g$, 8 min). The supernatant was discarded and samples were fixed for transmission electron microscopy 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% OsO_4 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 CM100 transmission electron microscope operating at 80 kV.

Genome sequencing and assembly

Cells and DNA were prepared as previously described (Schmitz-Esser *et al.*, 2010). Purified DNA was subjected to multiple displacement amplification (MDA) using the REPLI-g Midi kit (QIAGEN) to obtain sufficient amount of DNA required for library preparation. MDA was carried out according to the manufacturer's instructions. A standard paired-end library (Illumina TruSeq) was prepared from 3 μg amplified DNA with insert size of approximately 400 bp and was sequenced on an Illumina HiSeq2000 instrument, yielding 2×100 bp paired reads with a total of approximately 10.5 Gb of sequence data.

Prior to assembly, the quality of the sequence data was assessed with FastQC 0.10.1 (http://www.bioinformatics .babraham.ac.uk/projects/fastqc). The data passed the inspection (average Phred Score per read: 38) and could be used for downstream analyses. The reads were assembled *de novo* using SPAdes 2.4 (Nurk *et al.*, 2013) with default k-mer sizes (21, 33, 55), and using the 'error correction' and 'careful' flags. From the resulting initial assembly, contigs were removed in three steps. First, all contigs smaller than 200 bp and an average k-mer coverage of lower than 300× were removed. Then, all contigs in which more than two thirds of the predicted open reading frames (ORFs) were classified as non-bacterial were removed. ORFs were predicted with PRODI-GAL 2.50 (Hyatt *et al.*, 2010) and classified to a certain taxonomic domain by MEGAN 4.7 (Huson *et al.*, 2011) based on a BLASTP search (Altschul *et al.*, 1990) versus NCBI's non-redundant database.

Genome annotation and analysis

The draft genome of *Jidaibacter* was annotated with an in-house genome annotation pipeline. In brief, putative protein coding genes were identified with a combination of the gene predictors Genemark, Prodigal, Glimmer and Critica (Badger and Olsen, 1999; Besemer *et al.*, 2001; Hyatt *et al.*, 2010; Kelley *et al.*, 2012) with homology information derived from a BLAST search against NCBI non-redundant protein database (NCBI Resource Coordinators, 2014). RNAs were annotated with tRNAScan, RNAmmer and Rfam (Lowe and Eddy, 1997; Griffiths-Jones *et al.*, 2005; Lagesen *et al.*, 2007). Functional prediction was performed by a BLAST against the UniProt/SwissProt database (The UniProt Consortium 2014) and domains were predicted with INTERPROSCAN 5 (Jones *et al.*, 2014). Genes involved in metabolic pathways were assessed with KEGG (Kanehisa *et al.*, 2014) and BIOCYC (Caspi *et al.*, 2014). Furthermore, we screened for putative effector proteins with secret4 (Bi *et al.*, 2013) and effective (Jehl *et al.*, 2011), and by performing an HMM-search from the HMMER3 package against the Pfam database (Bateman, 2002; Eddy, 2011). Identical repeats were detected using the MUMmer repeat-match algorithm (Kurtz *et al.*, 2004) and visualized with CIRCOS (Krzywinski *et al.*, 2009). The completeness of the genome was estimated based on the presence or absence of 31 universal bacterial marker proteins integrated in AMPHORA2 (Wu and Scott, 2012). This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession NZ_JSWE00000000.

Phylogenomics

Clustering of orthologous genes was performed using ORTHOMCL 2.0 on a data set of 78 mainly alphaproteobacterial genomes (Table S1; Li *et al.*, 2003). We applied the following settings; a BLASTP E-value cut-off of 1×10^{-5} and the default MCL inflation parameter of 1.5, which were previously shown to detect orthologues with the highest accuracy (Salichos and Rokas, 2011). We used a perl script included in ORTHOMCL to detect the most central protein from each cluster and annotated each COG by performing a search against the eggNOG database (Powell *et al.*, 2014).

For each COG, sequences were aligned with MAFFT-7.050b (Katoh and Standley, 2013) using the local pair option (mafft-linsi) and subsequently trimmed with trimAl 1.4 (Capella-Gutiérrez *et al.*, 2009), selecting sites for which less than half of the taxa contain a gap. To mitigate the effects of horizontal gene transfer on phylogeny inference, 34 (20%) COGs that displayed the most divergent phylogenies were removed using a discordance filter (Williams *et al.*, 2010; Guy *et al.*, 2014). The single gene trees required for the discordance filter were inferred using RAxML 7.2.8 (Stamatakis, 2006) with 100 non-parametric bootstraps, the GAMMA model for rate heterogeneity among sites and the LG substitution matrix (Le and Gascuel, 2008). The remaining COGs were then concatenated, resulting in an alignment of 48 721 sites. To account for compositional bias in the data set, a χ^2-filter was applied (Viklund *et al.*, 2012) that split the alignment into two halves: one half which contained sites that positively and the other half containing the sites that negatively contributed to the compositional bias respectively. For each half, phylogenies were inferred using maximum likelihood and Bayesian methods. For maximum likelihood RAxML 7.2.8 was used, with 100 rapid bootstraps, the

GAMMA model for rate heterogeneity and the LG substitution matrix. For Bayesian analysis, PHYLOBAYES-MPI 1.4F (Lartillot *et al.*, 2013) was used, using 4 Markov Chain Monte Carlo (MCMC) chains under the CAT–Poisson model that ran for 20 000 generations. The final Bayesian consensus tree was based on all four chains, removing the first 1000 generations (burn-in), and sampling every 100 generations.

Phylogeny and synteny of type IV secretion systems and flagellar genes

Protein sequences were extracted from COGs with annotated functions in either flagellar assembly or type IV secretion. Data sets were checked for completeness and proteins from additional reference genomes were added using HMM-search from the HMMER3 package (Eddy, 2011) based on profile HMMs built for the respective data set. Sequences in each data set were aligned with MAFFT-7.050B (Katoh and Standley, 2013) using the local pair option (mafft-linsi) and the 'maxiterate 1000' setting and subsequently trimmed with TRIMAL 1.4 (Capella-Gutiérrez *et al.*, 2009), selecting sites for which less than half of the taxa contain a gap. If in a COG, more than one protein was present per taxon, then only the protein with the longest sequence was retained in the alignment. Protein alignments were concatenated with FASCONCAT (Kück and Meusemann, 2010) and trees were calculated with PHYLOBAYES-MPI 1.4F (Lartillot *et al.*, 2013), using two MCMC chains under the CAT-GTR model that ran until convergence was reached (maxdiff < 0.1). A burnin was chosen as 1/5 of the chain's length. For maximum likelihood, RAxML 7.2.8 was used (Stamatakis, 2006), with 1000 rapid bootstraps, the GAMMA model for rate heterogeneity and the LG substitution matrix (Le and Gascuel, 2008). Synteny of gene clusters was assessed in GENOPLOTR (Guy *et al.*, 2010) based on comparative sequence analysis with TBLASTX (Altschul *et al.*, 1990).

Acknowledgements

References

Abby, S.S., and Rocha, E.P.C. (2012) The non-flagellar type III secretion system evolved from the bacterial flagellum and diversified into host-cell adapted systems. *PLoS Genet* **8:** e1002983.

Aistleitner, K., Heinz, C., Hörmann, A., Heinz, E., Montanaro, J., Schulz, F., *et al.* (2013) Identification and characterization of a novel porin family highlights a major difference in the outer membrane of chlamydial symbionts and pathogens. *PLoS ONE* **8:** e55010.

Al-Quadan, T., Price, C.T., and Abu Kwaik, Y. (2012) Exploitation of evolutionarily conserved amoeba and mammalian processes by *Legionella*. *Trends Microbiol* **20:** 299–306.

Altschul, S., Gish, W., and Miller, W. (1990) Basic local alignment search tool. *J Mol Biol* **215:** 403–410.

Amann, R., and Binder, B. (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* **56:** 1919–1925.

Andersson, S.G.E. (1998) Bioenergetics of the obligate intracellular parasite *Rickettsia prowazekii*. *Biochim Biophys Acta* **1365:** 105–111.

Andersson, S.G.E., Zomorodipour, A., Andersson, J.O., Sicheritz-Pontén, T., Alsmark, U.C.M., Podowski, R.M., *et al.* (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396:** 133–140.

Angot, A., Vergunst, A., Genin, S., and Peeters, N. (2007) Exploitation of eukaryotic ubiquitin signaling pathways by effectors translocated by bacterial type III and type IV secretion systems. *PLoS Pathog* **3:** e3.

Azad, F., and Traub, R. (1985) Transmission of murine typhus rickettsiae by *Xenopsylla cheopis*, with notes on experimental infection and effects of temperature. *Am J Trop Med Hyg* **34:** 555–563.

Backert, S., and Meyer, T.F. (2006) Type IV secretion systems and their effectors in bacterial pathogenesis. *Curr Opin Microbiol* **9:** 207–217.

Badger, J.H., and Olsen, G.J. (1999) CRITICA: coding region identification tool invoking comparative analysis. *Mol Biol Evol* **16:** 512–524.

Bateman, A. (2002) The Pfam protein families database. *Nucleic Acids Res* **30:** 276–280.

Besemer, J., Lomsadze, A., and Borodovsky, M. (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* **29:** 2607–2618.

Beyer, A.R., Truchan, H.K., May, L.J., Walker, N.J., Borjesson, D.L., and Carlyon, J.A. (2014) The *Anaplasma phagocytophilum* effector AmpA hijacks host cell SUMOylation. *Cell Microbiol* **17:** 504–519.

Bi, D., Liu, L., Tai, C., Deng, Z., Rajakumar, K., and Ou, H.-Y. (2013) SecReT4: a web-based bacterial type IV secretion system resource. *Nucleic Acids Res* **41:** D660–D665.

Blanc, G., Ogata, H., Robert, C., Audic, S., Claverie, J.-M., and Raoult, D. (2007a) Lateral gene transfer between obligate intracellular bacteria: evidence from the *Rickettsia massiliae* genome. *Genome Res* **17:** 1657–1664.

Blanc, G., Ogata, H., Robert, C., Audic, S., Suhre, K., Vestris, G., *et al.* (2007b) Reductive genome evolution from the mother of *Rickettsia*. *PLoS Genet* **3:** e14.

Boscaro, V., Petroni, G., Ristori, A., Verni, F., and Vannini, C. (2013a) 'Candidatus Defluviella procrastinata' and

'Candidatus Cyrtobacter zanobii', two novel ciliate endosymbionts belonging to the 'Midichloria Clade'. *Microb Ecol* **65:** 302–310.

Boscaro, V., Schrallhammer, M., Benken, K.A., Krenek, S., Szokoli, F., Berendonk, T.U., *et al.* (2013b) Rediscovering the genus *Lyticum*, multiflagellated symbionts of the order Rickettsiales. *Sci Rep* **3:** 3305.

Boussau, B., Karlberg, E.O., Frank, A.C., Legault, B.-A., and Andersson, S.G.E. (2004) Computational inference of scenarios for alpha-proteobacterial genome evolution. *Proc Natl Acad Sci USA* **101:** 9722–9727.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25:** 1972–1973.

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C.A., *et al.* (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* **42:** D459–D471.

Cho, N.-H., Kim, H.-R., Lee, J.-H., Kim, S.-Y., Kim, J., Cha, S., *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 USA* **104:** 7981–7986.

Collingro, A., Walochnik, J., Baranyi, C., Michel, R., Wagner, M., Horn, M., and Aspöck, H. (2004) Chlamydial endocytobionts of free-living amoebae differentially affect the growth rate of their hosts. *Eur J Protistol* **40:** 57–60.

Collingro, A., Tischler, P., Weinmaier, T., Penz, T., Heinz, E., Brunham, R.C., *et al.* (2011) Unity in variety – the pangenome of the Chlamydiae. *Mol Biol Evol* **28:** 3253–3270.

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

Daims, H., Stoecker, K., and Wagner, M. (2005) Fluorescence in situ hybridization for the detection of prokaryotes. *Mol Microb Ecol* **213:** 239.

Dale, C., Wang, B., Moran, N., and Ochman, H. (2003) Loss of DNA recombinational repair enzymes in the initial stages of genome degeneration. *Mol Biol Evol* **20:** 1188–1194.

Darby, A.C., Cho, N.-H., Fuxelius, H.-H., Westberg, J., and Andersson, S.G.E. (2007) Intracellular pathogens go extreme: genome evolution in the Rickettsiales. *Trends Genet* **23:** 511–520.

Daugherty, R.M., Linka, N., Audia, J.P., Urbany, C., Neuhaus, H.E., and Winkler, H.H. (2004) The nucleotide transporter of *Caedibacter caryophilus* exhibits an extended substrate spectrum compared to the analogous ATP/ADP translocase of *Rickettsia prowazekii*. *J Bacteriol* **186:** 3262–3265.

Dietrich, C., Heuner, K., Brand, B., Hacker, J., and Steiner, M. (2001) Flagellum of *Legionella pneumophila* positively affects the early phase of infection of eukaryotic host cells. *Infect Immun* **69:** 2116–2122.

Domman, D., Collingro, A., Lagkouvardos, I., Gehre, L., Weinmaier, T., Rattei, T., *et al.* (2014) Massive expansion of ubiquitination-related gene families within the Chlamydiae. *Mol Biol Evol* **31:** 2890–2904.

Dreher-Lesnick, S.M., Ceraul, S.M., Rahman, M.S., and Azad, A.F. (2008) Genome-wide screen for temperature-regulated genes of the obligate intracellular bacterium, *Rickettsia typhi*. *BMC Microbiol* **8:** 61.

Driscoll, T., Gillespie, J.J., Nordberg, E.K., Azad, A.F., and Sobral, B.W. (2013) Bacterial DNA sifted from the *Trichoplax adhaerens* (Animalia: Placozoa) genome project reveals a putative rickettsial endosymbiont. *Genome Biol Evol* **5:** 621–645.

Eddy, S.R. (2011) Accelerated profile HMM searches. *PLoS Comput Biol* **7:** e1002195.

Ettema, T.J.G., and Andersson, S.G.E. (2009) The alpha-proteobacteria: the Darwin finches of the bacterial world. *Biol Lett* **5:** 429–432.

Ferla, M.P., Thrash, J.C., Giovannoni, S.J., and Patrick, W.M. (2013) New rRNA gene-based phylogenies of the Alphaproteobacteria provide perspective on major groups, mitochondrial ancestry and phylogenetic instability. *PLoS ONE* **8:** e83383.

Ferrantini, F., Fokin, S.I., Modeo, L., Andreoli, I., Dini, F., Görtz, H.-D., *et al.* (2009) 'Candidatus Cryptoprodotis polytropus,' a novel Rickettsia-like organism in the ciliated protist *Pseudomicrothorax dubius* (Ciliophora, Nassophorea). *J Eukaryot Microbiol* **56:** 119–129.

Foster, P.G., and Hickey, D.A. (1999) Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *J Mol Evol* **48:** 284–290.

Frank, A.C., Alsmark, C.M., Thollesson, M., and Andersson, S.G.E. (2005) Functional divergence and horizontal transfer of type IV secretion systems. *Mol Biol Evol* **22:** 1325–1336.

Fritsche, T.R., Horn, M., Seyedirashti, S., Gautom, R.K., Schleifer, K.H., and Wagner, M. (1999) In situ detection of novel bacterial endosymbionts of *Acanthamoeba* spp. phylogenetically related to members of the order Rickettsiales. *Appl Environ Microbiol* **65:** 206–212.

Gaspar, A.H., and Machner, M.P. (2014) VipD is a Rab5-activated phospholipase A1 that protects *Legionella pneumophila* from endosomal fusion. *Proc Natl Acad Sci USA* **111:** 4560–4565.

Georgiades, K., Madoui, M.-A., Le, P., Robert, C., and Raoult, D. (2011a) Phylogenomic analysis of *Odyssella thessalonicensis* fortifies the common origin of Rickettsiales, *Pelagibacter ubique* and *Reclimonas americana* mitochondrion. *PLoS ONE* **6:** e24857.

Georgiades, K., Merhej, V., El Karkouri, K., Raoult, D., and Pontarotti, P. (2011b) Gene gain and loss events in *Rickettsia* and *Orientia* species. *Biol Direct* **6:** 6.

Gillespie, J.J., Brayton, K.A., Williams, K.P., Diaz, M.A.Q., Brown, W.C., Azad, A.F., and Sobral, B.W. (2010) Phylogenomics reveals a diverse Rickettsiales type IV secretion system. *Infect Immun* **78:** 1809–1823.

Gillespie, J.J., Nordberg, E., Azad, A.F., and Sobral, B.W. (2012a) Phylogeny and Comparative Genomics: The Shifting Landscape in the Genomics Era. In *Intracellular Pathogens II: Rickettsiales*. Azad, A.F., and Palmer, G.H. (eds.). American Society of Microbiology, pp. 84–141.

Gillespie, J.J., Joardar, V., Williams, K.P., Driscoll, T., Hostetler, J.B., Nordberg, E., *et al.* (2012b) 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.

Gillespie, J.J., Kaur, S.J., Rahman, M.S., Rennoll-Bankert, K., Sears, K.T., Beier-Sexton, M., and Azad, A.F. (2014) Secretome of obligate intracellular Rickettsia. *FEMS Microbiol Rev* **39:** 47–80.

Gray, M.W. (1998) Rickettsia, typhus and the mitochondrial connection. *Nature* **396:** 109–110.

Gray, M.W. (2012) Mitochondrial evolution. *Cold Spring Harb Perspect Biol* **4:** a011403.

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R., and Bateman, A. (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* **33:** D121–D124.

Guy, L., Kultima, J.R., and Andersson, S.G.E. (2010) genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26:** 2334–2335.

Guy, L., Saw, J.H., and Ettema, T.J.G. (2014) The archaeal legacy of Eukaryotes: a phylogenomic perspective. *Cold Spring Harb Perspect Biol* **6:** a016022.

Haferkamp, I., Schmitz-Esser, S., Wagner, M., Neigel, N., Horn, M., and Neuhaus, H.E. (2006) Tapping the nucleotide pool of the host: novel nucleotide carrier proteins of *Protochlamydia amoebophila*. *Mol Microbiol* **60:** 1534–1545.

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

Horn, M., Fritsche, T.R., Gautom, R.K., Schleifer, K.-H., and Wagner, M. (1999) Novel bacterial endosymbionts of *Acanthamoeba* spp. related to the *Paramecium caudatum* symbiont *Caedibacter caryophilus*. *Environ Microbiol* **1:** 357–367.

Huson, D.H., Mitra, S., Ruscheweyh, H.-J., Weber, N., and Schuster, S.C. (2011) Integrative analysis of environmental sequences using MEGAN4. *Genome Res* **21:** 1552–1560.

Hyatt, D., Chen, G.-L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11:** 119.

Ishida, K., Sekizuka, T., Hayashida, K., Matsuo, J., Takeuchi, F., Kuroda, M., *et al.* (2014) Amoebal endosymbiont *Neochlamydia* genome sequence illuminates the bacterial role in the defense of the host amoebae against *Legionella pneumophila*. *PLoS ONE* **9:** e95166.

Jehl, M.-A., Arnold, R., and Rattei, T. (2011) Effective – a database of predicted secreted bacterial proteins. *Nucleic Acids Res* **39:** D591–D595.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30:** 1236–1240.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* **42:** D199–D205.

Katoh, K., and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30:** 772–780.

Kawafune, K., Hongoh, Y., Hamaji, T., and Nozaki, H. (2012) Molecular identification of rickettsial endosymbionts in the non-phagotrophic volvocalean green algae. *PLoS ONE* **7:** e31749.

Kelley, D.R., Liu, B., Delcher, A.L., Pop, M., and Salzberg, S.L. (2012) Gene prediction with Glimmer for metagenomic sequences augmented by classification and clustering. *Nucleic Acids Res* **40:** e9.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., *et al.* (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* **19:** 1639–1645.

Ku, B., Lee, K.-H., Park, W.S., Yang, C.-S., Ge, J., Lee, S.-G., *et al.* (2012) VipD of *Legionella pneumophila* targets activated Rab5 and Rab22 to interfere with endosomal trafficking in macrophages. *PLoS Pathog* **8:** e1003082.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004) Versatile and open software for comparing large genomes. *Genome Biol* **5:** R12.

Kück, P., and Meusemann, K. (2010) FASconCAT: convenient handling of data matrices. *Mol Phylogenet Evol* **56:** 1115–1118.

Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H.-H., Rognes, T., and Ussery, D.W. (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35:** 3100–3108.

Lagkouvardos, I., Weinmaier, T., Lauro, F.M., Cavicchioli, R., Rattei, T., and Horn, M. (2013) Integrating metagenomic and amplicon databases to resolve the phylogenetic and ecological diversity of the Chlamydiae. *The ISME journal* **8:** 115–125.

Lartillot, N., Rodrigue, N., Stubbs, D., and Richer, J. (2013) PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst Biol* **62:** 611–615.

Le, P.T., Ramulu, H.G., Guijarro, L., Paganini, J., Gouret, P., Chabrol, O., *et al.* (2012) An automated approach for the identification of horizontal gene transfers from complete genomes reveals the rhizome of Rickettsiales. *BMC Evol Biol* **12:** 243.

Le, S.Q., and Gascuel, O. (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* **25:** 1307–1320.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* **13:** 2178–2189.

Li, P.L., Everhart, D.M., and Farrand, S.K. (1998) Genetic and sequence analysis of the pTiC58 trb locus, encoding a mating-pair formation system related to members of the type IV secretion family. *J Bacteriol* **180:** 6164–6172.

Linka, N., Hurka, H., Lang, B.F., Burger, G., Winkler, H.H., Stamme, C., *et al.* (2003) Phylogenetic relationships of non-mitochondrial nucleotide transport proteins in bacteria and eukaryotes. *Gene* **306:** 27–35.

Liu, H., Bao, W., Lin, M., Niu, H., and Rikihisa, Y. (2012) Ehrlichia type IV secretion effector ECH0825 is translocated to mitochondria and curbs ROS and apoptosis by upregulating host MnSOD. *Cell Microbiol* **14:** 1037–1050.

Lockwood, S., Voth, D.E., Brayton, K.A., Beare, P.A., Brown, W.C., Heinzen, R.A., and Broschat, S.L. (2011) Identification of *Anaplasma marginale* type IV secretion system effector proteins. *PLoS ONE* **6:** e27724.

Lowe, T.M., and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25:** 955–964.

McCutcheon, J.P., and Moran, N.A. (2012) Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* **10:** 13–26.

McFall-Ngai, M., Hadfield, M.G., Bosch, T.C.G., Carey, H.V., Domazet-Lošo, T., Douglas, A.E., *et al.* (2013) Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci USA* **110:** 3229–3236.

Mariconti, M., Epis, S., Sacchi, L., Biggiogera, M., Sassera, D., Genchi, M., *et al.* (2012) A study on the presence of flagella in the order Rickettsiales: the case of 'Candidatus Midichloria mitochondrii'. *Microbiology* **158:** 1677–1683.

Mavromatis, K., Doyle, C.K., Lykidis, A., Ivanova, N., Francino, M.P., Chain, P., *et al.* (2006) The genome of the obligately intracellular bacterium *Ehrlichia canis* reveals themes of complex membrane structure and immune evasion strategies. *J Bacteriol* **188:** 4015–4023.

Montagna, M., Sassera, D., Epis, S., Bazzocchi, C., Vannini, C., Lo, N., *et al.* (2013) 'Candidatus Midichloriaceae' fam. nov. (Rickettsiales), an ecologically widespread clade of intracellular alphaproteobacteria. *Appl Environ Microbiol* **79:** 3241–3248.

NCBI Resource Coordinators (2014) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **42:** D7– D17.

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A.A., Korobeynikov, A., Lapidus, A., *et al.* (2013) Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol* **20:** 714–737.

Ogata, H., La Scola, B., Audic, S., Renesto, P., Blanc, G., Robert, C., *et al.* (2006) Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genet* **2:** e76.

Pan, X., Lührmann, A., Satoh, A., Laskowski-Arce, M.A., and Roy, C.R. (2008) Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* **320:** 1651–1654.

Pilhofer, M., Aistleitner, K., Ladinsky, M.S., König, L., Horn, M., and Jensen, G.J. (2014) Architecture and host interface of environmental chlamydiae revealed by electron cryotomography. *Environ Microbiol* **16:** 417–429.

Powell, S., Forslund, K., Szklarczyk, D., Trachana, K., Roth, A., Huerta-Cepas, J., *et al.* (2014) eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res* **42:** D231–D239.

Price, C.T.D., Al-Quadan, T., Santic, M., Rosenshine, I., and Kwaik, Y.A. (2011) Host proteasomal degradation generates amino acids essential for intracellular bacterial growth. *Science* **334:** 1553–1557.

Renvoisé, A., Merhej, V., Georgiades, K., and Raoult, D. (2011) Intracellular Rickettsiales: insights into manipulators of eukaryotic cells. *Trends Mol Med* **17:** 573–583.

Rodríguez-Ezpeleta, N., and Embley, T.M. (2012) The SAR11 group of alpha-proteobacteria is not related to the origin of mitochondria. *PLoS ONE* **7:** e30520.

Rolando, M., and Buchrieser, C. (2014) *Legionella pneumophila* type IV effectors hijack the transcription and translation machinery of the host cell. *Trends Cell Biol* **24:** 771–778.

Salichos, L., and Rokas, A. (2011) Evaluating ortholog prediction algorithms in a yeast model clade. *PLoS ONE* **6:** e18755.

Santos, H.A., and Massard, C.L. (2014) The family Holosporaceae. In *The Prokaryotes*. Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., and Thompson, F. (eds). Berlin, Germany: Springer, pp. 237–246.

Sassera, D., Beninati, T., Bandi, C., Bouman, E.A.P., Sacchi, L., Fabbi, M., and Lo, N. (2006) 'Candidatus Midichloria mitochondrii', an endosymbiont of the tick *Ixodes ricinus* with a unique intramitochondrial lifestyle. *Int J Syst Evol Microbiol* **56:** 2535–2540.

Sassera, D., Lo, N., Epis, S., and D'Auria, G. (2011) Phylogenomic evidence for the presence of a flagellum and cbb3 oxidase in the free-living mitochondrial ancestor. *Mol Biol Evol* **28:** 3285–3296.

Sällström, B., and Andersson, S.G. (2005) Genome reduction in the α-Proteobacteria. *Curr Opin Microbiol* **8:** 579–585.

Schmitz-Esser, S., Linka, N., Collingro, A., Beier, C.L., Neuhaus, H.E., Wagner, M., and Horn, M. (2004) ATP/ADP translocases: a common feature of obligate intracellular amoebal symbionts related to Chlamydiae and Rickettsiae. *J Bacteriol* **186:** 683–691.

Schmitz-Esser, S., Haferkamp, I., Knab, S., Penz, T., Ast, M., Kohl, C., *et al.* (2008a) *Lawsonia intracellularis* contains a gene encoding a functional rickettsia-like ATP/ADP translocase for host exploitation. *J Bacteriol* **190:** 5746–5752.

Schmitz-Esser, S., Toenshoff, E.E.R., Haider, S., Heinz, E., Hoenninger, V.M., Wagner, M., *et al.* (2008b) Diversity of bacterial endosymbionts of environmental acanthamoeba isolates. *Appl Environ Microbiol* **74:** 5822–5831.

Schmitz-Esser, S., Tischler, P., Arnold, R., Montanaro, J., Wagner, M., Rattei, T., and 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.

Schulz, F., Lagkouvardos, I., Wascher, F., Aistleitner, K., Kostanjšek, R., and Horn, M. (2014) Life in an unusual intracellular niche: a bacterial symbiont infecting the nucleus of amoebae. *ISME J* **8:** 1634–1644.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688–2690.

UniProt Consortium (2014) Activities at the universal protein resource (UniProt). *Nucleic Acids Res* **42:** D191–D198.

Vannini, C., Petroni, G., Verni, F., and Rosati, G. (2005) A bacterium belonging to the Rickettsiaceae family inhabits the cytoplasm of the marine ciliate *Diophrys appendiculata* (Ciliophora, Hypotrichia). *Microb Ecol* **49:** 434–442.

Vannini, C., Ferrantini, F., Schleifer, K.-H., Ludwig, W., Verni, F., and Petroni, G. (2010) Candidatus anadelfobacter veles' and 'Candidatus cyrtobacter comes,' two new Rickettsiales species hosted by the protist ciliate *Euplotes harpa* (Ciliophora, Spirotrichea). *Appl Environ Microbiol* **76:** 4047–4054.

Vannini, C., Boscaro, V., Ferrantini, F., Benken, K.A., Mironov, T.I., Schweikert, M., *et al.* (2014) Flagellar movement in two bacteria of the family Rickettsiaceae: a re-evaluation of motility in an evolutionary perspective. *PLoS ONE* **9:** e87718.

Viklund, J., Ettema, T.J.G., and Andersson, S.G.E. (2012) Independent genome reduction and phylogenetic reclassification of the oceanic SAR11 clade. *Mol Biol Evol* **29:** 599–615.

Wang, Z., and Wu, M. (2015) An integrated phylogenomic approach toward pinpointing the origin of mitochondria. *Sci Rep* **5:** 7949.

Weinert, L.A., Werren, J.H., Aebi, A., Stone, G.N., and Jiggins, F.M. (2009) Evolution and diversity of Rickettsia bacteria. *BMC Biol* **7:** 6.

Werren, J.H., Baldo, L., and Clark, M.E. (2008) Wolbachia: master manipulators of invertebrate biology. *Nat Rev Microbiol* **6:** 741–751.

Williams, K.P., Sobral, B.W., and Dickerman, A.W. (2007) A robust species tree for the alphaproteobacteria. *J Bacteriol* **189:** 4578–4586.

Williams, K.P., Gillespie, J.J., Sobral, B.W.S., Nordberg, E.K., Snyder, E.E., Shallom, J.M., and Dickerman, A.W. (2010) Phylogeny of gammaproteobacteria. *J Bacteriol* **192:** 2305–2314.

Wu, M., and Scott, A.J. (2012) Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* **28:** 1033–1034.

Zug, R., and 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.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Circular plot of the *Jidaibacter acanthamoeba* genome. The outermost circle depicts the 108 largest contigs (> 2 kb) out of in total 249 contigs (> 200 nt) of the *Jidaibacter* draft genome; genes were assigned to eggNOG functional categories and are differentially colored. The GC skew of the genome is shown as a histogram (yellow/blue), and the AT content as a heatmap with dark red indicating highest AT content. Structural repeats with a length of at least 100 nt and a DNA sequence similarity of higher than 98% are connected with each other (orange lines; the higher the similarity the darker the connecting line). In addition, translated coding regions with an amino acid similarity of more than 70% are connected (purple lines; the higher the similarity the darker the connecting line).

Fig. S2. Influence of compositional bias, horizontal gene transfer and long-branch attraction on the phylogenetic placement of *Jidaibacter*.

A. Phylogenetic tree inferred by Maximum-likelihood with the LG model based on a concatenated alignment of 173 single-copy genes. Bootstrap support values lower than 99 are shown.

B. Ranking of the 173 single-copy genes by their discordance score; the cutoff for the selection of low biased genes is shown (dashed line).

Fig. S3. The genome of *Jidaibacter acanthamoeba* encodes for many unique features and the largest number of COGs with duplicated genes.

A. The uniqueness of *Rickettsiales* genomes is illustrated by the number of genes exclusively present in the respective genome (singletons) plotted against the total number of genes present in COGs which are not shared with other members of the order (unique COGs).

B. The total number of COGs containing duplicated genes is shown for members of the *Rickettsiales* in correlation with the total number of duplicated genes. The size of the circles correspond with the genome size of the respective organism. UWC36, *Jidaibacter acanthamoeba*; OrTs, *Orientia tsutsugamushi* str. Ikeda; AnPh, *Anaplasma phagocytophilum* HZ; MiMi, *Midichloria mitochondrii I*ricVA; EhCh, *Ehrlichia chaffeensis* str. Arkansas; Wolb, *Wolbachia* endosymbiont of *Drosophila melanogaster*; OrTB, *Orientia tsutsugamushi* str. Boryong; RiBe, *Rickettsia bellii* RML369-C; WoCu, *Wolbachia* endosymbiont of *Culex quinquefasciatus* Pel; NeSe, *Neorickettsia sennetsu* str. Miyayama; RiFe, *Rickettsia felis* URRWXCal2; RiRi, *Rickettsia rickettsii* str. Sheila Smith; NeRi, *Neorickettsia risticii* str. Illinois; AnMa, *Anaplasma marginale* str. Florida; RiCo, *Rickettsia conorii* str. Malish 7; EhCa, *Ehrlichia canis* str. Jake; EhRu, *Ehrlichia ruminantium* str. Welgevonden; WoBr, *Wolbachia* endosymbiont str. TRS of *Brugia malayi*; RiTy, *Rickettsia typhi* str. Wilmington.

Fig. S4. Genetic repertoire of *Jidaibacter*. Genes have been grouped in COGs and assigned to functional categories using eggNOG. Numbers of *Jidaibacter* genes per functional category are shown as light green bars. Numbers of COGs shared with *M. mitochondrii* (dark green bars) and all *Rickettsiales* (purple bars), respectively are indicated.

Fig. S5. The *Jidaibacter* genome encodes a high number of eukaryotic-like repeat domains. The distribution of proteins among *Rickettsiales* members containing at least one eukaryotic-like repeat domain out of major repeat superfamilies [Ankyrin (Pfam clan: CL0465), TPR (Pfam clan:CL0020), LRR (Pfam clan: CL0022), Pentapeptide (Pfam clan: CL0505)] is shown. The colored background represents the fraction these proteins make up in the respective organism's proteome.

Fig. S6. Metabolic capabilities of *Jidaibacter acanthamoeba* compared to other members of the *Rickettsiales*. Major metabolic categories are depicted in terms of the completeness of pathways predicted with BioCyc (dark color, complete; middle tone, one enzyme missing; light tone, two or more enzymes missing; white, pathway absent). Hierarchical clustering indicated at the top was performed based on the presence/absence of enzymes in the respective pathways.

Table S1. Genome sequences used in this study. Members of the *Rickettsiales* are in bold.

Table S2. Antibiotic treatment to cure *Acanthamoeba* sp. UWC36 from *Jidaibacter*. A combination of rifampicin (antibiotic 1) with a second antibiotic (antibiotic 2) was applied in various concentrations and under different incubation temperatures for up to 5 months.

Table S3. Selection of genes encoded in the *Jidaibacter* accessory genome which are either involved in metabolic processes or play a role in flagellar biosynthesis. Proteins were selected based on their presence in COGs which did not contain proteins of other *Rickettsiales* member. The table shows functional annotation of these proteins

classified in eggNOG functional categories (* indicates manual annotation). PFAM domains present in the respective protein are indicated. In addition, proteins possibly involved in biosynthesis of a flagellum and T4SS clusters are shown.

Table S4. Putative effector proteins encoded in the *Jidaibacter acanthamoeba* genome and their potential role in host interaction. Proteins in which at least one eukaryotic-like repeat domain together with one or more functional domains are indicated.

Table S5. The 15 largest COGs of *Jidaibacter*. PFAM domains found in the representative proteins are indicated.

Text S1. Description of '*Candidatus* Jidaibacter acanthamoeba'.

A *Rickettsiales* Symbiont of Amoebae with Ancient Features

Frederik Schulz, Joran Martijn, Florian Wascher, Ilias Lagkouvardos, Rok Kostanjšek, Thijs J. G. Ettema, Matthias Horn

Supplementary Information

Supplementary Text S1.

Description of 'Candidatus Jidaibacter acanthamoeba'. The name *Jidaibacter* pertains to the affiliation of the organism with members of the *Midichloriaceae*. The prefix "Jidai" is derived from Japanese "Era" and, like *Midichloria mitochondrii*, resembles character names (the Jedi and midi-chlorians, respectively) in George Lucas' Star Wars saga. The organism has not been cultured in host-free media. It was isolated together with its *Acanthamoeba* sp. UWC36 host from the cornea of a keratitis patient (Fritsche et al., 1999). The bacteria appear as rods of 0.6 to 4 μm in length and 0.3 to 0.4 μm in diameter and show a Gram-negative type cell wall. Its life cycle consists of an alternation between elongated and coccoid rods. The bacteria occur in host-derived vacuoles in the amoeba cytoplasm.

The organism is a member of the *Alphaproteobacteria*; its classification is based on its 16S rRNA sequence (Genbank acc. AF069962) and fluorescence in situ hybridization with the 16S rRNA-targeted oligonucleotide probe AcRic90 (5'-TGCCACTAGCAGAACTCC-3', Fritsche et al., 1999). A draft genome sequence is available (accession NZ_JSWE00000000). The organism in its amoeba host is deposited at the ATCC (ATCC PRA-6).

Fritsche TR, Horn M, Seyedirashti S, Gautom RK, Schleifer KH, Wagner M. 1999. In situ detection of novel bacterial endosymbionts of Acanthamoeba spp. phylogenetically related to members of the order Rickettsiales. Appl. Environ. Microbiol. **65**:206–12.
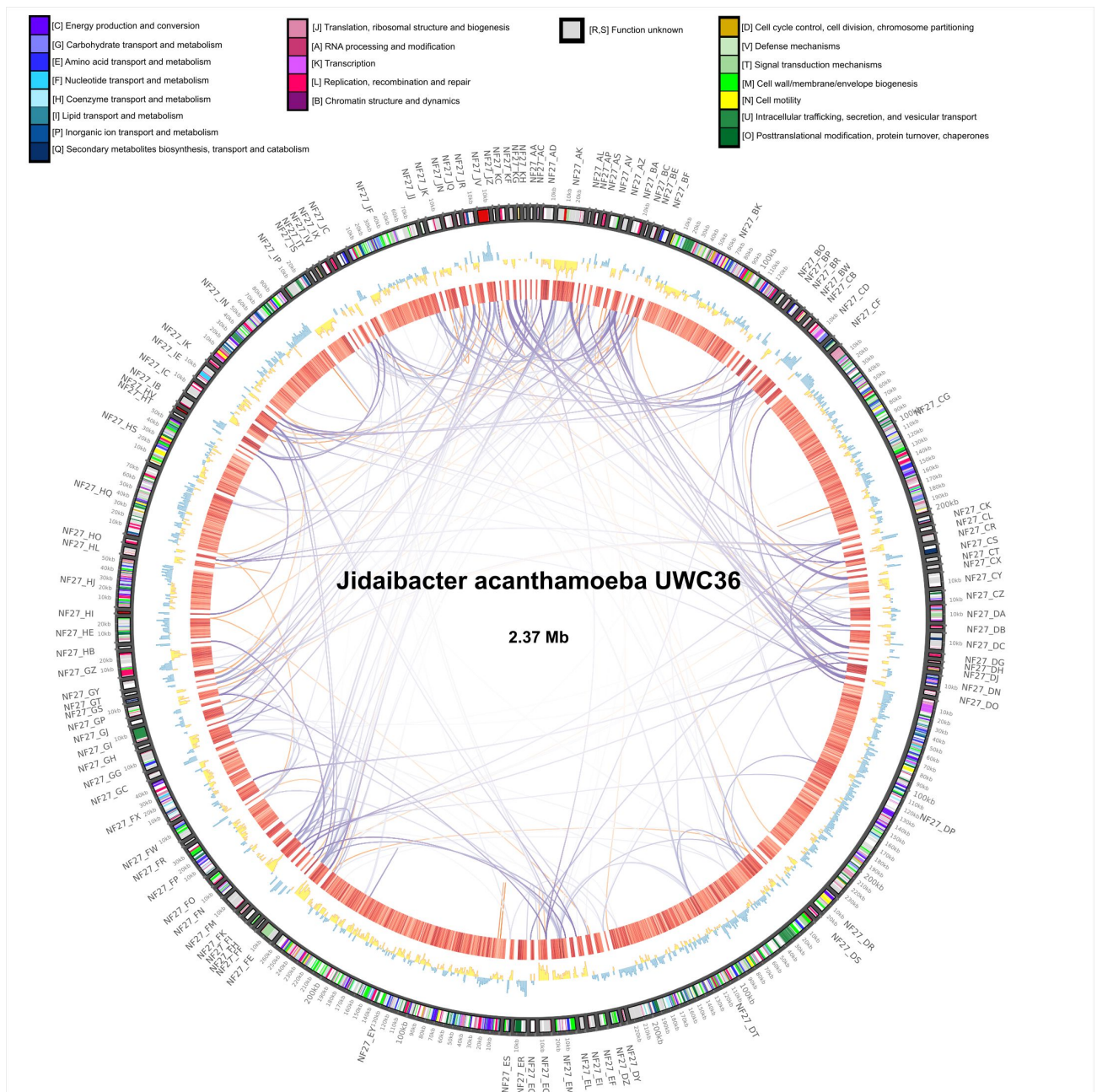
Supplementary Figure S1. Circular plot of the *Jidaibacter acanthamoeba* genome. The outermost circle depicts the 108 largest contigs (>2 kb) out of in total 249 contigs (>200 nt) of the *Jidaibacter* draft genome; genes were assigned to eggNOG functional categories and are differentially colored. The GC skew of the genome is shown as a histogram (yellow/blue), and the AT content as a heatmap with dark red indicating highest AT content. Structural repeats with a length of at least 100 nt and a DNA sequence similarity of higher than 98% are connected with each other (orange lines; the higher the similarity the darker the connecting line). In addition, translated coding regions with an amino acid similarity of more than 70% are connected (purple lines; the higher the similarity the darker the connecting line).

Supplementary Figure S2. Influence of compositional bias, horizontal gene transfer and long-branch attraction on the phylogenetic placement of *Jidaibacter*. (a) Phylogenetic tree inferred by Maximum-likelihood with the LG model based on a concatenated alignment of 173 single-copy genes. Bootstrap support values lower than 99 are shown. (b) Ranking of the 173 single-copy genes by their discordance score; the cutoff for the selection of low biased genes is shown (dashed line).
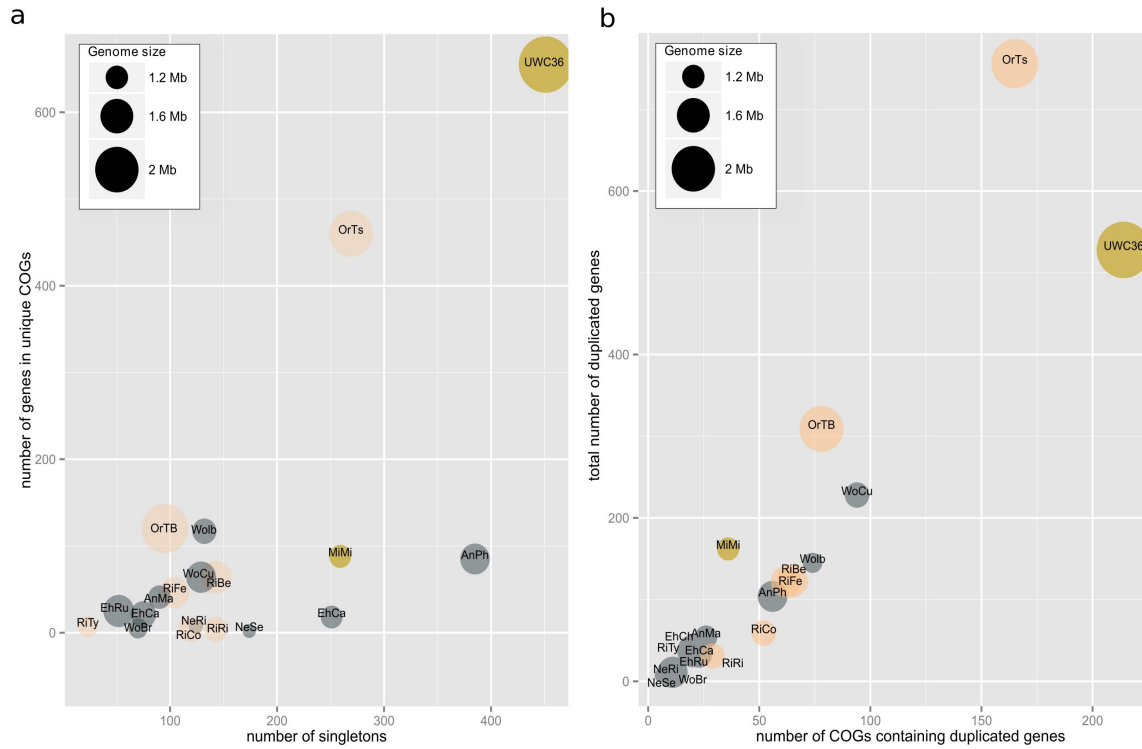
Supplementary Figure S3. The genome of *Jidaibacter acanthamoeba* encodes for many unique features and the largest number of COGs with duplicated genes. (a) The uniqueness of *Rickettsiales* genomes is illustrated by the number of genes exclusively present in the respective genome (singletons) plotted against the total number of genes present in COGs which are not shared with other members of the order (unique COGs). (b) The total number of COGs containing duplicated genes is shown for members of the *Rickettsiales* in correlation with the total number of duplicated genes. The size of the circles correspond with the genome size of the respective organism. UWC36, *Jidaibacter acanthamoeba*; OrTs, *Orientia tsutsugamushi* str. Ikeda; AnPh, *Anaplasma phagocytophilum* HZ; MiMi, *Midichloria mitochondrii I*ricVA; EhCh, *Ehrlichia chaffeensis* str. Arkansas; Wolb, *Wolbachia* endosymbiont of *Drosophila melanogaster*; OrTB, *Orientia tsutsugamushi* str. Boryong; RiBe, *Rickettsia bellii* RML369-C; WoCu, *Wolbachia* endosymbiont of *Culex quinquefasciatus* Pel; NeSe, *Neorickettsia sennetsu* str. Miyayama; RiFe, *Rickettsia felis* URRWXCal2; RiRi, *Rickettsia rickettsii* str. Sheila Smith; NeRi, *Neorickettsia risticii* str. Illinois; AnMa, *Anaplasma marginale* str. Florida; RiCo, *Rickettsia conorii* str. Malish 7; EhCa, *Ehrlichia canis* str. Jake; EhRu, *Ehrlichia ruminantium* str. Welgevonden; WoBr, *Wolbachia* endosymbiont str. TRS of *Brugia malayi*; RiTy, *Rickettsia typhi* str. Wilmington.

Supplementary Figure S4. Genetic repertoire of *Jidaibacter*. Genes have been grouped in COGs and assigned to functional categories using eggNOG. Numbers of *Jidaibacter* genes per functional category are shown as light green bars. Numbers of COGs shared with *M. mitochondrii* (dark green bars) and all *Rickettsiales* (purple bars), respectively are indicated.

	Jidaibacter acanthamoeba	Midichloria mitochondrii	Anaplasma marginale	Anaplasma phagocytophilum	Ehrlichia canis	Ehrlichia chaffeensis	Ehrlichia ruminantium	Wolbachia of Brugia malayi	Wolbachia of Culex quinquefasciatus	Wolbachia of Drosophila melanogaster	Neorickettsia risticii	Neorickettsia sennetsu	Rickettsia bellii	Rickettsia conorii	Rickettsia felis	Rickettsia rickettsii	Rickettsia typhi	Orientia tsutsugamush Boryong	Orientia tsutsugamushi Ikeda	Odyssella thessaloniciensis
Ank (CL0465)	132	7	3	4	4	4	4	6	56	22	3	3	23	3	16	4		33	45	10
TPR (CL0020)	54	6	3	3	3	3	3	5	5	6	2	2	16	10	15	10	4	14	22	56
LRR (CL0022)	10								2				8	1	2					12

Legend: > 5%, > 4%, > 2%, > 1%, > 0.5%, > 0%
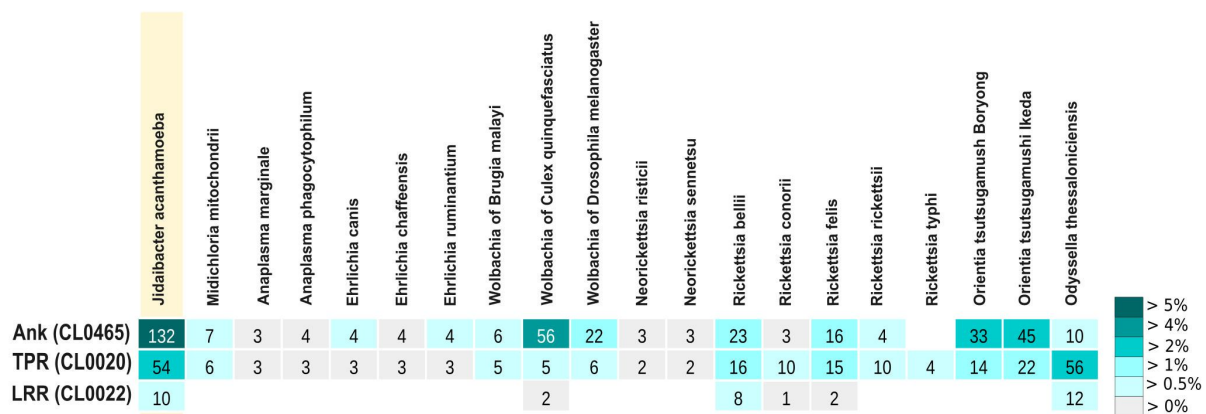
Supplementary Figure S5. The *Jidaibacter* genome encodes a high number of eukaryotic-like repeat domains. The distribution of proteins among *Rickettsiales* members containing at least one eukaryotic-like repeat domain out of major repeat superfamilies (Ankyrin (Pfam clan: CL0465), TPR (Pfam clan:CL0020), LRR (Pfam clan: CL0022), Pentapeptide (Pfam clan: CL0505)) is shown. The colored background represents the fraction these proteins make up in the respective organism's proteome.
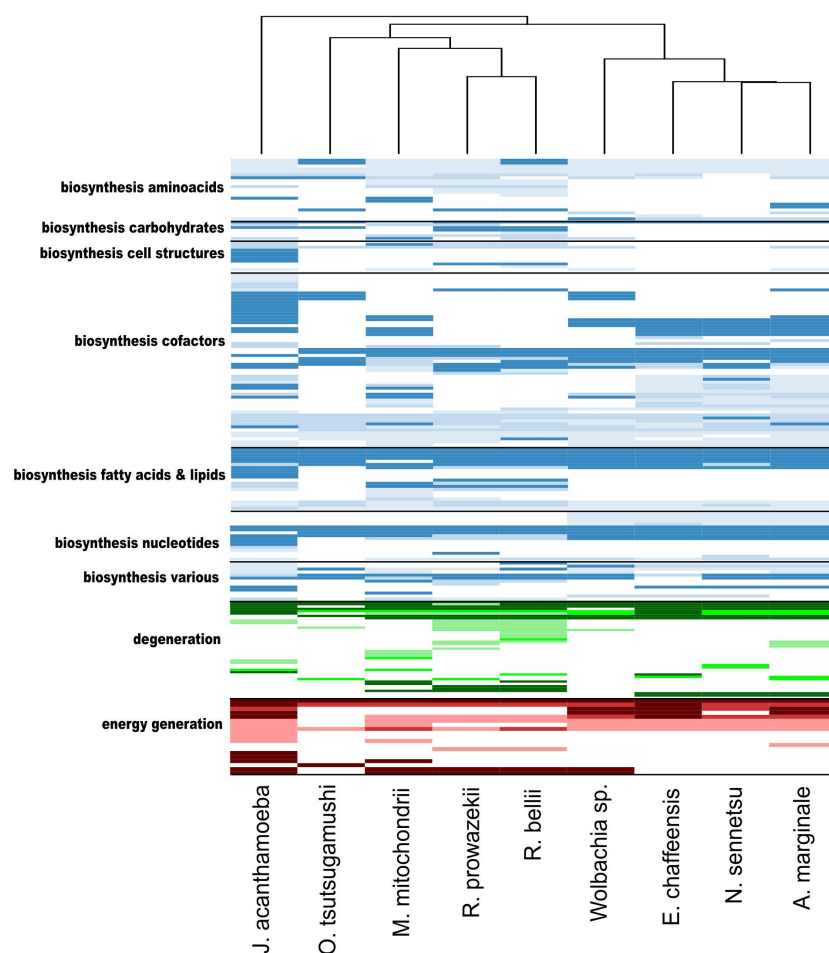
Supplementary Figure S6. Metabolic capabilities of *Jidaibacter acanthamoeba* compared to other members of the *Rickettsiales*. Major metabolic categories are depicted in terms of the completeness of pathways predicted with BioCyc (dark color, complete; middle tone, one enzyme missing; light tone, two or more enzymes missing; white, pathway absent). Hierarchical clustering indicated at the top was performed based on the presence/absence of enzymes in the respective pathways.

Supplementary Table S1. Genome sequences used in this study. Members of the *Rickettsiales* are in bold.

Organism	Accession	Genome size [bp]	Coding density [%]	GC [%]
Acetobacter pasteurianus IFO 3283-01-42C	NC_017150	2815241	91,1	53,09
Acidiphilium cryptum JF-5	NC_009484	3389227	93,1	67,99
Agrobacterium tumefaciens Ti plasmid pTiBo542	NC_002377	244978	83,3	54,62
Alpha proteobacterium HIMB59	NC_018644	1410127	98,1	32,26
Anaplasma marginale str. Florida	**NC_012026**	**1202435**	**86,9**	**49,77**
Anaplasma phagocytophilum HZ	**NC_007797**	**1471282**	**69,3**	**41,64**
Azospirillum brasilense Sp245	NC_016617	3023440	88,6	68,63
Azospirillum sp. B510	NC_013854	3311395	87,7	67,76
Bartonella australis Aust/NH1	NC_020300	1596490	80,2	41,84
Bartonella bacilliformis KC583	NC_008783	1445021	81,9	38,24
Beijerinckia indica subsp indica ATCC 9039	NC_010581	4170153	83,8	57,06
Bradyrhizobium japonicum USDA 110	NC_004463	9105828	88,4	64,06
Brevundimonas subvibrioides ATCC 15264	NC_014375	3445263	92,6	68,42
Brucella canis ATCC 23365 chromosome I sequence	NC_010103	2105969	86,2	57,21
Brucella suis 1330 chromosome I sequence	NC_004310	2107794	85,7	57,21
Campylobacter coli CVM N29710	NC_022347	1673221	91,8	31,44
Candidatus Liberibacter solanacearum CLso-ZC1	NC_014774	1258278	81,7	35,24
Candidatus Midichloria mitochondrii IricVA	**NC_015722**	**1183732**	**72,9**	**36,55**
Candidatus Pelagibacter sp IMCC9063	NC_015380	1284727	96,1	31,67
Candidatus Pelagibacter ubique HTCC1062	NC_007205	1308759	97,5	29,68
Candidatus Puniceispirillum marinum IMCC1322	CP001751	2753527	92,4	48,85
Caulobacter crescentus NA1000	NC_011916	4042929	92,8	67,17
Caulobacter segnis ATCC 21756	NC_014100	4655622	88,4	67,67
Ehrlichia canis str. Jake	**NC_007354**	**1315030**	**73,2**	**28,96**
Ehrlichia chaffeensis str. Arkansas	**NC_007799**	**1176248**	**80,8**	**30,1**
Ehrlichia ruminantium str. Welgevonden	**NC_005295**	**1516355**	**63,0**	**27,48**
Jidaibacter acanthamoeba UWC36	**JSWE00000000**	**2370652**	**88,7**	**33,85**
Geobacter sp. M21	NC_012918	4745806	87,9	60,47
Gluconacetobacter diazotrophicus PAl 5	NC_010125	3944163	91,6	66,43
Gluconacetobacter hansenii ATCC 23769	NZ_CM000920	3636659	84,1	60,5
Halothiobacillus neapolitanus c2	NC_013422	2582886	90,0	54,71
Helicobacter pylori B38	NC_012973	1576758	86,4	39,16
Legionella pneumophila subsp. pneumophila str. Philadelphia 1	NC_002942	3397754	90,0	38,27
Magnetospirillum magneticum AMB-1	NC_007626	4967148	89,7	65,09
Maricaulis maris MCS10	NC_008347	3368780	91,3	62,73
Mesorhizobium loti MAFF303099	NC_002678	7036071	87,0	62,75
Methylobacterium extorquens DM4	NC_012988	5943768	85,1	68,09
Methylobacterium nodulans ORS 2060	NC_011894	7772460	82,6	68,86
Methylovorus glucosetrophus	NC_012969	2995511	92,6	54,86
Methylovorus glucosetrophus SIP3-4 plasmid pMsip01	NC_012970	76680	91,8	45,71
Micavibrio aeruginosavorus EPB	NC_020812	2458610	91,9	54,96
Neisseria gonorrhoeae FA 1090	NC_002946	2153922	79,7	52,69
Neorickettsia risticii str Illinois	**NC_013009**	**879977**	**87,5**	**41,27**
Neorickettsia sennetsu str Miyayama	**NC_007798**	**859006**	**89,0**	**41,08**
Nitrobacter hamburgensis X14	CP000319	4406967	81,7	61,71
Nitrosomonas eutropha C91	CP000450	2661057	87,0	48,49

Novosphingobium sp. PP1Y	NC_015580	3911486	89,0	63,71
Ochrobactrum anthropi ATCC 49188	NC_009667	2887297	87,7	56,05
Odyssella thessaloniciensis L13	NZ_AEWF01000 000	2847648	85,5	41,97
Oligotropha carboxidovorans OM5	NC_015684	3595748	88,9	62,48
Orientia tsutsugamushi str. Ikeda	**NC_010793**	**2008987**	**76,1**	**30,51**
Orientia tsutsugamushi str. Boryong	**NC_009488**	**2127051**	**49,5**	**30,53**
Parvibaculum lavamentivorans DS-1	NC_009719	3914745	91,5	62,33
Polaromonas naphthalenivorans CJ2	NC_008781	4410291	91,6	62,53
Rhizobium etli CFN 42	NC_007761	4381608	87,5	61,27
Rhizobium leguminosarum bv trifolii WSM2304	NC_011369	4537948	88,2	61,53
Rhodobacter sphaeroides ATCC 17025	NC_009428	3217726	91,2	68,48
Rhodospirillum centenum SW	NC_011420	4355543	88,5	70,46
Rhodospirillum rubrum ATCC 11170	NC_007643	4352825	89,7	65,45
Rickettsia bellii RML369-C	**NC_007940**	**1522076**	**86,5**	**31,65**
Rickettsia conorii str. Malish 7	**NC_003103**	**1268755**	**81,9**	**32,44**
Rickettsia felis URRWXCal2	**NC_007109**	**1485148**	**85,1**	**32,45**
Rickettsia prowazekii str. NMRC Madrid E	**NC_020992**	**1111520**	**77,2**	**29**
Rickettsia rickettsii str. 'Sheila Smith'	**NC_009882**	**1257710**	**77,4**	**32,47**
Rickettsia typhi str. Wilmington	**NC_006142**	**1111496**	**76,7**	**28,92**
Ruegeria pomeroyi DSS-3	NC_003911	4109442	91,2	64,22
Ruegeria sp. TM1040	NC_008044	3200938	90,4	60,41
Shigella flexneri 2a str. 2457T	NC_004741	4599354	78,6	50,91
Sinorhizobium fredii NGR234	NC_012587	3925702	88,8	63,03
Sphingomonas wittichii RW1	NC_009511	5382261	93,3	68,4
Sphingopyxis alaskensis RB2256	NC_008048	3345170	91,9	65,5
Thiomonas intermedia K12	NC_014153	3396378	90,9	63,88
Tistrella mobilis KA081020-065	CP003236	3919492	90,4	68,15
Wolbachia endosymbiont of Culex quinquefasciatus Pel	**NC_010981**	**1482455**	**83,1**	**34,2**
Wolbachia endosymbiont of Drosophila melanogaster	**NC_002978**	**1267782**	**81,4**	**35,23**
Wolbachia endosymbiont str. TRS of Brugia malayi	**NC_006833**	**1080084**	**68,1**	**34,18**
Xanthomonas citri subsp. citri Aw12879	NC_020815	5321499	87,6	64,71
Xylella fastidiosa 9a5c	NC_002488	2679306	84,6	52,67
Zymomonas mobilis subsp pomaceae ATCC 29192	NC_015709	1989865	87,1	44,09

Supplementary Table S2. Antibiotic treatment to cure *Acanthamoeba* sp. UWC36 from *Jidaibacter*.
A combination of rifampicin (antibiotic 1) with a second antibiotic (antibiotic 2) was applied in various concentrations and under different incubation temperatures for up to 5 months.

Antibiotic 1	Class	Concentration	Antibiotic 2	Class	Concentration	Temperature	Duration
Rifampicin	Rifamycin	up to 80 µg/ml	Ampicillin	Aminopenicillin	up to 320 µg/ml	14/20/30°C	5 months
Rifampicin	Rifamycin	up to 80 µg/ml	Doxycyclin	Tetracyclin	up to 320 µg/ml	14/20/30°C	5 months
Rifampicin	Rifamycin	up to 80 µg/ml	Erythromycin	Macrolide	up to 320 µg/ml	14/20/30°C	5 months
Rifampicin	Rifamycin	up to 80 µg/ml	Gentamycin	Aminoglycoside	up to 320 µg/ml	14/20/30°C	5 months
Rifampicin	Rifamycin	up to 80 µg/ml	Phosphomycin	Epoxide	up to 320 µg/ml	14/20/30°C	4 months
Rifampicin	Rifamycin	up to 80 µg/ml	Kanamycin	Aminoglycoside	up to 320 µg/ml	14/20°C	3 months
Rifampicin	Rifamycin	up to 80 µg/ml	Ofloxacin	Fluoroquinolone	up to 320 µg/ml	14/20°C	3 months

Supplementary Table S3. Selection of genes encoded in the Jidaibacter accessory genome which are either involved in metabolic processes or play a role in flagellar biosynthesis. Proteins were selected based on their presence in COGs which did not contain proteins of other *Rickettsiales* member. The table shows functional annotation of these proteins (* indicates manual annotation) classified in eggNOG functional categories. PFAM domains present in the respective protein are indicated. In addition, proteins possibly involved in biosynthesis of a flagellum and T4SS clusters are shown.

Carbohydrate transport and metabolism [G]

Protein	Annotation	PFAM domains
NF27_DR00140	UDP-glucose 4-epimerase	RmlD_sub_bind,adh_short,Epimerase,3Beta_HSD,Polysacc_synt_2,NAD_binding_10
NF27_EY00230	UDP-glucuronate 5'-epimerase*	RmlD_sub_bind,Epimerase,Polysacc_synt_2,NAD_binding_10,3Beta_HSD
NF27_EY01320	Capsular polysaccharide ABC transporter permease protein KpsM*	ABC2_membrane
NF27_HE00170	Capsular polysaccharide biosynthesis protein-like protein*	DUF563

Energy production and conversion [C]

Protein	Annotation	PFAM domains
NF27_BK00540	NADH dehydrogenase	Pyr_redox_2,NAD_binding_8,Pyr_redox
NF27_CG00570	Inorganic pyrophosphatase/exopolyphosphatase	DHH
NF27_DP00620	Phosphoenolpyruvate carboxylase*	PEPcase,PEPcase
NF27_FN00070	Quinone reductase*	ADH_N,ADH_zinc_N,ADH_zinc_N_2
NF27_HQ00500	Na+-transporting NADH:ubiquinone oxidoreductase, subunit NqrB*	NQR2_RnfD_RnfE

Amino acid transport and metabolism [E]

Protein	Annotation	PFAM domains
NF27_BK00630	Oligoendopeptidase F*	Peptidase_M3_N,Peptidase_M3
NF27_BR00020	Carboxypeptidase Taq*	Peptidase_M32
NF27_DN00050	Glutamate decarboxylase*	Pyridoxal_deC,Aminotran_5
NF27_DN00090	FAD dependent oxidoreductase*	Thi4,DAO,FAD_binding_2
NF27_DT00730	CDGSH iron-sulfur domain-containing protein 3, mitochondrial	zf-CDGSH,zf-CDGSH
NF27_EY01140	tRNA 5-methylaminomethyl-2-thiouridine biosynthesis MnmC*	Methyltransf_30,DAO,Pyr_redox_2,FAD_binding_2,NAD_binding_8
NF27_FW00120	dTDP-3-amino-3,6-dideoxy-alpha-D- galactopyranose transaminase	FdtA,DegT_DnrJ_EryC1,Beta_elim_lyase,Aminotran_5
NF27_HJ00560	Shikimate kinase	AAA_18,AAA_17,AAA_33,Cytidylate_kin2,SKI

Coenzyme transport and metabolism [H]

Protein	Annotation	PFAM domains
NF27_DP01060	S-adenosyl-L-homocysteine hydrolase, NAD binding protein*	AdoHcyase_NAD,2-Hacid_dh_C,IlvN
NF27_DP01520	Cobaltochelatase, CobT subunit*	CobT,CobT_C
NF27_DT01000	Aerobic cobaltochelatase subunit CobS	AAA_5,AAA_3
NF27_EY00410	Ribosomal protein S6 modification protein	Zn_protease,ATP-grasp_4,RimK,ATP-grasp_3,GSH-S_ATP,Dala_Dala_lig_C
NF27_EY01160	Nicotinamide phosphoribosyltransferase	NAPRTase
NF27_EY01620	5,10-methenyltetrahydrofolate synthetase*	5-FTHF_cyc-lig
NF27_JF00230	Pyridoxal/pyridoxine/pyridoxamine kinase	Phos_pyr_kin,PfkB

Lipid transport and metabolism [I]

Protein	Annotation	PFAM domains
NF27_DT01330	Lipase/esterase*	COesterase,Chlorophyllase2,DUF2424,Abhydrolase_3
NF27_FX00130	2-deoxy-D-gluconate 3-dehydrogenase*	adh_short,KR,F420_oxidored,TrkA_N,adh_short_C2

Inorganic ion transport and metabolism [P]

Protein	Annotation	PFAM domains
NF27_BK00700	Auxin efflux carrier*	Mem_trans
NF27_DP00910	Iron(II) transport protein A*	FeoA
NF27_FX00170	Nitrate ABC transporter ATP-binding protein	ABC_tran,AAA_assoc_C
NF27_HS00170	Cation pump-linked membrane protein FixH*	FixH
NF27_JR00030	Silver exporting P-type ATPase	YHS,E1-E2_ATPase,Hydrolase,HADv,Hydrolase_3

Cell wall/membrane/envelope biogenesis [M]

Protein	Annotation	PFAM domains
NF27_CG00530	Glycosyl transferase	Glyco_tranf_2_3,Glycos_transf_2,Glyco_tranf_2_2
NF27_DP01490	Multidrug efflux membrane permease	HlyD_2,Biotin_lipoyl_2,HlyD
NF27_DR00140	UDP-glucose 4-epimerase	RmlD_sub_bind,adh_short,Epimerase,3Beta_HSD,Polysacc_synt_2,NAD_binding_10
NF27_DR00150	Alpha-D-QuiNAc alpha-1,3-galactosyltransferase	Glyco_trans_4_2,Glycos_transf_1,Glyco_trans_1_4,Glyco_trans_1_2
NF27_EY00230	UDP-glucuronate 5'-epimerase*	RmlD_sub_bind,Epimerase,Polysacc_synt_2,NAD_binding_10,3Beta_HSD
NF27_EY01350	Capsule polysaccharide modification protein LipA	Capsule_synth,Capsule_synth
NF27_EY01360	Capsule polysaccharide biosynthesis protein WcbO*	Capsule_synth
NF27_EY01930	Putative glycosyltransferase involved in capsule biosynthesis	Glyco_tranf_2_2,Glycos_transf_2,Capsule_synth
NF27_FW00080	Family 2 glycosyl transferase*	Glyco_trans_1_4,Glycos_transf_1,Glyco_trans_1_2
NF27_FW00100	dTDP-glucose 4,6-dehydratase	RmlD_sub_bind,adh_short,Epimerase,Polysacc_synt_2,3Beta_HSD,NAD_binding_4
NF27_FW00110	Glucose-1-phosphate thymidylyltransferase	NTP_transferase,NTP_transf_3
NF27_FW00120	dTDP-3-amino-3,6-dideoxy-alpha-D- galactopyranose transaminase	FdtA,DegT_DnrJ_EryC1,Beta_elim_lyase,Aminotran_5

Flagellar genes

Protein	Annotation	PFAM domains
NF27_BK0002	Flagellum-specific ATP synthase	ATP-synt_ab
NF27_BK0061	Flagellar motor switch protein FliM*	FliM,SpoA
NF27_BK0062	Flagellar basal body-associated protein FliL	DUF4366,FliL
NF27_CG0123	Flagellar basal-body rod modification protein FlgD*	FlgD,FlgD_ig
NF27_CG0124	Flagellar hook-length control protein FliK*	Flg_hook
NF27_CG0188	Flagellar basal-body rod protein FlgB*	Flg_bb_rod
NF27_CG0189	Flagellar basal-body rod protein FlgC	
NF27_CG0190	Flagellar hook-basal body complex protein FliE*	FliE
NF27_DP0052	Flagellar biosynthetic protein FliQ*	Bac_export_3
NF27_DP0069	Flagellar M-ring protein FliF	YscJ_FliF,YscJ_FliF_C
NF27_DP0070	Flagellar motor switch protein FliG	FliG_N,FliG_M,FliG_C
NF27_DP0071	Flagellar assembly protein FliH*	FliH
NF27_DP0072	Flagellar motor switch protein FliN	SpoA
NF27_DR0005	Flagellar P-ring protein FlgI	FlgI
NF27_DR0006	Flagellar protein FlgJ	Rod-binding
NF27_DR0008	Flagellar hook-associated protein FlgK*	Flg_bbr_C
NF27_DR0009	Flagellar hook-associated protein FlgL*	Flagellin_N
NF27_DR0010	Flagellar assembly protein FliW	FliW
NF27_DT0078	Flagellar hook protein flgE*	Flg_bbr_C
NF27_EY0060	Flagellar biosynthetic protein FlhB	Bac_export_2
NF27_EY0061	Flagellar biosynthesis protein FliR*	Bac_export_1
NF27_EY0155	Flagellar assembly protein FliX	FliX
NF27_FN00080	Flagellar export protein FliJ	FliJ
NF27_HJ0046	Flagellar motor stator protein MotA	MotA_ExbB

NF27_HQ0028	Flagellar motor protein MotA	MotA_ExbB
NF27_HQ0029	Flagellar motor protein MotB	MotB_plug,OmpA
NF27_HS0015	Flagellar hook-associated protein FliD*	FliD_N,Flagellin_IN,FliD_C
NF27_HS0016	Flagellin FliC	Flagellin_N,Flagellin_C
NF27_HS0030	Flagellar biosynthesis protein FlhA	FHIPEP
NF27_IN0004	Flagellar L-ring protein FlgH	FlgH
NF27_IN0005	Flagellar basal body P-ring formation protein FlgA*	SAF_2,SAF
NF27_IN0006	Flagellar basal-body rod protein FlgG	Flg_bb_rod,Flg_bbr_C
NF27_IN0007	Flagellar basal-body rod protein FlgF*	Flg_bb_rod,Flg_bbr_C
NF27_IN0082	Flagellar assembly protein FliO*	
NF27_IN0083	Flagellar biosynthetic protein FliP	FliP

Type IV secretion system clusters

Type Iva/P trb	Annotation	PFAM domains
NF27_DT00210	Probable conjugal transfer protein TrbB	T2SE,ResIII,DEAD,FtsK_SpoIIIE
NF27_DT00220	Conjugal transfer protein TrbC*	TrbC
NF27_DT00230	Conjugal transfer protein TrbD*	VirB3
NF27_DT00240	Conjugal transfer protein TrbE	CagE_TrbE_VirB,DUF87,FtsK_SpoIIIE,AAA_10,Zot
NF27_DT00250	P-type conjugative transfer protein TrbJ	
NF27_DT00270	P-type conjugative transfer protein TrbL*	TrbL
NF27_DT00280	Conjugal transfer protein TrbF*	VirB8
NF27_DT00290	Mating pair formation protein TrbG	CagX
NF27_DT00300	Conjugation TrbI family protein	TrbI
NF27_DT00310	Relaxase type IV secretory pathway VirD2*	DUF3363
NF27_DT00320	Conjugal transfer protein TraG	T4SS-DNA_transf,TrwB_AAD_bind,TraG-D_C
NF27_DT00330	Conjugative transfer signal peptidase TraF*	Peptidase_S26
NF27_DT00340	Conjugal transfer protein TraL*	SLT
NF27_DT00350	Conjugal transfer protein TraB*	CN_hydrolase

Type Iva/F tra	Annotation	PFAM domains
NF27_IP00010	Relaxase TrwC/TraI*	TrwC
NF27_IP00020	Relaxase TrwC/TraI*	TrwC
NF27_IP00030	Conjugal transfer protein TraD*	T4SS-DNA_transf,TrwB_AAD_bind,AAA_10,DUF87,TraG-D_C
NF27_IP00050	TraG-like N-terminal domain protein*	TraG_N
NF27_IP00060	Conjugal transfer protein TraH*	TraH
NF27_IP00070	Conjugal pilus assembly protein TraF*	TraF
NF27_IP00080	Mating pair stabilization protein TraN*	TraN,TraN
NF27_IP00090	IncF plasmid conjugative transfer protein TraN*	
NF27_IP00100	Conjugal transfer pilus assembly protein TraU	TraU
NF27_IP00110	Type-F conjugative transfer system protein TraW*	
NF27_IP00120	Type-IV secretion system protein TraC*	TraC_F_IV,AAA_10
NF27_IP00130	Conjugative transfer protein TraV*	
NF27_IP00140	Conjugal transfer pilus assembly protein TraB*	TrbI
NF27_IP00150	Conjugal transfer protein traK*	TraK
NF27_IP00160	Conjugal transfer protein TraE*	TraE
NF27_IP00170	Conjugal transfer pilus assembly protein TraL*	TraL
NF27_IP00180	TrbC/VirB2 family protein*	TrbC

Type Iva/P rvh	Annotation	PFAM domains
NF27_BK00140	Type IV secretion system protein VirB6*	TrbL
NF27_BK00160	Type IV secretion system protein VirB6*	TrbL
NF27_BK00170	Type IV secretion system protein VirB6*	TrbL
NF27_BK00180	Type IV secretion system ATPase VirB4	CagE_TrbE_VirB,AAA_10
NF27_BK00190	Type IV secretion system protein VirB3	VirB3
NF27_HQ00250	Type IV secretion system protein VirB4	CagE_TrbE_VirB,AAA_10
NF27_IN00260	Putative virD4 protein	T4SS-DNA_transf,TrwB_AAD_bind,TraG-D_C
NF27_IN00270	Type IV secretion system protein VirB11	T2SE
NF27_IN00280	Type IV secretion system protein VirB10	TrbI
NF27_IN00290	Type IV secretion system protein VirB9	CagX

NF27_IN00300	Type IV secretion system protein VirB8		VirB8
NF27_CG01440	Type IV secretion system protein VirB8*		VirB8
NF27_CG01450	Type IV secretion system protein virB9*		CagX
NF27_FX00020	Type IV secretion system protein VirB2		TrbC

Supplementary Table S4. Putative effector proteins encoded in the *Jidaibacter acanthamoeba* genome and their potential role in host interaction. Proteins in which at least one eukaryotic-like repeat domain together with one or more functional domains are indicated.

Protein	Length (aa)	PFAM domains	Comment
NF27_AE00010	102	Apc3, TPRs, TrwB_AAD_bind	possible manipulation of the cyclosome/mitosis
NF27_BJ00030	120	Apc3, TPRs	"
NF27_AC00030	120	Apc3, TPRs	"
NF27_GW00010	81	Apc3, TPRs	"
NF27_GI00070	3780	ANKs, TPRs, SecA_DEAD	possible function as a N-terminal ATP-dependent helicase
NF27_HL00120	1098	DUF294, DUF294_C, ANKs, HEAT_2	putative nucleotidyltransferase
NF27_DT02000	1966	DUF294, DUF294_C, LRRs, ANKs	"
NF27_AK00010	2300	TPRs, DUF294	"
NF27_AK00020	868	TPRs, DUF294, NTP_transf_2	"
NF27_ER00030	2165	DUF294, DUF294_C, 50xTPR, ANKs	"
NF27_GY00050	809	DUF294 , TPRs	"
NF27_AV00060	463	ANKs, Ribonucleas_3_3, Ribonuclease_3	possibly involved in rDNA transcription and rRNA processing
NF27_II00010	408	ANKs, 2xRibonuclease	"
NF27_JV00010	3713	F-box, TPRs	putative role in ubiquitination and cell signalling
NF27_GP00090	712	U-box, LRRs	possible role in ubiquitin mediated proteolysis
NF27_IE00170	434	U-box, ANKs	"
NF27_AK00050	961	AAA_22, AAA_18, TPR	putative role in DNA replication, protein degradation, membrane fusion,
			signal transduction and/or regulation of gene expression
NF27_FE00020	1370	3xPentapeptide, AAA_22	"
NF27_CG01850	620	Peptidase_C97, ANKs	predicted thiol peptidase
NF27_DT02000	662	Patatin, ANKs	patatin-like phospholipase
NF27_EI00020	649	Patatin, ANKs	"
NF27_GZ00040	659	DnaJ, ANKs	possibly involved in protein folding, ATPase activity

Supplementary Table S5. The 15 largest COGs of *Jidaibacter*. PFAM domains found in the representative proteins are indicated.

COG	# of proteins	pfam domain present	type
alpha962	24	DDE_Tnp_IS240, various IS transposases	transposase
alpha1741	19	F-box, TPRs, DUF294, DUF294_C, NTP_transf_2	eukaryotic-like
alpha2555	18	HEAT_PBS, HEAT, HEAT_2, HEAT_EZ	eukaryotic-like
alpha3052	15	ANKs	eukaryotic-like
alpha2339	14	ANKs	eukaryotic-like
alpha3051	14	DDE_Tnp_IS240	transposase
alpha2920	14	ANKs	eukaryotic-like
alpha2450	14	ANKs,HEAT_PBS, HEAT, HEAT_2, HEAT_EZ, NACHT	eukaryotic-like
alpha53	10	ANKs	eukaryotic-like
alpha4153	10	ANKs	eukaryotic-like
alpha3199	10	DDE_Tnp_1	transposase
alpha452	10	Resolvase, HTHs, CENP-B_N	transposase
alpha493	9	TPRs, LRRs, ANKs, SecA_DEAD, Apc3, DUF294_C	eukaryotic-like
alpha1680	8	DDE_Tnp_1_5	transposase
alpha4149	8	ANKs	eukaryotic-like

Chapter VII

Marine amoebae with cytoplasmic and perinuclear symbionts deeply branching in the *Gammaproteobacteria*

Marine amoebae with cytoplasmic and perinuclear symbionts deeply branching in the *Gammaproteobacteria*

Frederik Schulz[1†], Tomáš Tyml[2,3,4†], Ilaria Pizzetti[5], Iva Dyková[4], Stefano Fazi[5], Martin Kostka[2,3], Matthias Horn[1*]

[1]Department of Microbiology and Ecosystem Sience, University of Vienna, Althanstraße 14, A-1090 Vienna, Austria; [2]Faculty of Science, University of South Bohemia, Branišovská 31, 370 05 České Budějovice, Czech Republic; [3]Institute of Parasitology, Biology Centre CAS, Branišovská 31, 370 05 České Budějovice, Czech Republic; [4]Department of Botany and Zoology, Faculty of Science, Masaryk University, 61137 Brno, Czech Republic; [5]Water Research Institute, National Research Council of Italy (IRSA-CNR), Via Salaria Km 29,300, 00015 Monterotondo - Roma, Italy

Running title: Bacterial symbionts in marine amoebae

Keywords: symbiosis | endosymbiont | Amoebozoa | *Vannella* | *Legionella* | *Coxiella* |

perinuclear space | intranuclear bacteria

†these authors contributed equally to this work

*Corresponding author: Department of Microbiology and Ecosystem Science, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria, Phone: +43 1 4277 76608, horn@microbial-ecology.net

Abstract

Amoebae play an important ecological role as predators in microbial communities. They also serve as niche for bacterial replication, harbor endosymbiotic bacteria and have contributed to the evolution of major human pathogens. Despite their high diversity, marine amoebae and their association with bacteria are poorly understood. Here we describe the isolation and characterization of two novel marine amoebae and their bacterial endosymbionts, tentatively named '*Candidatus* Occultobacter vannellae' and '*Candidatus* Nucleophilum amoebae'. While one amoeba isolate is related to *Vannella*, a genus common in marine habitats, the other represents a novel lineage in the Amoebozoa. The endosymbionts showed only low similarity to known bacteria (85-88% 16S rRNA sequence similarity) but together with other uncultured marine bacteria form a sister clade to the *Coxiellaceae*. Using fluorescence *in situ* hybridization and transmission electron microscopy, identity and intracellular location of both symbionts could be confirmed; one was replicating in host-derived vacuoles, whereas the other was located in the perinuclear space of its amoeba host. This study sheds light on a so far neglected group of protists and their bacterial symbionts. The new isolates represent easily maintainable model systems and pave the way for further studies on marine associations between amoebae and bacterial symbionts.

Introduction

Unicellular eukaryotes, in particular free-living amoebae, are major players in the environment. Free-living amoebae are ubiquitous in soil, fresh- and seawater, but can also be found in anthropogenic environments, such as cooling towers, water pipes and waste-water treatment plants [1,2]. Taxonomically, free-living amoebae are scattered across the eukaryotic tree of life, with the supergroup Amoebozoa containing a substantial part of known free-living amoebae, such as naked lobose amoebae (gymnamoebae) [3,4]. In total, there are more than 200 described species of gymnamoebae classified into over 50 genera [5]. A substantial proportion of this diversity is found in marine environments, and some genera represent exclusively marine lineages. Yet our current knowledge of marine amoebae is still scarce.

Free-living amoebae shape microbial communities; they control environmental food webs by preying on bacteria, algae, fungi and other protists and contribute to elemental cycles in diverse ecosystems [1]. Free-living amoebae typically take up their food by phagocytosis. However, some bacteria have developed strategies to survive digestive processes and eventually use amoebae as niche for intracellular replication [6–9]. Free-living amoebae are thus considered to have served as evolutionary training ground for intracellular microbes. Bacteria such as *Legionella pneumophila*, *Francisella tularensis*, or *Mycobacterium* species transiently exploit these protists as a vehicle to reach out for higher eukaryotic hosts. Others engage in long-term, stable associations with free-living amoebae [7,8]. These obligate intracellular symbionts include a diverse assemblage of phylogenetically different bacterial groups [10–12]. Their analysis has facilitated unique insights into the evolution of the intracellular life style [13,14]. However, virtually nothing is known about bacterial symbionts in marine amoebae.

Here we report on the isolation and characterization of two novel marine amoeba strains harboring obligate intracellular bacterial symbionts. Both bacteria represent deeply branching novel lineages in the *Gammaproteobacteria* affiliated with the *Coxiellaceae*. While one of the symbionts thrives in the amoeba cytoplasm, the other exploits a highly unusual intracellular niche, its host's perinuclear space.

Methods

Amoeba isolation and cultivation

Lago di Paola is a meso-eutrophic lake located on the Tyrrhenian coast of Central Italy (Latium). Two narrow artificial channels at the northwestern and southeastern ends of the lake allow for a limited water exchange with the sea, sustaining a high degree of salinity throughout the year (33.7 during sampling). A surface water sample was collected on October 1, 2013, from station SAB2 [15]. The number of protist-sized particles per milliliter lake water was determined with a Neubauer counting chamber. Between one and ten protist-sized particles were placed in wells on a 96-well plate (Corning Costar, Sigma-Aldrich, Germany) containing 200 µl artificial seawater (ASW, DSMZ 607) and *E. coli tolC-* as well as ampicillin (200 ng/ml) [11]. The amoeba isolate A1, which was propagating on these plates was screened for the presence of bacterial endosymbionts with fluorescence *in situ* hybridization (FISH), and maintained in cell culture flasks (Nunclon delta-surface, Thermo Scientifc, Germany) containing ASW and *E. coli tolC-* as well as ampicillin (200 ng/ml).

Grains of wet sand collected on a sea shore (Montego Bay, Jamaica) were placed onto a MY75S agar plate and moistened daily with ASW (75 %) as described previously [16]. Two weeks later, a morphologically uniform group of cells was transferred onto a new MY75S plate, and the newly established amoeba isolate JAMX8 was sub-cultured either on plates or in liquid medium containing ASW and *E. coli tolC-* as well as ampicillin (200 ng/ml).

Trophozoites of both strains were observed in hanging drop preparations and documented using an Olympus BX51 microscope equipped with a Nomarski differential interference contrast (DIC) and an Olympus DP70 camera (Olympus Optical Co. Ltd, Japan). Trophozoites and cellular structures were analyzed using ImageJ software [17].

Transmission electron microscopy

For transmission electron microscopy (TEM), ASW in culture flasks containing isolate A1 was replaced with 3% glutaraldehyde in 0.1 M Na-cacodylate buffer. Trophozoites of isolate

JAMX8 were fixed *in situ* on MY75S plates with the same fixative. Pelleted trophozoites were rinsed in 0.1M Na-cacodylate buffer, post-fixed in 1% osmium tetroxide, dehydrated in graded acetone series, and embedded in Spurr's resin. Ultrathin sections were stained with 2% uranyl acetate in 50% methanol and Reynold's lead citrate and examined using a JEOL JEM 1010 electron microscope (Jeol Ltd, Japan) operating at 80 kV.

Fluorescence in situ hybridization

Amoeba cells were harvested by centrifugation (3000 x g, 8 min), washed with ASW and left to adhere on slides for 30 min prior to fixation with 4% formaldehyde (15 min at room temperature). The samples were hybridized for two hours at 46°C at a formamide concentration of 25% using standard hybridization and washing buffers [18] and a combination of the following probes: symbiont specific probes JAMX8_197 (5'-GAAAGGCCAAAACCCCCC-3') or A1_1033 (5'-GCACCTGTCTCTGCATGT-3'), together with EUK-516 (5′-ACCAGACTTGCCCTCC-3′,[19]) targeting most eukaryotes and the EUB338 I-III probe mix (5′-GCTGCCTCCCGTAGGAGT-3′, 5'-GCAGCCACCCGTAGGTGT-3', 5'-GCTGCCACCCGTAGGTGT-3′, [20]) targeting most bacteria. All probes were purchased from ThermoFisher Scientific (Germany). Cells were subsequently stained with DAPI (0.5 µg/ml in double distilled water, 3 min), washed once and embedded in Citifluor (Agar-Scientific, UK). Slides were examined using a confocal laser scanning microscope (SP8, Leica, Germany).

DNA extraction, PCR, cloning and sequencing

DNA was extracted from infected amoeba cultures using the DNeasy Blood and Tissue Kit (Qiagen, Austria). Amoebal 18S rRNA genes were amplified by PCR using primers 18e (5'-CTGGTTGATCCTGCCAGT-3') and RibB (5'-TGATCCTTCTGCAGGTTCACCTA-3') at an annealing temperature of 52°C [21,22]. Bacterial 16S rRNA genes were amplified using primers 616V (5′-AGAGTTTGATYMTGGCTCAG-3′) and 1492R (5′-GGYTACCTTGTTACGACTT-3′) at an annealing temperature of 52°C [23,24]. PCR reactions typically contained 100 ng template DNA, 50 pmol of each primer, 1 unit of Taq DNA polymerase (TopBio, Czech Republic for 18S rDNA; Fermentas, Germany for 16S rDNA), 10x Taq buffer with KCl and 2 µM MgCl2 and 0.2 µM of each deoxynucleotide in a

total volume of 50 µl. PCR products were purified using the PCR Purification Kit (Qiagen, Germany) and cloned using the TOPO TA Cloning Kit (Invitrogen, Germany) following the manufacturer's instructions. Nucleotide sequences were determined at Microsynth (Vienna, Austria) and Macrogen Europe (Amsterdam, Netherlands). Newly obtained rRNA gene sequences were deposited at Genbank/EMBL/DDBJ (accession numbers LC025958, LC025959, LC025974, LC025975).

Phylogenetic analysis

To infer the phylogenetic position of the isolated amoeba strains in the Amoebozoa a representative dataset of 18S rRNA sequences from a total of 57 taxa was compiled. The length of the final trimmed alignment was 1226 nt; alternative alignments obtained by altering taxon sampling and/or trimming stringency were also analyzed to check the stability of deeper nodes. A more detailed analysis of the position of strain A1 in the Vannellidae was performed, comprising in total 19 taxa, both nominal species and unnamed sequences assigned to morphologically characterized strains. The alignment was processed as described above and trimmed to a final length of 1440 nt. Both datasets were analyzed using with RAxML 8.0.20 [25], with the GTR gamma model of evolution and rapid bootstrapping (1000 replicates). Bayesian interference analysis was computed for both datasets in MrBayes 3.1.2 [26] with default options, GTR gamma model and 10^6 generations; burnin 25%.

For phylogenetic analysis of bacterial 16S rRNA sequences the sequence editor integrated in the software ARB was used to build alignments based on the current Silva ARB 16S rRNA database [27,28], which was updated with sequences from GenBank/EMBL/DDBJ obtained by sequence homology searches using BLASTn available at the NCBI web site (National Centre for Biotechnology Information; [29,30]). The alignment was trimmed to the length of the shortest sequence, manually curated and exported from ARB using a 50% conservation filter. The resulting alignment comprised 61 sequences and 1417 positions. For Bayesian analysis, PhyloBayes [31] was used with two independent chains under GTR and the CAT+GTR model. Both analyses ran until convergence was reached (maxdiff < 0.1) and as burnin 25% of the sampled trees were removed. Posterior predictive tests were performed in Phylobayes with the ppred program (sampling size 1000 trees).

Results

Two novel stenohaline amoebae containing bacterial symbionts

Two novel strains of marine amoebae, initially referred to as A1 and JAMX8, were isolated from samples taken from a costal lake in Italy and a sea shore in Jamaica, respectively. Both amoeba strains were truly stenohaline, they were successfully cultivated only in artificial seawater. In hanging drop preparations, strain A1 exhibited a flattened, oval to fan-shaped locomotive form (Fig. 1A) with an average length of 17.5 μm (S.D. 2.5, n = 50), width of 15.7 μm (S.D. 2.4, n = 50) and a length/width ratio of 0.8–1.6 (in average 1.1). The anterior hyaloplasm typically occupied about half the cell length. At the ultrastructure level, the cytoplasm contained a nucleus with a peripheral nucleolus or nucleoli, oval mitochondria and food vacuoles (Fig. 1Bi). The cell surface was covered with fine, hair-like filamented glycostyles (length of 71 ± 9 nm) (Fig. 1Bii). The mitochondria possessed branching tubular cristae (Fig. 1Biii). The partial 18S rRNA gene sequence (1889 nt) of strain A1 was most similar to *Vannella plurinucleolus* and other *Vannella* species (98% sequence similarity). In our phylogenetic analyses the placement of strain A1 in the family Vannellidae was highly supported (Figs. 2, S1) and further confirmed by a more detailed analysis focusing on the Vannellidae only, in which *V. plurinucleus* appeared as closest relative (Fig. 2 inset, S1).

The second isolate, strain JAMX8, showed flattened trophozoites with variable shape with an average length of 19.9 μm (S.D. 3.9, n = 26), width of 16.3 μm (S.D. 3.5, n = 26) and length/width ratio 0.88–1.84 (in average 1.25) (Fig. 3A). A frontal irregular hyaline zone occupied about one third of the cell length and was clearly separated from the granuloplasm containing a large quantity of spherical granules. The hyaloplasm possessed typically one to three longitudinal ridges. The locomotive cells often produced short dactylopodia (usually not more than 5 μm in length) that could freely move horizontally or vertically. No cysts were observed during subculturing. Floating forms consisted of a spherical central body with an average size of 4.8 μm in diameter (S.D. 0.7, n = 20) and thin radiating pseudopodia not longer than 10 μm (6.3 μm in average). A single, vesicular nucleus was located near the border of the granuloplasm (Figs. 3A, 3Bi, v). The cell surface was covered with a thin and amorphous cell coating (Fig. 3Bii). The cytoplasm contained numerous phagosomes (Fig. 3Bi), rounded or ovoid mitochondria (Figs. 3Bi, 3Biii) with tubular cristae (Fig. 3Biii) and a Golgi complex organized as dictyosome (Fig. 3Biv). Comparison of the partial 18S rRNA sequence (2081 nt)

of strain JAMX8 with known sequences revealed the absence of highly similar sequences in the NCBI nr/nt database. Taxa with moderate sequence similarity (<88 %) were scattered among various amoebozoan lineages. In our phylogenetic analyses the JAMX8 strain represented a deeply branching novel lineage in the Amoebozoa with no clear affiliation to described taxa (Figs. 2, S1). Electron microscopy and staining with the DNA dye DAPI readily revealed the presence of bacterial endosymbionts in both amoeba isolates (Figs. 1, 3).

Bacterial endosymbionts in the amoeba cytoplasm and perinuclear space

In addition to ingested bacteria in food vacuoles (Fig. 1Bi), amoeba isolate A1 harbored morphologically different rod-shaped bacteria with a diameter of about 0.44 μm and a maximum length of 1.2 μm (Figs. 1Bi, iv, v, vi). These bacteria were predominantly located enclosed in vacuoles (arrowheads in Figs. 1Biv, 1Bv), in which dividing cells were observed (Fig. 1Bv). The bacterial endosymbionts appeared to be few in numbers and scattered throughout the cytoplasm. However, nearly 100 % of all amoeba trophozoites were infected.

Ultrastructural analysis of amoeba isolate JAMX8 revealed bacterial symbionts at a conspicuous location within the cells (Figs. 3Bi, v, vi); rod-shaped bacteria of about 0.41 μm in diameter and a maximum length of 1.7 μm were found enclosed in the perinuclear space, between the inner and the outer nuclear membrane. Bacteria were never observed within the nucleoplasm. Nearly 100% of amoeba cells were infected. Both amoeba isolates showed no apparent signs of symbiont-induced stress or lysis; the symbiotic associations could be stably maintained non-axenically in the lab.

Novel gammaproteobacteria related to the Coxiellaceae

Sequencing of the 16S rRNA gene revealed that the endosymbiont of amoeba isolate A1 showed highest 16S rRNA sequence similarity to *Legionella longbeachae* (85%) in the NCBI RefSeq database, which contains only sequence information of well described organisms [32]. The bacterial symbiont was tentatively named "*Candidatus* Occultobacter vannellae A1" (referring to the hidden location of these bacteria inside their *Vannella* sp. host and their small cell size; hereafter: *Occultobacter*). The endosymbiont of amoeba isolate JAMX8 was most

similar to *Coxiella burnettii* (88% 16S rRNA sequence similarity), and is provisionally referred to as "*Candidatus* Nucleophilum amoebae JAMX8" (referring to the association of these bacteria with the amoeba nucleus; hereafter: *Nucleophilum*). The rRNA sequences of both endosymbionts had a similarity of 85% with each other.

For phylogenetic analysis we first calculated trees with the CAT+GTR and the GTR models in Phylobayes [31]. We then used posterior predictive tests to compare the fit of the models to the data (observed diversity: 2.866), indicating that CAT+GTR (posterior predictive diversity: 2.859 +/- 0.027, p-value: 0.57) was superior to the GTR (posterior predictive diversity: 3.138 +/- 0.027, p-value: 0). We therefore used the CAT+GTR model to assess the phylogenetic position of the two endosymbionts, demonstrating that both represent deeply branching lineages in the *Gammaproteobacteria* (Figs. 4, S2). In our analysis they grouped with several marine and freshwater clones, together forming a sister clade to the *Coxiellaceae* (Bayesian posterior probability = 0.78). *Occultobacter* and *Nucleophilum* were also moderately related to a clade comprising the two unclassified amoeba-associated bacteria CC99 and HT99 (Bayesian posterior probability = 0.98) [33]. We searched published 16S rRNA amplicon and metagenomic sequence datasets using an approach described recently (Lagkouvardos 2014), but did not find significant numbers of similar sequences to *Occultobacter* or *Nucleophilum* at a 97% similarity threshold.

In order to demonstrate the intracellular location of the bacterial symbionts FISH experiments were performed by combining symbiont-specific probes with a universal bacterial probe mix [20]. The positive hybridization reaction with both probes respectively confirmed the location of *Occultobacter* in the cytoplasm of its *Vannella* sp. A1 host (Fig. 1C) and the association of *Nucleophilum* with the nucleus of its JAMX8 amoeba host (Fig. 3C).

Discussion

Here we report on the recovery of two novel stenohalione amoeba from marine samples. Based on light microscopy, amoeba isolate A1 was readily identified as a member of the ubiquitous family Vannellidae whose members are also frequently found in marine environments [34]. Nuclear structure (laterally located nucleolus/nucleoli) and trophozoite size allow an

assignment of this isolate to *Vannella plurinucleolus* (Fig. 1). However, the shape of its cell surface is in conflict with the species diagnosis. Yet, in our phylogenetic analyses isolate A1 clustered together with *V. plurinucleolus* strain 50745 (Fig. 2). As species diagnosis/identity of this strain is under debate [35], we decided to leave the A1 strain undetermined at the species level.

By light microscopy, trophozoites of isolate JAMX8 showed a combination of morphological features typical of the genera *Mayorella* and *Korotnevella* [36]. However, neither essential diagnostic features, like a surface cuticle or surface microscales, nor any other distinct characteristics were found (Fig. 3). We were thus not able to assign isolate JAMX8 to any described gymnamoeba species or genus based on morphological criteria. Furthermore, the placement of JAMX8 within the Amoebozoa could not be unambiguously determined in our phylogenetic analyses (Fig. 2). The deeper nodes in our Amoebozoa tree are generally rather unstable, which is consistent with previous studies [37,38], and the tree topology was dependent on taxon sampling and alignment trimming stringency. However, isolate JAMX8 groups with low statistical support with *Vermistella antarctica* (Bayesian posterior probability 0.69, maximum likelihood bootstrap value 16%, and this sister taxa relationship was recovered repeatedly in different analyses. Taken together, the exact relationship of isolate JAMX8 to known amoebae remains elusive, but based on both morphological and phylogenetic analysis this isolate JAMX8 belongs to a completely new taxon within the Amoebozoa

To our knowledge this is the first molecular identification and characterization of bacterial symbionts of marine amoebae. In the last two decades numerous reports described the discovery of obligate intracellular amoeba symbionts [8,10]. However, these studies were often biased towards the isolation of *Acanthamoeba*, *Naegleria* or *Vermamoeba* (former *Hartmannella*) strains, usually from anthropogenic or freshwater habitats or clinical samples [7,39,40]. This is not surprising as these amoebae are in the focus of medical and parasitological research, and standard isolation protocols are available [2,16,41,42]. Interestingly, bacterial symbionts previously found in these amoebae were frequently very similar to each other; although isolated from geographically distant places, the endosymbionts belonged to symbiont clades either in the *Alpha-* or *Betaproteobacteria*, the *Bacteroidetes,* or the *Chlamydiae* [10,11,43,43–46]. In addition, *Gammaproteobacteria*, such as *Coxiella, Francisella, Legionella*, and

Legionella-like amoebal pathogens, may also be associated with amoebae [7,46–48], but these are mostly facultative associations. These bacteria show a parasitic life style, and many also infect higher eukaryotes [7]. This study for the first time reports on *Gammaproteobacteria* naturally living in a stable association with their amoeba hosts.

The two gammaproteobacterial symbionts *Occultobacter* and *Nucleophilum* represent novel phylogenetically deeply branching lineages, with only low 16S rRNA sequence similarity to known bacteria (85% and 88%, respectively; Fig. 4). In addition, no close relatives (>97% 16S rRNA sequence similarity) could be retrieved in any of the numerous 16S rRNA amplicon studies targeting biodiversity of marine environments, suggesting that both endosymbionts are rather rare and/or their hosts have not been captured during sampling. Our phylogenetic analysis showed that the exact position of the two endosymbionts in the *Gammaproteobacteria* is difficult to pinpoint. *Occultobacter* and *Nucleophilum* group together in a well-supported monophyletic clade with the *Coxiellaceae* and a cluster comprising the legionella-like amoebal pathogens HT99 and CC99 (Fig. 4). The topology within this clade is, however, not very robust. This is in agreement with previous observations that resolving the phylogeny of major groups of *Gammaproteobacteria* is a challenging task [49].

The closest relatives of *Occultobacter* and *Nucleophilum* are other symbionts and pathogens of eukaryotes. The *Coxiellaceae* mainly include bacteria infecting arthropods, which occasionally also invade mammalian or protozoan hosts [47,50,51]. The bacteria referred to as HT99 and CC99 were associated with amoebae found in a hot tub and a cooling tower, respectively [33]. Of note, while the *Coxiellaceae*, CC99, HT99 and related taxa mainly originate from freshwater, anthropogenic and non-marine habitats, many of the closest relatives of *Occultobacter* and *Nucleophilum* were detected in marine environments (Fig. 4).

The two symbionts described here colonize fundamentally different intracellular niches. Whereas, similar to many known intracellular bacteria, *Occultobacter* establishes replication inside host-derived vacuoles and is also occasionally found as single cell inside the cytoplasm, *Nucleophilum* is associated with its host cell's nucleus (Figs. 1, 3). The latter is a very unusual life style [52,53], but there are few reports on bacteria located in the nuclear compartment of other

amoebae, namely the chlamydial symbiont of *Naegleria* "Pn" [54,55], the two gammaproteobacteria HT99 and CC99 [33], and the alphaproteobacterium *Nucleicultrix amoebiphila* [56]. Bacteria capable to invade the nucleus possibly benefit from a nutrient-rich and safe environment for replication [53]. However, the pattern of how bacteria settle in this compartment shows striking differences; while *Nucleicultrix* is spread out in the nucleoplasm, Pn is associated with the nucleolus, and *Nucleophilum* is located in the perinuclear space [54–56]. Embedded in between the inner and outer nuclear membrane, the bacteria thus do not have direct access to the nucleoplasm. The perinuclear space, which is continuous with the endoplasmic reticulum, serves as a calcium storage [57] and has regulatory impact on processes in the nucleus, such as gene expression [58]. The exact physicochemical conditions of this compartment remains currently unknown [57] , however, in contrast to the nucleus or the cytoplasm it likely contains less substrates to support bacterial growth. We therefore expect *Nucleophilum* to have evolved unconventional strategies to target its peculiar perinuclear niche and to satisfy its nutritional needs. Future analysis of the symbiont's genome in combination with functional approaches, such as transcriptomics and proteomics, will help to clarify this enigma and to obtain insights into the infection process, interaction mechanisms, and the evolution of this unique lifestyle.

This is the first report on the concomitant isolation and characterization of marine amoebae and their bacterial endosymbionts. The low degree of relationship of the symbionts to known bacteria and the discovery of a symbiont thriving in the host perinuclear space, a niche not reported previously for intracellular microbes, indicates that marine habitats represent a rich pool of hidden symbiotic associations.

Author contributions statement

I.P., S.F., T.T. performed the sampling; F.S., T.T., I.D., M.K. and M.H. conceived the experiments; F.S., T.T., I.P. and M.K. conducted the experiments; F.S., T.T. and M.K. analyzed the results. All authors discussed and commented the results. F.S. and T.T. prepared figures; F.S., T.T. and M.H. wrote the main manuscript text. All authors reviewed the manuscript.

Competing financial interests

The author(s) declare no competing financial interests.

References

1. Rodríguez-Zaragoza, S. Ecology of free-living amoebae. *Crit. Rev. Microbiol.* **20,** 225–41 (1994).
2. Khan, N. A. *Acanthamoeba: Biology and Pathogenesis.* (Horizon Scientific Press, 2009).
3. Adl, S. M. *et al.* The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* **59,** 429–514 (2012).
4. Cavalier-Smith, T. *et al.* Multigene phylogeny resolves deep branching of Amoebozoa. *Mol. Phylogenet. Evol.* **83,** 293–304 (2015).
5. Smirnov, A. V. in *Encyclopedia of. Microbiology* 558–577 (Elsevier, 2008).
6. Barker, J. & Brown, M. Trojan horses of the microbial world: protozoa and the survival of bacterial pathogens in the environment. *Microbiology* **140,** 1253–1259 (1994).
7. Greub, G. & Raoult, D. Microorganisms resistant to free-living amoebae. *Clin. Microbiol. Rev.* **17,** 413–433 (2004).
8. Horn, M. & Wagner, M. Bacterial endosymbionts of free-living amoebae. *J. Eukaryot. Microbiol.* **51,** 509–14 (2004).
9. Molmeret, M., Horn, M., Wagner, M., Santic, M. & Abu Kwaik, Y. Amoebae as training grounds for intracellular bacterial pathogens. *Appl. Environ. Microbiol.* **71,** 20–28 (2005).
10. Schmitz-Esser, S. *et al.* Diversity of bacterial endosymbionts of environmental acanthamoeba isolates. *Appl. Environ. Microbiol.* **74,** 5822–31 (2008).

11. Lagkouvardos, I., Shen, J. & Horn, M. Improved axenization method reveals complexity of symbiotic associations between bacteria and acanthamoebae. *Environ. Microbiol. Rep.* **6,** 383–388 (2014).

12. Pagnier, I. *et al.* Babela massiliensis, a representative of a widespread bacterial phylum with unusual adaptations to parasitism in amoebae. *Biol. Direct* **10,** 1–17 (2015).

13. Horn, M. Chlamydiae as symbionts in eukaryotes. *Annu. Rev. Microbiol.* **62,** 113–31 (2008).

14. Tosetti, N., Croxatto, A. & Greub, G. Amoebae as a tool to isolate new bacterial species, to discover new virulence factors and to study the host–pathogen interactions. *Microb. Pathog.* **77,** 125–130 (2014).

15. Pizzetti, I., Fazi, S., Fuchs, B. M. & Amann, R. High abundance of novel environmental chlamydiae in a Tyrrhenian coastal lake (Lago di Paola, Italy). *Environ. Microbiol. Rep.* **4,** 446–452 (2012).

16. Dyková, I. & Kostka, M. *Illustrated guide to culture collection of free-living amoebae.* (Academia, 2013). at <http://www.muni.cz/research/publications/1091345>

17. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).

18. Daims, H., Stoecker, K. & Wagner, M. Fluorescence in situ hybridization for the detection of prokaryotes. *Mol. Microb. Ecol.* **213,** 239–239 (2005).

19. Amann, R. & Binder, B. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* **56,** 1919–25 (1990).

20. Daims, H., Brühl, A., Amann, R., Schleifer, K. H. & Wagner, M. The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Syst. Appl. Microbiol.* **22,** 434–44 (1999).

21. Medlin, L., Elwood, H. J., Stickel, S. & Sogin, M. L. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* **71,** 491–499 (1988).

22. Hillis, D. M. & Dixon, M. T. Ribosomal DNA: molecular evolution and phylogenetic inference. *Q. Rev. Biol.* **66,** 411–453 (1991).

23. Juretschko, S. *et al.* Combined molecular and conventional analyses of nitrifying bacterium diversity in activated sludge: Nitrosococcus mobilis and Nitrospira-like bacteria as dominant populations. *Appl. Environ. Microbiolgy* **64,** 3042–3051 (1998).

24. Loy, A., Lehner, A. & Lee, N. Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Appl. Environ. Microbiol.* **68,** 5064–5081 (2002).

25. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–1313 (2014).

26. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61,** 539–42 (2012).

27. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32,** 1363–1371 (2004).

28. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41,** D590–6 (2013).

29. Altschul, S., Gish, W. & Miller, W. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).

30. Wheeler, D. L. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **36,** D13–21 (2008).

31. Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinforma. Oxf. Engl.* **25,** 2286–2288 (2009).

32. Pruitt, K. D., Tatusova, T., Brown, G. R. & Maglott, D. R. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* **40,** D130–D135 (2012).

33. Farone, A. L., Berk, S. G., Farone, M. B. & Gunderson, J. H. *The isolation and characterization of naturally-occurring amoeba-resistant bacteria from water samples.* (2010). at <http://cfpub.epa.gov/ncer_abstracts/index.cfm/fuseaction/display.highlight/abstract/8114/report/F>

34. Page, F. C. Fine-structure of some marine strains of Platyamoeba (Gymnamoebia, Thecamoebidae). *Protistologica* **16,** 605–612 (1980).

35. Smirnov, A. V., Nassonova, E. S., Chao, E. & Cavalier-Smith, T. Phylogeny, evolution, and taxonomy of vannellid amoebae. *Protist* **158,** 295–324 (2007).

36. Patsyuk, M. New Gymnamoebae Species (Gymnamoebia) in the Fauna of Ukraine. *Vestn. Zool.* **46,** e–7–e–13 (2012).

37. Tekle, Y. I. *et al.* Phylogenetic placement of diverse amoebae inferred from multigene analyses and assessment of clade stability within 'Amoebozoa' upon removal of varying rate classes of SSU-rDNA. *Mol. Phylogenet. Evol.* **47,** 339–352 (2008).

38. Lahr, D. J. G., Grant, J., Nguyen, T., Lin, J. H. & Katz, L. A. Comprehensive phylogenetic reconstruction of amoebozoa based on concatenated analyses of SSU-rDNA and actin genes. *PloS One* **6,** e22780 (2011).

39. Corsaro, D., Pages, G. S., Catalan, V., Loret, J.-F. & Greub, G. Biodiversity of amoebae and amoeba-associated bacteria in water treatment plants. *Int. J. Hyg. Environ. Health* **213,** 158–166 (2010).

40. Delafont, V., Brouke, A., Bouchon, D., Moulin, L. & Héchard, Y. Microbiome of free-living amoebae isolated from drinking water. *Water Res.* **47,** 6958–6965 (2013).

41. Walochnik, J. *et al.* Discrimination between Clinically Relevant and NonrelevantAcanthamoeba Strains Isolated from Contact Lens- Wearing Keratitis Patients in Austria. *J. Clin. Microbiol.* **38,** 3932–3936 (2000).

42. Maghsood, A. H. *et al.* Acanthamoeba genotype T4 from the UK and Iran and isolation of the T2 genotype from clinical isolates. *J. Med. Microbiol.* **54,** 755–759 (2005).

43. Fritsche, T. R. *et al.* In situ detection of novel bacterial endosymbionts of Acanthamoeba spp. phylogenetically related to members of the order Rickettsiales. *Appl. Environ. Microbiol.* **65,** 206–12 (1999).

44. Horn, M., Fritsche, T. R., Gautom, R. K., Schleifer, K.-H. & Wagner, M. Novel bacterial endosymbionts of Acanthamoeba spp. related to the Paramecium caudatum symbiont Caedibacter caryophilus. *Environ. Microbiol.* **1,** 357–367 (1999).

45. Matsuo, J. *et al.* Survival and transfer ability of phylogenetically diverse bacterial endosymbionts in environmental Acanthamoeba isolates. *Environ. Microbiol. Rep.* **2,** 524–533 (2009).

46. Lamoth, F. & Greub, G. Amoebal pathogens as emerging causal agents of pneumonia. *FEMS Microbiol. Rev.* **34,** 260–280 (2010).

47. La Scola, B. & Raoult, D. Survival of Coxiella burnetii within free-living amoeba Acanthamoeba castellanii. *Clin. Microbiol. Infect.* **7,** 75–79 (2001).

48. La Scola, B. *et al.* Legionella drancourtii sp. nov., a strictly intracellular amoebal pathogen. *Int. J. Syst. Evol. Microbiol.* **54,** 699–703 (2004).

49. Williams, K. P. *et al.* Phylogeny of gammaproteobacteria. *J. Bacteriol.* **192,** 2305–14 (2010).

50. Cordaux, R. *et al.* Molecular Characterization and Evolution of Arthropod-Pathogenic Rickettsiella Bacteria. *Appl. Environ. Microbiol.* **73,** 5045–5047 (2007).

51. Smith, T. A., Driscoll, T., Gillespie, J. J. & Raghavan, R. A Coxiella-like Endosymbiont is a potential vitamin source for the Lone Star Tick. *Genome Biol. Evol.* (2015). doi:10.1093/gbe/evv016

52. Fujishima, M. & Kodama, Y. Endosymbionts in Paramecium. *Eur. J. Protistol.* **48,** 124–37 (2012).

53. Schulz, F. & Horn, M. Intranuclear bacteria: inside the cellular control center of eukaryotes. *Trends Cell Biol.* **0,** (2015).

54. Michel, R., Hauroder, B., Muller, K. & Zoller, L. An environmental Naegleria-strain, unable to form cysts-turned out to harbour two different species of endocytobionts. *Endocytobiosis Cell Res.* **118,** 115–118 (1999).

55. Walochnik, J., Muller, K., Aspock, H. & Michel, R. An endocytobiont harbouring Naegleria strain identified as N. clarki De Jonckheere, 1994. *Acta Protozool.* **44,** 301–310 (2005).

56. Schulz, F. *et al.* Life in an unusual intracellular niche: a bacterial symbiont infecting the nucleus of amoebae. *ISME J.* **8,** 1634–1644 (2014).

57. Matzke, A. J. M., Weiger, T. M. & Matzke, M. Ion channels at the nucleus: electrophysiology meets the genome. *Mol. Plant* **3,** 642–652 (2010).

58. Kim, M. C., Chung, W. S., Yun, D.-J. & Cho, M. J. Calcium and calmodulin-mediated regulation of gene expression in plants. *Mol. Plant* **2,** 13–21 (2009).
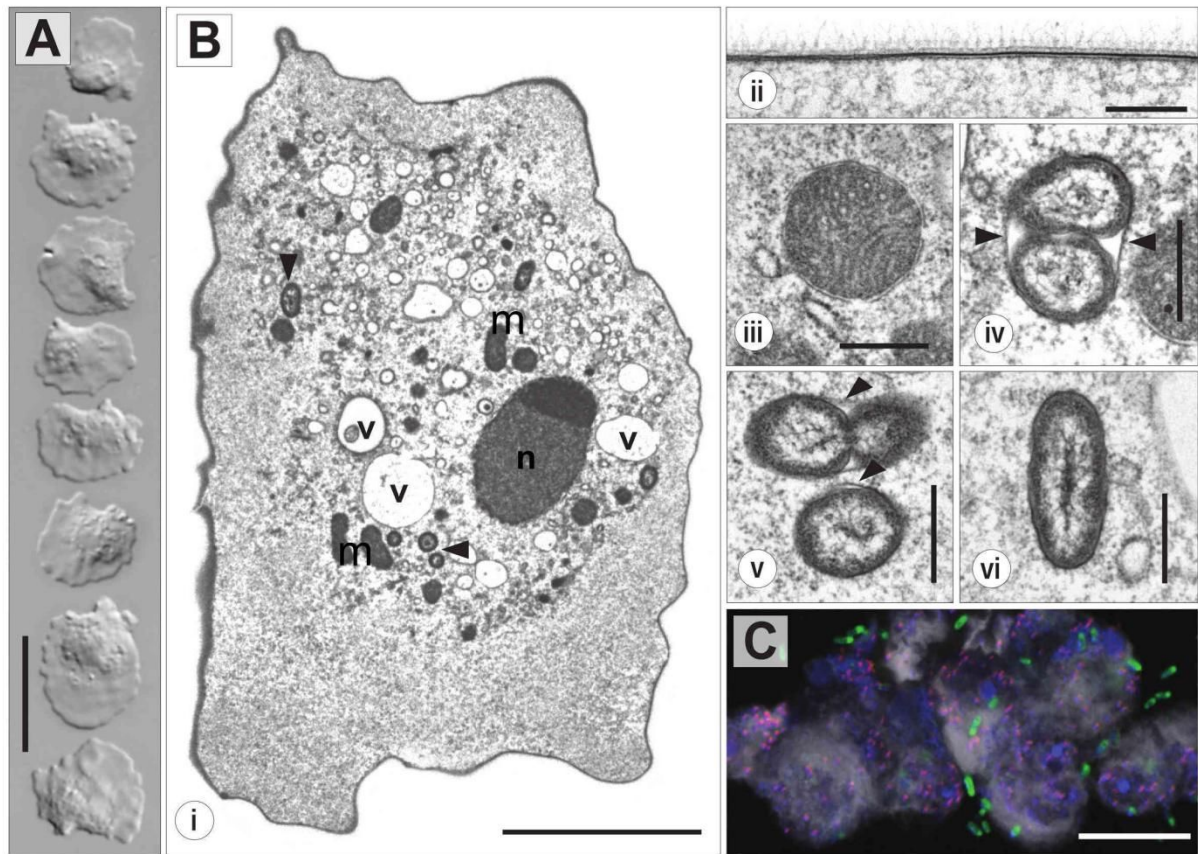
Figure 1. *Vannella* **sp. A1 and its bacterial endosymbiont '*Candidatus* Occultobacter vannellae'.**
(A) Trophozoites as seen in hanging drop preparations (scale bar = 20 μm). (B) Fine structure of
Vannella sp. A1 and its bacterial symbiont. (i) Section of an amoeba trophozoite: cell organelles located
within granuloplasm; nucleus (n) with laterally located nucleolus, mitochondria (m), vacuoles (v),
bacterial endosymbionts (arrowheads) (scale bar = 5 μm). (ii) Cell surface of trophozoite with
amorphous glycocalyx (scale bar = 200 nm). (iii) Mitochondria with tubular cristae (scale bar = 500 nm).
(iv-vi) Bacterial endosymbionts in detail: (iv, v) host-derived vacuolar membranes (arrowheads)
enclosing endosymbionts undergoing cell division, (vi) longitudinal section through an endosymbiont
(scale bar = 500 nm). (C) Fluorescence *in situ* hybridization image showing the intracytoplasmic
location of '*Candidatus* Occultobacter vannellae' (*Occultobacter*-specific probe A1_1033, pink) in its
Vannella sp. A1 host (probe EUK516, grey) with DAPI stained nuclei (blue) and food bacteria (general
bacterial probe EUB338-mix, green) ; scale bar indicates 10 μm.
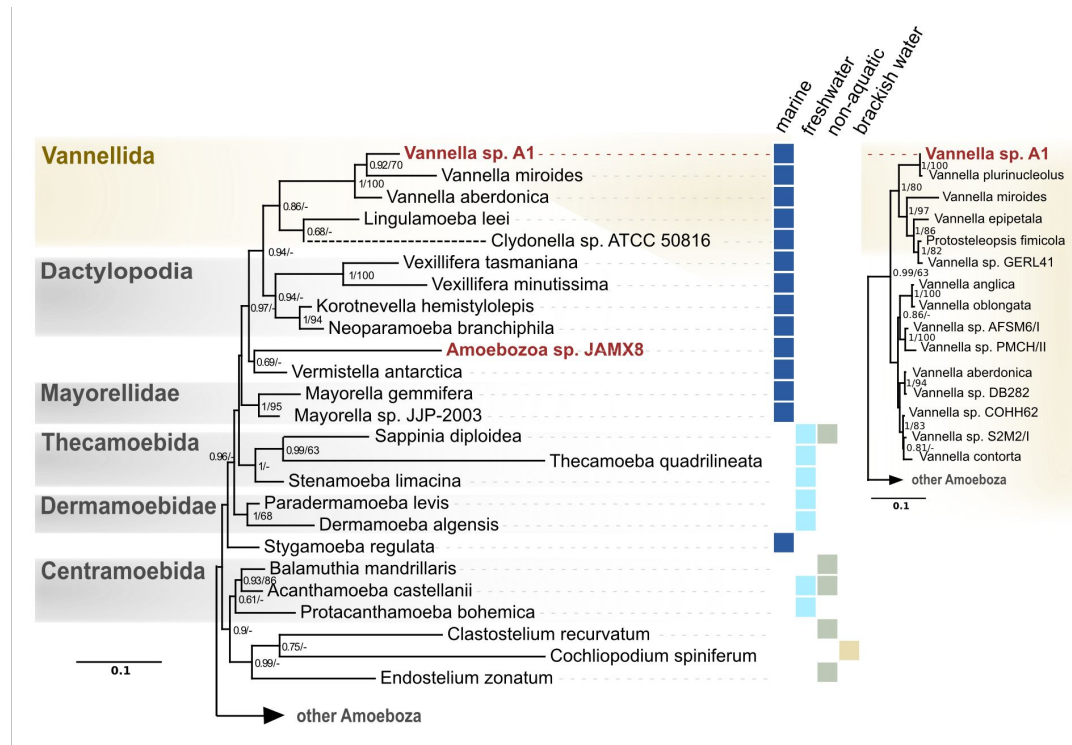
Figure 2. Phylogenetic relationships of *Vannella* sp. A1 and amoeba isolate JAMX8 within the Amoebozoa. Phylogenetic 18S rRNA-based trees of the Amoebozoa (left panel) and Vannellidae (right panel) constructed using the Bayesian inference method. Bayesian posterior probabilities (> 0.6) and RaxML bootstrap support values (> 60%) are indicated at the nodes; the dashed line indicates a branch shortened by 50% to enhance clarity. Colored squares indicate the typical habitat of the respective amoeba species (left panel). A detailed version of the trees including accession numbers is available as supplementary Figure S1.
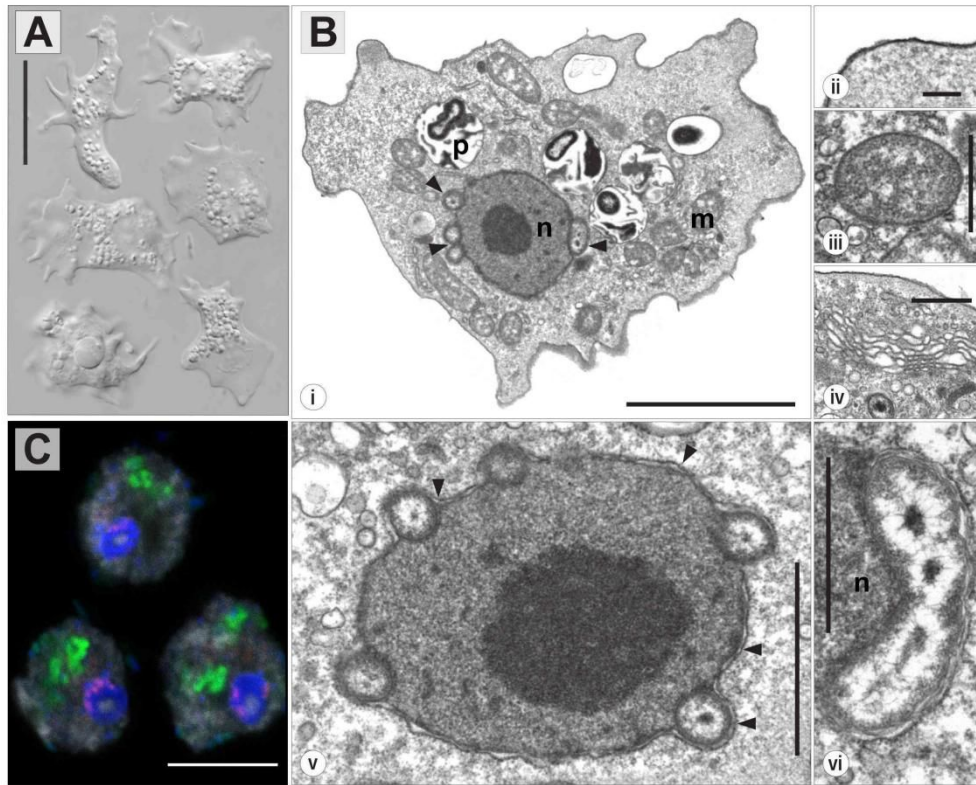
Figure 3. Amoeba isolate JAMX8 and its bacterial endosymbiont 'Candidatus Nucleophilum amoebae'. (A) Trophozoites as seen in hanging drop preparations (scale bar = 20 µm). (B) Fine structure of JAMX8 and its bacterial endosymbiont inhabiting the perinuclear space. (i) Section of an amoeba trophozoite: vesicular nucleus (n), phagosomes (p), mitochondria (m), bacterial endosymbionts associated with the nuclear envelope (arrowheads) (scale bar = 5 µm). (ii) Amorphous and tenuous cell coat (scale bar = 200 nm). (iii) Mitochondria with tubular cristae (scale bar = 1 µm). (iv) The Golgi complex organized as dictyosome (scale bar = 1 µm). (v,vi) Bacterial endosymbionts located within the perinuclear space, between inner and outer nuclear membrane. (v) Nucleus in detail with numerous endosymbiotic bacteria in transverse section (scale bar = 2 µm). Arrowheads indicate outer nuclear membrane. (vi) Longitudinal section through a rod-shaped bacterial endosymbiont, nucleus (n) (scale bar = 1 µm). (C) Fluorescence *in situ* hybridization image showing the co-localization of 'Candidatus Nucleophilum amoebae' (*Nucleophilum*-specific probe JAMX8_197, pink) with its host nucleus (DAPI, blue); food bacteria (general bacterial probe EUB338-mix, green) enclosed in the amoeba cytoplasm (probe EUK516, grey); scale bar indicates 10 µm.
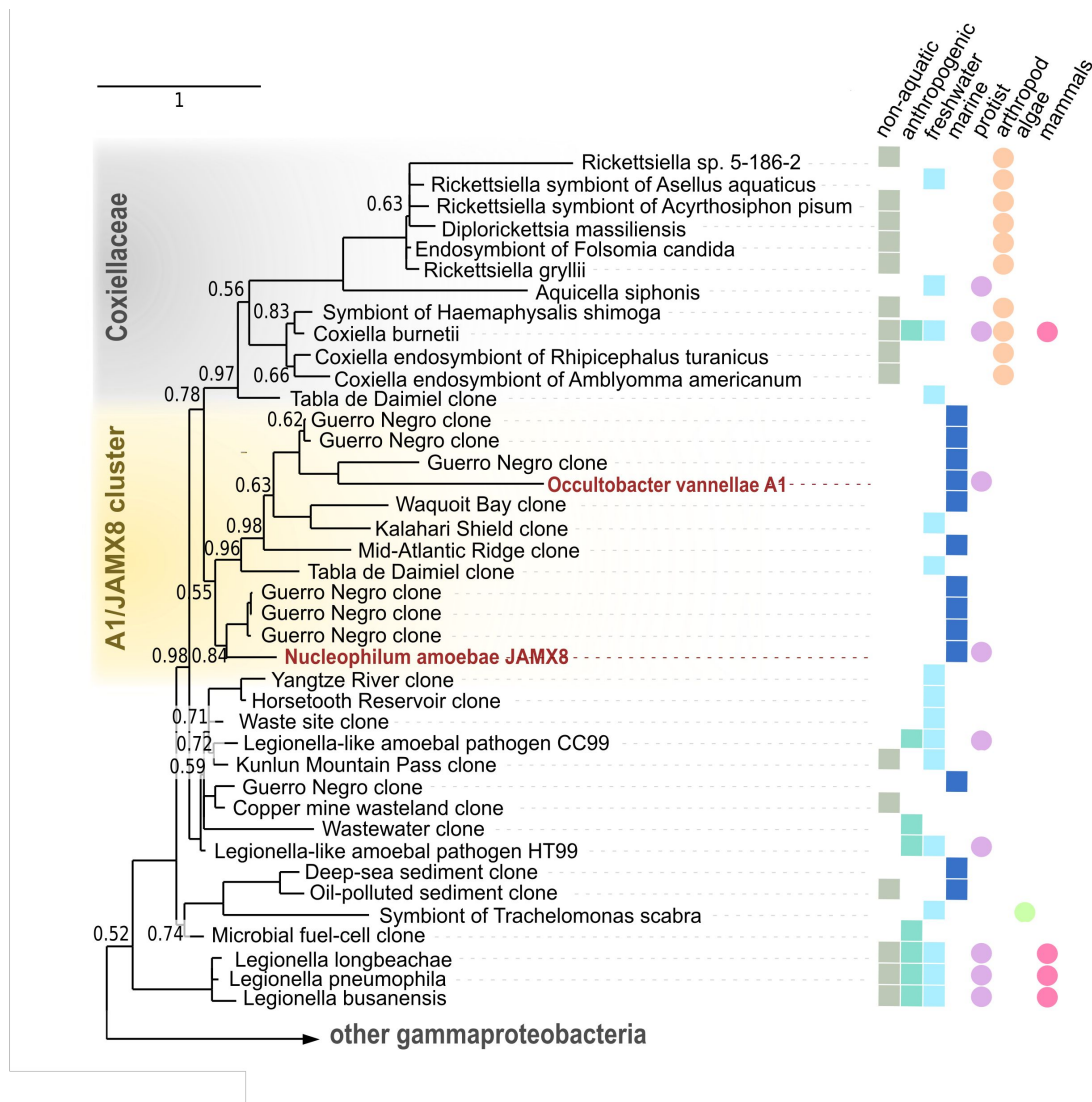
Figure 4. Phylogenetic relationship of 'Candidatus Occultobacter vannellae' and 'Candidatus Nucleophilum amoebae' with the Gammaproteobacteria. The phylogenetic tree (Phylobayes, CAT+GTR) is based on the 16S rRNA sequences, Bayesian posterior probabilities are indicated at the nodes (only values < 0.99 are shown). Colored squares indicate the environmental origin of the respective sequence; colored circles indicate host association. A detailed version of this tree including accession numbers is available as supplementary Figure S2.

Marine amoebae with cytoplasmic and perinuclear symbionts deeply branching in the *Gammaproteobacteria*

Frederik Schulz[1†], Tomáš Tyml[2,3,4†], Ilaria Pizzetti[5], Iva Dyková[4], Stefano Fazi[5], Martin Kostka[2,3], Matthias Horn[1*]

[1]Department of Microbiology and Ecosystem Sience, University of Vienna, Althanstraße 14, A-1090 Vienna, Austria

[2]Faculty of Science, University of South Bohemia, Branišovská 31, 370 05 České Budějovice, Czech Republic

[3]Institute of Parasitology, Biology Centre CAS, Branišovská 31, 370 05 České Budějovice, Czech Republic

[4]Department of Botany and Zoology, Faculty of Science, Masaryk University, 61137 Brno, Czech Republic

[5]Water Research Institute, National Research Council of Italy (IRSA-CNR), Via Salaria Km 29,300–, 00015 Monterotondo - Roma, Italy

†these authors contributed equally to this work

*Corresponding author: Department of Microbiology and Ecosystem Science, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria, Phone: +43 1 4277 76608, horn@microbial-ecology.net

Supplementary Material
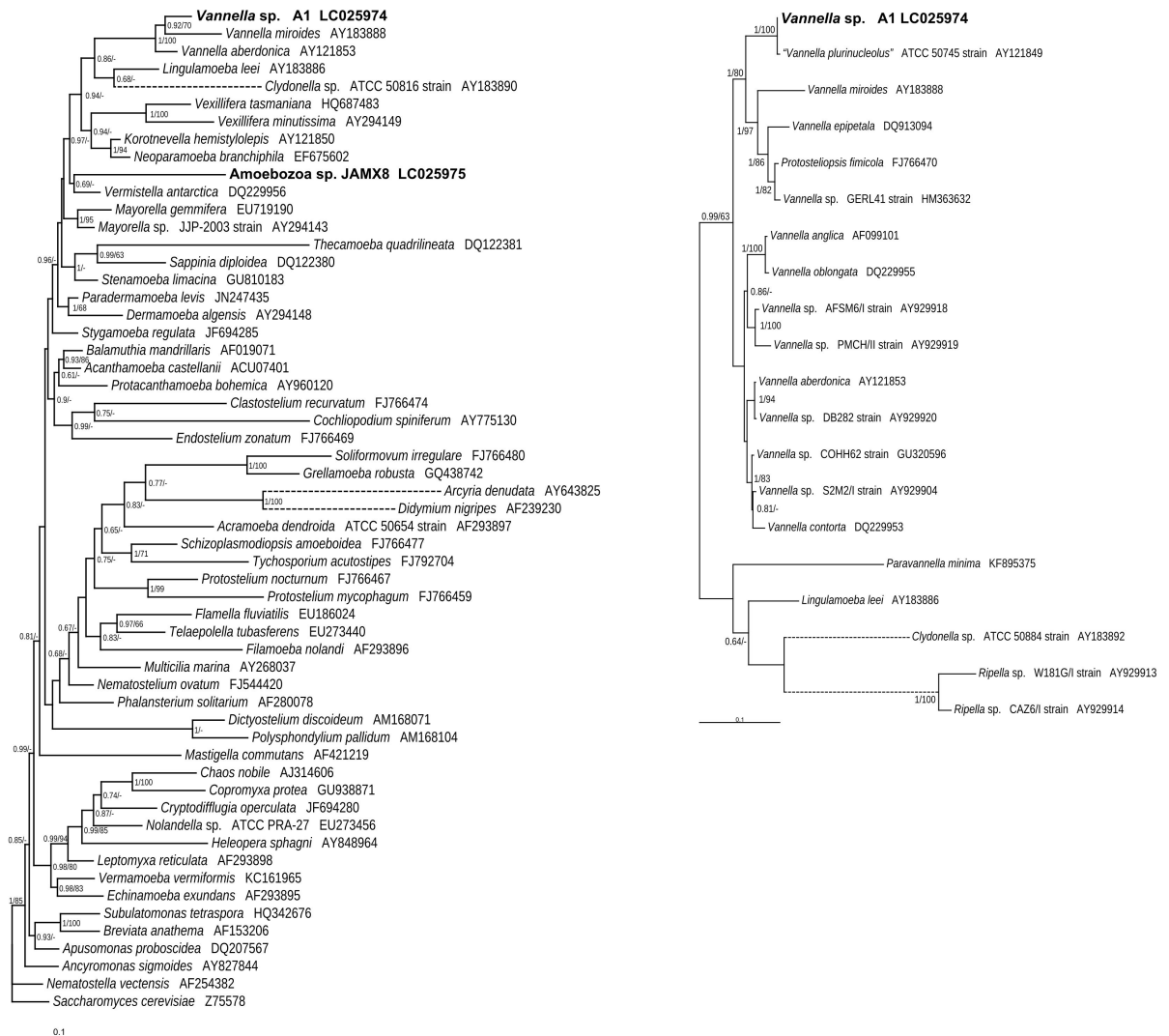
Figure S1. Phylogenetic relationships of *Vannella* sp. A1 and amoeba isolate JAMX8 within the Amoebozoa. Phylogenetic 18S rRNA-based trees of the Amoebozoa (left panel) and Vannellidae (right panel) constructed using the Bayesian inference method. Bayesian posterior probabilities 8 (> 0.6) and RaxML bootstrap support values (> 60%) are indicated at the nodes; the dashed line indicates a branch shortened by 50% to enhance clarity.
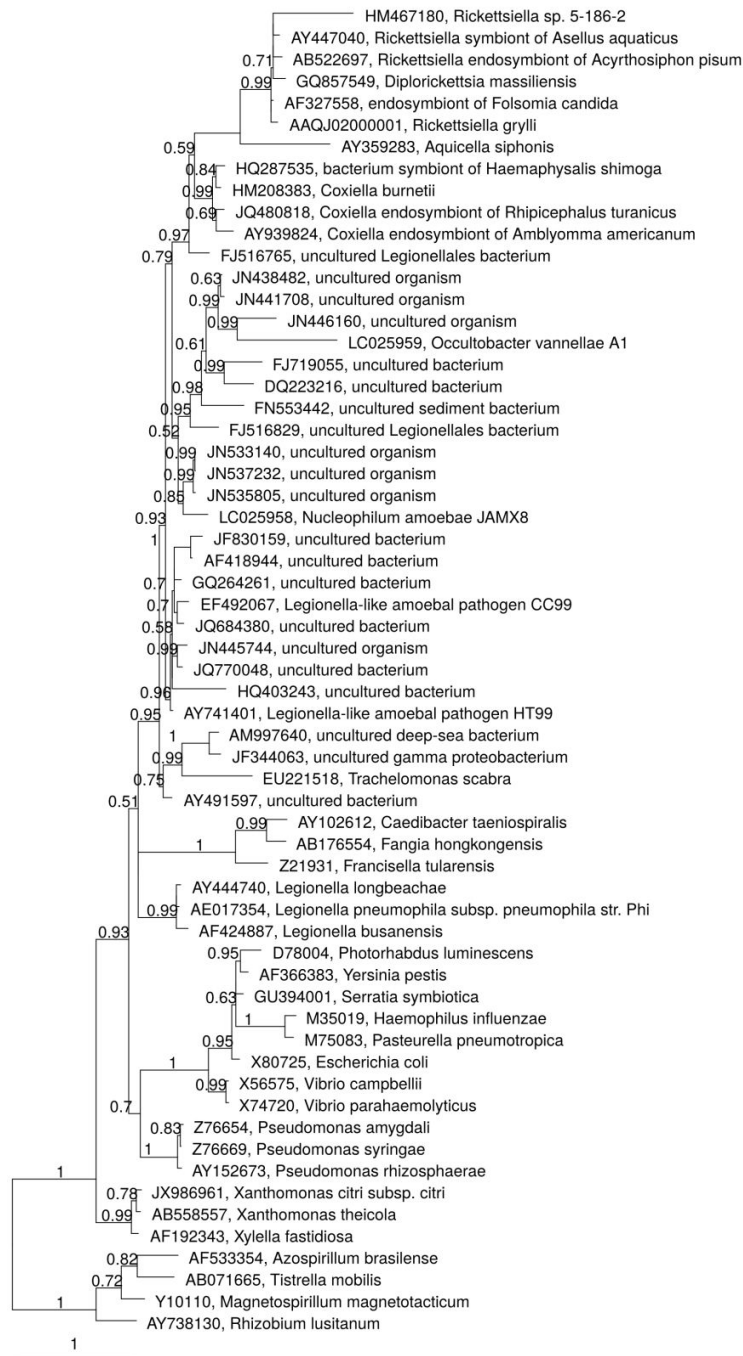
Figure S2. Phylogenetic relationship of 'Candidatus Occultobacter vannellae' and 'Candidatus Nucleophilum amoebae' with the Gammaproteobacteria. The phylogenetic tree (Phylobayes, CAT+GTR) is based on the 16S rRNA sequences, Bayesian posterior probabilities are indicated at the nodes.

Chapter VIII

Synthesis

Synthesis

With this work, we show that the eukaryotic nucleus represents a lucrative niche for bacterial replication. While reviewing the literature of 150 years of research on intranuclear bacteria (Chapter III), it became clear that from the early days of symbiosis research on, scientists have been fascinated by organisms invading this special compartment.

In this thesis, we provide evidence that the intranuclear lifestyle is rather widespread (Chapter III) and new lineages of intranuclear bacteria may readily be retrieved from environmental amoeba isolates (IV, VII). Importantly, even our focus was primarily put on intranuclear symbionts of bacterial origin, it is worth to consider that also eukaryotic microbes, such as microsporidiae have been found to nest in their hosts' nuclei (Stentiford et al., 2013, Corsaro et al. 2014). It should be noted that most intranuclear organisms observed in the early days have never been characterized at a molecular level and many isolates have likely been lost. A targeted sampling of their hosts, such as dinoflagellates, sponges, planthoppers and termites, will help to retrieve these symbionts. Moreover, we expect that besides the so far identified intranuclear bacteria many more lineages are yet to be found.

Our experimental effort provides detailed insights into the life cycle of *Nucleicultrix* (Chapter IV). Based on these findings we then analyzed genomic and proteomic data sets, further deepening our understanding of this unique symbiosis (Chapter V). In particular, we obtained molecular details on the host-symbiont nutritional interplay and the infection process, confirming observations and extending hypotheses made in Chapter IV. Furthermore, we found evidence for the assumption that intranuclear bacteria use effectors to target processes in the host nucleus (Chapter III); the genetic repertoire of *Nucleicultrix* comprises an array of candidate nucleomodulins, of which several are expressed in the proteome (Chapter V). An

obvious advantage of living inside the nucleus is the close association with the host chromatin, which likely eases the delivery of these effector molecules. Downstream experimental characterization of predicted candidate nucleomodulins will help to confirm their function. Ultimately, with additional genomic and functional data becoming available, a systematic analysis of potential nucleomodulins including other intranuclear and cytoplasmic bacteria alike, will contribute to our understanding of the evolution of nucleomodulins in different bacterial lineages and their impact on pathogenicity in general.

At the moment, only four genome sequences of bacteria exclusively thriving in their host's nucleus are available: three *Holospora* species (Dohra et al., 2014) and the genome of *Nucleicultrix* (Chapter V). However, several intranuclear bacteria are currently maintained in the lab together with their hosts; the intranuclear mussel symbiont *Endonucleobacter bathimodioli* (Zielinski et al., 2009), the chlamyidal symbiont "Pn" (Michel et al., 1999), the verrucomicrobial *Nucleococcus* species (Sato et al., 2014), and the alphaproteobacterial paramecium symbiont *Caedibacter caryophilus* (Schrallhammer & Schweikert, 2009). Considering the rapid advancement of single cell genomics and high-throughput sequencing, it is only a matter of time until these and additional genomic data of intranuclear bacteria will become available. Taking a larger set of genomes into account, a comprehensive analysis will be helpful to test another hypothesis made in chapter III, namely to pinpoint "key" genes, which are shared by phylogenetically different intranuclear bacteria and potentially promote this lifestyle. So far, the comparative analysis of the genomes of *Nucleicultrix* and the three *Holospora* species has delivered 20 proteins uniquely shared among these two alphaproteobacterial symbiont lineages (Chapter V). An additional strategy will be to compare the genomes of the intranuclear symbionts with the ones of closely related intracytoplasmic bacteria. However, currently there are no very close relative of the already

sequenced intranuclear bacteria available. In either way, without additional data from functional studies, such as transcriptome and proteome analyses during the course of infection, it will remain difficult to correlate the function of potential key genes with the symbionts' lifestyle.

Phylogenetic analysis has shown that *Nucleicultrix* is a member of a clade of protist associated *Alphaproteobacteria*, which was originally described as family in the order *Rickettsiales* (Chapter V, VI). This clade contains bacteria which are mainly associated with protists, mainly cytoplasmic amoeba symbionts, but also bacteria replicating in the nuclei of ciliates, and pathogens of shrimps (Nunan et al., 2013, Santos and Massard, 2014). Our work has delivered evidence that this symbiont clade does not share a last common ancestor with the *Rickettsiales*, but instead evolved independently out of the *Rhodospirillales* (Chapter V). A distinct feature of most members of this clade is a type VI secretion system (T6SS), in contrast to the *vir*-type IV secretion system of rickettsiae (Chapter V, VI). Interestingly, the T6SS found in the symbiont clade is phylogenetically not related to the T6SS found in free-living alphaproteobacteria but more similar to gammaproteobacterial T6SS and has likely been acquired by horizontal gene transfer (Chapter V). By delivering effector molecules into the host cell, the acquisition of this secretion system might have played an important role in the successful adaptation to the intracellular life.

Besides the obligate intranuclear bacterium *Nucleicultrix*, we investigated a related rickettsial intracytoplasmic symbiont of amoebae, which we named "*Jidaibacter acanthamoeba*" (Chapter VI). Rickettsiae are well-known pathogens of humans and animals, affecting millions of people annually worldwide. This success is based on an intracellular lifestyle and a broad host range, which facilitates spread through arthropod vectors. Rickettsiae are also

found as symbionts in unicellular eukaryotes such as amoebae and thus represent ideal models for investigating the adaptation to eukaryotes and the evolution of an intracellular lifestyle (Darby et al., 2007). Analysis of the genome of *Jidaibacter* has revealed unique features, such as the presence of three type IV secretion systems alongside with a so far unparalleled number of potential effector proteins, a nearly complete set of flagellar genes and in general and a high overlap in gene content with free-living bacteria. When taking together these findings and the insights obtained from the more reduced *Nucleicultrix* genome (Chapter V, VI), it becomes obvious that in particular symbionts of amoebae possess a rich genetic repertoire to enable symbiotic interactions which are likely to be polished in the course of the adaptation to higher eukaryotic hosts or to special niches.

Finally, the work presented in chapter VII nicely illustrates that we are still far away from knowing the full diversity of intracellular bacteria and host-symbiont interactions. Our attempt to isolate amoebae from so-far often neglected marine habitats could extend the known range of strictly host associated intracellular bacteria. Even only two symbionts were identified, their phylogenetically deep-branching nature in the *Gammaproteobacteria* together with their unique features is remarkable. The best example is one of the symbionts which uses the perinuclear space of its host for intracellular replication, which opens up intriguing questions about infection process, host-symbiont nutritional interplay and evolution of this unique lifestyle. We assume that the two isolates only represent the tip of the iceberg and that in particular marine habitats clearly represent a rich pool of yet to be discovered bacterial symbionts.

References

Birtles, R.J., Rowbotham, T.J., Michel, R., Pitcher, D.G., Lascola, B., Alexiou-Daniel, S., and Raoult, D. (2000). "*Candidatus* Odyssella thessalonicensis" gen. nov., sp. nov., an obligate intracellular parasite of *Acanthamoeba* species. Int J Syst Evol Microbiol *50*, 63–72.

Corsaro, D., Walochnik, J., Venditti, D., Müller, K.-D., Hauröder, B., and Michel, R. (2014). Rediscovery of *Nucleophaga amoebae*, a novel member of the *Rozellomycota*. Parasitol Res *113*, 4491–4498.

Darby, A.C., Cho, N.-H., Fuxelius, H.-H., Westberg, J., and Andersson, S.G.E. (2007). Intracellular pathogens go extreme: genome evolution in the *Rickettsiales*. Trends Genet. *23*, 511–520.

Dohra, H., Tanaka, K., Suzuki, T., Fujishima, M., and Suzuki, H. (2014). Draft genome sequences of three *Holospora* species (*Holospora obtusa* , *Holospora undulata* , and *Holospora elegans*), endonuclear symbiotic bacteria of the ciliate *Paramecium caudatum*. FEMS Microbiology Letters *359(1)*, 16-8.

Michel, R., Hauroder, B., Muller, K., and Zoller, L. (1999). An environmental *Naegleria*-strain, unable to form cysts-turned out to harbour two different species of endocytobionts. Endocytobiosis Cell Research *118*, 115–118.

Nunan, L.M., Pantoja, C.R., Gomez-Jimenez, S., and Lightner, D.V. (2013). "Candidatus Hepatobacter penaei," an intracellular pathogenic enteric bacterium in the hepatopancreas of the marine shrimp *Penaeus vannamei* (Crustacea: Decapoda). Appl. Environ. Microbiol. *79*, 1407–1409.

Santos, H.A., and Massard, C.L. (2014). The Family *Holosporaceae*. In The Prokaryotes, E. Rosenberg, E.F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson, eds. (Springer Berlin Heidelberg), pp. 237–246.

Sato, T., Kuwahara, H., Fujita, K., Noda, S., Kihara, K., Yamada, A., Ohkuma, M., and Hongoh, Y. (2014). Intranuclear verrucomicrobial symbionts and evidence of lateral gene transfer to the host protist in the termite gut. The ISME Journal *8*, 1008–1019.

Schmitz-Esser, S., Toenshoff, E.E.R., Haider, S., Heinz, E., Hoenninger, V.M., Wagner, M., Horn, M., and Isolates, A. (2008). Diversity of bacterial endosymbionts of environmental acanthamoeba isolates. Applied and Environmental Microbiology *74*, 5822–5831.

Schrallhammer, M., and Schweikert, M. (2009). The killer effect of *Paramecium* and its causative agents. In Endosymbionts in *Paramecium*, (Springer Berlin Heidelberg), pp. 227–246.

Stentiford, G.D., Feist, S.W., Stone, D.M., Bateman, K.S., and Dunn, A.M. (2013). *Microsporidia*: diverse, dynamic, and emergent pathogens in aquatic systems. Trends in Parasitology *29*, 567–578.

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

Chapter IX

Abstract

Abstract

Most eukaryotes are associated with bacteria which are either beneficial for their hosts or appear as pathogens. Amoebae represent an ideal model system to study such host-symbiont interactions; they frequently harbor diverse intracellular bacteria, which almost exclusively replicate within the cytoplasm. Here we deliver evidence that bacteria infecting their host cell's nucleus are surprisingly widespread, are affiliated with four major bacterial groups and infect a wide-range of eukaryotes. We report on the discovery of a novel alphaproteobacterial symbiont which occurs in freshwater and soil habitats and specifically targets the nucleus of amoebae. Although the bacterium has no impact on the fitness of its native *Hartmannella* sp. host, it is lytic for *Acanthamoeba castellanii*. The symbiont's 1.84 Mb genome is enriched in factors facilitating host interaction and virulence; three nucleotide transporters, type IV pili, a polysialic acid capsule, a flagellum, a type VI secretion system and myriads of candidate nucleomodulins. Analysis of the proteome of infected and uninfected amoeba nuclei confirmed the expression of the predicted macromolecular structures and effectors and illustrated the strong impact of an intranuclear infection on host cellular processes, such as transcription, translation and signal transduction. Besides this obligate intranuclear bacterium, we studied the infection process and genome of a related intracytoplasmic symbiont of amoebae, revealing unique features, such as the presence of three type IV secretion systems along with a so far unparalleled number of potential effector proteins, and a nearly complete set of flagellar genes. At last, we report on two novel gammaproteobacteria which infect amoebae in so-far often neglected marine environments. We show that one of the two symbionts colonizes the perinuclear space of its host cell, representing yet another example of the nucleus as a niche for bacterial replication. Overall, this work significantly contributes to our current knowledge of evolution, underlying mechanisms and adaptation of bacteria to intracellular life in eukaryotes.

Zusammenfassung

Die meisten Eukaryonten sind eng mit Bakterien assoziiert, die entweder als Mutualisten oder als Pathogene auftreten. Amöben stellen ein ideales Modellsystem dar, um das Zusammenspiel zwischen Wirt und Symbiont zu untersuchen. Diese Einzeller tragen häufig Bakterien in sich, welche sich nahezu ausschließlich im Wirtszytoplasma vermehren. In der vorliegenden Doktorarbeit zeigen wir, dass auch Bakterien die den Zellkern ihres Wirtes infizieren überraschend häufig vorkommen, sie zu vier großen bakteriellen taxonomischen Gruppen gehören und eine Vielzahl an verschiedenen Eukaryonten infizieren können. Wir beschreiben die Entdeckung eines neuartigen Alphaproteobakteriums, welches im Süßwasser und im Boden vorkommt und gezielt den Zellkern von Amöben infiziert. Obwohl das Bakterium keinen schädigenden Einfluss auf Hartmannellen als den originalen Wirt hat, lysiert es Acanthamöben. Das 1.84 Megabasenpaare große Genom des Symbionten enthält eine Ansammlung an Genen die im Zusammenhang mit Wirtsinteraktion und Virulenz gebracht werden können, wie zum Beispiel drei Nukleotidtransporter, Typ IV Pili, eine Kapsel bestückt mit Sialinsäuren, ein Flagellum, ein Typ VI Sekretionssystem und eine Vielzahl an möglichen Nukleomodulinen. Die Analyse des Proteoms infizierter und uninfizierter Zellkerne konnte des Weiteren zeigen, dass viele der vorhergesagten makromolekularen Strukturen und Effektoren tatsächlich exprimiert werden und dass die Infektion des Zellkerns einen starken Einfluss auf grundlegende Vorgänge in der Wirtszelle hat, wie zum Beispiel auf Transkription, Translation und Signaltransduktion. Neben diesem obligaten Kernsymbionten wurden im Rahmen der Doktorarbeit auch der Infektionsprozess und das Genom eines intrazytoplasmatischen Amöbensymbionten untersucht. Dieses Bakterium weist ebenfalls viele einzigartige Merkmale auf, wie zum Beispiel drei verschiedene Typ IV Sekretionssysteme, eine nie zuvor beobachtete Vielzahl an möglichen Effektorproteinen und ein Flagellum. Darüber hinaus beschreiben wir zwei neuartige Gammaproteobakterien die aus marinen Habitaten stammen und Amöben infizieren. Einer der beiden Symbionten vermehrt sich im perinukleären Raum seines Wirtes und stellt dadurch ein weiteres Beispiel für Bakterien dar, die sich den Zellkern als Nische ausgesucht haben. Insgesamt gesehen trägt diese Arbeit zu unser Verständnis über Evolution und Anpassung von Bakterien an das intrazelluläre Leben bei.

Appendix

Acknowledgments

Curriculum vitae

Acknowledgments

Curriculum vitae

Frederik Schulz

Division of Microbial Ecology • Department of Microbiology and Ecosystem Science

University of Vienna • Althanstr. 14 • 1090 Vienna • Austria

schulz@microbial-ecology.net • +43 1 4277 76609

Education

- M.Sc. Molecular Microbiology, University of Vienna, Austria, 2011
- B.Sc. Biochemistry, University of Applied Sciences Mannheim, Germany, 2009

Current Position

- 2012-2015: Graduate student, Division of Microbial Ecology, University of Vienna, Austria

University Positions

- 2011-2012: Research assistant, Division of Microbial Ecology, University of Vienna, Austria
- 2010-2011: Master thesis (12 months), Division of Microbial Ecology, University of Vienna, Austria
- 2008-2009: Research internship (6 months), Centre for Medical Research, University of Western Australia, Australia
- 2006-2007: Research internship (5 months), Department of Clinical Chemistry, University of Heidelberg, Germany

Fellowships

- 2012-2015: OeAW ("Austrian Academy of Sciences") DOC scholarship for PhD studies
- 2008-2009: DAAD ("German Academic Exchange Service") scholarship for a research project abroad
- 2007-2008: VDI ("The Association of German Engineers") scholarship to cover tuition fees

Teaching Experience

- 2015: International FISH course, University of Vienna, Austria
- 2012-2013: Supervision of a Master student (12 months), University of Vienna, Austria
- 2011: Supervision of a Bachelor student's final research project (3 months), University of Vienna, Austria
- 2011,2013: Practical course "Fluorescence in situ hybridization (FISH) - Identification of uncultivated microorganisms", University of Vienna, Austria

Services to the University and the Scientific Community

- 2014-2015: IT support for the Division of Microbial Ecology, University of Vienna, Austria
- 2014: Co-organizer of the 15th International Meeting on the Biology and Pathogenicity of Free-Living Amoebae (FLAM2014) at the Natural History Museum Vienna, Austria
- 2013: Organization of workshop "Let's talk about symbiosis", University of Vienna, Austria

Publications

1. **Schulz, F.**, Tyml, T., Pizzetti, I., Dykova, I., Kostka, M., & Horn, M. (2015) "Marine amoebae with cytoplasmic and perinuclear symbionts deeply branching in the *Gammaproteobacteria*.", *submitted* to Scientific Reports 30/04/2015
2. Na, H., Lever, M. A., Kjeldsen, K.U., **Schulz, F.**, & Jørgensen, B. B. (2015). "Uncultured *Desulfobacteraceae* and Crenarchaeotal group C3 incorporate [13C]acetate in coastal marine sediment."Environmental Microbiology Reports, DOI: 10.1111/1758-2229.12296, *in press*
3. **Schulz, F.**, Martijn, J., Wascher, F., Rok, K., Ettema, T., & Horn, M. (2015). "A *Rickettsiales* symbiont of amoebae with ancient features." Environmental Microbiology, DOI: 10.1111/1462-2920.12881, *in press*
4. Martijn, J., **Schulz, F.**, Zaremba-Niedzwiedzka, K., Viklund, J., Stepanauskas, R., … & Ettema, T. (2015)."Single cell genomics of a rare environmental alphaproteo-bacterium provides unique insights into *Rickettsiaceae* evolution." The ISME journal, DOI: 10.1038/ismej.2015.46, *in press*
5. **Schulz, F.**, & Horn, M. (2015). "Intranuclear bacteria: inside the cellular control center of eukaryotes." Trends in Cell Biology, 25 (6): 339-346.
6. **Schulz, F.**, Lagkouvardos, I., Wascher, F., Aistleitner, K., & Horn, M. (2014). "Life in an unusual intracellular niche: a bacterial symbiont infecting the nucleus of amoebae." The ISME journal, 8(8), 1634-1644.
7. Aistleitner, K., Heinz, C., Hörmann, A., Heinz, E., Montanaro, J., **Schulz, F.**, ... & Horn, M. (2013)."Identification and characterization of a novel porin family highlights a major difference in the outer membrane of chlamydial symbionts and pathogens." PloS one, 8(1), e55010.
8. Zhang, A., London, R., **Schulz, F.M.**, Giguère-Simmonds W.R., Delriviere, L., Chandraratana H., Hardy K., Zheng, S., Olynyk, J.K., & Yeoh, G. (2010). "Human liver progenitor cell lines are readily established from non-tumorous tissue adjacent to hepatocellular carcinoma." Stem cells and development 19, 1277-1284.

in preparation

- **Schulz, F.**, Wascher, F., Martijn, J., Ettema, T., Becher, D., Markert, S., Schweder, T.,Weinmaier, T., Rattei, T., & Horn, M. "What it takes to live inside the nucleus: an '-omics' perspective on the intranuclear lifestyle.", *in prep*
- **Schulz, F.**, Volland, J.M., Szabo, G., Bright, M., & Horn, M. "Life at the minimum: the genome of a ciliate-associated dark matter bacterium.", *in prep*
- **Schulz, F.**, Martijn, J., Horn, M., Ettema, T. "Multiple origin of obligate intracellular symbiosis in the Alphaproteobacteria.", *in prep*
- **Schulz, F.**, Horn, M. & Haferkamp I. "Highly specific UTP uptake by an intranuclear bacterium", *in prep*
- Pizzetti I., **Schulz, F.**, Fuchs, B., Tomas, T., Amann, R., Horn, M., & Fazi, S. "Chlamydial seasonal dynamics and isolation of marine chlamydiae from a Tyrrhenian coastal lake.", *in prep*
- Szabo, G., **Schulz, F.**, & Horn, M. "Evolutionary recent co-obligatory symbionts in the Adelges laricis/tardus species complex (Insecta: Hemiptera: Adelgidae).", *in prep*
- Szabo, G., **Schulz, F.**, Volland, J.M., & Horn, M. "Convergent patterns in the evolution of mealybug symbioses involving different endobacterial symbionts.", *in prep*

Oral Presentations at Scientific Meetings

- **Schulz, F.**, Wascher, F., Martijn, J., Ettema, T., Markert, S., Schweder, T.,Weinmaier, T., Rattei, T., & Horn, M. "What it takes to live inside the nucleus: an '-omics' perspective on the intranuclear lifestyle." Let's talk about symbiosis workshop, Vienna, Austria
- 2013: **Schulz F.**, Wascher, F., Weinmaier, T., Kostanjšek, R., Horn, M. "A Journey to the Nucleus." 15th International Meeting on the Biology and Pathogenicity of Free-Living Amoebae at the Natural History Museum Vienna, Austria
- 2013: **Schulz, F.**, Weinmeier, T., Lagkouvardos, I., Wascher, F., Kostanjšek, R., Rattei,T., Horn, M. "An Enigmatic Bacterial Symbiont Residing in the Nucleus of Amoebae." American Society of Microbiology (ASM): 113th general meeting, Denver, Colorado, USA
- 2013: **Schulz, F.**, Weinmeier, T., Lagkouvardos, I., Wascher, F., Kostanjšek, R., Rattei, T., Horn, M. "An Enigmatic Bacterial Symbiont Residing in the Nucleus of Amoebae." Let's talk about symbiosis workshop, Vienna, Austria
- 2012: **Schulz, F.**, Kuroll, M., Aistleitner, K., Horn, M. "Life Inside the Nucleus – An Unusual Symbiont of Amoebae Related to Rickettsiae." 5. Gemeinsame Jahrestagung der Deutschen Gesellschaft für Hygiene und Mikrobiologie (DGHM) und der Vereinigung für Allgemeine und Angewandte Mikrobiologie (VAAM), Tuebingen, Germany
- 2011: **Schulz, F.**, Horn, M. "Life inside the nucleus - an unusual symbiosis." INTIME workshop 2011, Slettestrand, Denmark

Poster Presentations at Scientific Meetings

- 2014: **Schulz, F.**, Martijn, J., Wascher, F., Rok, K., Ettema, T., & Horn, M. "A Rickettsiales symbiont of amoebae with ancient features." EMBO conference: Microbiology after the genomics revolution: Genomes 2014, Institut Pasteur, Paris, France
- 2014: Szabo, G., **Schulz, F.**, Domman, D., Turaev, D., Weinmaier, T., Rattei, T., Horn, M. "Evolutionary young co-obligatory symbionts in the Adelges laricis/tardus complex (Insecta: Hemiptera: Adelgidae)" EMBO conference: Microbiology after the genomics revolution: Genomes 2014, Institut Pasteur, Paris, France
- 2014: Martijn, J., **Schulz, F.**, Stepanauskas, R., Andersson, S., Horn, M., Guy, L., Ettema, T. "Single cell genomics analyses of an ultra-rare, deeply branching alphaproteobacterium provides unique insights in the evolution of the Rickettsiales" EMBO conference: Microbiology after the genomics revolution: Genomes 2014, Institut Pasteur, Paris, France
- 2012: **Schulz, F.**, Kostanjšek, R., Horn, M. "An Enigmatic Bacterial Symbiont Residing in the Nucleus of Amoebae." 14th International Symposium on Microbial Ecology (ISME), Copenhagen, Denmark

Press Coverage

- 2014: "Der Zellkern als Lebensraum" Frankfurter Allgemeine Zeitung (FAZ), 30/03/2014