Molecular biology and evolution of the bacterial intranuclear parasite *Ca.* Endonucleobacter

Dissertation

Zur Erlangung des Grades eines

Doktors der Naturwissenschaften

- Dr. rer. nat. -

dem Fachbereich Biologie/Chemie der

Universität Bremen

vorgelegt von

Miguel Ángel González Porras

Bremen

September 2020

Die Untersuchungen zur vorliegenden Doktorarbeit wurden in der Abteilung Symbiose am Max-Planck-Institut für Marine Mikrobiologie unter der Leitung von Prof. Dr. Nicole Dubilier durchgeführt.

Gutachter

1. Prof. Dr. Nicole Dubilier

2. Prof. Dr. Matthias Horn

Prüfer

3. Dr. Bernhard Fuchs

4. Prof. Dr. Michael W. Friedrich

Tag des Promotionskolloquiums

24. November. 2020

A mi padre (to my father).

“En la ciencia, como en la vida, el fruto viene siempre después del amor.”

(“In science, like in life, fruit comes always after love.”)

- Santiago Ramón y Cajal

A los 32 años, yo sabía algo de biología y muy poco de la vida. Me llevó 4 meses aprender los Principios de Bioquímica de Lehninger y 32 años aprender que debía respetarme a mí mismo.

- Nota a mí mismo

(When I was 32, I knew a couple of things about biology, but very little about life. It took me 4 months to learn the Principles of Biochemistry of Lehninger, and 32 years to learn that I must respect myself.)

- (Note to myself)

Summary

Few bacteria are able to colonize eukaryotic host cells, and even fewer can invade organelles. Among the latter, those which colonize the nucleus (intranuclear bacteria) are a majority. The nucleus is considered as a replication niche that offers bacteria shelter, the opportunity to manipulate nuclear processes and nutrients in form of chromatin. Previous studies have suggested that intranuclear bacteria consume host chromatin, but this would inevitably lead to a collapse of their replication niche. However, nuclear colonization has intrinsic risks for bacteria, as eukaryotic cells are able to sense nuclear deformation and respond to it activating apoptosis. Therefore, intranuclear bacteria must count on molecular mechanisms to prevent their host cells to shut down. Although many intranuclear bacteria have been visually characterized, little is known about how they colonize the nucleus, how they thrive within, and which effects they have on their host cells. This is because cultivation of intranuclear bacteria has the immense additional difficulty of having to be co-cultivated with their host cells. In this thesis, I therefore aimed to deepen our current knowledge about the molecular principles behind intranuclear lifestyle in bacteria by studying the biology of *Ca.* Endonucleobacter, an intranuclear parasite of bathymodiolin mussels. To do so, I performed genomic, transcriptomic, proteomic and visualization analyses of *Ca.* Endonucleobacter. *Ca.* Endonucleobacter has a characterized infectious cycle with clearly defined stages, which allow me to relate omics data with the progression of the infection.

"Bathymodiolus" childressi infected by *Ca.* Endonucleobacter is an ideal system to study intranuclear bacteria-host cell interactions using cultivation-independent technologies. In *"B". childressi*, *Ca.* Endonucleobacter is limited to the ciliated edge of the gill filaments. It was therefore possible to retrieve high amounts of microbial biomass in single sectioning planes. I took advantage of this biological feature to profile the transcriptomes of *Ca.* Endonucleobacter and its host cell along the infectious cycle. To do so, I developed a pipeline that coupled laser-capture microdissection with ultra-low-input RNAseq (chapter II). The gills of infected *"B." childressi* were analyzed with fluorescence *in situ* hybridization (FISH) targeting *Ca.* Endonucleobacter and different infection stages were microdissected and sequenced. My analyses revealed that *Ca.* Endonucleobacter does not consume host chromatin but chitin, an extracellular matrix component of *"B." childressi* gill filaments. Further, my results revealed that *Ca.* Endonucleobacter might avoid the collapse of the host cell by disentangling the host cell cytoskeleton and expressing inhibitors of apoptosis. The host cell was transcriptionally active during the whole infectious cycle, upregulating sugar import, glycolysis, lipid droplets synthesis and nuclear deformation sensors. My results suggested that *Ca.* Endonucleobacter brings the host cell to starvation state by consuming its energy budget. Consequently, the host cell tries to compensate the metabolic burden of a "hungry nucleus" by importing chitobiose disaccharides that are produced during extracellular chitin digestion. The results from this study lead us to conclude that *Ca.* Endonucleobacter uses the host cell as the living funneling agent that creates a constant flow of nutrients between the extracellular matrix and the nucleus.

Ca. Endonucleobacter is the sister clade of the genus *Endozoicomonas* (*Oceanospirillales*, *Hahellaceae*), a ubiquitous group of symbiotic bacteria that stablish mutualistic relationships with a great diversity of marine metazoans. These two genera have drastically different symbiotic lifestyles (parasitism vs. mutualism), and so far it is unclear if their different strategies are associated with a specific set of genes. Moreover, the evolutionary events that might have caused the divergence of both genera remain elusive. Comparative genomic and orthology analyses (chapter III) revealed that *Ca.* Endonucleobacter spp. dedicated more than 20% of their genes to "replication, recombination and repair", a functional category that includes mobile elements. Further, our analyses revealed that inhibitors of apoptosis were expanded in Ca. Endonucleobacter spp., while virtually absent in the genus *Endozoicomonas*. These results suggested that genomic plasticity might have played a role in the switching from a mutualistic to a parasitic lifestyle, when the genus *Ca.* Endonucleobacter diverged from

Endozoicomonas. We conclude that inhibitors of apoptosis might be the ultimate genomic innovation that allowed *Ca.* Endonucleobacter to set the nucleus of its host cell as a permanent residence.

Bathymodiolin mussels dominate hydrothermal vents and cold seeps in the deep sea in terms of biomass. These marine invertebrates live in symbiosis with chemosynthetic sulfur-oxidizing (SOX) or methane-oxidizing (MOX) symbionts, or both, harboring them in specialized gill cells called bacteriocytes. Up to 16 different functional strains of the SOX symbionts can coexist in a single bathymodiolin mussel. Despite these significant numbers, little is known about how SOX strains organize spatially within the gill filaments of bathymodiolin mussels. This is because metagenomic approaches cannot resolve the spatial structuring of symbiotic communities within the symbiotic organs of their hosts. Therefore, I tackle the limitations of metagenomic approaches by adapting the Direct-geneFISH protocol on host-microbiome systems, which co-localizes the 16S rRNA of a bacterium and a gene of interest (chapter IV). Using fluorescently labeled probes, I targeted the strain-specific markers methanol dehydrogenase (MDH) and hydrogenase (HYD) gene clusters in the SOX population of the gill filaments of *B. azoricus*, and localized them using confocal laser scanning and super-resolution microscopy. My results revealed that bacteriocytes are quantitatively dominated by a single SOX strain. Further, single SOX strains co-occurred in neighboring bacteriocytes forming patches. My results suggested biological exclusion mechanisms at a single bacteriocyte level that might prevent the colonization of bacteriocytes by more than one SOX strain. Further, these results made me conclude that the colonization of newly formed aposymbiotic bacteriocytes by adjacent and symbiotic bacteriocytes of the same gill filament might create patches in a phenomenon of intra-filament colonization.

By targeting the transcriptomic profiles of specific life stages of an intranuclear parasite *in situ*, my combined FISH – LCM – ultra-low-input RNAseq pipeline closed the gap between the visual, molecular and temporal characterization of an intranuclear parasite-host interaction. My findings challenge previous hypotheses about nutritional and survival strategies of intranuclear bacteria, and point out molecular mechanisms and genomic innovations that might define intranuclear lifestyle in bacteria.

Zusammenfassung

Nur wenige Bakterien können eukaryotische Wirtszellen besiedeln, und noch weniger können in Organellen eindringen. Unter den letzteren machen Bakterien, die den Zellkern besiedeln (intranukleäre Bakterien), die Mehrheit aus. Der Zellkern wird als Replikationsnische angesehen, der den Bakterien Schutz bietet und die Möglichkeit bietet Prozesse innerhalb des Zellkerns zu manipulieren und an Nährstoffe in Form von Chromatin zu gelangen. Frühere Studien haben gezeigt, dass intranukleäre Bakterien sich von Wirtschromatin ernähren, was jedoch unweigerlich zu einem Zusammenbruch ihrer Replikationsnische führen würde. Die Besiedlung des Zellkerns birgt jedoch wesentliche Risiken für Bakterien, da eukaryotische Zellen in der Lage sind, die Verformung des Zellkerns wahrzunehmen und darauf zu reagieren, indem die Apoptose eingeleitet wird. Daher müssen intranukleäre Bakterien auf molekulare Mechanismen zurückgreifen, um zu verhindern, dass ihre Wirtszellen herunterfahren. Obwohl viele intranukleäre Bakterien optisch charakterisiert wurden, ist wenig darüber bekannt, wie sie den Zellkern besiedeln, wie sie sich im Inneren entwickeln und welche Auswirkungen sie auf ihre Wirtszellen haben. Dies liegt daran, dass die Kultivierung von intranukleären Bakterien die große zusätzliche Schwierigkeit birgt, sie zusammen mit ihren Wirtszellen zu kultivieren. Ziel dieser Arbeit war daher unser derzeitiges Wissen über die molekularen Prinzipien des intranukleären Lebensstils von Bakterien vertiefen, indem ich die Biologie von *Ca*. Endonucleobacter, einem intranukleären Parasiten von Bathymodiolus-Muscheln, untersuchte. Zu diesem Zweck führte ich genomische, transkriptomische, proteomische und mikroskopische Analysen von *Ca*. Endonucleobakter durch. *Ca*. Endonucleobacter hat einen charakteristischen Infektionszyklus mit klar definierten Stadien, die es mir ermöglichten, Omics-Daten mit dem Voranschreiten der Infektion in Beziehung zu bringen.

"*Bathymodiolus*" *childressi* infiziert mit *Ca*. Endonucleobacter ist ein höchst geeignetes System zur Untersuchung der Wechselwirkungen zwischen intranukleären Bakterien und deren Wirtszellen unter Verwendung von Kultivierungsunabhängigen Techniken. In "*B*". *childressi*, ist die Verbreitung von *Ca*. Endonucleobacter auf die ziliierten Ränder der Kiemenfilamente beschränkt. Es war daher möglich, eine große Menge mikrobieller Biomasse von einzelnen Schnittebenen zu gewinnen. Ich nutzte diese biologische Besonderheit, um die Transkriptome von *Ca*. Endonukleobakter und seiner Wirtszelle während des Infektionszyklus zu untersuchen. Zu diesem Zweck habe ich eine Pipeline entwickelt, die Laser-Capture-Mikrodissektion mit ultra-low-input RNAseq koppelt (Kapitel II). Die Kiemen infizierter "*B*". *childressi* wurden nach Anwendung von Fluoreszenz *in situ* Hybridisierung (FISH) auf *Ca*. Endonucleobacter untersucht und mit verschiedene Infektionsstadien herausgeschnitten und sequenziert. Meine Analysen ergaben, dass sich *Ca*. Endonucleobacter nicht von Wirtschromatin, sondern Chitin, einer extrazelluläre Matrixkomponente der Kiemenfilamente von "*B*". *childressi*, ernährt. Ferner zeigten meine Ergebnisse, dass *Ca*. Endonucleobacter den Zerfall der Wirtszelle vermeiden könnte, indem es Vernetzungen des Zytoskeletts der Wirtszelle aufhebt und Apoptoseinhibitoren exprimiert. Die Wirtszelle war während des gesamten Infektionszyklus transkriptionell aktiv und regulierte den Zuckerimport, die Glykolyse, die Lipidtröpfchensynthese und die Kerndeformationssensoren. Meine Ergebnisse legen nahe, dass *Ca*. Endonucleobacter die Wirtszelle in einen Hungerzustand bringt, indem es ihr Energiebudget verbraucht. Folglich versucht die Wirtszelle, die Stoffwechselbelastung eines „hungrigen Kerns“ durch den Import von Chitobiose-Disacchariden zu kompensieren, die während der extrazellulären Chitinverdauung entstehen. Die Ergebnisse dieser Studie lassen den Schluss zu, dass *Ca*. Endonucleobacter die Wirtszelle als Zwischenglied benutzt, welches einen konstanten Nährstofffluss zwischen der extrazellulären Matrix und dem Kern erzeugt.

Ca. Endonucleobacter ist die Schwesterklade der Gattung *Endozoicomonas* (*Oceanospirillales*, *Hahellaceae*), einer überall vorkommenden Gruppe symbiotischer Bakterien, die mutualistische Beziehungen mit einer großen Vielfalt mariner Metazoen entwickeln. Diese beiden Gattungen haben einen erheblich unterschiedlichen symbiotischen Lebensstil (Parasitismus vs. Mutualismus), und es ist bisher unklar, ob ihre unterschiedlichen Strategien mit bestimmten Genen verbunden sind. Darüber hinaus bleiben die evolutionären Ereignisse, die die Divergenz beider Gattungen verursacht haben könnten, schwer nachvollziehbar. Vergleichende genomische und orthologische Analysen (Kapitel III) ergaben, dass *Ca*. Endonucleobacter spp. mehr als 20% ihrer Gene der „Replikation, Rekombination und Reparatur“ zugeordnet haben, einer Funktionskategorie, die mobile Elemente umfasst. Ferner zeigten unsere Analysen, dass Apoptose-Inhibitoren in *Ca*. Endonucleobacter spp. zunahmen, während diese in der Gattung *Endozoicomonas* praktisch nicht vorhanden sind. Diese Ergebnisse legen nahe, dass die genomische Veränderlichkeit eine Rolle beim Übergang von einem mutualistischen zu einem parasitären Lebensstil gespielt haben könnte, als sich die Gattung *Ca*. Endonucleobacter von *Endozoicomonas* abgespalten hat. Wir schließen daraus, dass Apoptose-Inhibitoren die maßgebliche genomische Erfindung sein könnten, die *Ca*. Endonucleobacter dazu ermöglicht, den Kern seiner Wirtszelle als permanenten Aufenthaltsort/Wohnsitz festzulegen.

Bathymodiolus-Muscheln dominieren hinsichtlich ihrer Biomasse Hydrothermalquellen und kalte Quellen in der Tiefsee. Diese marinen Invertebraten leben in Symbiose mit entweder chemosynthetischen schwefeloxidierenden (SOX) oder methanoxidierenden (MOX) Symbionten oder beiden und beherbergen diese in spezialisierten Kiemenzellen, die als Bakteriozyten bezeichnet werden. Bis zu 16 funktionell verschiedene Stämme der SOX-Symbionten können in einer einzigen Bathymodiolus-Muschel zusammen vorkommen. Trotz dieser erheblichen Anzahl ist wenig darüber bekannt, wie sich SOX-Stämme räumlich innerhalb der Kiemenfilamente von Bathymodiolus-Muscheln organisieren. Dies liegt daran, dass metagenomische Methoden die räumliche Strukturierung symbiotischer Gemeinschaften innerhalb der symbiotischen Organe ihrer Wirte nicht auflösen können. Daher gehe ich die Einschränkungen metagenomischer Ansätze an, indem ich das „Direct-geneFISH“-Protokoll an Wirt-Mikrobiom-Systeme anpasse, welches die 16S-rRNA eines Bakteriums und ausgewählter Gene gleichzeitig lokalisiert (Kapitel IV). Mit fluoreszenzmarkierten Sonden versah ich die Stamm-spezifischen Genmarker Methanolhydrogenase (MDH) und Hydrogenase (HYD) in der SOX-Population der Kiemenfilamente von *B. azoricus* und lokalisierte sie mithilfe von konfokaler Laserscanning Mikroskopie und „super-resolution“ Mikroskopie. Meine Ergebnisse zeigten, dass Bakteriozyten quantitativ von einem einzelnen SOX-Stamm dominiert werden. Ferner traten einzelne SOX-Stämme in benachbarten Bakteriozyten auf, die ungleichmäßige Flecken bildeten. Meine Ergebnisse deuteten auf biologische Ausschlussmechanismen einzelner Bakteriozyten hin, welche die Besiedlung von Bakteriozyten durch mehr als einen SOX-Stamm verhindern könnten. Ferner ließen diese Ergebnisse den Schluss zu, dass die Besiedlung neu gebildeter aposymbiotischer Bakteriozyten durch benachbarte und symbiotische Bakteriozyten desselben Kiemenfilaments stellenweise in einem Phänomen der Intrafilamentkolonisation auftreten könnte.

Durch die gezielte Erfassung der Transkriptomprofile bestimmter Lebensstadien eines intranukleären Parasiten *in situ* schloss meine kombinierte FISH - LCM – ultra-low-input RNAseq-Pipeline die Lücke zwischen der visuellen, molekularen und zeitlichen Charakterisierung einer intranukleären Parasiten-Wirt-Beziehung. Meine Ergebnisse stellen frühere Hypothesen über Ernährungs- und Überlebensstrategien intranukleärer Bakterien in Frage und weisen auf molekulare Mechanismen und genomische Innovationen hin, die den intranukleären Lebensstil von Bakterien definieren könnten.

Chapter I | Introduction

1.1 Symbiosis

If a poet should define symbiosis, it would probably state that is “the art of living together”. Not a poet, but it was Heinrich Anton de Bary who first introduced the term as “the continuous and intimate living together of not-alike organisms” (de Bary, 1879). The term was originally used to describe lichens, and it was a debate during the last century if it should be limited to mutually beneficial relationships among organisms. This debate was abandoned long time ago, and it has been accepted that symbiosis should include all kind of relationships between different organisms, regardless of who gets the benefit. The word symbiosis can be decomposed into two parts derived from Greek - syn which means “together” and -bios which means “life”.

Symbiosis includes three types of interactions based on the benefit of the involved organisms. Two organisms can be mutually beneficial (mutualism), neutral (commensalism) or one partner can profit from the relationship while harming the other partner (parasitism). The line that separates mutualism or neutralism from parasitism has long been object of study by the scientific community. Several microorganisms can be beneficial for their hosts, only to become detrimental under certain circumstances. Such is the case of *Streptococcus aureus*, which becomes pathogenic when the immune system of its human host gets compromised (Schenck et al. 2016). According to the flexible definition of symbiosis, *S. aureus* would be *sensu lato* considered as a human symbiont, but *sensu stricto* a mutualistic or a parasite microorganism according to the restricted definitions mentioned above. Time is another factor that plays a role in symbiosis. It is required that for two organisms being considered symbionts, their relationship must have occurred during a significant evolutionary time (Douglas, 2010).

This highlights the complexity of the concept of symbiosis itself, as it depends not only of the nature of the relationship but also on the common evolutionary history of the organisms involved.

Despite the complexity of the term, it is evident that symbiosis has shaped life as it is actually known. It was through an endosymbiosis process how the ancestors of mitochondria and plastids were kidnapped into a proto-eukaryotic cell (Margulis & Fester, 1991; Margulis, 1970). The origin of mitochondria and plastids, known as symbiogenesis, sets the starting point for the diversification of all eukaryotic life forms that we know today. Eukaryotic organisms continued to stablish symbiotic relationships with prokaryotes, and it is widely accepted that all multicellular eukaryotes live in symbiosis with bacteria (Douglas, 2010). The establishment of symbiotic relationships between eukaryotes and bacteria can be a source of functional innovation (Margulis & Fester, 1991; Moran & Telang, 1998). For example, a great diversity of marine animals thrive in chemosynthetic environments independently of photosynthetic primary production thank to the chemosynthetic bacteria that live associated to them (Dubilier et al. 2008). In this type of mutualistic symbioses, the animal benefits on the chemosynthates delivered by the symbiont, while the symbiont benefits on shelter and protection provided by the host (Duperron et al., 2006). Throughout my study I will focus on a more pernicious type of symbiotic association: Parasitism. The consequences and regulatory mechanisms of that type of association in the host (bathymodiolin mussel) and the symbiont (*Ca.* Endonucleobacter) are the subject of this thesis.

1.2 Parasitism

"Parasites are predators that eat their prey in units of less than one." These are the words that the famous entomologist E. O. Wilson used to defined parasites (Wilson 2014). The word parasite can be decomposed in two parts derived from Greek - para which means "next to" and -sitos which means "food". Evolutionary ecology considers parasitism as a symbiotic relationship in which one organism (the parasite) thrives on another organism (the host) while causing a variable degree of harm, and being functionally adapted to this lifestyle (Poulin 2007). Parasites spread through the branches of the three of life, ranging from protozoans (e.g. *Plasmodium spp.,* causative agent of malaria), animals (e.g. *Cymothoa exigua*, an isopod that replaces the tongue of its fish host), plants (e.g. *Rafflesia* spp., non-photosynthetic plants that parasite vines from the genus *Tetrastigma*), fungi (e.g. *Armillaria mellea*, which are phytopathogens of trees) or even viruses (e.g. HIV) (**Figure 1**). Bacteria are not an exception, as many adopt parasitic lifestyle to replicate at the expenses of their hosts (e.g. *Ca.* Endonucleobacter, an intranuclear parasite of deep-sea mussels) (**Fig.1**) (Brusca, 2008; Coetzee et al., 2018; Deeks et al., 2015; Hidayati & Walck, 2016; Phillips et al., 2017; Zielinski et al., 2009). Parasitic bacteria have been traditionally referred as pathogens in medical microbiology, mainly because many bacterial parasites can produce disease in the process of replicating and overcoming host defenses. The line that separates pathogenesis from parasitism is blurred, and the subtle differences between both terms are opened to discussion.

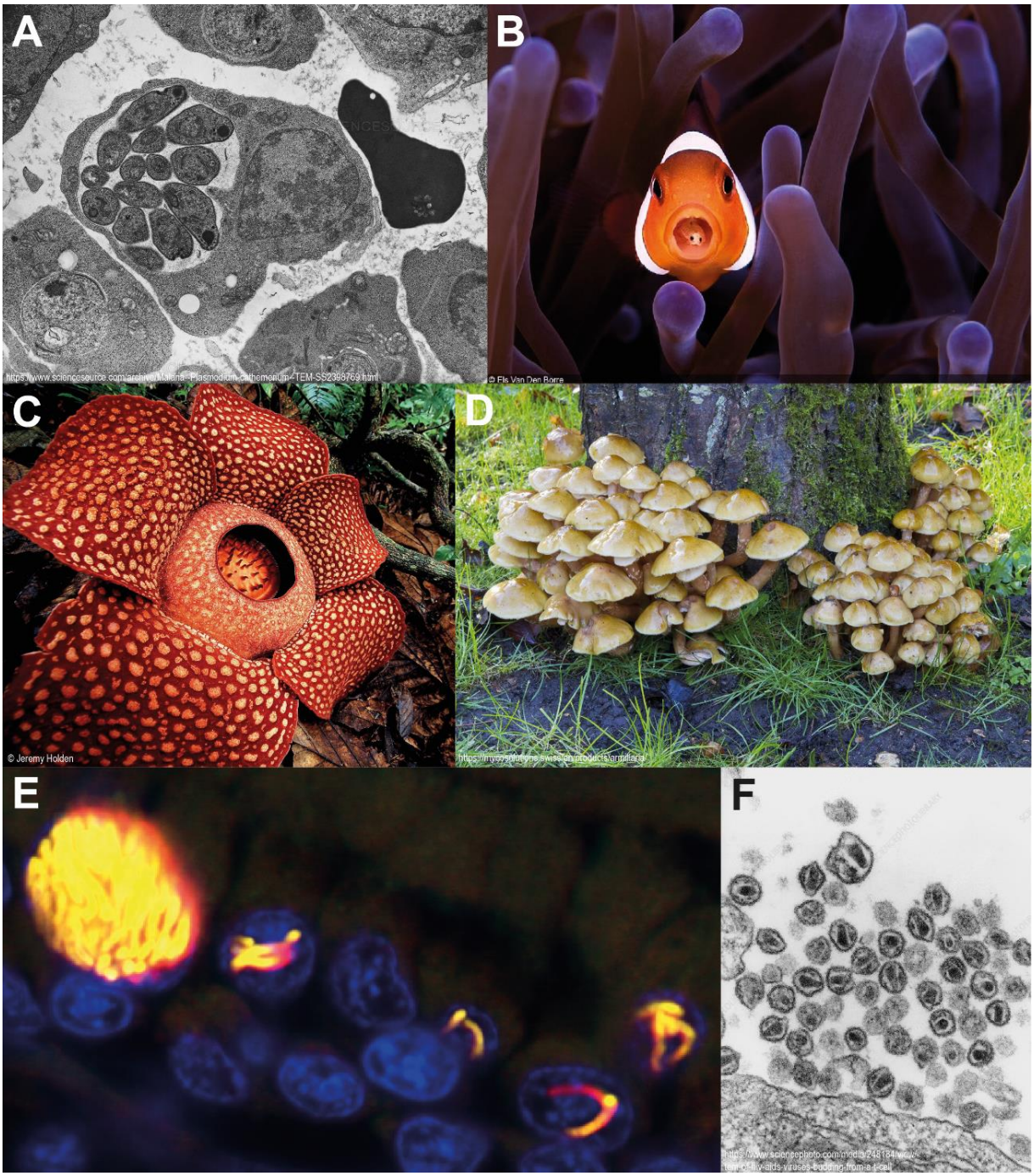


Figure 1. Parasitic lifestyle spreads across the branches of the three of life. A, The protozoan *Plasmodium cathemerium* infecting a bird erythrocyte. **B,** The parasitic copepod *Cymothoa exigua* replacing the tongue of a clown fish (*Amphiprion ocellaris*). **C,** *Rafflesia arnoldii,* a non-photosynthetic plant that parasites vines from the genus *Tetrastigma* to thrive in their photosynthates. **D,** The fungal pythophatogen *Armillaria mellea* infecting a tree. **E,** different infection stages of *Ca.* Endonucleobacter, an intranuclear bacterial parasite of bathymodiolin mussels. **F**, Human immunodeficiency viral particles budding from a T-type lymphocyte.

1.2.1 Deleterious bacteria: Pathogens or parasites?

In the early decades of the 20th century, the American microbiologist Hans Zinsser classified bacteria into three discrete categories: “Pure saprophytes” (unable to grow on host living tissues), “pure parasites” (able to rapidly overcome host defenses and replicate within) and “half parasites” (with low and conditional invasive power) (Zinsser, 1924). Things are not different today in medical microbiology, as most authorities still dividing bacteria into “pathogenic” and “non-pathogenic” bacteria (Casadevall 2012). The equivalent usage of the terms “pathogen” and “parasite” for deleterious bacteria might be the result of a narrow point of view of symbiotic relationships. The fact that both pathogens and parasites exert harm to their hosts for their own benefit leads the scientific community to use both terms indiscriminately. However, it is possible that the relationship between a bacterium and its host turns out pathogenic without meeting the temporal and functional adaptations requirements of a symbiotic partnership (Douglas, 2010; Poulin, 2007). E.g. *Vibrio cassostreae* is a benign colonizer of oysters that turns secondarily into a pathogen after acquisition of a virulence plasmid (Bruto et al. 2017). In this particular case, *V. cassostreae* could be consider a pathogen due to the harm it causes to its host, but not a parasite due to the short duration of the pathogenic relationship in evolutionary terms. The opposite case scenario occurs when over a long evolutionary time, a parasite modulates its virulence until the point that can profit on its host without causing any significant decrease in its fitness (Frank and Schmid-Hempel 2008). E.g. *Ca.* Nucleicultrix amoebiphila (*Alphaproteobacteria*) is a bacterium that invades the nucleus of *Hartmannella* sp., which is vertically transmitted during host cell division (Schulz et al., 2014). The authors discovered that infection by *Ca.* Nucleicultrix amoebiphila did not have any significant effect in the growth rate of *Hartmannella* when kept in culture. *Hartmannella* cells were rarely lysed by *Ca.* Nucleicultrix amoebiphila.

Still, this suggested that *Ca.* Nucleicultrix amoebiphila can be eventually deleterious for *Hartmannella*. *Ca.* Nucleicultrix amoebiphila could be considered a parasite of *Hartmannella* because it was potentially deleterious at an individual host scale. However, it is unclear if it can be considered a pathogen as it did not decrease *Hartmannella* fitness at a population scale.

For the sake of simplicity, I will refer to pathogenic bacteria as parasites of their hosts. In the course of evolution, parasitic bacteria have developed a myriad of strategies to thrive on their hosts. One of these strategies is the colonization of host cells, from which intracellular parasites can profit in terms of nutrients and shelter (Casadevall 2008).

1.2.2 Intracellular parasites

Parasitic bacteria from diverse phyla have adopted intracellular lifestyle using different strategies. Intracellular lifestyle presents the advantages of accessing host cell metabolites and being protected from the host immune system. The invasion of the host cell also opens the opportunity for the bacterium to manipulate eukaryotic processes by delivering factors called nucleomodulins that interfere with nuclear processes such as transcription (Bierne and Cossart 2012). Despite all its potential advantages, the evolutionary road towards intracellular lifestyle is usually a journey of niche specialization with no return: Genome reduction and loss of metabolic autonomy are common phenomena in bacteria that become highly dependent on their hosts (Casadevall 2008).

Intracellular parasites show different strategies to invade the host cell, stablishing the intracellular replication niche, interfering with eukaryotic processes and egressing the host cell (Friedrich et al. 2012). Although each bacterial lineage has developed unique

pathogenic strategies to thrive intracellularly, it is possible to identify certain generalities that define the virulence and pathogenicity of intracellular parasites.

Cell invasion is inevitably the starting point of intracellular parasitism. Intracellular parasites display a myriad of strategies for cell colonization, which vary according to whether the target cell is naturally phagocytic or not. Phagocytic cells can naturally internalize intracellular parasites, which is considered a passive invasion strategy. Such is the case of *Legionella pneumophila*, which is naturally phagocyted by macrophages. Active invasion strategies of non-phagocytic cells usually imply manipulation of the cortical actin cytoskeleton, altering the properties of the plasma membrane in the process (Dramsi and Cossart 1998). One example of active colonization of non-phagocytic cells is represented by the enteropathogen *Shigella flexneri*. *S. flexneri* induces the formation of ruffles in the plasma membrane of the target enterocyte by injecting *IpgD* through a type 3 secretion system, which results in the uptake of the pathogen (Niebuhr et al. 2000).

Once inside the cell, intracellular parasites are contained in plasma-membrane derived phagosomes. Intracellular parasites display diverse mechanisms to survive phagosome-lysosome fusion and consequent acidification of the digestive vacuole (Cain and Vázquez-Boland 2015). For example, once in the plasma-membrane derived phagosome, *L. pneumophila* uses a Dot/Icm type 4 secretion system to send a signal that inhibits phagosome-lysosome fusion (Roy and Tilney 2002). An alternative strategy is shown by the enteropathogen *Salmonella enterica*, which adapts to the acidified digestive vacuole by lowering its own cytoplasmic pH (Chakraborty et al., 2015).

Another survival strategy implies that the intracellular parasite escapes the phagosome prior to lysosome fusion. E.g. Certain strains of *Mycobacterium tuberculosis* can escape the macrophage phagosome by activating a host cytoplasmic phospholipase A_2, stablishing residence in the host cytoplasm (Jamwal et al. 2016). Intracellular parasites that manage to escape the phagosome are freed into the host cell cytoplasm, which can serve the bacteria as replication niche. A common strategy of cytoplasmic intracellular parasites such as *Shigella*, *Listeria* and *Rickettsias* is to capitalize on the host cell cytoskeleton to move across the cytoplasm. Movement is linked to a polarized actin polymerization event, with the creation of actin tails (Dramsi and Cossart 1998). Actin polymerization propels the bacteria inside the cytoplasm until physical contact with the plasma membrane is made. Then, the actin-based machinery pushes through cell protrusions that help the bacteria to aggressively spread to adjacent cells (Dramsi and Cossart 1998).

Actin-based movement can also aid intracellular parasites to navigate the host cytoplasm to reach subcellular compartments. Certain intracellular bacteria can colonize organelles such as mitochondria (*Ca.* Midichloria mitochondrii) or the nucleus (*Ca.* Endonucleobacter bathymodioli) (Sassera et al. 2006; Zielinski et al. 2009). Added to the potential advantages of intracellular lifestyle, the colonization of the nucleus allows bacteria to access a nutrient-rich subcellular compartment, scape from cytoplasmic defense mechanisms or the possibility to directly interact with host DNA (Schulz & Horn, 2015). Although many of them have been visually characterized, the molecular biology of intranuclear bacteria remains elusive. Little is known about how intranuclear bacteria colonize their host cell, how they thrive in the nucleus and how their host cell react to nuclear invasion. In this thesis, I unveiled the molecular principles

of intranuclear lifestyle in bacteria by studying the molecular interaction between *Ca.* Endonucleobacter and its host cell.

1.3 Intranuclear lifestyle

The vast majority of intracellular bacteria occur in the host cell cytoplasm or in host-derived vacuoles (Ray et al. 2009). Rarely, intracellular bacteria are found within cellular organelles, such as the endoplasmic reticulum, the Golgi apparatus, plastids, mitochondria or the nucleus (Bierne and Cossart 2012; Cho et al. 2011; Gruber-Vodicka et al. 2019; Sassera et al. 2006; Schmid 2003; Vogt 1992; Wilcox 1986). The latter bacteria are referred as intranuclear, and they represent the largest group of bacteria occupying especific cellular organelles (Schulz & Horn, 2015). Since the discovery of “enigmatic particles” in the nuclei of paramecia in the 19th century, the number of described intranuclear bacteria has grown: They can be found in nuclei of protists, insects, marine invertebrates and mammals (Schulz & Horn, 2015).

Before intranuclear bacteria can exploit the host cell nucleus as replication niche, they must colonize this subcellular compartment. To do so, intranuclear bacteria must cross the plasma membrane of the host cell, escape or hijack the secretory pathway, confront cytoplasmic defenses, migrate to the nucleus and cross the nuclear envelope (Schulz & Horn, 2015). Greatly differing in nature from the plasma membrane, the nuclear envelope consists of two membranes (outer- and inner-nuclear membrane) that contain the chromatin, a complex of DNA and protein (histones) (Kite 1913). Early ultrastructure studies revealed that the inner nuclear membrane and the outer nuclear membrane are continuous with the endoplasmic reticulum (Watson 1955). The nuclear envelope is the barrier that separates the nuclear and cytoplasmic media, but it needs of channels for bidirectional trafficking known as the nuclear pore complexes. To

penetrate this double barrier, intranuclear bacteria must apply sophisticated colonization strategies (**Figure 2**). For example, *Ca.* Nucleicultrix amoebiphila is thought to escape the phagolysosome prior acidification. After being freed in the host cell cytoplasm, the bacterium presumably invades the nucleus of its amoeba host during the disorganization of the nuclear envelope that characterizes open mitosis (Schulz et al., 2014) (**Fig.2, C**). *Holospora* spp. can escape the acidified phagosome, being freed in the cytoplasm (Iwatani et al. 2005; Sabaneyeva et al. 2009). There, it presumably uses its “invasion tip” to mobilize through the host cell cytoplasm through a polarized actin polymerization phenomenon (Sabaneyeva et al. 2009). Moreover, *Holospora* spp. uses the “invasion tip” to penetrate the nuclear envelope (Fujishima and Kodama 2012) (**Fig.2, A**). Intranuclear bacteria in *Euglena* show an alternative strategy. After surviving the phagosome-lysosome fusion, the bacterium-containing vacuole is translocated to the nucleus, where it fuses with the outer nuclear membrane. The bacterium is freed in the perinuclear space, and presumably enters the nucleus by invagination of the inner nuclear membrane, shedding the host membrane in the process (Shin et al. 2003) (**Fig.2, B**).

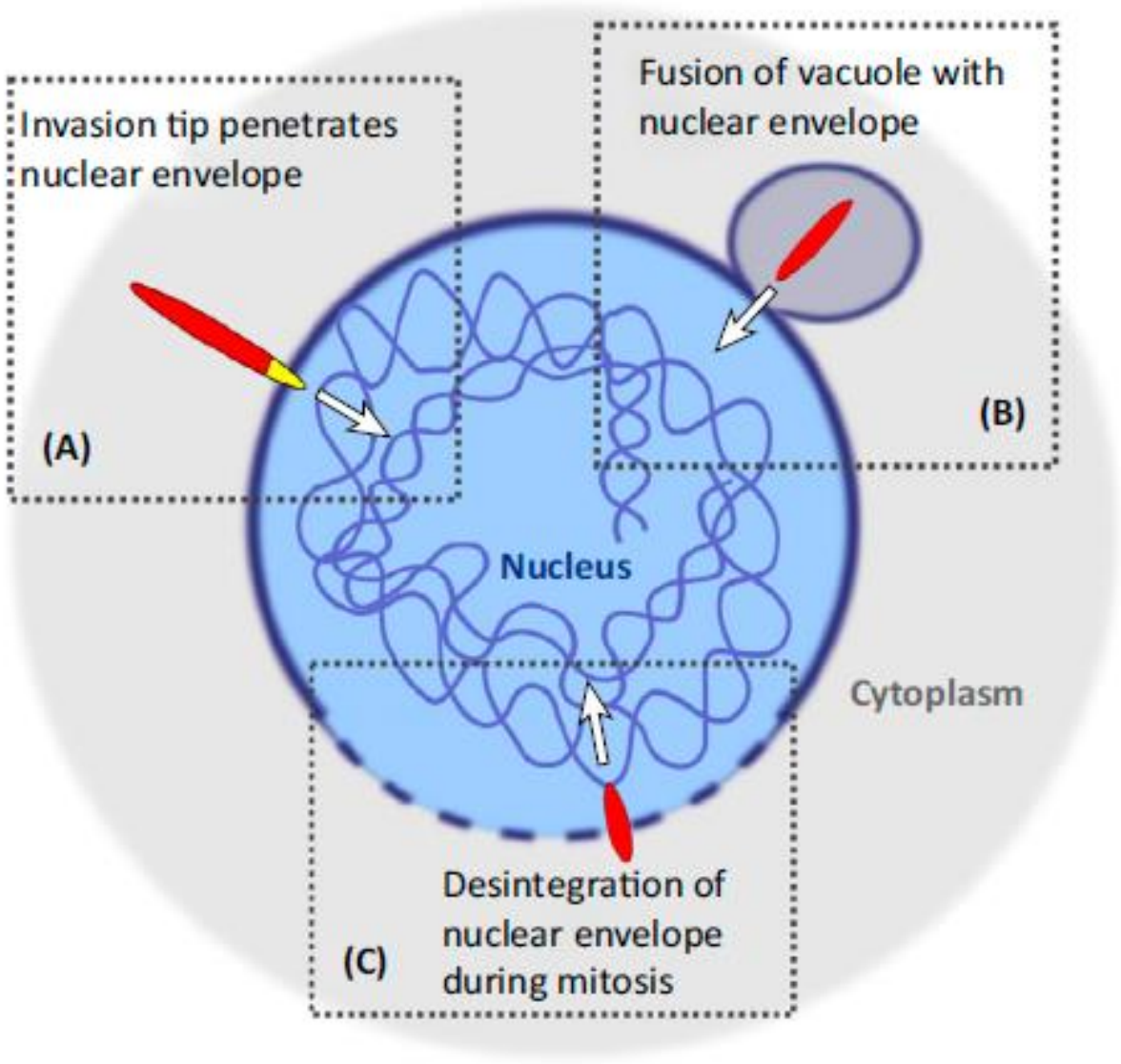


Figure 2. "**Invading the nucleus.** Different strategies of how intranuclear bacteria enter the nucleus have been proposed. (A) *Holospora* spp. employ a specialized macromolecular structure known as the invasion tip. This structure includes an 89 kDa protein with actin-binding activity, which is secreted 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) *Ca.* Nucleicultrix amoebiphila may use the disintegration of the nuclear envelope taking place during open mitosis to permit invasion of the nucleus." Taken from Schulz & Horn, 2015.

The nucleus contains most of the genetic material of the eukaryotic cell, being responsible for directing its functions and metabolism. Normal functioning of the nucleus is a perquisite for the viability of the eukaryotic cell. While sitting in the nucleus, bacteria take the risk of interfering with vital nuclear processes and destabilizing the host cell. Why a bacterium would then chose the nucleus as replication niche? Bacteria from phylogenetically distant phyla (*Verrucomicrobia*, *Chlamydiales*, *Proteobacteria*) have developed intranuclear lifestyle in a process of convergent evolution, suggesting that nuclear colonization might confer selective advantages (Schulz & Horn, 2015).

The advantages of intranuclear lifestyle are diverse. The nucleus can offer shelter against cytoplasmic defenses (e.g. autophagy) (Ray et al. 2009). Intranuclear lifestyle also allows direct interaction with host chromatin by delivering effectors (nucleomodulins) that interfere with nuclear processes (Bierne and Cossart 2012). The nucleus has been considered as a nutrient-rich replication niche: histones, small ribonucleotides, nucleic acids and nucleoside triphosphates are available for intracellular bacteria (Schulz and Horn 2015). For example, intranuclear rickettsiae, *Caedibacter caryophilus* and *Holospora* spp. codify for one or more nucleotide transporters, which allow them to pirate the energy budget of the host cell (Haferkamp et al. 2006; Schmitz-esser et al. 2004). A more aggressive use of the host genetic material (like digestion of DNA or histones) would lead to a destabilization of the replication niche. However, studies suggest that some intranuclear bacteria do not affect host cell fitness. E.g. As introduced previously, *Ca.* Nucleicultrix amoebiphila did not have any significant effect in the growth rate of *Hartmannella* amoeba when kept in culture (Schulz et al., 2014) (**see section 1.2.1**). A strategy to avoid DNA damage would be to digest other nucleic acids (such as mRNAs or intronic RNAs) or even histones. Zielinski et al., 2009 observed that the amount of heterochromatin in nuclei

infected by *Ca.* Endonucleobacter bathymodioli decreased towards the onset of the parasite replication. This suggested that *Ca.* Endonucleobacter bathymodioli might be using chromatin as a nutritional source. However, this would have major repercussions on host genetic expression and regulation, which would eventually also lead to a destabilization of the replication niche (Schulz & Horn, 2015). How *Ca.* Endonucleobacter can complete its intricate lifecycle while the host cell remains functional is one of the study subjects of this thesis.

1.4 *Ca.* Endonucleobacter

Between June 1983 and January 1984, a mass mortality event lead to a decrease of 95% of the population of razor clams (*Siliqua patula*) along coastal western North America within Washington State (Elston 1986). The authors described a *Rickettsia*-like microorganism that infected the nuclei of non-ciliated branchial cells of *S. patula* as the causative agent of the mass mortality. This intranuclear parasite was named NIX, standing for nuclear inclusion X, and it was the first reported bacterium infecting the nuclei of molluscs. After this discovery, more than 20 years passed until another intranuclear bacterium occurring in marine invertebrates was described as *Ca*. Endonucleobacter bathymodioli (Zielinski et al., 2009).

Ca. E. bathymodioli (*Oceanospirillales*, *Hahellaceae*) is a bacterial parasite that infects the nuclei of bathymodiolin mussels. During my doctoral studies, I visualized the lifecycle of *Ca.* Endonucleobacter using super resolution microscopy (**Figure 3**). After a single rod-shaped *Ca.* E. bathymodioli cell invades the host nucleus (**Fig. 3, A**), it elongates forming a coiled filament (**Fig.3, B**) and proliferates occupying the nuclear compartment (**Fig.3, C**). Continuous proliferation induces subtle nuclear deformation (**Fig.3, D**). Infected nuclei become dramatically deformed by the end of the infectious cycle (**Fig.3, E**), hosting up to 80,000 bacteria and having a volume up to 50 fold its

original size. As shown in **Fig.3**, *Ca.* E. bathymodioli presents a pronounced developmental cycle with morphologically distinct stages. However, little is known about how does infect the host cell, how does it colonize the nucleus, how does it thrive nutritionally within the nucleus, how does prevent the host cell from collapsing and how the host cell reacts to the infection.

Zielinski et al., 2009 observed a progressive decrease of heterochromatin in nuclei infected by *Ca.* E. bathymodioli towards the onset of the parasite replication. Consequently, the authors hypothesized that *Ca.* E. bathymodioli might be consuming host DNA. Among the close relatives of *Ca.* E. bathymodioli that have been cultured is *Endozoicomonas ascidiicola*, an extracellular symbiont of sea squirts (Schreiber et al., 2016a). The authors of Schreiber et al., 2016b observed that *E. ascidiicola* produced extracellular DNAse. However, *E. ascidiicola* was unable to digest high-molecular weight DNA, indicating an alternative type of interaction with its sea squirt host rather than chromatin consumption. The inability of *E. ascidiicola* to thrive in DNA bring questions about the chromatin-feeding hypothesis for *Ca.* E. bathymodioli. One alternative to this hypothesis is that Endonucleobacter might be accessing cytosolic resources, as nuclear enlargement along the infectious cycle might promote the transit of molecules between the cytosol and the nucleus through the nuclear pore complexes (Zielinski et al. 2009).

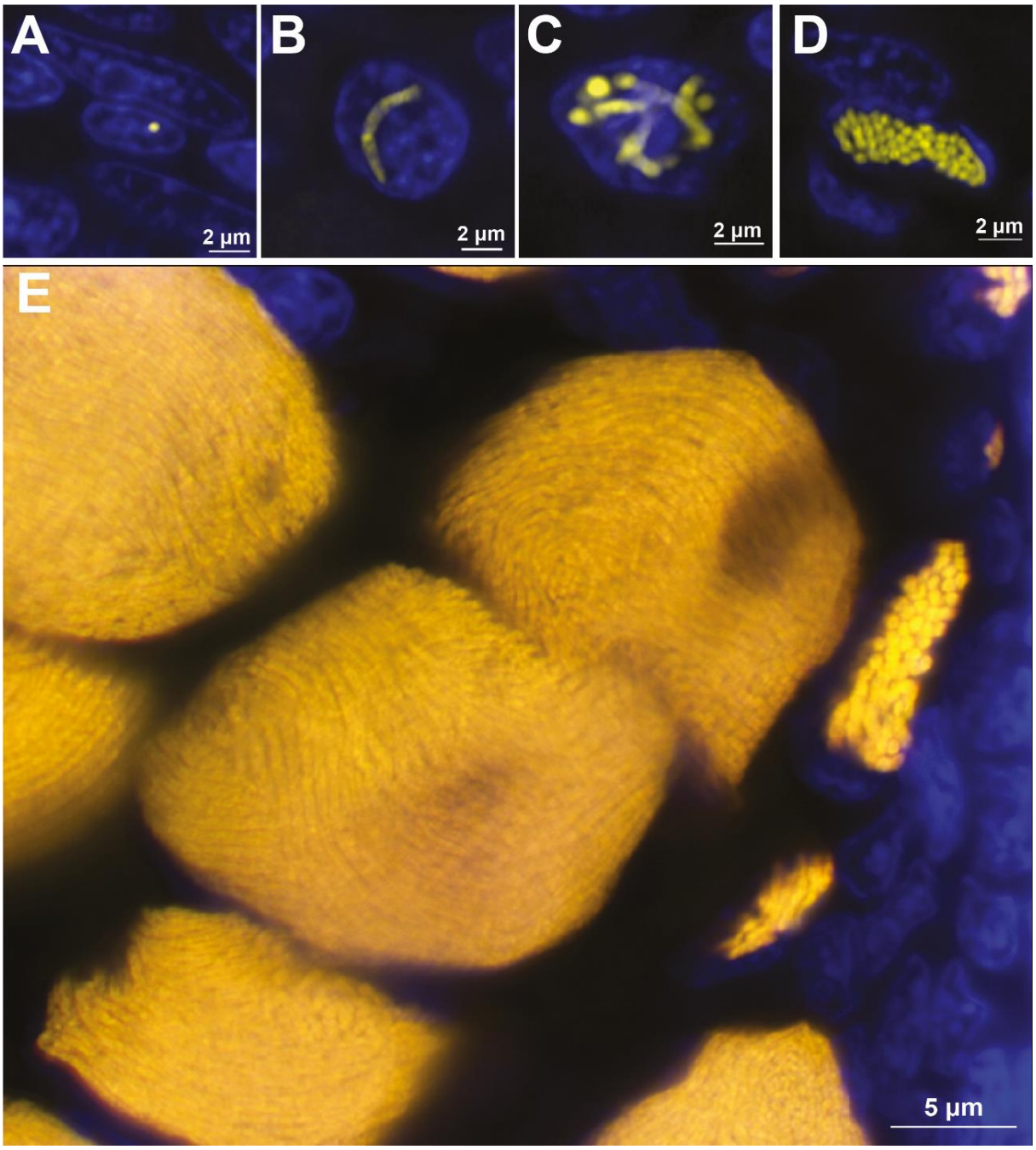


Figure 3. *Ca.* Endonucleobacter infecting *"B." childressi* has clearly defined life stages, as *Ca.* E. bathymodioli (Zielinski et al. 2009). Whole-mount Fluorescence in situ hybridization (Whole-mount FISH) on gill filaments of "*Bathymodiolus" childressi* gill filaments. Airyscan images of different life stages of *Ca.* Endonucleobacter. *Ca.* Endonucleobacter 16S rRNA (yellow: atto550), Host nuclei (blue: DAPI). **A,** A single rod-shaped bacterium colonizes the nucleus. **B,** The bacterium elongates until it adopts the shape of a coiled filament **C,** After several cycles of division, *Ca.* Endonucleobacter occupies the nuclear space. **D,** Progression of the lifecycle induces a subtle nuclear deformation. **E,** A nucleus in the last stage of infection increases its volume up to 50 fold its original size and can host up to 80,000 bacteria.

1.4.1 Bathymodiolin mussels: The host of *Ca.* Endonucleobacter

Deep-sea mussels thrive in chemosynthetic environments, independent of the photosynthetic primary production that takes place in the photic zone. These mussels depend on their chemosynthetic mutualistic symbionts for their nutrition, and although they are considered part of the *Mytilidae* family, they also form a monophyletic subfamily called *Bathymodiolinae*. Bathymodiolin mussels dominate in terms of biomass hydrothermal vents and cold seeps distributed worldwide forming mussel beds (**Figure 4, A**), but they have been also observed in whale falls and sunken wood transitory ecosystems (Distel et al., 2000; Duperron et al., 2013; Van Dover et al., 2002). Bathymodiolin mussels are thought to have evolved approximately 89 million years ago from their shallow water relatives using spatial stepping-stones such as whale falls and sunken wood (Distel et al. 2000). Bathymodiolin mussels are an evolutionary innovation within the *Mytilidae* family and they have undergone rapid divergence (Liu et al. 2018; Lorion et al. 2013). The most recent analyses have suggested nine genera forming the subfamily Bathymodiolinae: *Bathymodiolus*, *Benthomodiolus*, *Vulcanidas*, *Lignomodiolus*, *Idas*, *Terua*, *"Bathymdiolus"*, *Gigantidas* and *Nypamodiolus* (Liu et al. 2018).

Bathymodiolin mussels have a functional digestive system, buy they rely on their chemosynthetic symbionts for nutrition (Duperron, 2010). The symbiotic organ of bathymodiolin mussels are their gills (**brown structures in Fig.4, B**), which are greatly enlarged in comparison with the gills of their shallow water relatives (Duperron et al., 2016). The gills of bathymodiolin mussels are formed by thousands of gill filaments (**Fig.4, C**), in which two main different areas can be distinguished: A symbiotic region, where the chemosynthetic mutualistic symbionts occur, and the ciliated edge, which is deprived of mutualistic symbionts (Kenk and Wilson 1985). Bathymodiolin mussels live

in symbiosis with one or two types of gammaproteobacterial symbionts: a sulfur-oxidizing (SOX) and/or a methane-oxidizing (MOX) symbiont (**SOX in green; MOX in red; Fig.4, D**). The symbionts provide the host with chemosynthates while the host provides shelter and protection to the symbionts (Duperron et al., 2006). Bathymodiolin mussels symbionts are harbored within specialized gill epithelial cells that occur in the symbiotic region of gill filaments called bacteriocytes (**Fig.4, E**).

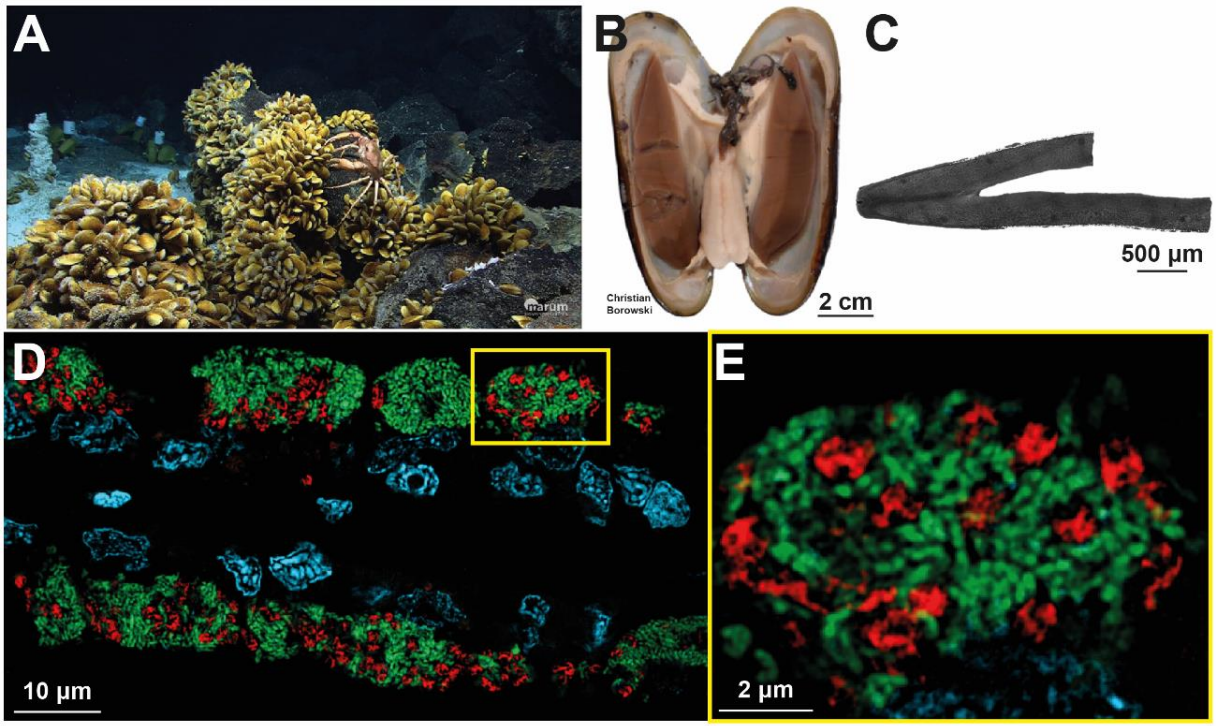


Figure 4. Bathymodiolin mussels dominate in term of biomass hydrothermal vents and cold seeps distributed worldwide. **A,** Mussel bed, courtesy of MARUM. **B,** Opened bathymodiolin mussel, showing greatly enlarged gills in brown. **C,** CLSM image of single gill filament. **D,** Fluorescence in situ hybridization (FISH) on gills' cross sections of *Bathymodiolus puteoserpentis* gill filament, showing bacteriocytes full of chemosynthetic symbionts. Super- resolution structured illumination microscopy image colocalizing MOX 16S rRNA (red: Alexa 647), SOX 16S rRNA (green: atto488) and host nuclei (cyan: DAPI). **E,** enlargement of single bacteriocyte.

The sequencing advances made during the last decade have allowed us to better understand the biology of bathymodiolin symbioses. Through high-resolution metagenomic studies, Ansorge et al. 2019 revealed that up to 16 functional strains of the SOX symbiont can co-occur in a single mussel. However, omics studies made on bulk extractions of DNA and/or RNA lead to loss of biological information. Analyses on DNA extracted from a whole *Bathymodiolus* gill could inform us about the total number of SOX strains, but not how they are spatially organized within the gill filaments. Following the same line of argument, analyses on RNA extracted from a whole *Bathymodiolus* gill infected by *Ca.* Endonucleobacter would not yield information about which transcripts is the parasite expressing along its infectious cycle. During my doctoral studies, I have made methodological advances on microscopy and laser-capture microdissection techniques that have allowed me to spatially and temporally structure omics information of bathymodiolin symbioses. Direct-geneFISH (Barrero-Canosa et al. 2017) is a microscopy technique originally developed on free-living bacteria that co-localizes the 16S rRNA of a bacterium and a gene marker. By optimizing Direct-geneFISH on bathymodiolin symbioses, I could demonstrate that: First, different SOX strains segregate spatially within the gill filaments of bathymodiolin mussels forming patches and second, bacteriocytes seem to be quantitatively dominated by a single SOX strain (**Fig.5**) (**chapter IV**). During my doctoral studies, I also development of a pipeline that coupled FISH, laser-capture microdissection and ultra-low-input RNAseq. This pipeline allowed me to characterize the expression profile of *Ca.* Endonucleobacter and its host cell along the infectious cycle, and to relate the progression of the infectious cycle to other omics datasets (**Fig.6**) (**chapter II**).

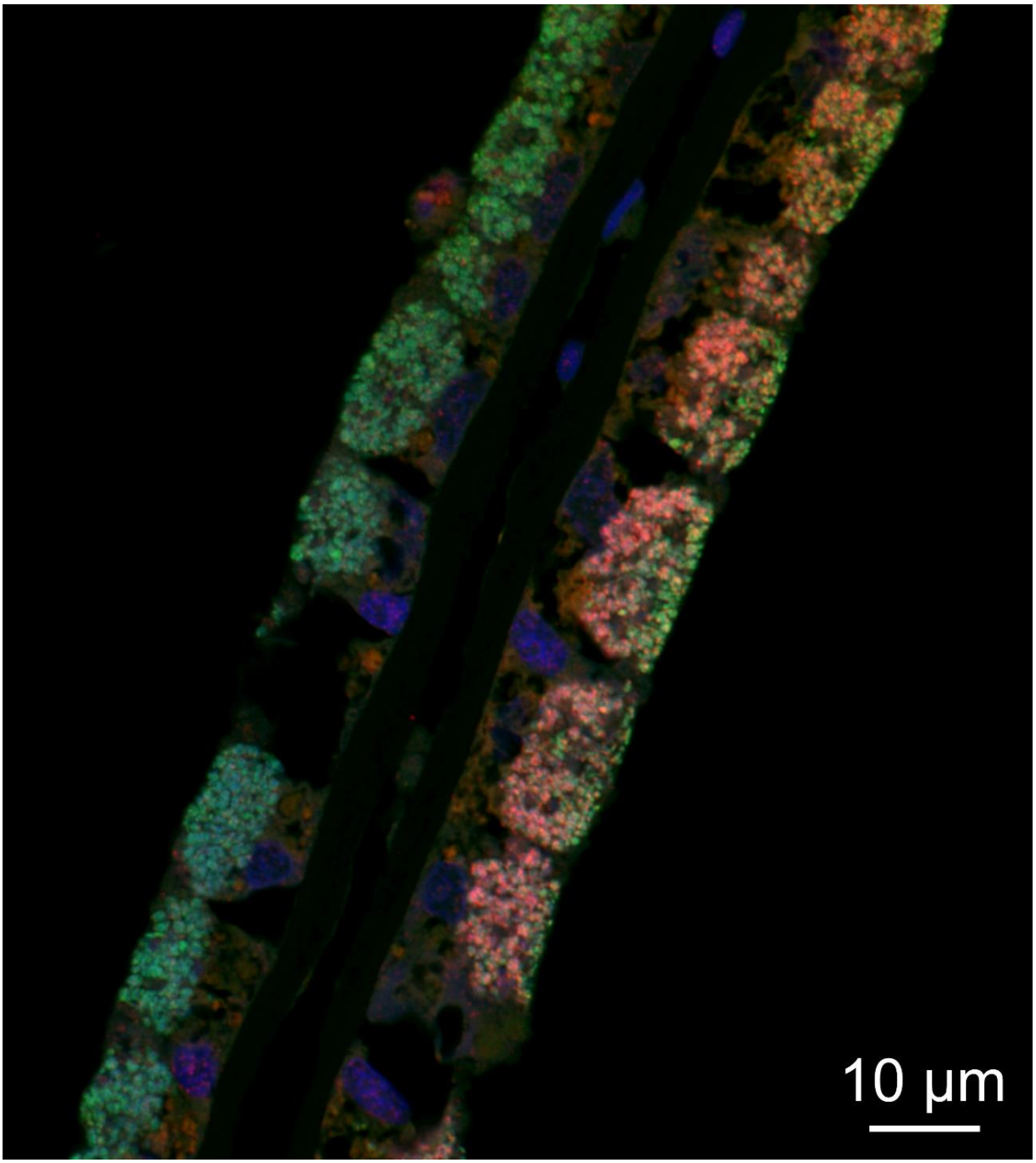


Figure 5. Different SOX strains co-occur in neighboring bacteriocytes forming patches (right side of the filament). Bacteriocytes seem to be quantitatively dominated by a single SOX strain. Direct-geneFISH hybridization on gills' cross sections of *Bathymodiolus azoricus* gill filament. Super-resolution Airyscan microscopy image colocalizing the hydrogenase gene cluster (red: Alexa 647), SOX 16S rRNA (green: atto488) and DNA (blue: DAPI).

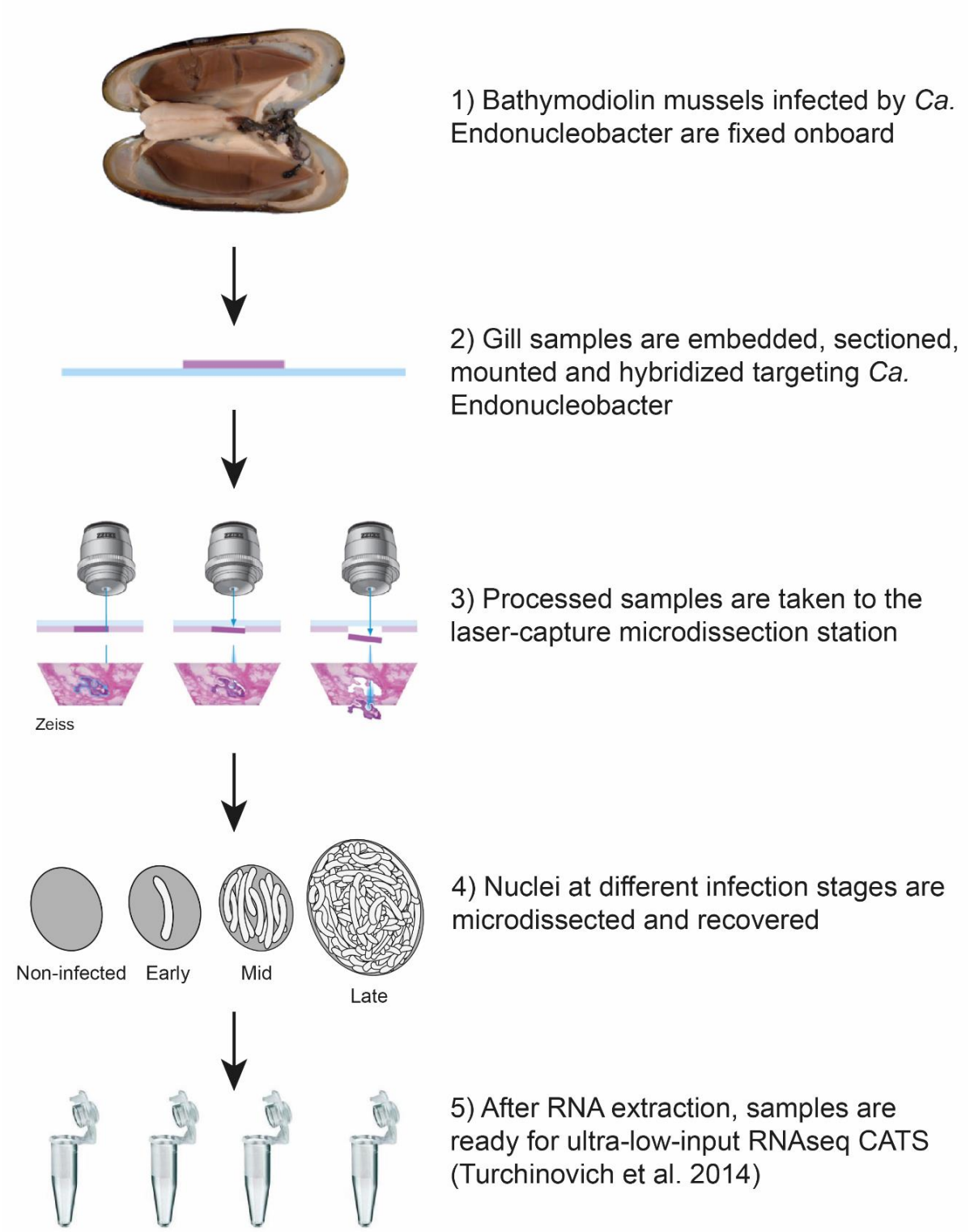


Figure 6. Developed pipeline to profile the transcriptome of *Ca.* Endonucleobacter and its host cell along the infectious cycle. This pipeline couples fluorescence *in situ* hybridization, laser-capture microdissection and ultra-low-input RNAseq CATS (Turchinovich et al. 2014).

The symbiotic region of gill filaments is also formed by intercalary cells, which are aposymbiotic (they lack chemosynthetic symbionts). Zielinski et al., 2009 observed that *Ca.* E. bathymodioli never occurs within bacteriocytes, where the SOX and MOX symbionts occur. Instead, Endonucleobacter infects intercalary cells (**yellow clusters in Figure 7**). Therefore, a mechanism of spatial exclusion between the mutualistic symbionts of bathymodiolin mussels (SOX & MOX) and *Ca.* E. bathymodioli has been proposed (Zielinski et al. 2009). What could prevent *Ca.* E. bathymodioli of infecting symbiont-containing cells? Intracellular pathogens such as the enteropathogenic *Escherichia coli* or *Vibrio parahaemolyticus* trigger an ultrastructural transformation in their host cell known as microvilli effacement (Kaper et al. 2004; Zhou et al. 2014). In all studied *Bathymodiolus* symbioses, bacteriocytes have smooth surfaces, while non-symbiotic intercalary cells have microvilli (Duperron 2010; Fiala-Medioni and Pennec 1987; Fisher et al. 1987). In addition, Franke et al., *in Prep* showed that the mutualists pioneers that colonize a bacteriocyte induce surface modifications that culminate in microvilli effacement, suggesting ultrastructural modifications of the bacteriocyte surface that might impede recognition between posterior colonizers (such as *Ca.* E. bathymodioli) and the plasma membrane of the bacteriocyte.

The genus *Ca.* Endonucleobacter infects bathymodiolin mussels that are distributed worldwide. 16S rRNA sequences from *Ca.* Endonucleobacter have been found in mussel species from the Gulf of México *("B". childressi*, *B. brooksi*, *B. heckerae*), the Mid-Atlantic Ridge (*B. azoricus*, *B. puteoserpentis, B. sp. Wideawake*, *B. sp. Liliput*) and the Pacific-Antarctic Ridge (*B. aff. thermophilus*) (Zielinski et al. 2009). Only two species (*B. aff. boomerang*, *B. brevior*) screened by Zielinski et al., 2009 did not show any *Ca.* Endonucleobacter 16S rRNA sequences. Phylogenetic analyses based on the 16S rRNA sequences extracted from the described samples placed *Ca.* Endonucleobacter as a monophyletic group. Moreover, the *Ca.* Endonucleobacter genus was formed by three subclades reflecting the mussels' sampling sites (Gulf of México, Mid-Atlantic Ridge and Pacific-Antarctic Rise), reflecting the biogeography of their hosts. This could suggest a co-speciation phenomenon between Endonucleobacter and bathymodiolin mussels (Zielinski et al. 2009), implying that *Ca.* E. bathymodioli might not be the only species of the genus.

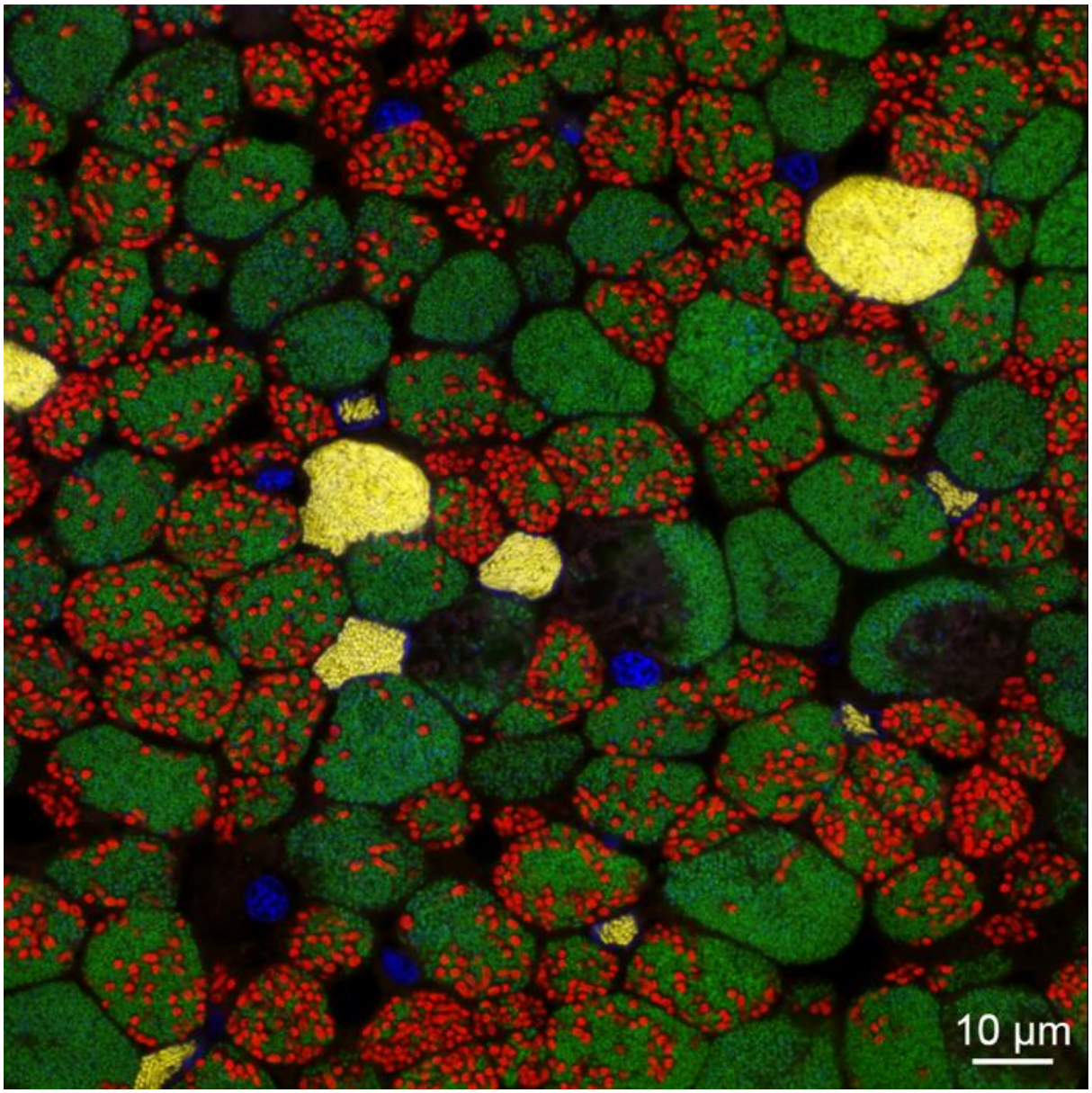


Figure 7. The symbiotic landscape of a gill filament of *B. puteoserpentis.* *Ca.* E. bathymodioli never occurs in bacteriocytes together with the SOX and the MOX symbionts. Whole-mount FISH on gill filament of *B. puteoserpentis*. CSLM image colocalizing MOX 16S rRNA (red: Alexa 647), *Ca.* E. bathymodioli 16S rRNA (yellow: atto550), SOX 16S rRNA (green: atto488) and host nuclei (blue: DAPI).

1.5. *Ca.* Endonucleobacter is the closest relative of *Endozoicomonas*

Phylogenetic studies of the 16S rRNA gene made by Zielinski et al., 2009 placed the genus *Ca.* Endonucleobacter as the sister clade of the genus *Endozoicomonas.* The placement of *Ca.* Endonucleobacter as sister clade of *Endozoicomonas* was also supported by the 16S rRNA phylogenetic analysis made by Schreiber et al. 2016a and Pike et al. 2013. These authors also demonstrated that, together with *Kistimonas* (Choi et al. 2010), the tandem *Ca.* Endonucleobacter and *Endozoicomonas* forms a monophyletic group of bacteria associated with marine metazoa (**orange frame in Figure 8**).

Endozoicomonas is a genus of marine *Gammaproteobacteria* within the family *Hahellaceae*. This genus was firstly described in 2007 after the isolation and physiological characterization of *Endozoicomonas elysicola*, symbiont of the marine slug *Elysia ornata* (Kurahashi and Yokota 2007). The word Endozoicomonas means “monad living inside an animal”. Honoring its name, *Endozoicomonas* establish symbiotic relationships with a huge variety of marine metazoans distributed worldwide, including cnidarians, poriferans, molluscs, annelids, tunicates and fish (Fiore et al. 2015; Forget and Kim Juniper 2013; Jensen et al. 2010; Katharios et al. 2015; Morrow et al. 2012).

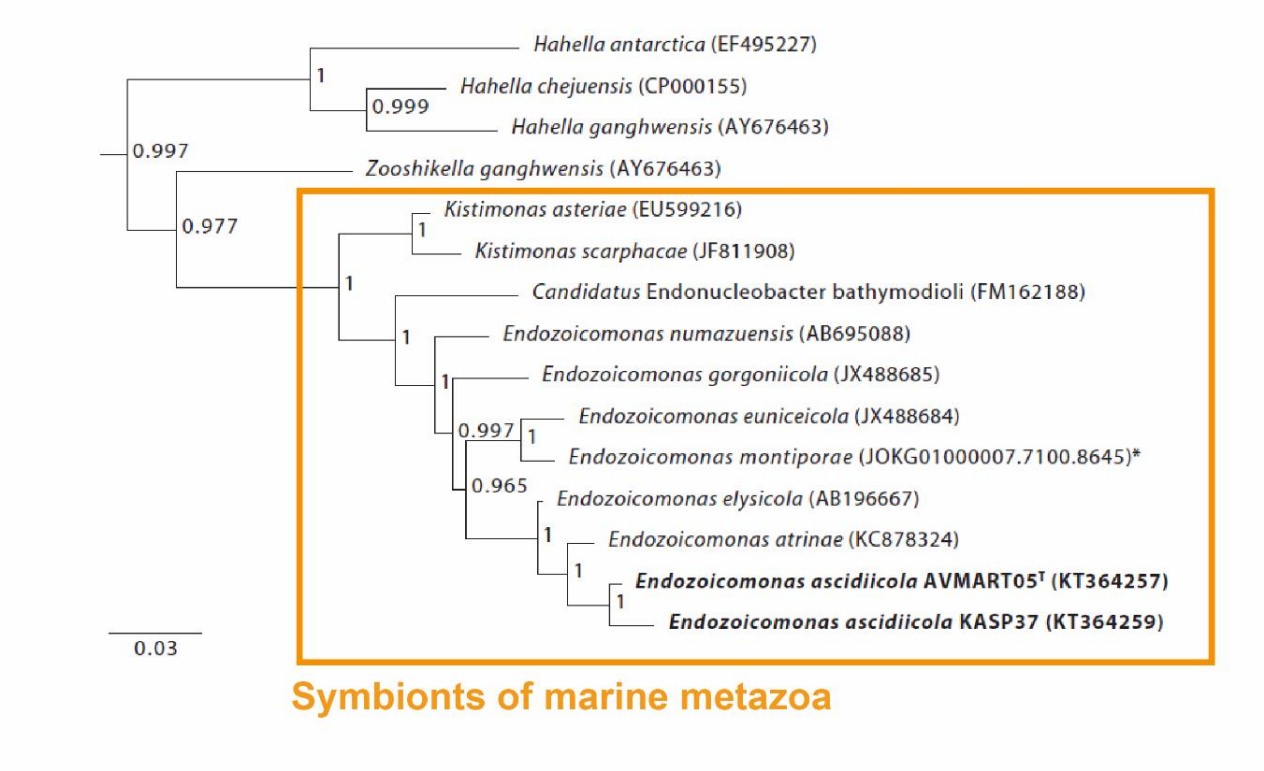


Figure 8. "Phylogeny of the Hahellaceae based on the 16S rRNA gene and Bayesian inference. Node support is expressed as posterior probabilities. Sequences of *E. ascidiicola* are shown in bold. Accession numbers are shown in parentheses. For *E. montiporae* (*) the sequence identifier of the genomic contig and the sequence coordinates of the 16S rRNA gene are shown in parentheses. Sequences of the genus *Marinobacter* were used as outgroup (not shown). Scale bar: 3% estimated sequence divergence." The orange frame indicates the monophyletic group of symbionts of marine metazoans, formed by the genera *Kistimonas*, *Endozoicomonas* and *Ca.* Endonucleobacter. Taken and modified from Schreiber et al. 2016a.

Due to its worldwide distribution and the variety of hosts that can colonize, the recognition of this genus has been progressively increasing over the last decade (Neave et al., 2016). Although *Endozoicomonas* can colonize a huge variety of hosts, they are frequently detected as symbionts of scleractinian corals *(Anthozoa, Hexacorallia, Scleractinia*), which are the primary engineers of coral reefs ecosystems. Coral reefs ecosystems are distributed worldwide (Neave et al., 2017b), including the Great Barrier Reef in Australia (Bourne et al., 2007; Lema et al., 2014), Papua New Guinea (Morrow et al. 2015), Indonesia and the Pacific (Neave et al., 2017b; Yang et al., 2010), the Red Sea (Bayer et al., 2013; Jessen et al., 2013; Neave et al., 2017b; Ziegler et al., 2016), Indian Ocean (Neave et al., 2017b) and the Caribbean (Morrow et al. 2012; Rodriguez-Lanetty et al. 2013). Despite coral reefs ecosystems account for less than 0.1% of the seabed surface, they are among the most diverse and valuable ecosystems on earth (Reaka-Kudla 1997). The ecological importance of coral reef ecosystems further explains the increasing attention that the scientific community has dedicated to *Endozoicomonas*.

Endozoicomonas also stablish symbiotic relationships with a huge variety of non-cnidarians marine invertebrates. In terms of adaptability, is worth to mention that *Endozoicomonas* can be associated to deep-sea metazoans occurring in hydrothermal vents, such as the tubeworm *Ridgeia piscesae* or the snail *Alviniconcha* (Beinart et al. 2014; Forget and Juniper 2013). *Endozoicomonas* also associate with poriferans, one of the oldest groups of metazoans. Fiore et al., 2015 not only found Endozoicomonas in the tissues of the giant barrel sponge *Xestospongia muta* (Caribbean), but also evidences of transcriptional activity of the symbionts. Other studies (Esteves et al. 2013; Gardères et al. 2015; Nishijima et al. 2013; Rua et al. 2014) have also found *Endozoicomonas* in sponges from Brazil, Japan and European seas. *Endozoicomonas*

are also associated to ascidians (*Chordata, Tunicata*) (Dishaw et al., 2014; Schreiber et al., 2016a; Schreiber et al., 2016b). Last, although *Endozoicomonas* has been reported occurring mainly in marine invertebrates, their symbiotic relationship with fish maintained in aquaculture has been described in the recent years. Remarkably, *Ca.* Endozoicomonas cretensis is the only known *Endozoicomonas* that shows pathogenicity to its host (Katharios et al. 2015; Qi et al. 2018).

The conclusion that emanates from the cited literature is that *Endozoicomonas* is an extremely flexible symbiont, able to adapt to a wide range of environments, hosts and symbiotic lifestyles. *Endozoicomonas* has the malleability to be a mutualistic symbiont of corals in warm coral reefs, a pathogen of fish in aquaculture facilities or a mutualistic symbiont of snails in cold deep-sea hydrothermal vents.

1.5.1 Genomic features of *Endozoicomonas*

Although *Endozoicomonas* present a mutualistic lifestyle, they usually have relatively large genomes (> 5 Mb), suggesting that they are not undergoing genome reduction (Neave et al. 2017a). One explanation to the big sizes of their genomes is that *Endozoicomonas* might have a free-living stage when dispersing between hosts. This would require the metabolic autonomy that gives a complete genomic repertoire (Neave et al. 2017a). The fact that *Endozoicomonas* are not dependent on their hosts for acquisition of metabolites suggested that they are facultative symbionts (Neave et al. 2017a).

Endozoicomonas genomes are enriched in mobile elements, which is one of the defining features of the genus (Katharios et al. 2015; Neave et al. 2017a; Qi et al. 2018). Transposition and DNA recombination are processes that might allow *Endozoicomonas* to rapidly evolve to adopt new hosts or to switch between symbiotic lifestyles (mutualistic, commensalistic or parasitic) (Neave et al. 2017a). The early stage in the transition of a free-living bacterium to a host-associated lifestyle can be associated to expansion of transposases, especially insertion sequences (Siguier et al. 2014).

Abundance of eukaryotic-like proteins such as ankyrin repeats proteins (ANKs) is another common genomic feature to the members of the genus *Endozoicomonas* (Alex and Antunes 2019). ANKs are one of the most abundant protein-protein interaction motifs in nature (Bork 1993), and symbiotic bacteria use them to modulate eukaryotic processes (Gomez-Valero et al. 2011; Nguyen et al., 2014). One possibility is that *Endozoicomonas* is actively modulating eukaryotic processes by secreting ANKs,

regardless of the type of symbiotic interaction that *Endozoicomonas* maintains with its host (mutualism, commensalism or parasitism) (Alex and Antunes 2019).

Overall, *Endozoicomonas* are facultative symbionts with metabolic autonomy. This has been hypothesized based on their relatively big genome sizes (> 5 Mb), implying that *Endozoicomonas* has a free-living stage. The flexibility of *Endozoicomonas* in terms of host, habitat and symbiotic lifestyle range emanates from its malleable genome. Hundreds of mobile elements might have played a role in their rapid evolution, opening new mutualistic and pathogenic niches in a myriad of hosts and habitats. Last, *Endozoicomonas* genomes are enriched in eukaryotic-like protein such as ANKs. *Endozoicomonas* might be using these protein-protein interaction motifs to modulate host processes for its own benefit (pathogenesis) or the benefit of the holobiont (mutualism).

1.5.2 Function and localization of *Endozoicomonas* within their host

Endozoicomonas might play a significant role in nutrient acquisition and provision within their hosts (Neave et al., 2016). Neave et al., 2017a found that Endozoicomonas genomes were enriched in genes from the phosphoenolpyruvate-dependent sugar phosphotransferase (PTS) system. PTS systems are sensitive to the nutritional demands of the cell, and they participate in the phosphorylation and uptake of sugars (Deutscher et al., 2006). This suggested that *Endozoicomonas* might play a role in carbon cycling within their hosts (Neave et al., 2017a; Neave et al., 2016). Interestingly, Neave et al., 2017a found that in Endozoicomonas, PTS systems codified for specific cellobiose uptake subunits. Cellobiose is a component of cellulose, which forms part of the extracellular matrix (cell wall) of plant cells, including algal cells (Brady et al. 2015). According to Neave et al., 2017a, this might indicate that *Endozoicomonas* could play a role recycling death or unwanted *Symbiodinium* algae that live in co-symbiosis with *Endozoicomonas* and scleractinian coral hosts. *Endozoicomonas* might also play a role in nitrogen cycling within their hosts (Neave et al., 2017a). Neave et al., 2017a showed that *Endozoicomonas* genomes were enriched in dicarboxylic acids transporters. In legume-*Rhizobium* symbioses, dicarboxylic acids transporters play a role in the interchanging of dicarboxylic acid (produced by the plant) and ammonia (produced by the bacterium) between host and symbiont (Poole and Allaway 2000). A similar case scenario is proposed by Neave et al., 2017a, in which *Endozoicomonas* might be providing ammonia to the coral host in exchange of dicarboxylic acids. Indeed, *E. elysicola*, *E. numazuensis* and *E. montiporae* showed the genetic potential for nitrate and nitrite reduction to ammonia (Neave et al., 2017a). Ammonia can be assimilated by *Endozoicomonas* through the synthesis of glutamine and glutamate (Neave et al., 2017a), which are the only precursors needed by *Buchnera* bacteria to

provide its pea aphid host with all essential amino acids (Hansen and Moran 2011). As *Endozoicomonas* genomes encoded for complete pathways for the synthesis of the amino acids alanine, aspartate, cysteine, glycine, homocysteine, homoserine, leucine, lysine, methionine, serine and threonine, the authors of Neave et al., 2017a hypothesized that *Endozoicomonas* might be using ammonia as the starting building block for essential amino acids for the host. Sulfur cycling has been also proposed as a potential role of *Endozoicomonas* within their hosts (Pike et al. 2013; Ransome et al. 2014). The authors of Bourne et al., 2013 suggested that *Endozoicomonas* symbionts might be relying on the carbon and sulfur contained in the large amounts of dimethylsulfopropionate (DMSP) produced by *Symbiodinium* algae (Bourne et al., 2013; Correa et al., 2013). Supporting the hypothesis that *Endozoicomonas* play a role in the coral sulfur cycle, the authors of Tandon et al., 2020 demonstrated that *E. acroporae* not only posses the genomic potential to metabolize DMSP into dimethylsulfide (DMS), but also to use DMS as a carbon source for growth. In addition to their role in nutrient cycling and essential amino acid delivery to their host, *Endozoicomonas* might be supplying their host with essential micronutrients such as vitamins. Neave et al., 2017a discovered that one Endozoicomonas strain was enriched for the production of riboflavin and folic acid, two B vitamins that might be important for the wellbeing of its host (Agostini et al., 2009).

The role of *Endozoicomonas* as shapers of the host microbiome has also been suggested (Neave et al., 2016). *Endozoicomonas* might be regulating bacterial colonization of the host by secreting bioactive secondary metabolites, or simply occupying the available space in the host tissues that could be otherwise occupied by opportunistic pathogens (Bayer et al., 2013; Jessen et al. 2013; Morrow et al. 2015; Rua et al. 2014). In addition, *Endozoicomonas* tend to be less abundant in corals which

wellbeing has been compromised. Thus, *Endozoicomonas* abundance might be an indicator of the health status of diverse coral species (Bayer et al., 2013; Morrow et al. 2012, 2015; Roder et al. 2015; Ziegler et al. 2016)

The precise localization of *Endozoicomonas* within their hosts' tissues can vary among different Endozoicomonas-host associations. For example, Schreiber et al., 2016a localized *E. ascidiicola* forming extracellular aggregations in the sea squirt *Ascidiella scabra*. *Ca.* Endozoicomonas cretensis also formed massive aggregations in fish (Katharios et al., 2015). However, the authors could not determine whether *Ca.* Endozoicomonas cretensis occurred intra- or extracellularly within its host. Ding et al., 2016 found an N-deglycosylation enzyme and eukaryotic ephrin ligand B2 in the genome of *E. montiporae*. The N-deglycosylation enzyme would allow *E. montiporae* to penetrate through its host's mucus layer, while the eukaryotic ephrin ligand B2 would allow *E. montiporae* to penetrate the host cell by binding to its ephrin receptors. Whether different *Endozoicomonas* localize intra- or extracellularly within their respective hosts' tissues, the formation of massive bacterial aggregates is a common feature of the group. The assembly of Endozoicomonas into bacterial aggregates might have functional implications in the interaction with their host (Neave et al., 2016). For example, *Ca.* Endozoicomonas cretensis aggregates were surrounded by a thin membrane, which might protect the pathogen against host defense mechanisms (Katharios et al. 2015). These bacterial aggregates might also act as factories of amino acid and vitamins synthesis that could benefit the host. Moreover, *Endozoicomonas* aggregates might be formed by complementary strains or even different species that work synergically to passage nutrients to the host (Neave et al., 2016). Last, the occurrence of *Endozoicomonas* in aggregates implies some sort of cell-to-cell communication such as quorum sensing (Waters and Bassler 2005).

Summarizing, *Endozoicomonas* can occur intra- or extracellularly within their hosts' tissues, but always forming bacterial aggregates. These bacterial aggregates might participate in carbon, nitrogen or sulfur cycling within the host, but also as production centers of amino acids and/or vitamins. Last, *Endozoicomonas* can structure the microbiome of its host, preventing the settlement of opportunistic pathogens.

Aims of this thesis

When I started my doctoral studies, little was known about the molecular biology of *Ca.* Endonucleobacter. It has been postulated that intranuclear lifestyle allows bacteria to access a nutrient-rich subcellular compartment. Yet, the utilization of the nucleus as a replication niche has intrinsic risks for intranuclear bacteria, as eukaryotic cells can sense nuclear deformation and react to it by triggering apoptosis. There is a lot of open questions related to the nutrition of intranuclear bacteria, but there is also gaps of knowledge about how intranuclear bacteria colonize and thrive in the nuclear compartment, how they prevent their host cell from undergoing apoptosis and how they alter the host cell biology. Another topic that I addressed during my doctoral studies was how the different strains of SOX symbionts organize spatially within the gill filaments of bathymodiolin mussels. Ikuta et al. 2016 demonstrated that SOX symbionts from the same strain co-occur in neighboring bacteriocytes in the gill filaments of *B. septemdierum* forming patches. However, we did not know if this spatial organization of the SOX symbionts was common to all bathymodiolin symbioses. Moreover, we did not know if more than one SOX strain can co-occur in the same bacteriocyte. Addressing these questions by studying non-cultivable bacteria like *Ca.* Endonucleobacter or the SOX symbionts is challenging. Together with my coauthors, we managed to bypass these challenges by doing genomic, transcriptomic, proteomic and visualization analyses of *Ca.* Endonucleobacter. *Ca.* Endonucleobacter has a characterized infectious cycle with clearly defined stages, which allow me to relate omics data with the progression of the infection. I developed a pipeline that coupled FISH, laser-capture microdissection (LCM) and ultra-low-input RNAseq (**Fig.6**) (**chapter II**). The gills of infected *"B." childressi* were analyzed with fluorescence *in situ* hybridization (FISH) targeting *Ca.* Endonucleobacter and different infection stages

were microdissected and sequenced. Using this pipeline and other culture-independent methods, we disentangled the molecular interactions of an intranuclear bacterium and its host cell (**chapter II**), as well as the evolutionary origin of this enigmatic genus (**chapter III**). During my doctoral studies, I also optimized the Direct-geneFISH protocol (Barrero-Canosa et al. 2017) on bathymodiolin symbioses. Together with my co-authors, I developed polynucleotide probes that targeted the SOX strain-specific markers methanol dehydrogenase- and hydrogenase-gene clusters. By coupling Direct-geneFISH and super-resolution microscopy techniques, I could demonstrate that bacteriocytes are quantitatively dominated by a single SOX strains. Moreover, I could demonstrate that SOX symbionts from the same strain co-occur in neighboring bacteriocytes forming patches in *B. azoricus* (**chapter IV**). This approach allowed me to visualize heterogeneous symbiont populations that so far could only be computationally distinguished.

The methodological advancements developed during my doctoral studies bypass the limitations of addressing biological questions in host-microbiome systems only with bioinformatic approaches. In one hand, my combined FISH – LCM – ultra-low-input RNAseq pipeline closes the gap between the visual, molecular and temporal characterization of a parasite-host interaction. Its potential could be extended to other intracellular bacteria-host cell interactions with immense value in medical microbiology. On the other hand, the optimization of Direct-geneFISH in host-microbiome systems will allow to visualize the spatial structuring of heterogeneous symbiotic populations *in situ* within their hosts. The biological questions that these methodologies allowed me to answer are listed next.

How does *Ca.* Endonucleobacter thrive in the nucleus?

Previous studies suggested that intranuclear bacteria might exploit nutritionally the nucleus (Schulz & Horn, 2015). However, the consumption of chromatin and/or nucleic acids would have major consequences for the viability of the host cell as replication niche. *Ca.* Nucleicultrix amoebiphila did not have a deleterious effect in the fitness of its host (Schulz et al., 2014), suggesting that was not digesting host chromatin and/or nucleic acids. It remains unknown how intranuclear bacteria can thrive nutritionally within the nucleus without disrupting the host cell. Moreover, it is not well understood whether the host cell defends itself against intranuclear colonization. To solve this questions would help to understand the molecular principles of intranuclear lifestyle, which might be common to most of intranuclear bacteria.

Together with my coauthors, I developed a pipeline that involved the microdissection of hybridized nuclei targeting *Ca.* Endonucleobacter at different infection stages using LCM. Collected nuclei were subjected to downstream ultra-low-input RNA-seq, and resulting single-stage transcriptomes were analyzed. We characterized the expression profile of *Ca.* Endonucleobacter and its *"B." childressi* host cell along the infectious cycle. My results suggested that *Ca.* Endonucleobacter infecting *"B." childressi* is thriving on host chitin, which occurs extracellularly. *Ca.* Endonucleobacter is also able to complete its life cycle by manipulating host apoptosis and cytoskeleton components. These findings are covered and discussed in **chapter II**.

How does the host cell react to *Ca.* Endonucleobacter infection?

Intranuclear bacteria are hidden from cytoplasmic defense mechanisms. Yet proliferation in the nucleus exerts mechanical distress to the host cell. Eukaryotic cells have mechanisms to report nuclear deformation (Zhang et al. 2009), which can trigger the caspase-mediated apoptotic cascade (Crawford et al. 2012; Kräter et al. 2018). Therefore, intranuclear bacteria must count on mechanisms to impair host cell apoptosis. In addition, a growing bacterial population within the nucleus has metabolic demands that must be supplied by the host cell, which most likely enters in starvation state. One of the ways in which nutrient-deprived cells can react to starvation is synthesizing lipid droplets, phenomenon that has been observed in cells infected by *Toxoplasma gondii* (Nolan et al. 2017). Little is known about the transcriptional changes that intranuclear bacteria can trigger in their host cell. By understanding the host cell reaction to the infection we might understand the nutritional and pathogenic strategies that intranuclear bacteria are displaying to thrive in the nucleus.

To fill these gaps of knowledge, together with my coauthors I analyzed the host cell fraction of the single-infection-stage transcriptomes described above. My results showed that the *"B." childressi* host cell is transcriptionally active along the infectious cycle, suggesting that *Ca.* Endonucleobacter does not digest chromatin. Moreover, the host cell reacted to *Ca.* Endonucleobacter infection by upregulating sugar import, glycolysis, lipid droplets synthesis and sensors of nuclear deformation. These results are also presented and discussed in **chapter II**.

Which is the evolutionary origin of *Ca.* Endonucleobacter?

It seems parsimonious that the intranuclear lifestyle has appeared secondarily in bacteria that already had intracellular lifestyle (Schulz & Horn, 2015). The genus *Ca.* Endonucleobacter is the sister clade of *Endozoicomonas* (*Oceanospirillales*, *Hahellaceae*), a group that is mainly known to stablish mutualistic relationships with diverse marine metazoans (Neave et al., 2016). One of the genetic features of *Endozoicomonas* is that they are enriched in mobile elements, which allow them to rapidly adapt to new environments, hosts and symbiotic lifestyles (Neave et al., 2017a). However, it is not well understood whether the presence of mobile elements might explain by itself the origin of intranuclear lifestyle. To understand the mechanisms that might have cause the appearance of intranuclear parasitism from a mutualistic group could broaden our scope about transition between symbiotic lifestyles in bacteria.

To shed light into the evolutionary origin of the genus *Ca.* Endonucleobacter, I compared the genomic capabilities of *Ca.* Endonucleobacter vs. *Endozoicomonas*. In addition, I included in the study non-symbiotic representatives of the family *Hahellaceae* (*Hahella* spp.) as negative control. My results suggested that *Ca.* Endonucleobacter genomes are also enriched in mobile elements. In addition, inhibitors of apoptosis were expanded in *Ca.* Endonucleobacter genomes, while almost absent in *Endozoicomonas* representatives. I hypothesize that mobile elements and inhibitors of apoptosis might have originated intranuclear lifestyle in the last common ancestor of *Ca.* Endonucleobacter and *Endozoicomonas*, causing the separation of both genera. These findings and their discussion are covered in **chapter III**.

How do different SOX strains organize spatially within the gill filaments of bathymodiolin mussels?

Bathymodiolin mussels harbor only a few mutualistic symbiont species (SOX and/or MOX). Symbiont diversity in bathymodiolin mussels might have been underestimated, as microbial diversity has been traditionally studied at the 16S rRNA gene level. Recent studies by Ansorge et al. 2019 demonstrated that up to 16 functional strains of the SOX symbiont can coexist in the same mussel. Ikuta et al., 2016 demonstrated that different SOX functional strains cluster separately within the gill filaments of *B. septemdierum*, forming patches. Despite these findings, it is unknown if the spatial exclusion between SOX strains is an extended feature to other bathymodiolin mussels. Moreover, we do not know if the spatial exclusion between SOX strains also occur at a single bacteriocyte level. Answering these questions would help us to understand a previously hidden level of spatial organization of bacterial communities within their hosts.

To tackle these questions, I optimized and applied the Direct-geneFISH protocol (Barrero-Canosa et al. 2017) in host-microbiome systems. This protocol, which was initially developed to work in free-living bacteria, allows the simultaneous localization of a microorganism 16S rRNA gene and a genetic marker of interest. My results confirmed that the formation of patches by SOX symbionts is not an exclusive feature of *B. septemdierum*. However, by simultaneously localizing two genetic markers, I could demonstrate that patches from different strains segregate spatially from each other. Last, by imaging of SOX strains using super resolution microscopy I demonstrated that bacteriocytes are quantitatively dominated by a single SOX strain,

suggesting biological mechanisms that lead to inter-strain exclusion at a single bacteriocyte level. These analyses are presented and discussed in **chapter IV**.

List of publications

1. The hungry nucleus: The nutritional demands of a chitinolytic intranuclear parasite trigger its host cell to upregulate sugar import

Miguel-Ángel González-Porras, Adrien Assié, Målin Tietjen, Marlene Jensen, Manuel Kleiner, Harald Gruber-Vodicka, Nicole Dubilier, Nikolaus Leisch

Manuscript *in preparation*

2. Conquering the nucleus: Inhibitors of apoptosis are the genomic innovation that originated intranuclear lifestyle in *Hahellaceae*

Miguel-Ángel González-Porras, Rebecca Ansorge, Harald Gruber-Vodicka, Nicole Dubilier, Nikolaus Leisch

Manuscript *in preparation*

3. Super-resolution localization of strain-specific markers in sulfur-oxidizing symbionts of *Bathymodiolus azoricus* suggests strain spatial segregation at a single bacteriocyte level

Miguel-Ángel González-Porras, Jimena Barrero-Canosa, Lizbeth Sayavedra, Rebecca Ansorge, Nicole Dubilier, Nikolaus Leisch

Manuscript *in preparation*

Contributed works not included in this thesis

4. Symbiont strain heterogeneity confers metabolic flexibility in deep-sea mussels: Methylotrophy in sulfur-oxidizing symbionts

Lizbeth Sayavedra, **Miguel-Ángel González-Porras**, Chakkiath Paul Antony, Anne-Christin Kreutzmann, Wolfgang Bach, Dolma Michellod, Manuel Kleiner, Jimena Barrero-Canosa, Nikolaus Leisch, Maxim Rubin-Blum, Manuel Liebeke, Jillian M. Petersen, Nicole Dubilier

Manuscript *in preparation*

5. Functional diversity enables multiple symbiont strains to coexist in deep-sea mussels

Rebecca Ansorge, Stefano Romano, Lizbeth Sayavedra, **Miguel-Ángel González-Porras**, Anne Kupczok, Halina E. Tegetmeyer, Nicole Dubilier, Jillian Petersen

Manuscript published on Nature Microbiology

https://doi.org/10.1038/s41564-019-0572-9

6. Correlative 3D anatomy and spatial chemistry in animal-microbe symbioses: developing sample preparation for phase-contrast synchrotron radiation based micro-computed tomography and mass spectrometry imaging

Benedikt Geier, Maximilian Franke, Bernhard Ruthensteiner, **Miguel-Ángel González-Porras**, Alexander Gruhl, Lars Wörmer, Julian Moosmann, Jörg U. Hammel, Nicole Dubilier, Nikolaus Leisch, Manuel Liebeke

Manuscript published on Developments in X-Ray Tomography XII

https://doi.org/10.1117/12.2530652

References for chapter I

Agostini, Sylvain, Yoshimi Suzuki, Beatriz E. Casareto, Yoshikatsu Nanako, Michio Hidaka, and Nesa Badrun. 2009. “Coral Symbiotic Complex: Hypothesis through Vitamin B12 for a New Evaluation.” *Galaxea, Journal of Coral Reef Studies* 11(1):1–11.

Alex, Anoop, and Agostino Antunes. 2019. “Comparative Genomics Reveals Metabolic Specificity of Endozoicomonas Isolated from a Marine Sponge and the Genomic Repertoire for Host-Bacteria Symbioses.” *Microorganisms* 7(12).

Anon. 1924. “Infection and resistance.” *Archives of Internal Medicine* 33(1):156.

Ansorge, Rebecca, Stefano Romano, Lizbeth Sayavedra, Miguel Ángel González Porras, Anne Kupczok, Halina E. Tegetmeyer, Nicole Dubilier, and Jillian Petersen. 2019. “Functional Diversity Enables Multiple Symbiont Strains to Coexist in Deep-Sea Mussels.” *Nature Microbiology* 4(12):2487–97.

Barrero-Canosa, Jimena, Cristina Moraru, Laura Zeugner, Bernhard M. Fuchs, and Rudolf Amann. 2017. “Direct-GeneFISH: A Simplified Protocol for the Simultaneous Detection and Quantification of Genes and rRNA in Microorganisms.” *Environmental Microbiology* 19(1):70–82.

Bary, A. de. 1879. *Die Erscheinung der Symbiose : Vortrag*. Strassburg: Verlag von Karl J. Trübner.

Bayer, Till, Matthew J. Neave, Areej Alsheikh-Hussain, Manuel Aranda, Lauren K. Yum, Tracy Mincer, Konrad Hughen, Amy Apprill, and Christian R. Voolstra. 2013. “The Microbiome of the Red Sea Coral *Stylophora pistillata* Is Dominated by Tissue-Associated Endozoicomonas Bacteria.” *Applied and Environmental Microbiology* 79(15):4759–62.

Beinart, R. A., S. V. Nyholm, N. Dubilier, and P. R. Girguis. 2014. “Intracellular Oceanospirillales Inhabit the Gills of the Hydrothermal Vent Snail *Alviniconcha* with Chemosynthetic, γ-Proteobacterial Symbionts.” *Environmental Microbiology Reports* 6(6):656–64.

Bierne, Hélène, and Pascale Cossart. 2012. “When Bacteria Target the Nucleus: The Emerging Family of Nucleomodulins.” *Cellular Microbiology* 14(5):622–33.

Bork, Peer. 1993. “Hundreds of Ankyrin-like Repeats in Functionally Diverse Proteins: Mobile Modules That Cross Phyla Horizontally?” *Proteins: Structure, Function, and Bioinformatics* 17(4):363–74.

Bourne, David G., Paul G. Dennis, Sven Uthicke, Rochelle M. Soo, Gene W. Tyson, and Nicole Webster. 2013. “Coral Reef Invertebrate Microbiomes Correlate with the Presence of Photosymbionts.” *ISME Journal* 7(7):1452–58.

Bourne, David, Y. Iida, Sven Uthicke, and Carolyn Smith-Keune. 2007. “Changes in Coral-Associated Microbial Communities during a Bleaching Event.” *ISME J* 2:350–63.

Brady, Sonia K., Sarangapani Sreelatha, Yinnian Feng, Shishir P. S. Chundawat, and Matthew J. Lang. 2015. “Cellobiohydrolase 1 from *Trichoderma reesei* Degrades Cellulose in Single Cellobiose Steps.” *Nature Communications* 6(May).

Brusca, Richard C. 2008. “A Monograph on the Isopoda Cymothoidae (Crustacea) of the Eastern Pacific.” *Zoological Journal of the Linnean Society* 73(2):117–99.

Bruto, Maxime, Adèle James, Bruno Petton, Yannick Labreuche, Sabine Chenivesse, Marianne Alunno-Bruscia, Martin F. Polz, and Frédérique Le Roux. 2017. “*Vibrio crassostreae*, a Benign Oyster Colonizer Turned into a Pathogen after Plasmid Acquisition.” *ISME Journal* 11(4):1043–52.

Cain, Rob, and José Vázquez-Boland. 2015. “Survival Strategies of Intracellular Bacterial Pathogens.” *Molecular Medical Microbiology 2nd edition*: Pp. 491–515.

Casadevall, Arturo. 2008. “Evolution of Intracellular Pathogens.” *Annual Review of Microbiology* 62(1):19–33.

Casadevall, Arturo. 2012. “The Future of Biological Warfare.” *Microbial Biotechnology* 5(5):584–87.

Chakraborty, Smarajit, Hideaki Mizusaki, and Linda J. Kenney. 2015. “A FRET-Based DNA Biosensor Tracks OmpR-Dependent Acidification of *Salmonella* during Macrophage Infection.” *PLoS Biology* 13(4):1–32.

Cho, Kyung Ok, Go Woon Kim, and Ok Kyung Lee. 2011. “*Wolbachia* Bacteria Reside in Host Golgi-Related Vesicles Whose Position Is Regulated by Polarity Proteins.” *PLoS ONE* 6(7).

Choi, Eun Ju, Hak Cheol Kwon, Young Chang Sohn, and Hyun Ok Yang. 2010. “*Kistimonas asteriae* Gen. Nov., Sp. Nov., a Gammaproteobacterium Isolated from *Asterias amurensis*.” *International Journal of Systematic and Evolutionary Microbiology* 60(4):938–43.

Coetzee, Martin P. A., Brenda D. Wingfield, and Michael J. Wingfield. 2018. “*Armillaria* Root-Rot Pathogens: Species Boundaries and Global Distribution.” *Pathogens* 7(4):1–18.

Correa, Hebelin, Bradley Haltli, Carmenza Duque, and Russell Kerr. 2013. “Bacterial Communities of the Gorgonian Octocoral *Pseudopterogorgia elisabethae*.” *Microbial Ecology* 66.

Crawford, E. D., J. E. Seaman, A. E. Barber, D. C. David, P. C. Babbitt, A. L. Burlingame, and J. A. Wells. 2012. “Conservation of Caspase Substrates across Metazoans Suggests Hierarchical Importance of Signaling Pathways over Specific Targets and Cleavage Site Motifs in Apoptosis.” *Cell Death and Differentiation*

19(12):2040–48.

Deeks, Steven G., Julie Overbaugh, Andrew Phillips, and Susan Buchbinder. 2015. “HIV Infection.” *Nature Reviews Disease Primers* 1(October).

Deutscher, Josef, Christof Francke, and Pieter W. Postma. 2006. “How Phosphotransferase System-Related Protein Phosphorylation Regulates Carbohydrate Metabolism in Bacteria.” *Microbiology and Molecular Biology Reviews* 70(4):939–1031.

Ding, Jiun Yan, Jia Ho Shiu, Wen Ming Chen, Yin Ru Chiang, and Sen Lin Tang. 2016. “Genomic Insight into the Host-Endosymbiont Relationship of Endozoicomonas Montiporae CL-33T with Its Coral Host.” *Frontiers in Microbiology* 7(MAR).

Distel, Daniel L., Amy R. Baco, Ellie Chuang, Wendy Morrill, Colleen Cavanaugh, and Craig R. Smith. 2000. “Do Mussels Take Wooden Steps to Deep-Sea Vents?” *Nature* 403(6771):725–26.

Douglas, E. A. 2010. “The Symbiotic Habit.” *Symbiosis* 51(2):197–98.

van Dover, Cindy L., C. R. German, K. G. Speer, L. M. Parson, and R. C. Vrijenhoek. 2002. “Evolution and Biogeography of Deep-Sea Vent and Seep Invertebrates.” *Science* 295(5558):1253–57.

Dramsi, S., and P. Cossart. 1998. “Intracellular Pathogens and the Actin Cytoskeleton.” *Annual Review of Cell and Developmental Biology* 14(1):137–66.

Dubilier, Nicole, Claudia Bergin, and Christian Lott. 2008. “Symbiotic Diversity in Marine Animals: The Art of Harnessing Chemosynthesis.” *Nature Reviews Microbiology* 6(10):725–40.

Duperron, S., S. M. Gaudron, C. F. Rodrigues, M. R. Cunha, C. Decker, and K. Olu. 2013. “An Overview of Chemosynthetic Symbioses in Bivalves from the North Atlantic and Mediterranean Sea.” *Biogeosciences* 10(5):3241–67.

Duperron, Sébastien. 2010. “The Diversity of Deep-Sea Mussels and Their Bacterial Symbioses BT - The Vent and Seep Biota: Aspects from Microbes to Ecosystems.” Pp. 137–67 in, edited by S. Kiel. Dordrecht: Springer Netherlands.

Duperron, Sébastien, Claudia Bergin, Frank Zielinski, Anna Blazejak, Annelie Pernthaler, Zoe P. McKiness, Eric DeChaine, Colleen M. Cavanaugh, and Nicole Dubilier. 2006. “A Dual Symbiosis Shared by Two Mussel Species, *Bathymodiolus azoricus* and *Bathymodiolus puteoserpentis* (*Bivalvia*: *Mytilidae*), from Hydrothermal Vents along the Northern Mid-Atlantic Ridge.” *Environmental Microbiology* 8(8):1441–47.

Duperron, Sébastien, Adrien Quiles, Kamil M. Szafranski, Nelly Léger, and Bruce Shillito. 2016. “Estimating Symbiont Abundances and Gill Surface Areas in Specimens of the Hydrothermal Vent Mussel Bathymodiolus Puteoserpentis Maintained in

Pressure Vessels." *Frontiers in Marine Science* 3(FEB):1–12.

Elston, R. A. 1986. "An Intranuclear Pathogen [Nuclear Inclusion X (NIX)] Associated with Massive Mortalities of the Pacific Razor Clam, *Siliqua patula*." *Journal of Invertebrate Pathology* 47(1):93–104.

Esteves, Ana I. S., Cristiane C. P. Hardoim, Joana R. Xavier, Jorge M. S. Gonçalves, and Rodrigo Costa. 2013. "Molecular Richness and Biotechnological Potential of Bacteria Cultured from Irciniidae Sponges in the North-East Atlantic." *FEMS Microbiology Ecology* 85(3):519–36.

Fialamedioni, A., and M. Pennec. 1987. "Trophic Structural Adaptations in Relation to the Bacterial Association of Bivalve Molluscs from Hydrothermal Vents and Subduction Zones."

Fiore, Cara L., Micheline Labrie, Jessica K. Jarett, and Michael P. Lesser. 2015. "Transcriptional Activity of the Giant Barrel Sponge, *Xestospongia muta* Holobiont: Molecular Evidence for Metabolic Interchange." *Frontiers in Microbiology* 6(APR).

Fisher, C. R., J. J. Childress, R. S. Oremland, and R. R. Bidigare. 1987. "The Importance of Methane and Thiosulfate in the Metabolism of the Bacterial Symbionts of Two Deep-Sea Mussels." *Marine Biology* 96(1):59–71.

Forget, Nathalie L., and S. Kim Juniper. 2013. "Free-Living Bacterial Communities Associated with Tubeworm (Ridgeia Piscesae) Aggregations in Contrasting Diffuse Flow Hydrothermal Vent Habitats at the Main Endeavour Field, Juan de Fuca Ridge." *MicrobiologyOpen* 2(2):259–75.

Frank, S. A., and P. Schmid-Hempel. 2008. "Mechanisms of Pathogenesis and the Evolution of Parasite Virulence." *Journal of Evolutionary Biology* 21(2):396–404.

Franke, Maximilian, Benedikt Geier, Jörg U. Hammel, Nicole Dubilier and Nikolaus Leisch. "Symbiont colonization and early development of bathymodiolin mussels." *In Preparation*.

Friedrich, N., M. Hagedorn, D. Soldati-Favre, and T. Soldati. 2012. "Prison Break: Pathogens' Strategies To Egress from Host Cells." *Microbiology and Molecular Biology Reviews* 76(4):707–20.

Fujishima, Masahiro, and Yuuki Kodama. 2012. "Endosymbionts in Paramecium." *European Journal of Protistology* 48(2):124–37.

Gardères, Johan, Gilles Bedoux, Vasiliki Koutsouveli, Sterenn Crequer, Florie Desriac, and Gaël Le Pennec. 2015. "Lipopolysaccharides from Commensal and Opportunistic Bacteria: Characterization and Response of the Immune System of the Host Sponge Suberites Domuncula." *Marine Drugs* 13(8):4985–5006.

Gomez-Valero, Laura, Christophe Rusniok, Christel Cazalet, and Carmen Buchrieser. 2011. "Comparative and Functional Genomics of *Legionella* Identified Eukaryotic like

Proteins as Key Players in Host-Pathogen Interactions." *Frontiers in Microbiology* 2(OCT):1–20.

Gruber-Vodicka, Harald R., Nikolaus Leisch, Manuel Kleiner, Tjorven Hinzke, Manuel Liebeke, Margaret McFall-Ngai, Michael G. Hadfield, and Nicole Dubilier. 2019. "Two Intracellular and Cell Type-Specific Bacterial Symbionts in the Placozoan Trichoplax H2." *Nature Microbiology* 4(9):1465–74.

Haferkamp, Ilka, Stephan Schmitz-Esser, Michael Wagner, Nadjeschka Neigel, Matthias Horn, and H. Ekkehard Neuhaus. 2006. "Tapping the Nucleotide Pool of the Host: Novel Nucleotide Carrier Proteins of *Protochlamydia amoebophila*." *Molecular Microbiology* 60(6):1534–45.

Hansen, Allison K., and Nancy A. Moran. 2011. "Aphid Genome Expression Reveals Host-Symbiont Cooperation in the Production of Amino Acids." *Proceedings of the National Academy of Sciences of the United States of America* 108(7):2849–54.

Hidayati, Siti Nur, and Jeffrey L. Walck. 2016. "A Review of the Biology of Rafflesia: What Do We Know and What's Next?" *Buletin Kebun Raya* 19(2):67–78.

Ikuta, Tetsuro, Yoshihiro Takaki, Yukiko Nagai, Shigeru Shimamura, Miwako Tsuda, Shinsuke Kawagucci, Yui Aoki, Koji Inoue, Morimi Teruya, Kazuhito Satou, Kuniko Teruya, Makiko Shimoji, Hinako Tamotsu, Takashi Hirano, Tadashi Maruyama, and Takao Yoshida. 2016. "Heterogeneous Composition of Key Metabolic Gene Clusters in a Vent Mussel Symbiont Population." *ISME Journal* 10(4):990–1001.

Iwatani, Koichi, Hideo Dohra, B. Franz Lang, Gertraud Burger, Manabu Hori, and Masahiro Fujishima. 2005. "Translocation of an 89-KDa Periplasmic Protein Is Associated with Holospora Infection." *Biochemical and Biophysical Research Communications* 337(4):1198–1205.

Jamwal, Shilpa V., Parul Mehrotra, Archana Singh, Zaved Siddiqui, Atanu Basu, and Kanury V. S. Rao. 2016. "Mycobacterial Escape from Macrophage Phagosomes to the Cytoplasm Represents an Alternate Adaptation Mechanism." *Scientific Reports* 6(May 2015):1–9.

Jensen, Sigmund, Sébastien Duperron, Nils Kåre Birkeland, and Martin Hovland. 2010. "Intracellular Oceanospirillales Bacteria Inhabit Gills of Acesta Bivalves." *FEMS Microbiology Ecology* 74(3):523–33.

Jessen, Christian, Javier Felipe Villa Lizcano, Till Bayer, Cornelia Roder, Manuel Aranda, Christian Wild, and Christian R. Voolstra. 2013. "In-Situ Effects of Eutrophication and Overfishing on Physiology and Bacterial Diversity of the Red Sea Coral *Acropora hemprichii.*" *PLoS ONE* 8(4).

Kaper, James B., James P. Nataro, and Harry L. T. Mobley. 2004. "Pathogenic *Escherichia coli*." *Nature Reviews Microbiology* 2(2):123–40.

Katharios, Pantelis, Helena M. B. Seth-Smith, Alexander Fehr, José M. Mateos, Weihong Qi, Denis Richter, Lisbeth Nufer, Maja Ruetten, Maricruz Guevara Soto, Urs Ziegler, Nicholas R. Thomson, Ralph Schlapbach, and Lloyd Vaughan. 2015. "Environmental Marine Pathogen Isolation Using Mesocosm Culture of Sharpsnout Seabream: Striking Genomic and Morphological Features of Novel *Endozoicomonas* Sp." *Scientific Reports* 5(August):17609.

Kenk, Carmen, and Barry R. Wilson. 1985. "New Mussel From Hydrothermal Vents." *Malacología* 26(253):253–71.

Kräter, Martin, Jiranuwat Sapudom, Nicole Bilz, Tilo Pompe, Jochen Guck, and Claudia Claus. 2018. "Alterations in Cell Mechanics by Actin Cytoskeletal Changes Correlate with Strain-Specific Rubella Virus Phenotypes for Cell Migration and Induction of Apoptosis." *Cells* 7(9):136.

Kurahashi, Midori, and Akira Yokota. 2007. "*Endozoicomonas elysicola* Gen. Nov., Sp. Nov., a γ-Proteobacterium Isolated from the Sea Slug *Elysia ornata*." *Systematic and Applied Microbiology* 30(3):202–6.

Lema, Kimberley A., Bette L. Willis, and David G. Bourne. 2014. "Amplicon Pyrosequencing Reveals Spatial and Temporal Consistency in Diazotroph Assemblages of the *Acropora millepora* Microbiome." *Environmental Microbiology* 16(10):3345—3359.

Liu, Jun, Helu Liu, and Haibin Zhang. 2018. "Phylogeny and Evolutionary Radiation of the Marine Mussels (*Bivalvia*: *Mytilidae*) Based on Mitochondrial and Nuclear Genes." *Molecular Phylogenetics and Evolution* 126:233–40.

Lorion, Julien, Steffen Kiel, Baptiste Faure, Masaru Kawato, Simon Y. W. Ho, Bruce Marshall, Shinji Tsuchida, Jun Ichi Miyazaki, and Yoshihiro Fujiwara. 2013. "Adaptive Radiation of Chemosymbiotic Deep-Sea Mussels." *Proceedings of the Royal Society B: Biological Sciences* 280(1770).

Margulis 1938-2011 (viaf)108159261, Lynn, and René Fester. 1991. *Symbiosis as a Source of Evolutionary Innovation : Speciation and Morphogenesis*. Cambridge (Mass.) : MIT press.

Margulis, Lynn. 1970. *Origin of Eukaryotic Cells; Evidence and Research Implications for a Theory of the Origin and Evolution of Microbial, Plant, and Animal Cells on the Precambrian Earth.* New Haven: Yale University Press.

Moran, Nancy A., and Aparna Telang. 1998. "Bacteriocyte-Associated Symbiotic of Insects: A Variety of Insect Groups Harbor Ancient Prokaryotic Endosymbionts." *BioScience* 48(4):295–304.

Morrow, Kathleen M., David G. Bourne, Craig Humphrey, Emmanuelle S. Botté, Patrick Laffy, Jesse Zaneveld, Sven Uthicke, Katharina E. Fabricius, and Nicole S. Webster. 2015. "Natural Volcanic CO2 Seeps Reveal Future Trajectories for Host-

Microbial Associations in Corals and Sponges." *ISME Journal* 9:894–908.

Morrow, Kathleen M., Anthony G. Moss, Nanette E. Chadwick, and Mark R. Liles. 2012. "Bacterial Associates of Two Caribbean Coral Species Reveal Species-Specific Distribution and Geographic Variability." *Applied and Environmental Microbiology* 78(18):6438–49.

Neave, Matthew J., Amy Apprill, Christine Ferrier-Pages, and Christian R. Voolstra. 2016. "Diversity and Function of Prevalent Symbiotic Marine Bacteria in the Genus *Endozoicomonas*." *Applied Microbiology and Biotechnology* 1–10.

Neave, Matthew J., Craig T. Michell, Amy Apprill, and Christian R. Voolstra. 2017a. "*Endozoicomonas* Genomes Reveal Functional Adaptation and Plasticity in Bacterial Strains Symbiotically Associated with Diverse Marine Hosts." *Scientific Reports* 7(January):1–12.

Neave, Matthew J., Rita Rachmawati, Liping Xun, Craig T. Michell, David G. Bourne, Amy Apprill, and Christian R. Voolstra. 2017b. "Differential Specificity between Closely Related Corals and Abundant *Endozoicomonas* Endosymbionts across Global Scales." *ISME Journal* 11(1):186–200.

Nguyen, Mary T. H. D., Michael Liu, and Torsten Thomas. 2014. "Ankyrin-Repeat Proteins from Sponge Symbionts Modulate Amoebal Phagocytosis." *Molecular Ecology* 23(6):1635–45.

Niebuhr, Kirsten, Noureddine Jouihri, Abdelmounaaim Allaoui, Pierre Gounon, Philippe J. Sansonetti, and Claude Parsot. 2000. "IpgD, a Protein Secreted by the Type III Secretion Machinery of *Shigella flexneri*, Is Chaperoned by IpgE and Implicated in Entry Focus Formation." *Molecular Microbiology* 38(1):8–19.

Nishijima, Miyuki, Kyoko Adachi, Atsuko Katsuta, Yoshikazu Shizuri, and Kazuhide Yamasato. 2013. "*Endozoicomonas numazuensis* Sp. Nov., a Gammaproteobacterium Isolated from Marine Sponges, and Emended Description of the Genus *Endozoicomonas* Kurahashi and Yokota 2007." *International Journal of Systematic and Evolutionary Microbiology* 63(PART2):709–14.

Nolan, S. J., Romano, J. D., & Coppens, I. (2017). "Host lipid droplets: An important source of lipids salvaged by the intracellular parasite *Toxoplasma gondii*". PLoS pathogens, 13(6), e1006362.

Phillips, Margaret A., Jeremy N. Burrows, Christine Manyando, Rob Hooft Van Huijsduijnen, Wesley C. Van Voorhis, and Timothy N. C. Wells. 2017. "Malaria." *Nature Reviews Disease Primers* 3.

Pike, Rebecca E., Brad Haltli, and Russell G. Kerr. 2013. "Description of *Endozoicomonas euniceicola* Sp. Nov. and *Endozoicomonas gorgoniicola* Sp. Nov., Bacteria Isolated from the Octocorals *Eunicea fusca* and *Plexaura* Sp., and an Emended Description of the Genus *Endozoicomonas*." *International Journal of*

Systematic and Evolutionary Microbiology 63(PART 11):4294–4302.

Poole, P., and D. Allaway. 2000. “Carbon and Nitrogen Metabolism in *Rhizobium*.” *Advances in Microbial Physiology* 43:117–63.

Poulin, R. 2007. “The Evolutionary Ecology of Parasites.” in *Evolutionary Ecology of Parasites: (Second Edition)*.

Qi, Weihong, Maria Chiara Cascarano, Ralph Schlapbach, Pantelis Katharios, Lloyd Vaughan, and Helena M. B. Seth-Smith. 2018. “*Ca.* Endozoicomonas cretensis: A Novel Fish Pathogen Characterized by Genome Plasticity.” *Genome Biology and Evolution* 10(6):1363–74.

Ransome, Emma, Sonia J. Rowley, Simon Thomas, Karen Tait, and Colin B. Munn. 2014. “Disturbance to Conserved Bacterial Communities in the Cold-Water Gorgonian Coral *Eunicella verrucosa*.” *FEMS Microbiology Ecology* 90(2):404–16.

Ray, Katrina, Benoit Marteyn, Philippe J. Sansonetti, and Christoph M. Tang. 2009. “Life on the inside: The Intracellular Lifestyle of Cytosolic Bacteria.” *Nature Reviews Microbiology* 7(5):333–40.

Reaka-Kudla, Marjorie L. 1997. “The Global Biodiversity of Coral Reefs: A Comparison with Rain Forests.” *Biodiversity II: Understanding and Protecting Our Biological Resources* (October):83–108.

Roder, Cornelia, Till Bayer, Manuel Aranda, Maren Kruse, and Christian R. Voolstra. 2015. “Microbiome Structure of the Fungid Coral *Ctenactis echinata* Aligns with Environmental Differences.” *Molecular Ecology* 24(13):3501–11.

Rodriguez-Lanetty, Mauricio, Camila Granados-Cifuentes, Albert Barberan, Anthony J. Bellantuono, and Carolina Bastidas. 2013. “Ecological Inferences from a Deep Screening of the Complex Bacterial Consortia Associated with the Coral, *Porites astreoides*.” *Molecular Ecology* 22(16):4349–62.

Roy, Craig R., and Lewis G. Tilney. 2002. “The Road Less Traveled: Transport of Legionella to the Endoplasmic Reticulum.” *Journal of Cell Biology* 158(3):415–19.

Rua, Cintia P. J., Amaro E. Trindade-Silva, Luciana R. Appolinario, Tainá M. Venas, Gizele D. Garcia, Lucas S. Carvalho, Alinne Lima, Ricardo Kruger, Renato C. Pereira, Roberto G. S. Berlinck, Rogério A. B. Valle, Cristiane C. Thompson, and Fabiano Thompson. 2014. “Diversity and Antimicrobial Potential of Culturable Heterotrophic Bacteria Associated with the Endemic Marine Sponge *Arenosclera brasiliensis*.” *PeerJ* 2014(1):1–14.

Sabaneyeva, Elena V, M. E. Derkacheva, K. A. Benken, Sergei I. Fokin, Seppo Vainio, and Ilya N. Skovorodkin. 2009. “Actin-Based Mechanism of Holospora Obtusa Trafficking in Paramecium Caudatum.” *Protist* 160(2):205–19.

Sassera, Davide, Tiziana Beninati, Claudio Bandi, Edwin A. P. Bouman, Luciano

Sacchi, Massimo Fabbi, and Nathan Lo. 2006. "'*Candidatus* Midichloria mitochondrii', an Endosymbiont of the Tick *Ixodes ricinus* with a Unique Intramitochondrial Lifestyle." *International Journal of Systematic and Evolutionary Microbiology* 56(Pt 11):2535–40.

Schenck, Louis Patrick, Michael G. Surette, and Dawn M. E. Bowdish. 2016. "Composition and Immunological Significance of the Upper Respiratory Tract Microbiota." *FEBS Letters* 590(21):3705–20.

Schmid, Anna-Maria M. 2003. "Endobacteria in the diatom Pinnularia (Bacillariophyceae). I. 'Scattered Ct-Nucleoids' explained: DAPI–DNA complexes stem from exoplastidial bacteria boring into the chloroplasts1,2." *Journal of Phycology* 39(1):122–38.

Schmitz-esser, Stephan, Nicole Linka, Astrid Collingro, L. Beier, H. Ekkehard Neuhaus, Michael Wagner, Matthias Horn, and Cora L. Beier. 2004. "ATP / ADP Translocases : A Common Feature of Obligate Intracellular Amoebal Symbionts Related to Chlamydiae and Rickettsiae." *Journal of Bacteriology* 186(3):683–91.

Schreiber, Lars, Kasper Urup Kjeldsen, Matthias Obst, Peter Funch, and Andreas Schramm. 2016a. "Description of *Endozoicomonas ascidiicola* Sp. Nov., Isolated from Scandinavian Ascidians." *Systematic and Applied Microbiology* 39(5):313–18.

Schreiber, Lars, Kasper U. Kjeldsen, Peter Funch, Jeppe Jensen, Matthias Obst, Susanna López-Legentil, and Andreas Schramm. 2016b. "*Endozoicomonas* Are Specific, Facultative Symbionts of Sea Squirts." *Frontiers in Microbiology* 7(JUL):1–15.

Schulz, Frederik, and Matthias Horn. 2015. "Intranuclear Bacteria: Inside the Cellular Control Center of Eukaryotes." *Trends in Cell Biology* 25(6):339–46.

Schulz, Frederik, Ilias Lagkouvardos, Florian Wascher, Karin Aistleitner, Rok Kostanjšek, and Matthias Horn. 2014. "Life in an Unusual Intracellular Niche: A Bacterial Symbiont Infecting the Nucleus of Amoebae." *ISME Journal* 8(8):1634–44.

Shin, Woongghi, Sung Min Boo, and Lawrence Fritz. 2003. "Endonuclear Bacteria in *Euglena hemichromata* (Euglenophyceae): A Proposed Pathway to Endonucleobiosis." *Phycologia* 42(2):198–203.

Siguier, Patricia, Edith Gourbeyre, and Mick Chandler. 2014. "Bacterial Insertion Sequences: Their Genomic Impact and Diversity." *FEMS Microbiology Reviews* 38(5):865–91.

Tandon, Kshitij, Chih Ying Lu, Pei Wen Chiang, Naohisa Wada, Shan Hua Yang, Ya Fan Chan, Ping Yun Chen, Hsiao Yu Chang, Yu Jing Chiou, Ming Shean Chou, Wen Ming Chen, and Sen Lin Tang. 2020. "Comparative Genomics: Dominant Coral-Bacterium *Endozoicomonas acroporae* Metabolizes Dimethylsulfoniopropionate (DMSP)." *ISME Journal* 14(5):1290–1303.

Turchinovich, Andrey, Harald Surowy, Andrius Serva, Marc Zapatka, Peter Lichter, and

Barbara Burwinkel. 2014. “Capture and Amplification by Tailing and Switching (CATS), Anultrasensitive Ligation-Independent Method for Generation of DNA Libraries for Deep Sequencing from Picogram Amounts of DNA and RNA.” *RNA Biology* 11(7):817–28.

Vogt, G. 1992. “Enclosure of Bacteria by the Rough Endoplasmic Reticulum of Shrimp Hepatopancreas Cells.” *Protoplasma* 171(3):89–96.

Waters, Christopher, and Bonnie Bassler. 2005. “Waters CM, Bassler BL.. Quorum Sensing: Cell-to-Cell Communication in Bacteria. Annu Rev Cell Dev Biol 21: 319-346.” *Annual Review of Cell and Developmental Biology* 21:319–46.

WATSON, M. L. 1955. “The Nuclear Envelope; Its Structure and Relation to Cytoplasmic Membranes.” *The Journal of Biophysical and Biochemical Cytology* 1(3):257–70.

Wilcox, L. W. 1986. “Prokaryotic Endosymbionts in the Chloroplast Stroma of the Dinoflagellate *Woloszynskia pascheri*.” *Protoplasma* 135(2):71–79.

Wilson, Edward O. 2014. *The Meaning of Human Existence.* New York, NY, US: W W Norton & Co.

Yang, Cho Song, Ming Hui Chen, A. B. Arun, Chaolun Allen Chen, Jih Terng Wang, and Wen Ming Chen. 2010. “*Endozoicomonas montiporae* Sp. Nov., Isolated from the Encrusting Pore Coral *Montipora aequituberculata*.” *International Journal of Systematic and Evolutionary Microbiology* 60(5):1158–62.

Zhang, Jianlin, Amanda Felder, Yujie Liu, Ling T. Guo, Stephan Lange, Nancy D. Dalton, Yusu Gu, Kirk L. Peterson, Andrew P. Mizisin, G. Diane Shelton, Richard L. Lieber, and Ju Chen. 2009. “Nesprin 1 Is Critical for Nuclear Positioning and Anchorage.” *Human Molecular Genetics* 19(2):329–41.

Zhou, Xiaohui, Ramiro H. Massol, Fumihiko Nakamura, Xiang Chen, Benjamin E. Gewurz, Brigid M. Davis, Wayne Lencer, and Matthew K. Waldor. 2014. “Remodeling of the Intestinal Brush Border Underlies Adhesion and Virulence of an Enteric Pathogen.” *MBio* 5(4):1–9.

Ziegler, Maren, Anna Roik, Adam Porter, Khalid Zubier, Mohammed S. Mudarris, Rupert Ormond, and Christian R. Voolstra. 2016. “Coral Microbial Community Dynamics in Response to Anthropogenic Impacts near a Major City in the Central Red Sea.” *Marine Pollution Bulletin* 105(2):629–40.

Zielinski, Frank U., Annelie Pernthaler, Sebastien Duperron, Luciana Raggi, Olav Giere, Christian Borowski, and Nicole Dubilier. 2009. “Widespread Occurrence of an Intranuclear Bacterial Parasite in Vent and Seep Bathymodiolin Mussels.” *Environmental Microbiology* 11(5):1150–67.

Chapter II | Molecular biology of *Ca*. Endonucleobacter

The hungry nucleus: The nutritional demands of a chitinolytic intranuclear parasite trigger its host cell to upregulate sugar import

Miguel-Ángel González-Porras[1], Adrien Assié[2], Målin Tietjen[1], Marlene Jensen[3], Manuel Kleiner[3], Harald Gruber-Vodicka[1], Nicole Dubilier[1], Nikolaus Leisch[1].

1 Department of Symbiosis, Max Planck Institute for Marine Microbiology, Bremen, Germany

2 Alkek Center for Metagenomics and Microbiome Research, Baylor College of Medicine, Houston, TX, USA

3 Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC, USA

The manuscript is in preparation and it has not been revised by all authors.

Author contributions

MAGP conceived the study, performed the fluorescence microscopy experiments, developed the laser-capture microdissection pipeline, analyzed the data, prepared the figures/tables and wrote the manuscript. AA helped to conceive the study and contributed to the conceptual design of all genomic analyses. MT contributed to the conceptual design and data interpretation of all transcriptomic analyses of the study. MJ and MK performed the proteomic analyses of the study. HGV contributed to the conceptual design and data interpretation of genomic and transcriptomic analyses of the study, as well as to the sequencing strategies for data acquisition. ND helped to

conceive the study. NL helped to conceive the study and performed the ultrastructure experiments.

Abstract

Intracellular bacteria commonly occur in the cytoplasm of host cells. Only very few bacteria are able to exploit intraorganellar niches such as the nucleus. *Ca*. Endonucleobacter is a gammaproteobacterial parasite related to *Endozoicomonas* that infests the nuclei of deep-sea mussels. After a single *Ca*. Endonucleobacter cell invades the host nucleus, it proliferates massively (up to 80,000 bacteria per nucleus). By doing so, the volume of the nucleus increases up to 50 fold and the host cell is dramatically deformed. It is hypothesized that intranuclear parasites could use host chromatin as a nutritional source, but this would lead to a destabilization of the replication niche. We wanted to investigate how *Ca*. Endonucleobacter thrives in the nucleus and how it affects host cell. To address these questions, we coupled laser capture microdissection of infected nuclei with ultra-low-input RNA-sequencing. The gills of infected *"Bathymodiolus" childressi* were analyzed with fluorescence *in situ* hybridization targeting *Ca*. Endonucleobacter and different infection stages were microdissected and sequenced. Data revealed that the host cell was transcriptionally active along the infectious cycle, showing upregulation of genes involved in lipid droplets biosynthesis. The expression of chitinase by *Ca*. Endonucleobacter progressively increased until the onset of parasite replication, indicating host chitin hydrolysis. Therefore, promotion of host lipid droplets synthesis and chitin hydrolysis suggests that *Ca*. Endonucleobacter access nutritional host resources rather than chromatin. We hypothesize that *Ca*. Endonucleobacter stimulates lipid droplets synthesis by bringing the host cell to a starvation state and expressing the *Shigella*-like factor *IpgD*. *IpgD* might also disentangle actin stress fibers, reducing host cell

stiffness. This would allow *Ca*. Endonucleobacter to increase the volume of the nucleus up to 50 fold without collapsing the host cell in a process of microniche engineering. Last, we hypothesize that *Ca*. Endonucleobacter is able to complete its life cycle by expressing inhibitors of apoptosis, which prevent the shutdown of the host cell. This study shows by the first time how an intranuclear parasite can prevent destabilization of its replication niche by avoiding chromatin consumption, disentangling host cell cytoskeleton and inhibiting apoptosis. Therefore, we expect this study to be an inflection point in the understanding of the molecular basis of intranuclear parasitism.

Keywords: *Endozoicomonas*; intranuclear parasite; chromatin; T3SS; *IpgD*; lipid droplets; chitin; inhibitors of apoptosis; laser capture microdissection; cytoskeleton manipulation.

Introduction

Few bacterial pathogens can live intracellularly, and when they do, they normally reside in the host cell cytoplasm or in host-derived vacuoles (Ray et al. 2009). A molecular crossfire between pathogen and host is stablished when intracellular parasitism occurs. Those few cases open the possibility to study inter-kingdom molecular communication in the frame of intracellular parasitism. Intracellular pathogens are rare, but even fewer can occupy specialized intracellular compartments such as the nucleus (Bierne and Cossart 2012). *Ca*. Endonucleobacter is a gammaproteobacterial parasite that infests the nuclei of deep-sea mussels (Zielinski et al. 2009). After a single *Ca*. Endonucleobacter cell invades the host nucleus, it proliferates massively (up to 80,000 bacteria per nucleus) and the volume of the nucleus increases up to 50 fold. At the end

of the infectious cycle the host cell bursts, releasing a bacterial progeny which is ready to infect new nuclei.

The intranuclear lifestyle has appeared several times in phylogenetically unrelated phyla (*Verrucomicrobia*, *Chlamydiales* and *Proteobacteria*) in a process of convergent evolution, suggesting selective advantages of this lifestyle. It seems parsimonious that the intranuclear lifestyle has appeared secondarily in already intracellular bacteria (Schulz and Horn 2015). Most of intranuclear parasites infect a wide range of unicellular eukaryotic hosts, including paramecia (Fujishima and Kodama 2012), flagellates isolated from termite guts (Sato et al. 2014) and amoeba (Schulz et al. 2014). There are few intranuclear parasites that can infect multicellular eukaryotic hosts like arthropods and mammalian cells (Burgdorfer et al. 1968) or molluscs (Elston 1986). *Ca*. Endonucleobacter is one of the latter, exclusively infecting bathymodiolin mussels (Zielinski et al. 2009).

Ca. Endonucleobacter is phylogenetically related to *Endozoicomonas* (*Oceanospirillales*, *Hahellaceae*) (Cano et al. 2018; Zielinski et al. 2009). *Endozoicomonas* is known to stablish symbiotic relationships with a great diversity of marine metazoans, including cnidarians, poriferans, molluscs, annelids, tunicates and fish (Fiore et al. 2015; Forget and Kim Juniper 2013; Jensen et al. 2010; Katharios et al. 2015; Morrow et al. 2012). Within the host, *Endozoicomonas* plays a role in carbohydrate transport and cycling and protein secretion (Neave et al. 2017). *Endozoicomonas* typically reside in bacterial aggregates within their host tissues (Katharios et al. 2015; Schuett et al. 2007). While certain *Endozoicomonas* have been described as extracellular symbionts (Schreiber et al. 2016), others have been

described to occur intracellularly (Beinart et al. 2014; Ding et al. 2016). To date, only one *Endozoicomonas* species has been described as pathogen (Katharios et al. 2015; Qi et al. 2018), while the rest have been described as mutualists. Contrarily, *Ca*. Endonucleobacter is a monophyletic group of parasites specialized in occupying the nucleus of its host cell (Zielinski et al. 2009).

The nucleus is enclosed by the nuclear envelope, a highly organized double membrane that contains the chromatin, a complex of DNA and protein (histones) (Kite 1913). Early ultrastructure studies revealed that the inner nuclear membrane and the outer nuclear membrane are continuous with the endoplasmic reticulum (Watson 1955). The nuclear envelope is the barrier that separates the nuclear and cytoplasmic media, but it needs of channels for bidirectional trafficking known as the nuclear pore complexes. The nucleus contains the nuclear genetic material, which is ultimately responsible for controlling and directing most of the activities of the cell. Is intriguing why an intranuclear parasite would choose this life strategy, as it might be easy to interfere with the viability of the replication niche.

The life strategy of an intranuclear parasite has many potential advantages. The nucleus is a cellular compartment that offers shelter against cytoplasmic defense mechanisms such as autophagy (Ray et al. 2009). It also opens the possibility for the parasite to directly interact with the host chromatin by delivering bacterial effectors called nucleomodulins that interfere with nuclear processes (Bierne and Cossart 2012). The nucleus is filled with a rich pool of proteins (up to 90%), small ribonucleotides, nucleic acids (up to 30%) and nucleoside triphosphates that are available to be used by the intranuclear parasite (Schulz and Horn 2015). Therefore, the nucleus has been

considered as a nutrient-rich replication niche for intranuclear parasites. The consumption of host genetic material would be dramatically deleterious for the host cell, but studies of several intranuclear parasites suggest that they do not reduce their host cell fitness. E.g. Intranuclear rickettsiae, *Caedibacter caryophilus* and *Holospora* spp. codify for one or more nucleotide importers, which allow them to pirate the energy budget of the host cell (Haferkamp et al. 2006; Schmitz-esser et al. 2004). A strategy to avoid DNA damage would be to digest other nucleic acids (such as mRNAs or intronic RNAs) or histones. However, this would have major repercussions on host genetic expression and regulation (Schulz and Horn 2015). To ensure that the host cell still functional during the infectious cycle, an intranuclear parasite might be accessing nutritional sources beyond the nucleus, like lipids or storage or structural polysaccharides.

Lipid droplets (LD) are organelles which main function is to store and provide energy to the cell. Cellular stresses such as starvation or invasion are phenomena which provoke changes in LD metabolism and dynamics (Henne et al. 2018). In fact, host-LD utilization by infectious agents, be they viral, bacterial or protozoan, is an common parasitic strategy (Henne et al. 2018). The intracellular parasite *Toxoplasma gondii* induces the formation of host-LD by bringing the host cell to a starvation state (Nolan et al. 2017). LD biosynthesis can be also promoted by increasing the intracellular levels of phosphatidylinositol 5-monophosphate (PtdIns5P). PtdIns5P is a phosphoinositide that participates in intracellular signaling events involved in cytoskeleton remodeling and cell morphology alteration (Czech 2000; Toker 1998). *IpgD*, a virulence factor firstly described in *Shigella flexneri*, is a potent inositol 4-phosphatase that specifically dephosphorylates phosphatidylinositol 4,5-biphosphate into PtdIns5P that then

accumulates intracellularly (Niebuhr et al. 2002). It has been demonstrated that PtdIns5P interacts directly with the cytoskeleton component Septin 9, promoting growth and perinuclear accumulation of LD (Akil et al. 2016).

The biosynthesis of LD through the glycerol phosphate pathway is a metabolically-demanding process (Lee et al. 2001). Martano et al. 2016 demonstrated that an upregulation of the glycolysis-TCA axis was correlated with the biosynthesis of glycerol 3-phosphate, first intermediary of the glycerol phosphate pathway. The glycolysis-TCA axis requires the catabolism of a myriad of molecules such as glycols. Monomeric glycols like glucose or N-acetylglucosamine (GlcNAc) could be imported into the host cell through sugar transporters (Navale and Paranjape 2016). Alternatively, monomeric glycols could be released intracellularly through the hydrolysis of storage or structural polysaccharides such as glycogen or chitin. Chitin, the second most abundant polymer on the planet after cellulose, is a polymer of β-linked GlcNAc. Chitin synthesis and utilization is a widespread trait among invertebrates, from Poriferans (Ehrlich et al. 2007) through the invertebrate chordates (Sannasi and Hermann 1970). In bivalves such as deep-sea mussels, chitin synthesis usually takes place in the extracellular medium of epithelial mantle cells during shell formation. There, chitin fibers play a role as extracellular matrix template for crystallization and biomineralization (Weiss 2012). However, the presence of chitin in bivalves' gills is not unknown: Jemaa et al. 2014 reported the existence of chitinous rods in the gill filaments of *Crassostrea gigas*. Wherever it occurs, chitin is synthesized by highly conserved enzymes named chitin synthases (Weiss 2012).

In this study, the transcriptomes of *Ca*. Endonucleobacter and its host cell were analyzed at three infection stages (early, mid, late) along the infectious cycle. The data demonstrated that *Ca*. Endonucleobacter infection does not silence the genetic expression of its host cell. Instead, the host cell upregulated genes involved in LD synthesis, sugar import, glycolysis and sensors of nuclear deformation. Simultaneously, *Ca*. Endonucleobacter expresses chitinase and lipase, which suggests nutritional access to host chitin and lipids. The results of this transcriptional study suggested that *Ca*. Endonucleobacter microengineers its replication niche by disentangling the host actin stress fibers using the *Shigella*-like factor *IpgD*. Last, *Ca*. Endonucleobacter also manipulates the host cell life cycle by arresting apoptosis using inhibitors of apoptosis. Remarkably, the intranuclear parasite is able to complete its life cycle while the host cell still transcriptionally and metabolically active.

Results

Phylogeny

Based on the sequence analysis of 43 rRNA marker genes (**Figure 1**) of 80 gammaproteobacterial genomes, *Ca*. Endonucleobacter clusters within the order *Oceanospirillales*, a monophyletic group of marine *Gammaproteobacteria*. Within the order *Oceanospirillales*, *Ca*. Endonucleobacter forms a monophyletic group together with *Endozoicomonas* (100% support) (**Fig.1, green box**). Both genera cluster together within a monophyletic group of symbionts of marine invertebrates that include the genus *Kistimonas* (100% support) (**Fig.1, blue box**). The phylogenetic analysis based on 16S rRNA sequences (**Supplementary Figure 1**) also placed *Ca*. Endonucleobacter as the sister clade of *Endozoicomonas*, although the support did not reach 80%. The 16S rRNA analysis (**Supp. Fig.1**) also revealed that the *Ca*. Endonucleobacter clade was formed by three subclades reflecting the mussels'

sampling sites as previously reported by Zielinski et al., 2009. In the phylogenomic analysis (**Fig.1**), *Ca*. Endonucleobacter was only placed as sister clade of Endozoicomonas in 71% of the iterations. The *Ca*. Endonucleobacter clade showed the longest evolutionary distance when compared with the members of the *Endozoicomonas* clade, suggesting a high substitution rate.

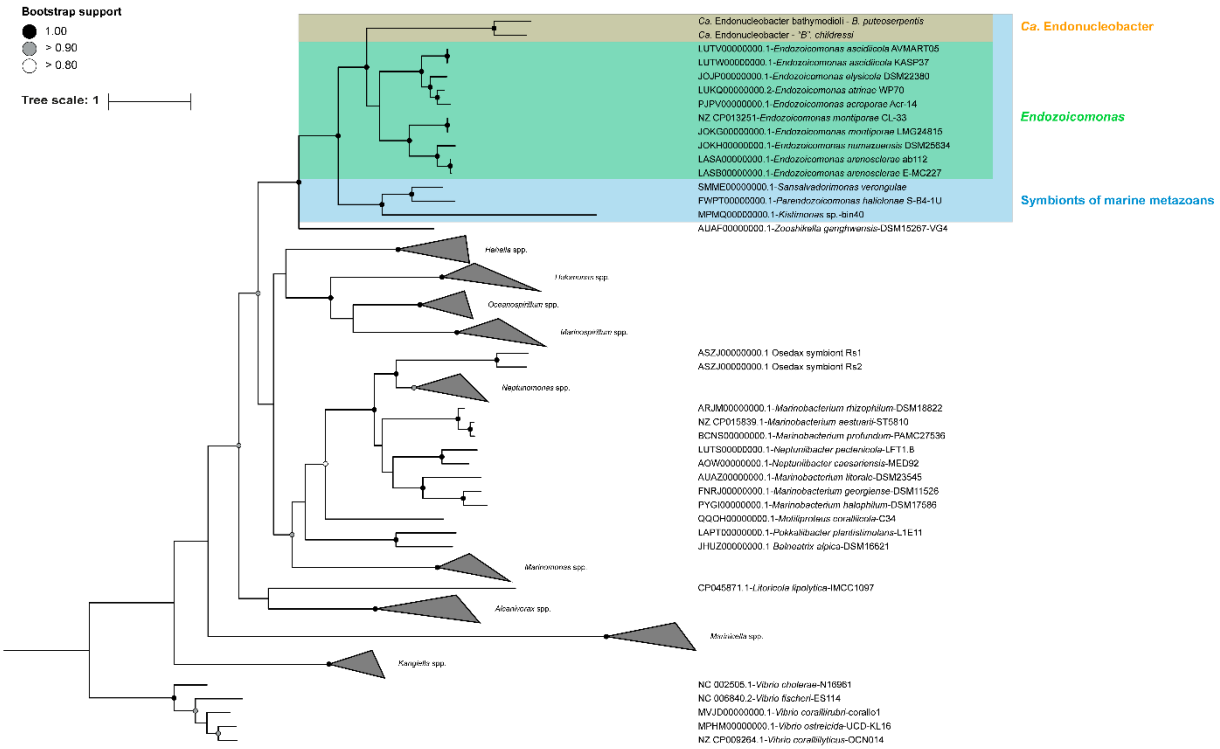


Figure 1. Phylogenomic tree of representatives of *Oceanospirillales.* The 43 rRNA single copy marker genes used in this analysis were chosen and aligned by CheckM (Parks et al. 2015). Five Vibrionales species were used to root the tree. Framed in blue is a monophyletic group of symbionts of marine invertebrates. Framed in green is the genus *Endozoicomonas*. Both *Ca.* Endonucleobacter species are framed in orange.

Localization pattern

We determined the localization pattern of *Ca*. Endonucleobacter in the deep-sea mussel "*Bathymodiolus" childressi* from the Gulf of Mexico (**Figure 2, a**). Like other bathymodiolin mussels, "*B". childressi* has huge brown gills (**Fig.2, b**) formed by thousands of gill filaments (**Fig.2, c**). Base on the presence or absence of mutualistic

symbionts, two anatomically distinct areas can be distinguished in the gill filaments: The symbiotic region where the mutualistic symbionts occur, and the ciliated edge where there are not mutualistic symbionts (**Fig.2, d**). Fluorescence *in situ* hybridization (FISH) targeting *Ca*. Endonucleobacter on whole filaments revealed that *Ca*. Endonucleobacter is limited to ciliated edge in "*B". childressi* (**yellow rim in Fig.2, c**). A closer look to the infected area confirmed that *Ca*. Endonucleobacter never occurs in the symbiotic region, neither in the adhesion patches (**Fig.2, d**). Higher resolution revealed that *Ca*. Endonucleobacter forms massive bacterial clusters (**Fig.2, e**).

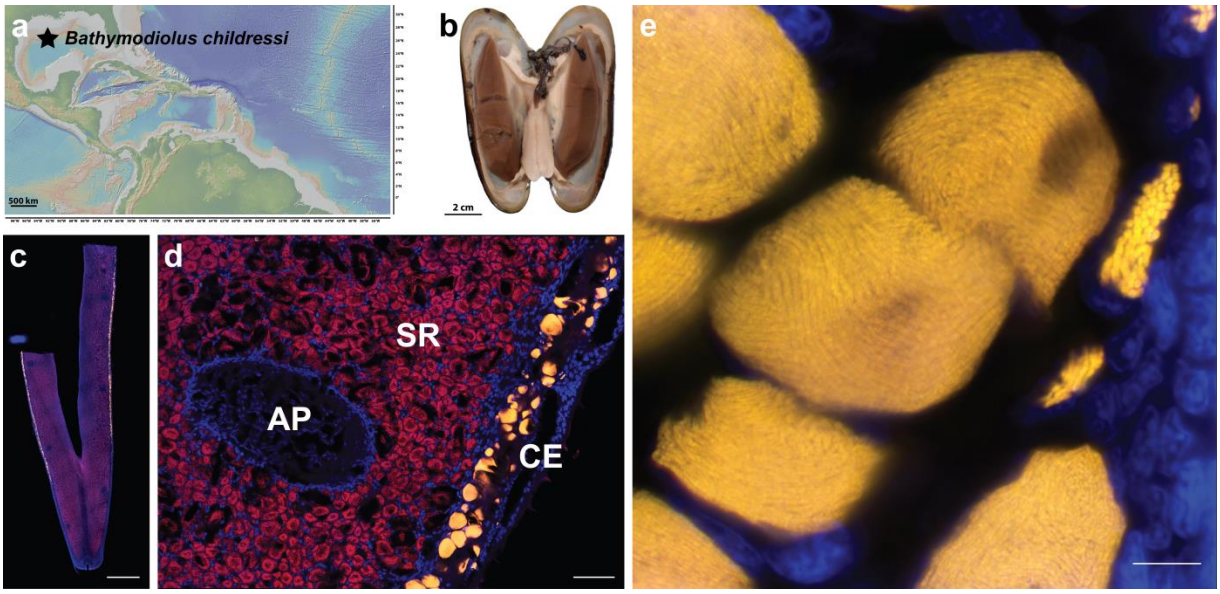


Figure 2. ***Ca. Endonucleobacter*** **is limited to the ciliated edge in "*B". childressi*.** Mussels were sampled in the Gulf of Mexico (a). The mussels' gills were dissected (brown structures in b) and single gill filaments (c) were subjected to Whole-mount FISH. Eubacteria 16S rRNA (red: Alexa Fluor 647), *Ca*. Endonucleobacter 16S rRNA (yellow: Atto550), DNA (blue: DAPI). ***c,*** epifluorescence overview of whole filament, scalebar: 2 mm. ***d,*** CLSM image of adhesion patch surroundings, scalebar: 50 µm. ***e,*** SR-Airyscan image of infection area, scalebar: 5 µm. (**SR**: Symbiotic area, **CE**: Ciliated edge, **AP**: Adhesion patch).

Physiological modelling

To model *Ca*. Endonucleobacter physiology, we analyzed its genomic potential and support it with bulk transcriptomic and proteomic data (**Figure 3**). The expression of every gene has been normalized to *RecA* (a single copy housekeeping gene) and

expressed as a ratio. The main findings are described next, but the numeric values of *Ca*. Endonucleobacter metabolism expression can be found can be found in **Supplementary table 3**.

Ca. Endonucleobacter had a chemoorganoheterotrophic metabolism, with complete glycolysis pathway, pentose phosphate pathway, TCA cycle and an aerobic transport chain. ATP synthase alpha subunit (ATP_{α}) and an electron transport chain thioredoxin (THIO) had both high expression values. Both factors had representation in the first quartile of the proteomic data. *Ca*. Endonucleobacter fuels oxidative phosphorylation importing glycols and lipids from the host cell, relying on PTS transporters (*PtsN*, PTSG, PTSP) and fatty acid transporters (FATP) to do so. Overall, all these transporters had medium expression values. In contrast, the competence system factors for DNA import *ComEC* and *ComF* were not expressed. Additionally, all pathways for *de novo* synthesis of purines and pyrimidines synthesis were transcriptionally active in *Ca*. Endonucleobacter (**Supplementary table 6**). *Ca*. Endonucleobacter genome did not encode for the *de novo* biosynthesis pathways of the amino acids Valine, Leucine, Isoleucine, Threonine, Asparagine, Cysteine, Methionine, Phenylalanine, Tyrosine, Arginine and Histidine. Remarkably, *Ca*. Endonucleobacter encoded for the oligopeptide importers oligopeptide permease operon (*OppB*, *OppC)* and the ABC putrescine importer (*PotF*, *PotG*), which were both represented in the transcriptome with low expression values. *Ca*. Endonucleobacter encoded for ABC detoxification systems for Co^{2+}, Zn^{2+} and Cd^{2+} (*CzcD*) and Cu^{2+} (CUCDT), which were not expressed. *Ca*. Endonucleobacter also encoded for metal importers, concretely for ABC importers for Zn2+ (*ZnuB*, *ZnuC*), Mn2+ (*SitD*) and Fe2+ (*FeoA*, *FeoB*). Remarkably, only the *FeoA* factor had high expression values, while the

rest had no expression. *Ca*. Endonucleobacter encoded for two porines: An aquaporin (*AqpZ*) and an undetermined porin (UPorin). Both factors showed high expression values. Remarkably, both Uporin and *AqpZ* were present in the third quartile of *Ca*. Endonucleobacter proteome.

Pathogenic modelling

The pathogenic capabilities of *Ca*. Endonucleobacter have been analyzed as in the metabolic modelling (see previous section). The main findings are described next, but the numeric values of can be found in **Supplementary table 3**.

Ca. Endonucleobacter encoded for a type 1 secretion system (T1SS) and for a type 3 secretion system (T3SS), the later also known as injectisome. The factors *LapC* and *LapE* (T1SS) and the factors *YscQ*, *YscO* and *YscR* (T3SS) had all high expression values. *Ca*. Endonucleobacter encoded for virulence factors which are T1SS-secretion dependent, concretely the *Vibrio*-like repeats-in-toxin adhesin (RTX-a). RTX-a had high expression values, and was also found in the third quartile of *Ca*. Endonucleobacter proteome. *Ca*. Endonucleobacter also encoded for virulence factors which are T3SS-secretion dependent, and presumably delivered through host cellular membranes. Concretely, *Ca*. Endonucleobacter encoded for 10 baculoviral apoptosis inhibitors (IAP). In **Fig.3** we represented one of the 10 IAP, which contained a T3SS-dependent signal peptide and had low expression values. *Ca*. Endonucleobacter encoded for two copies of the *Shigella*-like factor *IpgD* (*IpgD*). The amino acid sequence of both copies included a T3SS-dependent signal peptide and both had low expression values. In **Fig.3**, we represented one of the two copies of *IpgD*. *Ca*. Endonucleobacter encoded for hydrolytic enzymes which had T3SS-dependent signal

peptides: a lipase and a chitinase (*ChiA*). Both hydrolytic enzymes had high expression values. Moreover, chitinase was one of the top 10 most expressed factors in *Ca*. Endonucleobacter bulk transcriptome. Chitinase was also among the top most abundant 15 proteins in *Ca*. Endonucleobacter proteome.

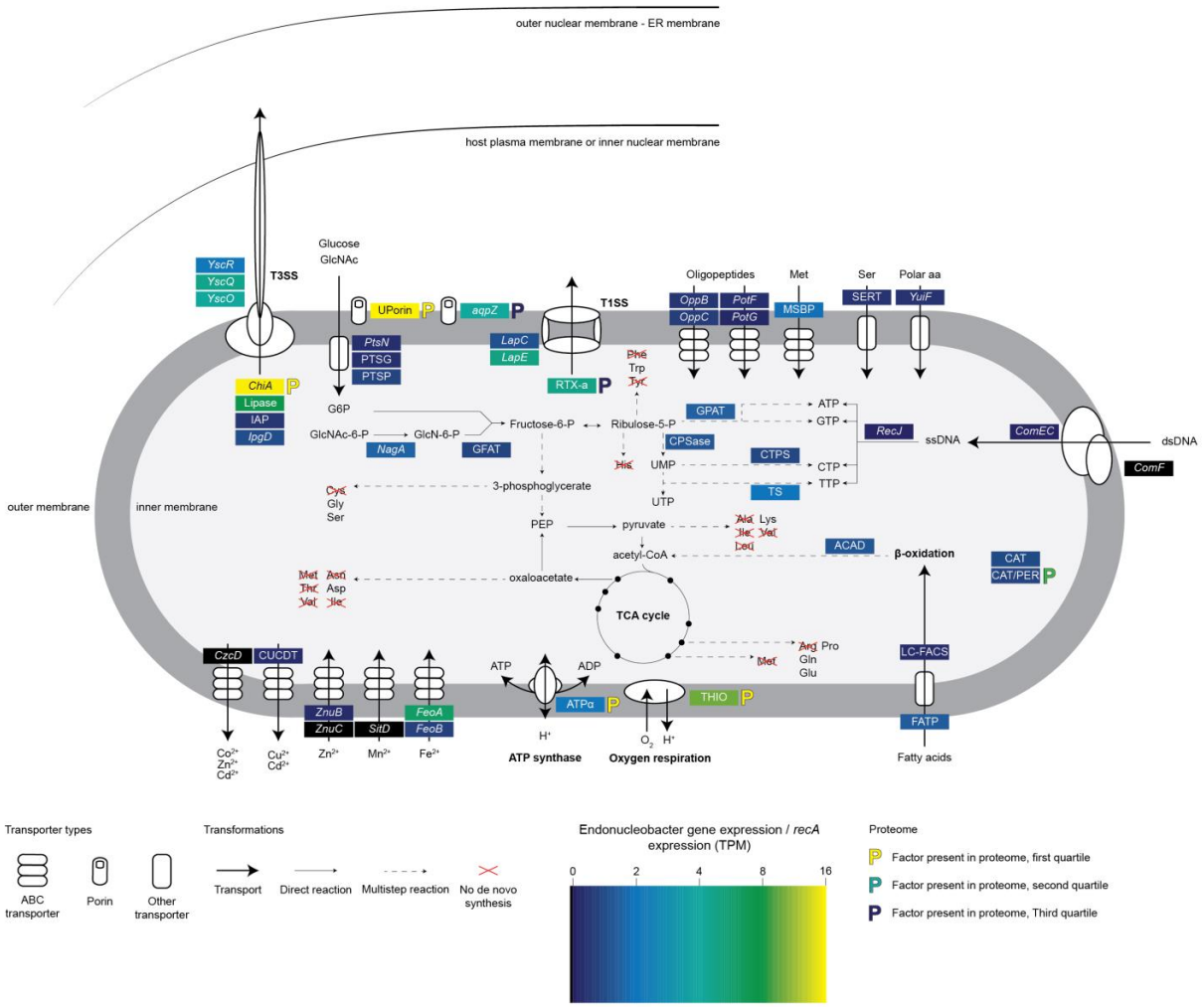


Figure 3. *Ca. Endonucleobacter* has a chemoorganoheterotroph lifestyle, in which chitin and lipids seem to be major nutritional sources. Physiological reconstruction based on RAST annotation and Pathway Tools metabolic modelling. Functions with indicated expression levels are discussed in the text. ***YscR***, T3SS inner membrane protein *YscR*. ***YscQ***, T3SS inner membrane protein *YscQ*. ***YscO***, T3SS bacterial envelope protein *YscO*. ***ChiA***, chitinase. **Lipase**, lipase. **IAP**, Inhibitor of apoptosis. ***IpgD***, Shigella-like Inositol phosphate phosphatase. **Uporin**, undetermined porin. ***agpZ***, aquaporin. ***LapC***, T1SS membrane fusion protein. ***LapE***, T1SS outer membrane component. **RTX-a**, *Vibrio*-like repeats-in-toxin adhesin. ***NagA***, N-acetylglucosamine-6-phosphate deacetylase. **GFAT**, Glucosamine-fructose-6-phosphate aminotransferase. ***CzcD***, Cobalt-zinc-cadmium resistance protein *CzcD*. **CUCDT**, Copper-translocating P-type ATPase. ***ZnuB***, Zinc ABC transporter inner membrane permease protein *ZnuB*. ***ZnuC***, Zinc ABC transporter ATP-binding protein *ZnuC*. ***SitD***, Manganese ABC transporter inner membrane permease protein *SitD*. ***FeoA***, Ferrous iron transport protein A. ***FeoB***, Ferrous iron transport protein B. **ATPα**, ATP

synthase alpha chain. **THIO**, Thioredoxin. **FATP**, Long-chain fatty acid transport protein. **LC-FACS**, Long-chain-fatty-acid-CoA ligase. **ACAD**, Acyl-CoA dehydrogenase. **CAT/PER**, Catalase / Peroxidase. **CAT**, Catalase. ***ComF***, Competence protein F homolog. ***ComEC***, DNA internalization-related competence protein ComEC. ***RecJ***, Single-stranded-DNA-specific exonuclease *RecJ*. ***OppB***, Oligopeptide transport system permease protein *OppB*. ***OppC***, Oligopeptide transport system permease protein *OppC*. ***PotF***, Putrescine ABC transporter putrescine-binding protein PotF. ***PotG***, Putrescine transport ATP-binding protein *PotG*. **MSBP**, Methionine ABC transporter substrate-binding protein. **SERT**, Serine transporter. ***YuiF***, Histidine permease *YuiF*. **GPAT**, Amidophosphoribosyltransferase. **CPSase**, Carbamoyl-phosphate synthase large chain. **CTPS**, CTP synthase. **TS**, Thymidylate synthase.

Phylogenetic and domain characterization of chitinase

The phylogenetic analysis of 45 chitinase protein sequences showed that *Ca*. Endonucleobacter chitinase clusters within the GH18 family of chitinases (**Supplementary figure 5, A**). Chitin from *Ca*. Endonucleobacter showed the typical structural elements of enzymatically active endochitinases from GH18 family: A secretion signal peptide, two chitin-binding domains and a catalytic domain with a glutamate residue in the catalytic center (**Supp. Fig.5, B**). Chitinase from *Ca*. Endonucleobacter from *"B". childressi* had a 26.36% sequence similarity with chitinase of *Ca*. Endonucleobacter bathymodioli - *B. puteoserpentis*, and 15.05%, 15.93% and 16.76% sequence similarity with chitinases from *Hahella chejuensis* and *Kistimonas* sp.40 (isoform 2 and 3, respectively). The chitinases from *H. chejuensis* and Kistimonas sp.40 clustered as a sister group of the chitinases of the genus *Ca*. Endonucleobacter, and presented as well the domains that characterize the enzymatically active chitinases from GH18 family of glycosidases.

Transcriptomic profiling of* Ca. *Endonucleobacter and its host cell along the infectious cycle

In this study, we wanted to investigate the expression changes of *Ca*. Endonucleobacter and its host cell along the infectious cycle. To do so, we analyzed

the transcriptomes of *Ca*. Endonucleobacter and its host cell obtained from microdissected single-cycle stages samples (**non-infected (not included in the figure), early, mid and late; Figure 4**). Like in **Fig.3**, *Ca*. Endonucleobacter expression has been normalized to *RecA* and expressed as a ratio (see "Physiological modelling" section). The host cell expression has been represented as log2 fold changes in comparison to the previous treatment (the expression changes in early stage treatment are compared to the non-infected treatment, mid stage vs. early stage, late stage vs. mid stage). The numeric values of transcriptomic changes of all factors represented in **Fig.4** has been included in **Supplementary table 4** (*Ca.* Endonucleobacter) and **Supplementary table 5** (host cell).

Early phase: Ca. *Endonucleobacter brings the host cell to a starvation state and triggers the synthesis of host lipid droplets*

The early stage of infection was characterized by *Ca*. Endonucleobacter colonization of the host cell, the expression of the T3SS, and the expression of hydrolytic enzymes (*ChiA* and lipase). In this stage, *Ca*. Endonucleobacter also expressed nutritional importers. The host cell reacted to the infection upregulating sugar importers (SWEET), chitobiase (CTBS), glycolysis and oxidative phosphorylation factors, as well as lipid droplets (LD) biosynthesis. *Ca*. Endonucleobacter expressed the T3SS factors *YscQ* and *YscR*, which had both high expression values. *Ca*. Endonucleobacter expressed chitinase (*ChiA*) and lipase. Both hydrolytic enzymes had high expression values. *Ca*. Endonucleobacter expressed the glucose PTS (PTS), fatty acid (FATP) and oligopeptide (*OppB*, *OppC, PotF*, *PotG*) importers. Overall, all importers had low expression values. *Ca.* Endonucleobacter also expressed the virulence factors RTX-a, *lpgD* and IAP. While *lpgD* and IAP had low levels of expression, RTX-a had high

levels of expression. The host cell reacted to *Ca*. Endonucleobacter infection upregulating chitobiase (CTBS), enzyme responsible for chitobiose hydrolysis. The enzymes from the glycolysis pathway fructose-bisphosphate aldolase (FBPA) and pyruvate kinase (PKM) were also upregulated in comparison with the non-infected treatment. The host cell expressed the mitochondrial factors for oxidative phosphorylation ATP-synthase alpha and beta subunits (ATP_{α} and ATP_{β}). While ATP_{α} expression remained constant in comparison with the non-infected treatment, the expression of ATP_{β} increased. The host cell also reacted to *Ca*. Endonucleobacter infection upregulating enzymes from the glycerolphosphate pathway for triacylglycerides biosynthesis (GPDH, FAS, PAP, DGAT). Only two enzymes from this pathway (GPAT and AGPAT) were not upregulated. While the expression of GPAT slightly decreased in comparison with the non-infected treatment, the expression of AGPAT remained constant. Remarkably, the host cell also upregulated the LD marker perilipin-2 (PLIN2). The host cell slightly upregulated the expression of chitin synthase (*Chs2*) in comparison with the non-infected treatment. Last, infection by *Ca*. Endonucleobacter also triggered host upregulation of the nuclear deformation sensor Nesprin-1 (SYNE1), the apoptosis promoter caspase-2 (CASP2), the GTPase IMAP family member 4 (GIMAP) and the phosphatidylinositol 4,5-bisphosphate phosphodiesterase (PLCG).

Mid phase: *The hydrolysis of chitin fuels the metabolisms of* Ca. *Endonucleobacter and its host cell*

In mid stage of infection, the hydrolysis of host chitin and lipids by *Ca*. Endonucleobacter reached its zenith through the upregulation of *ChiA* and lipase. *ChiA* and lipase upregulation was accompanied by the upregulation of PTS and FATP transporters, as well as by the upregulation of factors involved in GlcNAc (*NagA*, GFAT) and fatty acid (LC-FACS, ACAD) catabolism. In mid stage, *Ca*. Endonucleobacter also upregulated the T3SS components, as well as several virulence factors: RTX-a, *IpgD* and IAP. The host cell upregulated SWEET and CTBS, which was accompanied by the upregulation of the enzymes *NagA* and GFAT. In general, the factors involved in glycolysis and oxidative phosphorylation were constantly expressed or upregulated in comparison to the early stage of infection. Last, the host cell upregulated most of the enzymes involved in LD biosynthesis. *Ca*. Endonucleobacter *ChiA* and lipase were dramatically upregulated, both showing high levels of expression. The PTS and FATP transporters, as well as the factors *NagA*, GFAT, LC-FACS and ACAD were upregulated. LC-FACS showed low levels of expression, while PTS, FATP, *NagA*, GFAT and ACAD showed high levels of expression. *Ca*. Endonucleobacter upregulated the expression of *YscQ* and *YscR* (T3SS). YscQ and YscR had both high levels of expression. *Ca*. Endonucleobacter upregulated the expression of the virulence factors RTX-a, *IpgD* and IAP, being RTX-a the factor that was upregulated more pronouncedly. All virulence factors had high levels of expression except IAP, which showed medium expression. The host cell upregulated the sugar transporter SWEET in comparison to the early stage of infection. This was accompanied by a slight upregulation of CTBS and the upregulation of the enzymes *NagA* and GFAT. The glycolysis pathway remained transcriptionally active,

with an upregulation of phosphofructokinase-1 (PFK-1) and no changes in the expression of PKM. Only the expression of FBPA was slightly reduced in comparison to the early stage of infection. FAS and PLIN2 expression remained virtually unchanged, while all the other enzymes involved in LD biosynthesis (GPDH, GPAT, AGPAT, PAP and DGAT) were upregulated. The host cell slightly upregulated the expression of *Chs2* in comparison with the early-stage of infection. The host cell also upregulated the expression of SYNE1, which correlated with the upregulation of CASP2. Last, the host expression of GIMAP remained unaltered in comparison with the mid stage of infection.

Late phase: *The host cell enters in metabolic decline*

In late phase, *Ca*. Endonucleobacter downregulated the expression of the hydrolytic enzymes *ChiA* and lipase. However, most of the expression changes that characterized the late phase of infection occurred on the host cell side. The host cell experienced a general downregulation of chitin synthesis (*Chs2*), lipid droplets biosynthesis (GPDH, GPAT, AGPAT and DGAT) and reaction to mechanical nuclear deformation (SYNE1, CASP2). Intriguingly, the host cell upregulated GlcNAc catabolism (*NagA*, GFAT) and glycolysis (PFK-1, FBPA, PKM) in comparison to the mid-stage of infection. *Ca*. Endonucleobacter *ChiA* and lipase were downregulated. However, both glycolytic enzymes still had high levels of expression. The expression of the virulence factors *IpgD* and IAP remained unchanged, showing high and medium levels of expression, respectively. *Ca*. Endonucleobacter downregulated the expression of RTX-a, although its expression value was still high. Last, the expression of *YscQ* and *YscR* (T3SS) remained virtually constant at high expression values. The

host cell downregulated the sugar transporter SWEET and CTBS in comparison to the mid-stage of infection. Intriguingly, the downregulation of SWEET and CTBS was accompanied by a slight upregulation of *NagA* and GFAT. The host expression of PAP remained virtually unchanged. However, all the other enzymes involved in the glycerolphosphate pathway for TAG biosynthesis (GPAT, AGPAT and DGAT) were downregulated. Intriguingly, PLIN2 was slightly upregulated in comparison with the mid-stage of infection. The upregulation of PLIN2 was correlated with an upregulation of GIMAP. Last, the host downregulated the expression of CASP2, while the expression levels of SYNE-1 remained unchanged in comparison with the mid-stage of infection.

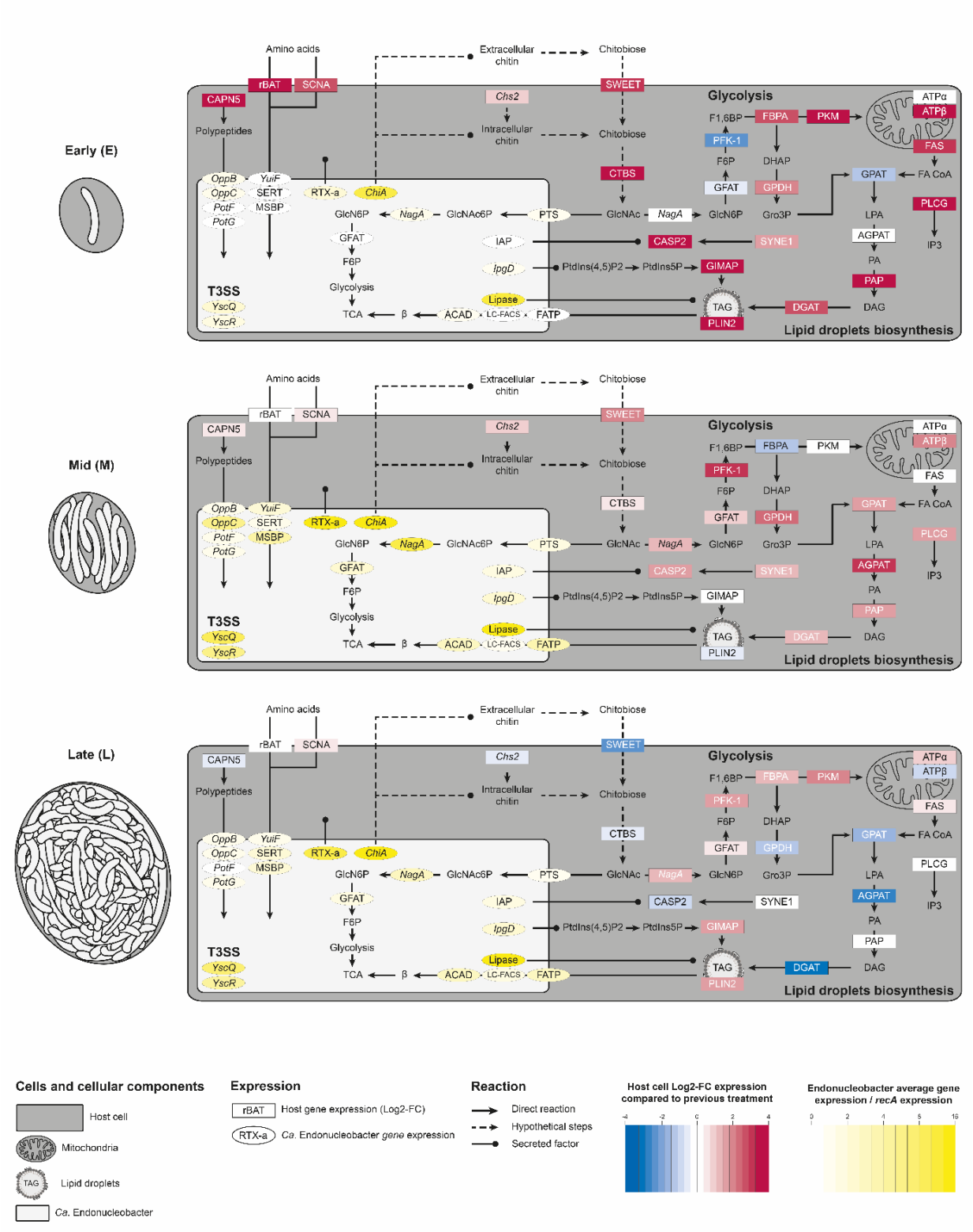


Figure 4. *Ca.* Endonucleobacter brings the host cell to a starvation state, which triggers lipid droplets synthesis. Transcriptomic profiling of the interaction between *Ca*. Endonucleobacter and *"B". childressi* host cell along the infectious cycle. ***Ca*. Endonucleobacter factors: *OppB***, Oligopeptide transport system permease protein *OppB*. ***OppC***, Oligopeptide transport system permease protein *OppC*. ***PotF***, Putrescine ABC transporter putrescine-binding protein PotF. ***PotG***, Putrescine transport ATP-binding protein *PotG*. **MSBP**, Methionine ABC transporter substrate-binding protein. **SERT**, Serine transporter. ***YuiF***,

Histidine permease *YuiF*. **RTX-a**, *Vibrio*-like repeats-in-toxin adhesin. ***ChiA***, chitinase. **PTS**, Phosphoenolpyruvate-protein phosphotransferase of PTS system. ***NagA***, N-acetylglucosamine-6-phosphate deacetylase. **GFAT**, Glucosamine-fructose-6-phosphate aminotransferase. **IAP**, Inhibitor of apoptosis. ***IpgD***, *Shigella*-like Inositol phosphate phosphatase. ***YscQ***, T3SS inner membrane protein *YscQ*. ***YscR***, T3SS inner membrane protein *YscR*. **Lipase**, lipase. **FATP**, Long-chain fatty acid transport protein. **LC-FACS**, Long-chain-fatty-acid-CoA ligase. **ACAD**, Acyl-CoA dehydrogenase. ***"B." childressi* factors: CAPN5**, calpain-5. **rBAT**, Neutral and basic amino acid transport protein rBAT. **SCNA**, Sodium-coupled neutral amino acid transporter. ***Chs2***, Chitin synthase chs-2. **SWEET**, SWEET sugar transporter 1. **CTBS**, Chitobiase. ***NagA***, N-acetylglucosamine-6-phosphate deacetylase. **GFAT**, Glucosamine-fructose-6-phosphate aminotransferase. **PFK-1**, Phosphofructokinase-1. **FBPA**, Fructose-bisphosphate aldolase. **PKM**, Pyruvate kinase PKM. **ATPα**, ATP synthase alpha chain**. ATPβ**, ATP synthase beta chain**. FAS**, Fatty acid synthase. **GPDH**, Glycerol-3-phosphate dehydrogenase. **GPAT**, glycerol-3-phosphate O-acyltransferase. **AGPAT**, 1-acyl-sn-glycerol-3-phosphate acyltransferase alpha. **PAP**, Phosphatidate phosphatase. **DGAT**, Diacylglycerol O-acyltransferase. **PLIN2**, Perilipin-2. **GIMAP**, GTPase IMAP family member 4. **SYNE1**, Nesprin-1. **CASP2**, Caspase-2. **PLCG**, phosphatidylinositol 4,5-bisphosphate phosphodiesterase. Not all factors from the glycolysis, TCA and β-oxidation of fatty acids (β) have been represented for space and clarity reasons.

Cytological changes

Ribosomal integrity

The digestion of DNA by *Ca*. Endonucleobacter would have major repercussions on the host cell genetic expression and regulation, affecting the integrity of ribosomes. We used the fluorescence intensity of host 18S rRNA as an indicator of DNA integrity along the infectious cycle. Although the non-infected (NI) cell showed the highest 18S rRNA fluorescent intensity (**Supplementary figure 2**), the cytoplasmic region surrounding nuclei at different infection stages (early; E, mid; M, late; L) showed comparable 18S rRNA fluorescent intensity. Remarkably, the cell at late stage of infection showed higher integrated density of 18S rRNA particles than the cell at mid stage of infection. This data suggested that the progression of *Ca*. Endonucleobacter life cycle does not have an effect on the amount of host cell 18S rRNA particles.

Ultrastructure

Ca. Endonucleobacter replication within the nucleus has major repercussions in cell morphology. To evaluate these morphological changes along the infectious cycle, we imaged *"B". childressi* infected cells using transmission electron microscopy. A *"B". childressi* infected cell in late stage revealed an apparently functional cell (**Supplementary figure 3**). A closer look to the outer nuclear membrane – mitochondria area interface confirmed that the mitochondria are intact in the late stage of infection (**Supp. Fig.3, b**). Additionally, the host chromatin seemed compressed against the inner-nuclear membrane.

Discussion

Endonucleobacter, the intranuclear "Endozoicomonas"

Our phylogenomic analysis of 43 rRNA markers genes placed *Ca*. Endonucleobacter within a monophyletic group together with the genus *Endozoicomonas* in 100% of the calculations (**Fig.1**). Intriguingly, only 70% of the calculations placed *Ca*. Endonucleobacter genus as sister clade of *Endozoicomonas*. Based in our phylogenomic analysis, we cannot discard the possibility that the genus *Ca*. Endonucleobacter originated from within the genus *Endozoicomonas*. As introduced previously, certain *Endozoicomonas* have been described as intracellular symbionts (Beinart et al. 2014). One could think about a hypothetical scenario in which *Ca*. Endonucleobacter derived from an *Endozoicomonas* subclade with a specialized intracellular (or intranuclear) lifestyle. Our results question the monophyly of the genus Endozoicomonas, and calls to revisit the phylogeny of the *Endozoicomonas* – *Ca*. Endonucleobacter tandem. Further studies in that direction should tackle the precise localization of *Endozoicomonas* members within their hosts' tissues. This highlights the importance of complementing phylogenetic analyses with imaging data. This might be critical in host-microbe systems, as describing the physical interactions between the holobiont members might shed light on the evolutionary history of the group under study.

Oxygen depletion by methane-oxidizing symbionts might limit **Ca.** *Endonucleobacter to the ciliated edge of gill filaments*

Our microscopy data revealed that *Ca*. Endonucleobacter is limited to the ciliated edge in the gill filaments of *"B". childressi* (**Fig.2**). We wondered whether this organized

distribution pattern was biologically and/or environmentally driven. As previously described for *Ca*. Endonucleobacter bathymodioli in Zielinski et al. 2009, *Ca*. Endonucleobacter can only infect mutualistic-symbionts free cells. One possible explanation to the distribution pattern of *Ca*. Endonucleobacter in the gills of *"B". childressi* is the higher abundance of symbiont-free cells in the ciliated edge. However, that would not explain the complete absence of *Ca*. Endonucleobacter in the symbiotic region (**Fig.2, d**), as it could hypothetically infect the symbiont-free intercalary cells present in that area. An alternative explanation to the limited distribution pattern of *Ca*. Endonucleobacter is that ciliated edge cells are "energy sinks". Ciliated edge cells present abundant mitochondria, as they require vast amounts of ATP to maintain the movement of cilia (Satir 1980). Because ciliated edge cells lack mutualistic symbionts, they need to acquire metabolites from the extracellular medium to fuel oxidative phosphorylation. We found in the *de novo* transcriptome of *"B". childressi* ciliated edge four copies of sodium-glucose linked transporters and two copies of facilitative glucose transporters, monosaccharides importers required for glucose uptake (Thorens 1996; Wright et al. 1994). We hypothesize that the ciliated edge cells are net sugar importers. If we consider that *Ca*. Endonucleobacter can grow up to 80,000 bacteria per nuclei (Zielinski et al. 2009), a cell that is continuously importing sugars seems a suitable infection target. Limitation in oxygen might also be a determinant factor in the distribution pattern of *Ca*. Endonucleobacter in *"B". childressi*. *"B". childressi* is known to occur in cold seeps from the Gulf of Mexico, where most of the oxygen is consumed by methane-oxidizing microbes (Boetius and Wenzhöfer 2013). One hypothesis that might explain the distribution pattern of *Ca*. Endonucleobacter is that the parasite is competing with the methane-oxidizing symbionts of *"B". childressi* for oxygen. This would limit *Ca*. Endonucleobacter to the ciliated edge, where oxygen concentration is

higher. *"B." childressi* specimens maintained in aquaria under different methane concentrations lost their methane-oxidizing symbionts when they were completely deprived of methane (**Supplementary Figure 8**). Interestingly, *Ca*. Endonucleobacter was localized in the symbiotic region in those *"B." childressi* specimens that lost their methane-oxidizing symbionts. This suggests that the methane-oxidizing symbionts of *"B." childressi* might prevent *Ca*. Endonucleobacter from colonizing symbiont-free cells in the symbiotic region, most likely due to competition for oxygen.

Ca*. Endonucleobacter expresses the T3SS injectisome along its life cycle*

Our bulk transcriptomic data revealed that *Ca*. Endonucleobacter expressed a T3SS injectisome (**Fig.3**), which was also expressed along its infectious cycle (**Fig.4**). The T3SS injectisome is a complex molecular nanomachine that allows bacteria to deliver virulence factors across eukaryotic cellular membranes (Cornelis 2006). The delivery of virulence factors into the eukaryotic cytoplasm results in the modulation of several host cell functions. One of the virulence factors which secretion is T3SS-dependent is the *Shigella*-like factor *IpgD*, an inositol phosphate phosphatase. The enteropathogen bacterium *Shigella flexneri* uses *IpgD* to induce the formation of plasma membrane ruffles by disassembling the cortical actin cytoskeleton (Niebuhr et al. 2000). *IpgD* plays a major role during cell invasion, as the plasma membrane ruffles are used by *S. flexneri* as the entry site to the enterocyte. Our bulk (**Fig.3**) and LMD (**Fig.4**) transcriptomic data showed that *Ca*. Endonucleobacter expressed *IpgD*. Moreover, the T3SS components and *IpgD* were co-expressed by *Ca*. Endonucleobacter during the early stage of its life cycle (**Fig.4**). We hypothesize that the T3SS-*IpgD* tandem might play a major role during colonization of the ciliated edge cells by *Ca*. Endonucleobacter. Our bulk transcriptomic data and proteomic data revealed that *Ca*.

Endonucleobacter expressed a RTX-adhesin (**Fig.3**), which was also expressed along its infectious cycle (**Fig.4**). RTX-adhesins are megaproteins displayed on the bacterial surface, which secretion is dependent of the T1SS (**Fig.3**). The ligand-binding region of these proteins plays a major role during adhesion and cohesion to the host cell (Guo et al. 2019). We hypothesize that RTX-adhesin aids *Ca*. Endonucleobacter to adhere to its host cell, while the tandem T3SS-*IpgD* acts as the active invasion executioner.

Like the enterocytes infected by *S. flexneri*, the cells from the ciliated edge of *"B." childressi* are non-phagocytic. One possibility is that *Ca*. Endonucleobacter injects *IpgD* into the ciliated edge cells to force phagocytic behavior and posterior uptake. Remarkably, the limited distribution pattern of *Ca*. Endonucleobacter to the ciliated edge has not been described for *Ca.* Endonucleobacter bathymodioli (Zielinski et al. 2009). *Ca.* Endonucleobacter bathymodioli infects the non-symbiotic intercalary cells present in the symbiotic region of *B. puteoserpentis* gill filaments, which do have endocytic behavior. Intriguingly, our preliminary analysis on the genomic capabilities of *Ca*. Endonucleobacter bathymodioli revealed that this parasite lacks *IpgD* (Gonzalez-Porras et al. *in Prep)*. As discussed previously, oxygen depletion in the symbiotic area of the gill filaments of *"B." childressi* might be the cause of the confinement of *Ca*. Endonucleobacter to the ciliated edge. However, there might be a selective pressure over *Ca*. Endonucleobacter to codify and express a molecular key (*IpgD*) to access cells that do not normally show endocytic behavior.

Ca. Endonucleobacter also expressed the T3SS during mid and late stage of infection (**Fig.4**). What could be the function of a membrane-piercing T3SS once the parasite has stablished itself the nucleus? One possible answer to this question is that *Ca*. Endonucleobacter expresses the T3SS before the mechanical rupture of the host cell

to ensure that the released progeny is ready to colonize new cells. However, that would only explain the expression of the T3SS during late stage of infection. An alternative answer to this question would be that *Ca*. Endonucleobacter maintains the secretory activity of the T3SS during its whole life cycle. Contact with the plasma membrane of the target host cell is considered the main trigger for T3SS gene expression (Hayes et al. 2010). It is tempting to speculate that the contact with the inner nuclear membrane induces the expression of the T3SS by *Ca*. Endonucleobacter. If that holds true, it is unclear whether *Ca.* Endonucleobacter is using the T3SS to deliver virulence factor directly to the nuclear space (no membrane-piercing activity) or used to secrete factors to the perinuclear space across the inner nuclear membrane (membrane-piercing activity). The perinuclear space is a continuum with the endoplasmic reticulum lumen. The hypothetical ability of *Ca*. Endonucleobacter to deliver virulence factors to the perinuclear space – endoplasmic reticulum lumen continuum would have implications in terms of host cell secretion manipulation. In theory, *Ca*. Endonucleobacter could hijack the secretory pathway of its host cell to secrete virulence factors to the extracellular space while sitting in the nucleus, phenomenon previously undescribed in intranuclear parasitism.

Ca. Endonucleobacter does not digest chromatin and prevents host cell apoptosis

If *Ca*. Endonucleobacter did live off the DNA and/or RNA present in the nucleus, one would expect an early collapse of the host cell in the worst case scenario, or a massive transcriptomic deregulation of the host cell in the best case scenario. Our transcriptomic profiling of the host cell along the infectious cycle (**Fig.4**) revealed that

the host cell did not suffer a generalized downregulation of expression between stages, even when compared with non-infected cells. This indicates that host cell is transcriptionally active along the infectious cycle, and that host chromatin remains transcriptionally competent until the end of the infection. *Ca*. Endonucleobacter inhabits the nucleus of its host cell, and it has been proposed that the parasite could be using host chromatin as carbon and energy source (Zielinski et al. 2009). The disappearance of host heterochromatin in infected eukaryotic nuclei has been previously reported for several intranuclear bacteria such as *Ca*. Endonucleobacter bathymodioli (Zielinski et al. 2009) or *Ca.* Nucleicultrix amoebiphila (Schulz et al. 2014). The first possible explanation to the reduction of host heterochromatin would be the use of DNA as nutritional source, which would lead to an early collapse of the host cell. The second possible explanation to heterochromatin disappearance would be the use of histones as carbon an energy source. However, both explanations would imply major consequences in host cell transcriptional regulation (Schulz and Horn 2015), which is not supported by our transcriptomic profiling of the host cell. We found that *Ca*. Endonucleobacter does not codify for nucleotide transporters, which have been found in another intranuclear parasites such as intranuclear rickettsiae, *Caedibacter caryophilus* and *Holospora* spp. (Haferkamp et al. 2006; Schmitz-esser et al. 2004). *Ca*. Endonucleobacter barely expressed its competence system genes (**Fig.3**), and expressed most of the enzymes of the *de novo* routes for nucleotides biosynthesis (**Suppl. Table 6**). Our ultrastructural data indicated that host cell mitochondria were functional at late stage of infection, and that host heterochromatin formed a thin layer right underneath the inner nuclear membrane (**Supp. Fig.3**). Additionally, we did not find significant differences in 18S rRNA fluorescence intensity between stages when localizing host ribosomes, which suggest that host rRNA still transcribed until the end

of the infectious cycle (**Supp. Fig.2**). Taken together, all these data indicate that host chromatin is not digested *Ca*. Endonucleobacter. We propose that the observed disappearance of the electron dense chromatin might just be an optical artifact of the chromatin being compressed against the inner nuclear membrane during bacterial growth. This optical effect would be magnified due to the fact that an infected nucleus in late stage is 50 times fold in volume compared with a non-infected nucleus. That *Ca*. Endonucleobacter does not digest chromatin might have the biological sense of maintaining the host cell functional to complete its life cycle.

This hypothesis is supported by the 10 copies of inhibitors of apoptosis (IAP) in *Ca*. Endonucleobacter genome. IAP have been described as physiologic caspase inhibitors which are able to arrest the apoptosis cascade (Deveraux et al. 1997; Deveraux and Reed 1999). Our transcriptomic profiling of the host cell along the infectious cycle (**Fig.4**) indicated that the infection by *Ca*. Endonucleobacter triggered the expression of caspase-2 (CASP2), an initiator caspase considered a master regulator of the apoptosis cascade (Troy and Ribe 2008). One of the intrinsic triggers of apoptosis cascade is cytoskeleton deformation (Kräter et al. 2018). Nesprin-1 (SYNE1) is an outer nuclear membrane protein which anchors the nucleus to the actin cytoskeleton. SYNE1 is responsible to transmit nuclear deformation information to the actin cytoskeleton (Zhang et al. 2009). The host expression of SYNE1 was upregulated after *Ca*. Endonucleobacter infection (**Fig.4**), suggesting host awareness of nuclear deformation. We hypothesize that the actin cytoskeleton deformation due to nuclear expansion triggers the apoptotic cascade by expressing CASP2. *Ca*. Endonucleobacter expressed one of the IAP along the infectious cycle (**Fig.4**). We suggest that *Ca*. Endonucleobacter interfere with CASP2 by secreting IAP, which

would allow the parasite to complete the infectious cycle by blocking apoptosis. To our knowledge, *Ca.* Endonucleobacter is one of the few intranuclear parasites of multicellular eukaryotes described up to date. Apoptosis in some form is found in all multicellular eukaryotes, and metazoans are not an exception (Koonin and Aravind 2002). In metazoans, caspases are the major regulators of apoptosis (Crawford et al. 2012). Sun et al. 2017 demonstrated that caspases were expanded and highly expressed in the gills of the deep-sea mussel *B. platifrons*. Thus, we hypothesize that IAP might have a great biological significance for *Ca*. Endonucleobacter lifestyle.

Our findings have several implications on how intranuclear parasitism has been previously interpreted, considering the nucleus merely as a subcellular compartment to profit from nutritionally. As introduced before, the direct use of chromatin as a nutritional source would lead to a collapse of the host cell. This would not have major repercussions for an intracellular parasite of a unicellular eukaryote, where the whole host system is limited to one cell. However, an intranuclear parasite of a multicellular eukaryote might exploit its host beyond the infected cell. As part of a tissue, the infected cell relies on the sugars and amino acids present in its extracellular medium for nutrition. By using the nucleus of its host cell in a non-destructive way, *Ca*. Endonucleobacter can turn the infected host cell into the funneling agent that creates a constant flow of metabolites from the extracellular medium to the nuclear compartment. We suggest that *Ca*. Endonucleobacter intentionally avoids chromatin consumption and arrests host cell apoptosis to complete its life cycle. *Ca*. Endonucleobacter might use the nucleus as a replication niche from where it hides from cytoplasmic molecular defenses and access alternative nutritional resources rather than chromatin.

Chitin seems to be an important nutritional source for* Ca. *Endonucleobacter

Our phylogenetic and domain analyses of *Ca*. Endonucleobacter chitinase (*ChiA*) suggested that this protein is an enzymatically active endochitinase from the GH18 family (**Supp. Fig.5**). *ChiA* was one of the top 10 factors most expressed in *Ca*. Endonucleobacter bulk transcriptome, being also represented in the first quartile of its proteome (**Fig.3**). *ChiA* was also the only *Ca.* Endonucleobacter gene to be differentially expressed along its life cycle (**Fig.4**). This data suggested that chitin might be an important nutritional source for *Ca.* Endonucleobacter. Chitin is a homopolymer of β-linked GlcNAc residues synthesized by many marine invertebrate groups, from Poriferans (Ehrlich et al. 2007) through the invertebrate chordates (Sannasi and Hermann 1970). In bivalves, chitin synthesis usually takes place in the extracellular medium of the epithelial mantle cells during shell formation (Weiss 2012). However, chitinous rod-like structures that give mechanical support to the gill filaments can be found in certain bivalves such as *C. gigas* (Jemaa et al. 2014). Is there chitin in the gill filaments of *"B." childressi*? And if so, where does chitin occur? Our ultrastructural images did not show any chitinous rod-like structure in the gill filaments of *"B." childressi*, neither any sort of chitin inclusions in the cells infected by *Ca*. Endonucleobacter (**Supp. Fig.6, a**). However, we were able to localize chitin-containing vesicles in secretory cells placed in the ciliated edge-symbiotic area interface (**cyan frames in Supp. Fig.6, b, c**). In addition, Tietjen, *pers. comm.* found differential expression of chitin synthase 2 (*Chs2*) when comparing ciliated edge (high expression) vs. symbiotic region (low expression) transcriptomes of *"B." childressi* gill filaments. This data suggested that chitin is produced in the ciliated edge of the gill filaments of *"B." childressi,* mainly by specialized secretory cells. The polarity and function of these secretory cells suggest that chitin is secreted as a component of the

extracellular mucus layer. Our transcriptomic characterization of the infectious cycle (**Fig.4**) revealed that the host cell infected by *Ca*. Endonucleobacter also expressed *Chs2*. We hypothesize that although the ciliated edge cells infected by *Ca*. Endonucleobacter synthesize chitin, the secretory cells present in the ciliated edge-symbiotic region interface seem to be specialized in chitin synthesis and secretion. Because *Ca*. Endonucleobacter does not infect secretory cells, chitin hydrolyzed by *Ca*. Endonucleobacter could be located within its host ciliated edge cell (intracellular chitin) and/or within the mucus layer (extracellular chitin), as pointed out by dashed arrows in **Fig.4**. How can *Ca*. Endonucleobacter hydrolyze chitin when this polysaccharide occurs outside the nucleus?

The most conservative model in metazoans predicts that chitin synthesis begins once *Chs2*-containing vesicles fuse with the plasma membrane (Muthukrishnan et al. 2012). Merzendorfer and Zimoch 2003 described a more speculative model in insects in which chitin has already formed before the vesicles have reached the plasma membrane. Our localization of chitin-containing vesicles in secretory cells confirmed that chitin synthesis occurs intracellularly in the gill filaments of *"B." childressi* (**Supp. Fig.6, b, c**). Moreover, chitin synthesis must occur within the endoplasmic reticulum-Golgi continuum as is synthesized through the secretory pathway. As previously discussed, *Ca*. Endonucleobacter expressed a membrane-piercing TS33 in all stages of its life cycle (**Fig.4**). While sitting in the nucleus, the T3SS would hypothetically allow *Ca*. Endonucleobacter to pierce the inner nuclear membrane and deliver virulence factors such as *ChiA* to the perinuclear space-ER lumen continuum. We hypothesize that *Ca*. Endonucleobacter delivers *ChiA* to the secretory pathway of its host cell by piercing the inner nuclear membrane with a T3SS. Once in the secretory pathway, *ChiA* could

hydrolyze chitin that is being synthesized by the host cell of *Ca*. Endonucleobacter or be secreted to the extracellular medium to digest chitin present in the mucus layer.

Our bulk transcriptomic data revealed that *Ca*. Endonucleobacter expressed the sugar phosphotransferase system PTS, as well as *NagA* and GFAT (**Fig.3**). The three factors were also expressed along the infectious cycle (**Fig.4**). Although the PTS system of *Ca.* Endonucleobacter was annotated as specific for glucose import, there are reports of PTS transporters that can import glucose and GlcNAc (Al Makishah and Mitchell 2013). NagA and GFAT participate in the transformation of GlcNAc into D-fructose-6-phosphate, which is ready to enter in glycolysis. This data suggested that *Ca*. Endonucleobacter is not only hydrolyzing host chitin, but also profiting nutritionally from it. Still, *Ca*. Endonucleobacter sits in the nucleus, while chitin digestion takes place in the secretory pathway of its host cell or in the extracellular mucus layer. How does *Ca*. Endonucleobacter access the saccharides resulting from chitin hydrolysis? Chitinase from *Ca*. Endonucleobacter was classified as endochitinase (3.2.1.14) by RAST annotation. Endochitinases hydrolyze the chitin internal chain generating soluble chitooligosaccharides such as chitobiose (Le and Yang 2018). Whether chitin hydrolysis takes place in the secretory pathway of the host cell, it escapes our understanding how chitobiose is translocated to the cytoplasm. Remarkably, the host cell encoded and expressed the SWEET sugar transporter 1 (SWEET), which was upregulated after infection by *Ca*. Endonucleobacter (**Fig.4**). SWEET transporters internalize disaccharides into the eukaryotic cell, and they are rarely found in animal cells (Meyer et al. 2011). We envision the possibility that the host cell uses the SWEET transporters to import chitobiase disaccharides that result from the hydrolysis of extracellular chitin. Chitobiose is hydrolyzed into GlcNAc monomers by the enzyme

chitobiase. Although *Ca*. Endonucleobacter encoded and expressed the molecular machinery for the uptake and metabolism of GlcNAc (PTS, *NagA*, GFAT), we did not find chitobiase in its genome. Intriguingly, the host cell encoded and expressed chitobiase (CTBS), which was upregulated after infection by *Ca*. Endonucleobacter (**Fig.4**). It is tempting to speculate that *Ca*. Endonucleobacter relies on the enzymatic activity of the host CTBS to incorporate GlcNAc, the basic component of chitin. For *Ca*. Endonucleobacter being able to import GlcNAc, this molecule must enter into the nucleus. Nuclear pore complexes (NPC) are macromolecular complexes that mediate selective and passive transport of molecules between the nucleus and the cytoplasm in all eukaryotic cells (Christie et al. 2016; Knockenhauer and Schwartz 2016). NPC are formed by hundreds of conserved proteins collectively referred as nups, from which the subtype FG nups give the transport selectivity to the NPC (Timney et al. 2016). *In vivo* studies done over the last two decades have determined a size range of 30-60 kDa, under which a molecule can freely diffuse across the NPC (Keminer and Peters 1999; Ma et al. 2012; Mohr et al. 2009; Ribbeck and Görlich 2001). With a size of approximately 220 Da, GlcNAc can diffuse passively between the cytoplasm and the nucleus. *Ca*. Endonucleobacter can then access a continuous supply of GlcNAc that is being produced in the host cell cytoplasm through the hydrolysis of chitobiose by CTBS. After internalization through the PTS system, *Ca*. Endonucleobacter incorporates GlcNAc to the glycolysis-TCA axis, previous deacetylation and isomerization by *NagA* and GFAT, respectively.

Ca. *Endonucleobacter stimulates the formation of host lipid droplets*

Infection by *Ca*. Endonucleobacter triggered the host cell upregulation of most of the enzymes from the glycerolphosphate pathway for triacylglycerides (TAG) synthesis

(GPAT, AGPAT, PAP, DGAT), as well as the upregulation of the lipid droplet (LD) marker perilipin-2 (PLIN2) (**Fig.4**). LD are the main cellular organelles for lipid storage, and they are formed by a hydrophobic core of neutral lipids (TAG) surrounded by a phospholipid bilayer with associated proteins like PLIN2 (Murphy and Vance 1999; Tauchi-Sato et al. 2002). Is *Ca.* Endonucleobacter actively stimulating the formation of LD, or is it a side effect reaction to the infection? One of the cellular stresses that can trigger LD formation in infected eukaryotic cells is depravation of certain nutrients, state known as starvation state (Henne et al. 2018). Nolan et al. 2017 demonstrated that the amount of LD in cells infected by the intracellular parasite *T. gondii* progressively increases until the onset of parasite replication. *Ca*. Endonucleobacter expressed sugar (PTS), lipids (FATP), polypeptides (*OppB*, *OppC*, *PotF*, *PotG*) and amino acid (*YuiF*, SERT, MSBP) importers in early stage of infection (**Fig.4**), suggesting immediate access to the host cell energy budget after nuclear colonization. Moreover, infection by *Ca*. Endonucleobacter triggered the host cell upregulation of sugar (SWEET) and amino acid importers (rBAT, SCNA) (**Fig.4**), indicating that the host cell started craving for nutrients. Our data suggested that *Ca*. Endonucleobacter brings the host cell to starvation state by consuming its energy budget during early stage of infection. Thus, *Ca*. Endonucleobacter might passively promote the formation of LD by bringing the host cell to starvation state. Noteworthy, *Ca*. Endonucleobacter expressed *IpgD* along its life cycle (**Fig.4**). As discussed previously, the *IpgD*-T3SS tandem might play an important role during host cell colonization. Which could be the function of *IpgD* during mid and late stage of infection? *IpgD* increases the host intracellular levels of phosphatidylinositol 5-phosphate (PtdIns5P), which has major repercussions in cytoskeleton dynamics (Viaud et al. 2014). Remarkably, the host cell upregulated the expression of a GTPase IMAP family member 4 (GIMAP) as a response of *Ca*.

Endonucleobacter infection (**Fig.4**). GIMAP is a nucleotide-binding protein from the GTP-ase family involved in cytoskeleton dynamics and evolutionary related to the septin family (Schwefel et al. 2010). Akil et al. 2016 demonstrated that PtdIns5P interacts with septin 9 promoting perinuclear accumulation of LD in a cytoskeleton-dependent manner. It is tempting to speculate that the increasing levels of intracellular PtdIns5P due to the expression of *IpgD* interact with GIMAP, resulting in the formation of LD. If this holds true, *Ca.* Endonucleobacter would not only stimulate the formation of LD passively through the consumption of the host energy budget, but also actively through the manipulation of cytoskeleton components. But, can *Ca*. Endonucleobacter benefit nutritionally from host LD?

Ca. Endonucleobacter bulk transcriptome showed high levels of expression of lipase and expression of factors involved in fatty acid import and metabolism (FATP, LC-FACS, ACAD) (**Fig.3**). These factors were also expressed in all stages of *Ca.* Endonucleobacter life cycle (**Fig.4**). This suggested that *Ca*. Endonucleobacter might be nutritionally exploiting host LD. The enzymes from the glycerolphosphate pathway for TAG synthesis occur primarily in the ER, the organelle where LD are synthesized (Olzmann and Carvalho 2019). How can *Ca*. Endonucleobacter profit nutritionally from LD when they occur outside the nuclear compartment? Secretion prediction on the amino acid sequence of *Ca*. Endonucleobacter lipase revealed that it was probably secreted through T3SS. It is again tempting to speculate that *Ca*. Endonucleobacter might be using the T3SS to deliver virulence factors like lipase to the endoplasmic reticulum lumen, resulting in the digestion of TAG. Our localization of LD in the gill filaments of *"B." childressi* did not show accumulation of LD in infected cells by *Ca*. Endonucleobacter (**Supp. Fig.7**). If the host cell upregulated most of the enzymes involved in LD biosynthesis after *Ca*. Endonucleobacter infection, why infected cells

do not accumulate LD? At low concentration, TAG are dispersed between the leaflets of the endoplasmic reticulum bilayer. Only an increase in TAG concentration leads to the formation of oil lenses that will finally grow through incorporation of TAG, forming LD (Olzmann and Carvalho 2019). Remarkably, *Ca*. Endonucleobacter lipase had already high expression values in all stages of its life cycle (**Fig.4**). The absence of LD-like structures in **Supp. Fig.7** could be explained by our transcriptomic data: Host LD synthesis is upregulated after *Ca*. Endonucleobacter infection, while lipase is highly expressed by *Ca*. Endonucleobacter from early stage on. We hypothesize that TAG cannot accumulate to form LD, as they are being digested by *Ca*. Endonucleobacter as soon as they are synthesized by the host cell.

In early stage of infection, the host cell upregulated the sugar importer SWEET, the factors from the glycolysis pathway FBPA and PKM, and maintained and upregulated the expression of the oxidative phosphorylation factors ATP_{α} and ATP_{β}, respectively (**Fig.4**). This data suggested that the host cell upregulated sugar import and oxidative metabolism after *Ca*. Endonucleobacter infection. Why would the host cell upregulate sugar import and oxidative metabolism as a reaction to the infection? As discussed previously, *Ca*. Endonucleobacter brings the host cell to starvation state by consuming its energy budget in the early stage of the infectious cycle. Animal cells employ a myriad of intracellular nutrient sensors that are sensitive to glucose, amino acids or lipids shortage. Once the shortage is reported by the intracellular nutrient sensors, a signaling cascade is triggered, resulting in upregulation of transporters or activation of certain biosynthesis routes that depend on the glycolysis-TCA axis (Efeyan et al. 2015). A plausible explanation to the host cell upregulation of sugar import and oxidative metabolism are the nutritional demands of *Ca*. Endonucleobacter. We envision the possibility that *Ca*. Endonucleobacter hides in the nucleus while

continuously depleting the host cell energy budget, setting the host cell in a perpetual starvation state. Then, the host cell would unsuccessfully try to answer the metabolic demands of a “hungry nucleus” by importing sugars and activating the glycolysis-TCA axis.

The host cell also upregulated GPDH and FAS in early stage of infection (**Fig.4**). GPDH and FAS synthesize glycerol-3-phosphate (Gro3P) and mono acyl CoA (FA CoA) respectively, which are the initial building blocks of the glycerolphosphate pathway for TAG biosynthesis. Noteworthy, FAS uses 14 NADPH and 7 ATP to synthesize a single C_{16} palmitic acid (Carta et al. 2017). Thus, it is safe to assume that TAG biosynthesis is an energy-demanding process. We hypothesize that the host cell might also upregulate sugar import and its oxidative metabolism to pay the energetic costs of LD biosynthesis.

As discussed previously, the host cell can metabolize chitobiose disaccharides into GlcNAc by expressing CTBS, which was upregulated as a reaction to *Ca.* Endonucleobacter infection (**early stage, Fig.4**). The host cell upregulated again CTBS in mid stage of infection, as well as *NagA* and GFAT (**Fig.4**). This data indicated that the host cell might be incorporating GlcNAc into the glycolysis-TCA axis. Intriguingly, the host cell upregulation of *NagA* and GFAT in mid stage correlated in time with the highest values of *ChiA* expressed by *Ca*. Endonucleobacter. It is tempting to speculate that *Ca*. Endonucleobacter hydrolyzes host chitin into chitobiose, which can simultaneously fuel both *Ca*. Endonucleobacter and the host cell oxidative metabolisms.

In late stage of infection, the host cell downregulated sugar transport (SWEET), chitin synthase (*Chs2*) and LD biosynthesis (GPAT, AGPAT, PAP, DGAT) (**Fig.4**).

Intriguingly, the host cell upregulated GlcNAc metabolism (*NagA* and GFAT) and the glycolysis pathway (PFK-1, FBPA, PKM) (**Fig.4**). This data suggested that the host cell ceased sugar import, chitin and LD synthesis towards the end of the infectious cycle. We hypothesize that the host cell enters in metabolic decline at late stage of infection, downregulating the SWEET transporter. This would make the host cell unable to sustain the energetically costly processes of chitin and LD biosynthesis. The host cell upregulated GlcNAc metabolism as well as the glycolysis pathway at the end of the infectious cycle. Why would the host cell upregulate the oxidative metabolism when entering in metabolic decline? One possible answer to this question is that the host cell is trying to maintain its own metabolism regardless of the nutritional demands of *Ca*. Endonucleobacter. We hypothesize that the growing bacterial population turns the host cell into a funnel of nutrients from the extracellular medium that stays alive and metabolically functional even at late stage of infection.

Ca. *Endonucleobacter microengineers its replication niche by interfering with cytoskeleton components*

The host cell stays alive and metabolically functional under a huge mechanical distress, as *Ca*. Endonucleobacter increases the volume of the infected nucleus up to 50 fold. How does *Ca*. Endonucleobacter manage to bring the host cell under such a mechanical deformation without compromising its viability?

As discussed previously, *Ca*. Endonucleobacter expressed *IpgD* along its life cycle (**Fig.4**). Niebuhr et al. 2002 demonstrated that in HeLa cells transfected with *IpgD*, the production of intracellular PtdIns5P lead to disappearance of actin stress fibers. Tavares et al. 2017 showed that the organization of actin stress fibers in a human

breast cell line lead to cell stiffening. This results are in consonance with the results from Niebuhr et al. 2002, which showed that the disappearance of actin stress fibers lead to a decrease in membrane tension in HeLa transfected cells, which adopted rounded morphology. It is tempting to speculate that *Ca*. Endonucleobacter is actively decreasing the host cell stiffness by increasing the intracellular levels of PtdIns5P. Additionally, the host cell upregulated the expression of phosphatidylinositol 4, 5-biphosphate phosphodiesterase (PLCG) as a reaction to *Ca*. Endonucleobacter infection (**Fig.4**). PLCG is a lipolytic enzyme that plays a major role in the intracellular IP3/Ca^{2+} signaling pathway, as it increases the intracellular levels of IP3 (Meldrum et al. 1991). The release of IP3 results in the release of Ca^{2+} ions from the endoplasmic reticulum, activating calcium-sensitive proteins. The infection by *Ca*. Endonucleobacter triggered the host upregulation of the cysteine protease calpain-5 (CAPN5), a calcium-dependent protease known to exist widely in animal tissue (Melloni and Pontremoli 1989). The activated CAPN5 acts on the host endogenous proteins, preferentially on the cytoskeleton component spectrin (Boivin et al. 1990). As cytoskeleton component, spectrin plays a major role in maintaining the morphology and organization of cell membranes (Morrow et al. 2011). Moreover, spectrin plays a critical role connecting the cytoskeleton elements to each other, the cell membrane and the nucleus (Liem 2016). Although the previous mechanisms have been described in mammalian cells, one could envision the possibility that *Ca*. Endonucleobacter interferes with the host cell cytoskeleton by expressing *IpgD* and triggering the proteolytic activity of CAPN5. The subsequent loss of host cell stiffness would allow *Ca*. Endonucleobacter to subject the host cell to mechanical distress without compromising its viability. Can *Ca*. Endonucleobacter profit in another way from the disentanglement of the host cell cytoskeleton?

Najm et al. 1991 demonstrated that the activation of calpain increases the intracellular levels of oligopeptides such as spermine, spermidine and putrescine, which result from the proteolysis of spectrin. The upregulation of CAPN5 by the host cell was followed by the upregulation of the oligopeptides importers *OppB, OppC, PotF* and *PotG* by *Ca*. Endonucleobacter along the infectious cycle (**Fig.4**). As presented in the results section, *Ca*. Endonucleobacter depends on the host for the acquisition of several amino acids. We hypothesize that the production of oligopeptides that results from the pathogenic activation of CAPN5 supplies *Ca*. Endonucleobacter with the amino acids that cannot synthesize *de novo*.

Conclusion and outlook

This work describes the molecular biology of the intranuclear parasite *Ca*. Endonucleobacter and how its host cell reacts to the infection. The intranuclear lifestyle of *Ca.* Endonucleobacter might have originated when certain *Endozoicomonas* acquired the capacity to invade the nucleus of their host cell. *Ca*. Endonucleobacter colonizes non-phagocytic cells from the ciliated edge of *"B". childressi* gill filaments by forcing its uptake using the T3SS-*IpgD* tandem. Once in the nucleus, one would expect that *Ca*. Endonucleobacter would make use of host chromatin as a nutritional source. Instead, the intranuclear parasite shows the Machiavellian intentions of avoiding chromatin digestion and keeping the host cell alive using inhibitors of apoptosis. This contrasts with was has been described for other intranuclear parasites such as *C. caryophilus,* which nutritionally exploits the nucleus of its host cell. *Ca*. Endonucleobacter stimulates the host cell to form LD, from which it might profit nutritionally. Though, the massive expression of chitinase by *Ca*. Endonucleobacter leads us to think that host chitin is its main nutritional source. This is especially interesting considering that *Ca*. Endonucleobacter sits in the nucleus while chitin is mainly found in the extracellular medium. It is tempting to speculate that *Ca*. Endonucleobacter hijacks the secretory pathway of its host cell by delivering chitinase into the endoplasmic reticulum lumen through its T3SS, resulting in the hydrolysis of extracellular chitin into chitobiose. We hypothesize that *Ca*. Endonucleobacter sits in the nucleus while hiding from cytoplasmic molecular defenses and consuming host resources rather than chromatin. The host cell tries then to satisfy the nutritional demands of a "hungry nucleus" by importing chitobiose from the extracellular medium. Thus, *Ca*. Endonucleobacter uses the host cell as the living funneling agent that

creates a constant flow of nutrients between the extracellular medium and the nucleus. Whether *Ca.* Endonucleobacter is hijacking the secretory pathway of its host cell has major repercussions. The hypothesized strategy is fascinating per se, but also intriguing if we consider how a *Ca*. Endonucleobacter colony organizes in late stage of infection. In theory, only the *Ca.* Endonucleobacter cells in direct contact with the inner nuclear membrane would be able to deliver virulence factors to the ER lumen through a T3SS, implying a division of labor strategy. By using transmission electron microscopy, we intend to resolve the ultrastructure of a *Ca*. Endonucleobacter-infected nucleus. We contemplate the imaging of a T3SS piercing the inner nuclear membrane as a strong support to our transcriptomic-based hypotheses.

Acknowledgements

We thank the captain, crew and ROV team on the cruise RV Meteor Nautilus NA-58 to the Gulf of México in May 2015 and the chief scientist Dr. Eric Cordes. Thank you Silke Wetzel for the sample treatment and fixation on board. We also thank Andreas Ellrott and Daniela Tienken from the Max Planck Institute for Marine Microbiology for technical assistance. We thank Prof. Dr. Matthias Horn for the discussions and input during the development of the project. Thank you to Dr. Bruno Hüttel and Lisa Czaja-Hasse from the Max Planck-Genome-Centre of Cologne for all the sequencing services and the optimization of the CATS protocol.

This study was funded by the Max Planck Society, an ERC Advanced Grant (BathyBiome, 340535), and a Gordon and Betty Moore Foundation Marine Microbial Initiative Investigator Award to ND (Grant GBMF3811).

Materials and methods

Sample collection

Mussels were collected with the remote operated vehicle (ROV) Hercules 4000m during the RV Meteor Nautilus NA-58 cruise to the Gulf of Mexico in May 2015 (**Fig.2, a**). "*B". childressi* mussels were collected during five dives from the Mississippi Canyon site (MC853, 28º07' N; -089º08' W) and the Green Canyon site (GC234, 27º45' N; -091º13' W) at a water depth of 1,070 and 540 m, respectively.

DNA extraction and infection screening

To look for *Ca*. Endonucleobacter infection, 14 "*B". childressi* gill samples were PCR screened (**Supp. Table. 1**). DNA was extracted from gill samples using the DNeasy Blood and Tissue Kit (QIAGEN, Germany) and used as template in PCR reactions. The *Ca*. Endonucleobacter 16S rRNA gene was amplified by PCR using the following conditions: Initial denaturation for 3 min at 95 °C, followed by 30 cycles at 95 °C for 30 s, 55 °C for 30 s and 72 °C for 2 min, followed by a final elongation step at 72 °C for 10 min. The *Ca*. Endonucleobacter 16S rRNA gene was amplified using the specific forward primer BNIX64 (AGCGGTAACAGGTCTAGC) (Zielinski et al. 2009) and the general reverse primer GM4 (TACCTTGTTACGACTT) (Muyzer et al. 1995). Taq DNA Polymerase (5 PRIME, Hamburg, Germany) was used in all PCR reactions. PCR products were loaded in agarose gels (2%) and stained with ethidium bromide for 30 min. Band thickness and intensity were considered as indicators of the degree of infection.

DNA library preparation and sequencing

The DNA from the three most infected specimens of the 14 "*B". childressi* gill samples screened was sequenced. DNA quality was assessed with an Agilent 2100 Bioanalyzer. Genomic DNA was sequenced using Illumina HiSeq 2500 technology as follows: 33 million 250 bp paired-end reads were generated. Additionally, the DNA from the most infected specimen out of the previous three was further sequenced using Illumina HiSeq 2500 technology and PacBio technology as follows: 333 million long reads were generated.

Genome analysis

To estimate the abundance of *Ca*. Endonucleobacter within the three DNA libraries, we quantify its 16S rRNA gene sequences using phyloFlash v3.3 (Gruber-Vodicka et al. 2019). Three individual metagenomes were assembled using Spades v3.7 (Bankevich et al. 2012), after decontamination, quality filtering (trimq = 2) and adapter-trimming of the short illumina reads with the BBDuk tool from the BBMap suite (Bushnell B, https://sourceforge.net/projects/bbmap/). The three *Ca*. Endonucleobacter genomes were binned based on genome coverage, GC content and taxonomic affiliation using Gbtools v2.6.0 (Seah and Gruber-Vodicka 2015). To reassemble the genomes, we re-mapped Illumina short reads to the bins using BBMap from the BBMap suite (Bushnell B, https://sourceforge.net/projects/bbmap/) with 0.98 minimum identity. The reads were reassembled with Spades v3.7, using a maximum k-mer length of 127. Following manual removal of contigs shorter than 1 kB and contamination screening using Bandage v0.8.1 (Wick et al. 2015), quality metrics were calculated with CheckM v1.0.18 (Parks et al. 2015). Among the three draft genomes, the one with higher completeness and lower fragmentation (number of contigs) was

chosen for additional Illumina HiSeq 2500 sequencing (see previous section). A new short-reads draft genome was assembled from the additional Illumina sequencing library and chosen to proceed further. The PacBio long reads were mapped against the chosen *Ca*. Endonucleobacter Illumina draft genome using the long read mapper ngmlr v.0.2.7 (Sedlazeck et al. 2018). Mapped long-reads were assembled using the long-read assembler CANU v2.0 (Koren et al. 2017). Following an hybrid assembly strategy, the CANU bin was loaded in Unicycler v0.4.8 (Wick et al. 2017) and supplemented with the illumina short-reads from the Illumina draft genome. The resulting hybrid genome was checked for quality metrics with CheckM v1.0.18 and contigs shorter than 1 kB were manually removed. The hybrid genome was annotated with RAST v2.0 (Aziz et al. 2008). The annotations were manually cross-checked and the annotations for the main genes discussed here were verified using v.2.10.1 NCBI's BLAST. The protocol described above was also applied to obtain a draft genome of the methane-oxidizing symbiont. This draft genome was used posteriorly to clean the host *de novo* transcriptome of bacterial reads.

Bulk transcriptomics: RNA extraction, library preparation and sequencing

The ciliated edges from 8-10 gill filaments of RNAlater-preserved samples from the three most infected "*B". childressi* specimens previously screened for infection were dissected. Total RNA was extracted from the ciliated edge samples using the RNeasy Mini Kit (QIAGEN, Germany) following manufacturer protocol. After extraction, RNA quantity was assessed with a QUANTUS Fluorometer (Promega, Germany) and kept at -80 °C until library preparation and sequencing. Library preparation and sequencing have been performed at the Max Planck Genome Center Cologne, Germany (https://mpgc.mpipz.mpg.de/home/). 50 ng genomic DNA have initially been

fragmented via sonication (Covaris S2, Covaris), followed by library preparation with NEBNext Ultra DNA v2 Library Prep Kit for Illumina (New England Biolabs). Library preparation included 7 cycles of PCR amplification. Quality and quantity were assessed at all steps via capillary electrophoresis (TapeStation, Agilent Technologies) and fluorometry (Qubit, ThermoFisher Scientific). Libraries were immobilized and processed onto a flow cell with cBot (Illumina) and subsequently sequenced on HiSeq3000 system (Illumina) with 2 x 150 bp paired end reads. 33 million paired-end reads were generated.

Laser-capture microdissection of host infected nuclei

To investigate the transcriptomic profile of "*B". childressi* cells infected by *Ca*. Endonucleobacter at different infection stages, the formalin-fixed gills of a "*B". childressi* specimen (H1423/002/N6) were embedded in polyester wax and sectioned using a conventional microtome. Sections were mounted on thermoexitable POL-membranes (No. 115005191; Leica, Germany) and hybridized with specific *Ca*. Endonucleobacter 16S rRNA oligonucleotide probes labeled with Atto550. The hybridization protocol was identical to the one described for the Whole-mountFISH (see "Whole-mount fluorescence *in situ* hybridization" section) with the following exceptions: Hybridization buffer did not contained formamide, only *Ca*. Endonucleobacter 16S rRNA was hybridized, sections were not DAPI-stained, sections were left air-exposed after washing step (no mounting). Hybridized samples were taken to the microdissection station Leica LMD6500 (Leica, Germany). We defined three infection stages (treatments) along *Ca*. Endonucleobacter life cycle: Early (E), Mid (M) and Late (L). Per treatment, 100 nuclei were microdissected and pooled in a single tube prefilled with 30 µl of extraction buffer (AllPrep DNA/RNA FFPE

kit; Qiagen, Germany). As negative control for host expression, an additional treatment of non-infected nuclei was included in the study. Per treatment, samples were done in triplicate (12 samples in total).

LCM transcriptomics: RNA extraction, library preparation and sequencing

Following microdissection, RNA extraction was done using the AllPrep DNA/RNA FFPE kit (Qiagen, Germany) following manufacturer protocol with the following modifications. Samples were incubated in proteinase K overnight at 37°C. Elution buffer was pre-warmed at 37°C and added to the column membrane. Elution buffer incubation time was doubled. After a first elution step, the eluent was loaded again in the membrane and incubated for 2 additional minutes prior second and final elution. After extraction, RNA quantity was assessed with a QUANTUS Fluorometer (Promega, Germany) and kept at -80 °C until library preparation and sequencing. Library preparation and sequencing have been performed at the Max Planck Genome Center Cologne, Germany (https://mpgc.mpipz.mpg.de/home/). Total RNA has been used for library preparation with CATS (Capture and Amplification by Tailing and Twitching, CATS) (Turchinovich et al. 2014) RNA-seq Kit v2 (Diagenaode) according to the manufacturer's specifications. Library preparation included 16 cycles of PCR amplification. Quality and quantity were assessed at all steps via capillary electrophoresis (TapeStation, Agilent Technologies) and fluorometry (Qubit, Thermo Fisher Scientific). Libraries were immobilized and processed onto a flow cell with cBot (Illumina) and subsequently sequenced on HiSeq3000 system (Illumina) with 1 x 150 bp single end reads. Total RNA libraries were sequenced using the Illumina HiSeq 2500 generating single-end reads with 150 bp length (**detailed in Supp. Table. 2**).

Host* de novo *transcriptome: RNA extraction, library preparation and sequencing

To study the host response to the infection, we assembled a "*B". childressi de novo* transcriptome. To do so, a single RNAlater preserved gill sample from the same specimen used for microdissection (H1423/002/N6) was used. The ciliated edge of 15-20 gill filaments were dissected and pulled in equal amounts into two samples considered technical replicates. RNA was extracted using the RNeasy Mini Kit (QIAGEN, Germany) following manufacturer protocol. After extraction, RNA quantity was assessed with a QUANTUS Fluorometer (Promega, Germany) and kept at -80 °C until library preparation and sequencing. Library preparation and sequencing have been performed at the Max Planck Genome Center Cologne, Germany (https://mpgc.mpipz.mpg.de/home/). 1 µg of total RNA was used for polyA enrichment with NEBNext poly(A) mRNA Magnetic Isolation Module (New England Biolabs) and subsequent library preparation with NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs). Library preparation included 11 cycles of PCR amplification. Quality and quantity were assessed at all steps via capillary electrophoresis (TapeStation, Agilent Technologies) and fluorometry (Qubit, Thermo Fisher Scientific). Libraries were immobilized and processed onto a flow cell with cBot (Illumina) and subsequently sequenced on HiSeq3000 system (Illumina) with 2 x 150 bp paired end reads. 33 million paired-end reads were generated.

Host* de novo *transcriptome: Assembly, curation and annotation

Quality trimming of reads and adapters removal were done with BBDuk tool from the BBMap suite (Bushnell B, https://sourceforge.net/projects/bbmap/). The biggest fraction of bacterial contaminants were removed mapping the reads against the *Ca*. Endonucleobacter and methane-oxidizing symbiont draft genomes. Nonspecific

bacterial contaminants were removed mapping the reads against the rRNA and tRNA Silva database (Quast et al. 2013). Both mapping steps were done with BBMap tool from the BBMap suite (Bushnell B, https://sourceforge.net/projects/bbmap/) using a minimum identity value of 0.85. Before assembling, reads were normalized using the bbnorm tool from the BBMap suite (Bushnell B, https://sourceforge.net/projects/bbmap/). Completeness and quality of the three assemblies was checked with Trinity Stats package from Trinity v.2.10.0 (Grabherr et al. 2013) and BUSCO v.4.1.2 (metazoan database) (Simão et al. 2015). All reads from the assembly were assigned to a taxonomic category using blast and loaded into MEGAN (Huson et al. 2007). Non-eukaryotic reads were removed from the assembly. The assembly was mapped again against the *Ca*. Endonucleobacter and the methane-oxidizing symbiont draft genomes to ensure minimal contamination. This mapping step was done with BBMap tool from the BBMap suite (Bushnell B, https://sourceforge.net/projects/bbmap/) using a minimum identity value of 0.98. The resulting assembly was annotated using the trinotate package from Trinity v.2.10.0 (Grabherr et al. 2013).

Ca. Endonucleobacter transcriptomic analysis

The analyses from both bulk and LCM transcriptomic reads are contemplated in this section. Quality trimming of reads and adapters removal were done with BBDuk tool from the BBMap suite (Bushnell B, https://sourceforge.net/projects/bbmap/). Non-mRNA contaminants were removed mapping the reads against the rRNA and tRNA Silva database v132 (Quast et al. 2013) with BBMap tool from the BBMap suite (Bushnell B, https://sourceforge.net/projects/bbmap/) using a minimum identity value of 0.85. Bulk transcriptomic reads were pseudoaligned using Kallisto v.0.44.0 (Bray et

al. 2016) for TPM expression quantification. LCM reads were mapped against *Ca*. Endonucleobacter hybrid genome with BBMap using a minimum identity value of 0.85. After mapping, LCM reads were counted with FeatureCounts v1.6.3 (Liao et al. 2014) for differential expression analysis. The differential expression analysis was done with Aldex2 v3.11 (Fernandes et al. 2013) at a level of significance of 0.05.

Proteomics

Proteins were extracted from gill filaments of 12 of the 14 specimens of "*B". childressi* screened for *Ca.* Endonucleobacter infection. We conducted a tryptic protein digestion following the filter-aided sample preparation (FASP) protocol, adapted from Wiśniewski et al. 2009 for all samples. Depending on the amount of available tissue, 100 or 150 µl of SDT-lysis buffer (4% (w/v) SDS, 100 mM Tris-HCl pH 7.6, 0.1 M DTT) were added and boiled samples at 95 °C for 10 min. To minimize sample loss, we did not do the 5 minutes centrifugation step at 21,000g as described in the original FASP protocol Wiśniewski et al. 2009. Instead, only a short spin down was conducted. Subsequently, lysate and UA solution were mixed in a 3:10 ratio (8 M urea in 0.1 M Tris/HCl pH 8.5) in a 10 kDa MWCO 500 µL centrifugal filter unit (VWR International) and the mixture was centrifuged at 14,000g for 40 min. This step was repeated multiple times until the whole lysate was loaded onto the filter unit. Next, we added 200 µL of UA solution and centrifugal filter spun again at 14,000g for 40 min. Subsequently, 100 µL of IAA solution (0.05 M iodoacetamide in UA solution) were added and samples were incubated at 22 °C for 20 min in the dark. We removed the IAA solution by centrifugation following three washing steps with 100 µL of UA solution. Subsequently, filters were washed three times with 100 µL of ABC buffer (50 mM ammonium bicarbonate). We added 1.05 µg of Pierce MS grade trypsin (Thermo Fisher Scientific, Germany) in 40 µL of ABC buffer

to each filter. Filters were incubated overnight in a wet chamber at 37 °C. The next day, we eluted the peptides by centrifugation at 14,000g for 20 min followed by addition of 50 µL of 0.5 M NaCl, shaking at 600 rpm and another centrifugation step. Peptides were quantified using the Pierce MicroBCA Kit (Thermo Fisher Scientific, Germany) following the instructions of the manufacturer. For each run, 1500 ng of peptide solution were loaded onto a 5 mm, 300 µm ID C18 Acclaim® PepMap100 pre-column (Thermo Fisher Scientific, Germany) using an UltiMate™ 3000 RSLCnano Liquid Chromatograph (Thermo Fisher Scientific, Germany) and desalted on the pre-column. After desalting the peptides, the pre-column was switched in line with a 75µm x 75 cm analytical EASY-Spray column packed with PepMap RSLC C18, 2 µm material (Thermo Fisher Scientific, Germany), which was heated to 55 °C. The analytical column was connected via an Easy-Spray source to a Q Exactive HF-X Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, Germany). Peptides were separated on the analytical column using a 460 min gradient as described by Kleiner et al. 2017 with the modification that the gradient went from 98% buffer A (0.1% formic acid) to 31% buffer B (0.1% formic acid, 80% acetonitrile) in 200 min, then from 31% to 50% buffer B in 40 min and ending with 20 min at 99% buffer B. Mass spectra were acquired in the Orbitrap as described by Hinzke et al. 2019 with some modifications. In brief, eluting peptides were ionized via electrospray ionization (ESI) and analyzed in Q Exactive HF-X. Full scans were acquired in the Orbitrap at 60,000 resolution. The 15 most abundant precursor ions were selected in a data dependent manner, isolated with the quadrupole with a 1.2 m/z isolation window size, fragmented in the HCD with 25 NCE, and measured in the Orbitrap at 7,500 resolution. The mass (m/z) 445.12003 was used as lock mass as described in Olsen et al. 2005 with the modification that lock mass was detected in the full scan rather than by

separate selected ion monitoring scan injection. Lock mass use was set to 'best'. Singly charged ions were excluded from MS/MS analysis. Dynamic exclusion was set to 30 s. On average, 258,842 MS/MS spectra were acquired per sample with the 460 min gradient.

Proteomics data processing

A database containing protein sequences from the host "*B". childressi* host as well as from *Ca.* Endonucleobacter was used. The cRAP protein sequence database (http://www.thegpm.org/crap/) containing protein sequences of common laboratory contaminants was appended to the database. The final database contained 38,281 protein sequences. Searches of the MS/MS spectra against this database were performed with the Sequest HT node in Proteome Discoverer version 2.2.0.388 (Thermo Fisher Scientific, Germany) as described previously in Gruber-Vodicka et al. 2019. Only proteins identified with medium or high confidence were retained resulting in an overall false discovery rate of <5%. For protein quantification, normalized spectral abundance factors (NSAFs, Zybailov et al. 2006, Nucleic Acids Research, Volume 44, Issue D1) were calculated per species and multiplied by 100, to give the relative protein abundance in %.

Phylogenomic analysis

To get a better resolution of the phylogenetic relationships of *Ca*. Endonucleobacter with related taxa, we aligned 43 unique genetic markers from 78 genomes using CheckM v1.0.18 (Parks et al. 2015). A total of 2 Ca. Endonucleobacter spp. genomes, 73 *Oceanospirillales* genomes were analyzed together with 5 Vibrionales genomes (outgroup). The 43 marker genes alignment was analyzed using IQ-TREE v.1.6.12

(Nguyen et al. 2015). The tree with the highest likelihood log value was chosen. How the associated taxa clustered together was determined based on 1000 bootstraps.

Phylogenetic analysis

Prior to phylogenetic analysis, a total of 102 16S rRNA sequences (*Ca*. Endonucleobacter and related taxa) were aligned using MAFFT v7.471 (Katoh 2002). The evolutionary history of the 16S rRNA gene was inferred using the TIM3 substitution model in IQ-TREE v.1.6.12 (Nguyen et al. 2015). The tree with the highest likelihood log value was chosen. How the associated taxa clustered together was determined based on 1000 bootstraps.

Chitinase phylogeny and protein domain analyses

45 chitinase amino acid sequences from the G18 glycosidases family were chosen to conduct the analysis. Sequences included chitinases from two *Ca.* Endonucleobacter species and other different representatives of the order *Oceanospirillales*. All sequences were aligned using MAFFT v7.471 (Katoh 2002). Phylogenetic relationship of the studied chitinases was inferred using the TIM3 substitution model in IQ-TREE v.1.6.12 (Nguyen et al. 2015). The tree with the highest likelihood log value was chosen. How the associated taxa clustered together was determined based on 1000 bootstraps. The protein domains of *Ca.* Endonucleobacter chitinase were predicted using the NCBI online service for protein domain prediction (Marchler-Bauer et al. 2011).

Whole-mount fluorescence* in situ *hybridization

The distribution pattern of *Ca*. Endonucleobacter in "*B". childressi* was determined using Whole-mount fluorescence *in situ* hybridization (Whole-mount FISH). To that end, a formalin-fixed "*B". childressi* sample H1423/002-N9-002 was chosen. Three gill filaments were dissected and hybridized with 16S rRNA-targeting probes. The 16 rRNA-targeting probes were in solution within the hybridization buffer (forming the hybridization mixture) containing 35% formamide, 80 mM NaCl, 400 mM Tris HCl, 0.4% blocking reagent for nucleic acids (Roche, Switzerland)), 0.08% SDS (v/v), 0.08 dextran sulfate (w/v), 5 $ng \cdot \mu l^{-1}$ of the Eubacterial 16S rRNA probe EUB 388 labeled with the dye Alexa Fluor 647 (Amann et al. 1990) and 5 $ng \cdot \mu l^{-1}$ of the *Ca*. Endonucleobacter 16S rRNA probe BNIX64 labeled with the dye atto550. The gill filaments were hybridized in 50 µL of hybridization mixture at 46°C for 3 h. Following hybridization, the gill filaments were washed in pre-warmed 48°C washing buffer (0.07 M NaCl, 0.02 M Tris-HCl pH 7.8, 5 mM EDTA pH 8, and 0.01% SDS (v/v)) for 15 min. After washing, the gill filaments were counterstained with DAPI for 10 min at room temperature, transferred to polylysine glass slides and mounted using the ProLong® Gold antifade mounting media (Thermo Fisher Scientific, MA, USA), cured overnight at room temperature and stored -20°C until visualization. All steps during and after hybridization were done in dark conditions.

Lipid droplets and chitin localization

The formalin-fixed paraffin-embedded gills of a *"B." childressi* specimen (H1423/001-N5-002) were cross sectioned at 10 µm using a conventional microtome and mounted on poly-L-lysine-coated glass slides (Sigma-Aldrich, MO, USA) using a water bath. Four sections were left to dry in vertical position at RT for 4 h. Prior to dewaxing,

sections were baked at 60°C for 1 h in vertical position for tissue adherence improvement. The gill sections were dewaxed with Roti®-Histol (Carl-Roth, Germany) in three consecutive steps for 10 min each followed by decreasing ethanol series of 96, 80, 70 and 50% (v/v) for 10 min each. Then, tissue sections were washed in milliQ water for 10 min. The *Ca*. Endonucleobacter 16S rRNA probe BNIX64 labeled with the dye Atto 550 was in solution within the hybridization buffer (forming the hybridization mixture) as explained in the previous section (see "Whole-mount fluorescence in situ hybridization"). A final volume of 500 µL of hybridization mixture was prepared and divided equally between the four *"B." childressi* sections. Before pipetting the hybridization mixture over the samples they were surrounded by adhesive silicon isolators (Sigma-Aldrich, MO, USA) to avoid buffer leaking during hybridization. The samples were then placed in a hybridization chamber. To maintain a humid atmosphere within the hybridization chamber during incubation, KIMTECHScience precision wipes (Kimberly-Clark, TX, USA) partially soaked in formamide 35% were located below the samples. The samples were then hybridized for 3 h at 46°C. Following the hybridization, the samples were rinsed in pre-warmed 48°C washing buffer (see formulation in "Whole-mount fluorescence in situ hybridization") and transferred to fresh pre-warmed washing buffer for 15 min followed by 20 min in 1X PBS, 1 min in milliQ water, a quick wash in ethanol 96% (v/v) and dried at 37°C for 30 min. From this point, two sections followed the lipid droplets localization treatment, and the other two sections the chitin localization treatment. For the lipid droplets localization, we pipetted 50 µL of Nile Red solution (0.5 $\mu g \cdot mL^{-1}$ in PBS) over two dry sections and incubated for 15 min at room temperature in dark conditions. Following incubation, sections were washed during 1 min in milliQ water, followed by a quick wash in ethanol 96% (v/v) and dried at 37°C for 30 min. For the chitin localization, we

used the Fungi-Fluor® Kit (Polysciences, PA, USA) following the manufacturer protocol. After incubation, sections were washed during 1 min in milliQ water, followed by a quick wash in ethanol 96% (v/v) and dried at 37°C for 30 min. After drying, only the lipid droplet localization sections were counterstained with DAPI for 10 min at room temperature followed by a quick wash in miliQ water, a quick wash in ethanol 96% (v/v) and dried at 37°C for 30 min. After drying, all sections were mounted using the ProLong® Gold antifade mounting media (Thermo Fisher Scientific, MA, USA), cured overnight at room temperature and stored -20°C until visualization.

Host 18S rRNA localization

Two sections from the *"B." childressi* specimen H1423/001-N5-002 were treated as described in the previous section (see "Lipid droplets and chitin localization") until hybridization. The *Ca*. Endonucleobacter 16S rRNA probe BNIX64 labeled with the dye Atto 550 and the 18S rRNA probe targeting host rRNA labeled with the dye Alexa 647 were dissolved in hybridization to a final volume of 250 µL of hybridization mixture as described in the "Whole-mount fluorescence in situ hybridization" section, both at final concentrations of 5 ng·µl^{-1}. Sections were isolated by silicon isolators as described in the previous section. The 250 µL of hybridization mixture were divided equally between the two *"B." childressi* sections. The samples were then placed in a hybridization chamber. To maintain a humid atmosphere within the hybridization chamber during incubation, KIMTECHScience precision wipes (Kimberly-Clark, TX, USA) partially soaked in formamide 35% were located below the samples. The samples were then hybridized for 3 h at 46°C. Following the hybridization, the samples were rinsed in pre-warmed 48°C washing buffer (see formulation in "Whole-mount fluorescence in situ hybridization") and transferred to fresh pre-warmed washing buffer

for 15 min followed by 20 min in 1X PBS, 1 min in milliQ water, a quick wash in ethanol 96% (v/v) and dried at 37°C for 30 min. After drying, sections were counterstained with DAPI for 10 min at room temperature followed by a quick wash in miliQ water, a quick wash in ethanol 96% (v/v) and dried at 37°C for 30 min.After drying, all sections were mounted using the ProLong® Gold antifade mounting media (Thermo Fisher Scientific, MA, USA), cured overnight at room temperature and stored -20°C until visualization.

Transmission electron microscopy

Upon recovery, *"B". childressi* specimens were fixed in 2.5% glutaraldehyde (GA) in PHEM buffer (piperazine-N, N′-bis , 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, ethylene glycol-bis(β-aminoethyl ether and MgCl2 (Montanaro et al. 2016)). After fixation, samples were stored in buffer for transport. Samples were post fixed with 1% (v/v) osmium tetroxide (OsO4) for 2 h at 4 °C, washed three times with PHEM and dehydrated with an ethanol series (30%, 50%, 70%, 80%, 90% and 100% (v/v)) at -10 °C, each step lasting 10 min. The sample was then transferred into 50:50 ethanol and acetone, followed by 100% acetone and infiltrated with low-viscosity resin (Agar Scientific, UK) using centrifugation embedding (McDonald 2014). Samples were centrifuged for 30s in resin: acetone mixtures of 25%, 50%, 75% and twice in 100%. Finally they were transferred into fresh resin in embedding molds and polymerized at 60–65 °C for 48 h. Ultra-thin (70 nm) sections were cut on a microtome (Ultracut UC7 Leica Microsystem, Austria), mounted on formvar-coated slot grids (Agar Scientific, United Kingdom) and contrasted with 0.5% aqueous uranyl acetate (Science Services, Germany) for 20 min and with 2% Reynold's lead citrate for 6 min. Sections were imaged at 20–30 kV with a Quanta FEG 250 scanning electron microscope (FEI

Company, USA) equipped with a STEM detector using the xT microscope control software ver. 6.2.6.3123.

Fluorescent microscopy

Whole-filament overviews (**Fig.2, c**) were visualized with the epifluorescence microscope Olympus BX53 (Olympus, Germany). The adhesion-patch overviews (**Fig.2, d**), the ciliated edge snapshots (**Fig.2, e**), the lipid droplets- (**Supp. Fig.7**), the chitin-localization (**Supp. Fig.6, b**) and the 18S rRNA intensity sections (**Supp.Fig.2**) were visualized with the dual system Zeiss LSM 780 & Airyscan detector (Carl Zeiss Microscopy GmbH, Germany). The samples were continuously illuminated using different excitation sources depending on the fluorophore used. Epifluorescence images were taken with the objective Olympus UCPlanFL 20X/0.70 air transmission using an Orca Flash 4.0 camera (Hamamatsu, Japan). LSM and Airyscan images were taken with two different objectives: A plan-APROCHROMAT 63X/1.4 oil immersion objective and a plan-APROCHROMAT 100X/1.46 DIC M27 Elyra oil immersion objective using an Andor iXon Ultra 897 High Speed EMCCD Camera (Andor, UK). Beam selection and modulation of the laser intensities were controlled in several ways, depending on the laser wavelength and the sample of study. Epifluorescence images were obtained and post-processed using the Olympus cellSens Dimension software v. 1.18 (Olympus, Germany). LSM and Airyscan images were obtained and post-processed using ZEN software (black edition, 64bits, version: 14.0.1.201) (Carl Zeiss Microscopy GmbH, Germany). Prior to image exporting, histograms were slightly modified to increase the contrast between channels. Exported images were brightness and levels-corrected using the software Adobe Photoshop (version: 12.0) (Adobe Systems, CA, USA).

References for chapter II

Akil, Abdellah, Juan Peng, Mohyeddine Omrane, Claire Gondeau, Christophe Desterke, Mickaël Marin, Hélène Tronchère, Cyntia Taveneau, Sokhavuth Sar, Philippe Briolotti, Soumaya Benjelloun, Abdelaziz Benjouad, Patrick Maurel, Valérie Thiers, Stéphane Bressanelli, Didier Samuel, Christian Bréchot, and Ama Gassama-Diagne. 2016. "Septin 9 Induces Lipid Droplets Growth by a Phosphatidylinositol-5-Phosphate and Microtubule-Dependent Mechanism Hijacked by HCV." *Nature Communications* 7.

Aziz, Ramy K., Daniela Bartels, Aaron Best, Matthew DeJongh, Terrence Disz, Robert A. Edwards, Kevin Formsma, Svetlana Gerdes, Elizabeth M. Glass, Michael Kubal, Folker Meyer, Gary J. Olsen, Robert Olson, Andrei L. Osterman, Ross A. Overbeek, Leslie K. McNeil, Daniel Paarmann, Tobias Paczian, Bruce Parrello, Gordon D. Pusch, Claudia Reich, Rick Stevens, Olga Vassieva, Veronika Vonstein, Andreas Wilke, and Olga Zagnitko. 2008. "The RAST Server: Rapid Annotations Using Subsystems Technology." *BMC Genomics* 9:1–15.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, Sergey I. Nikolenko, Son Pham, Andrey D. Prjibelski, Alexey V. Pyshkin, Alexander V. Sirotkin, Nikolay Vyahhi, Glenn Tesler, Max A. Alekseyev, and Pavel A. Pevzner. 2012. "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." *Journal of Computational Biology* 19(5):455–77.

Beinart, R. A., S. V. Nyholm, N. Dubilier, and P. R. Girguis. 2014. "Intracellular Oceanospirillales Inhabit the Gills of the Hydrothermal Vent Snail *Alviniconcha* with Chemosynthetic, γ-Proteobacterial Symbionts." *Environmental Microbiology Reports* 6(6):656–64.

Bierne, Hélène, and Pascale Cossart. 2012. "When Bacteria Target the Nucleus: The Emerging Family of Nucleomodulins." *Cellular Microbiology* 14(5):622–33.

Boetius, Antje, and Frank Wenzhöfer. 2013. "Seafloor Oxygen Consumption Fuelled by Methane from Cold Seeps." *Nature Geoscience* 6(9):725–34.

Boivin, P., C. Galand, and D. Dhermy. 1990. "In Vitro Digestion of Spectrin, Protein 4.1 and Ankyrin by Erythrocyte Calcium Dependent Neutral Protease (Calpain I)." *International Journal of Biochemistry* 22(12):1479–89.

Bray, Nicolas L., Harold Pimentel, Páll Melsted, and Lior Pachter. 2016. "Near-Optimal Probabilistic RNA-Seq Quantification." *Nature Biotechnology* 34(5):525–27.

Burgdorfer, W., R. L. Anacker, R. G. Bird, and D. S. Bertram. 1968. "Intranuclear Growth of *Rickettsia rickettsii*." *Journal of Bacteriology* 96(4):1415–18.

Cano, I., van Aerle, R., Ross, S., Verner-Jeffreys, D. W., Paley, R. K., Rimmer, G. S. E., Ryder, D., Hooper, P., Stone, D., & Feist, S. W. (2018). Molecular characterization of an *Endozoicomonas*-like organism causing infection in the king scallop (*Pecten*

maximus L.). *Applied and Environmental Microbiology*, *84*(3), 1–14.

Carta, Gianfranca, Elisabetta Murru, Sebastiano Banni, and Claudia Manca. 2017. “Palmitic Acid: Physiological Role, Metabolism and Nutritional Implications.” *Frontiers in Physiology* 8(NOV):1–14.

Christie, Mary, Chiung Wen Chang, Gergely Róna, Kate M. Smith, Alastair G. Stewart, Agnes A. S. Takeda, Marcos R. M. Fontes, Murray Stewart, Beáta G. Vértessy, Jade K. Forwood, and Bostjan Kobe. 2016. “Structural Biology and Regulation of Protein Import into the Nucleus.” *Journal of Molecular Biology* 428(10):2060–90.

Cornelis, Guy R. 2006. “The Type III Secretion Injectisome.” *Nature Reviews Microbiology* 4(11):811–25.

Crawford, E. D., J. E. Seaman, A. E. Barber, D. C. David, P. C. Babbitt, A. L. Burlingame, and J. A. Wells. 2012. “Conservation of Caspase Substrates across Metazoans Suggests Hierarchical Importance of Signaling Pathways over Specific Targets and Cleavage Site Motifs in Apoptosis.” *Cell Death and Differentiation* 19(12):2040–48.

Czech, Michael P. 2000. “PIP2 and PIP3: Complex Roles at the Cell Surface.” *Cell* 100(6):603–6.

Deveraux, Quinn L., and John C. Reed. 1999. “IAP Family Proteins - Suppressors of Apoptosis.” *Genes and Development* 13(3):239–52.

Deveraux, Quinn L., Ryosuke Takahashi, Guy S. Salvesen, and John C. Reed. 1997. “X-Linked IAP Is a Direct Inhibitor of Cell-Death Proteases.” *Nature* 388(6639):300–304.

Ding, Jiun Yan, Jia Ho Shiu, Wen Ming Chen, Yin Ru Chiang, and Sen Lin Tang. 2016. “Genomic Insight into the Host-Endosymbiont Relationship of Endozoicomonas Montiporae CL-33T with Its Coral Host.” *Frontiers in Microbiology* 7(MAR).

Efeyan, Alejo, William C. Comb, and David M. Sabatini. 2015. “Nutrient-Sensing Mechanisms and Pathways.” *Nature* 517(7534):302–10.

Ehrlich, Hermann, Manuel Maldonado, Klaus-dieter Spindler, Carsten Eckert, Thomas Hanke, René Born, Caren Goebel, Paul Simon, Sascha Heinemann, and Hartmut Worch. 2007. “First Evidence of Chitin as a Component of the Skeletal Fibers of Marine Sponges. Part I. Verongidae (*Demospongia*: *Porifera*).” *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 308B(4):347–56.

Elston, R. A. 1986. “An Intranuclear Pathogen [Nuclear Inclusion X (NIX)] Associated with Massive Mortalities of the Pacific Razor Clam, *Siliqua patula*.” *Journal of Invertebrate Pathology* 47(1):93–104.

Fernandes, Andrew D., Jean M. Macklaim, Thomas G. Linn, Gregor Reid, and Gregory B. Gloor. 2013. “ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq.” *PLoS ONE* 8(7).

Fiore, Cara L., Micheline Labrie, Jessica K. Jarett, and Michael P. Lesser. 2015.

"Transcriptional Activity of the Giant Barrel Sponge, *Xestospongia muta* Holobiont: Molecular Evidence for Metabolic Interchange." *Frontiers in Microbiology* 6(APR).

Forget, Nathalie L., and S. Kim Juniper. 2013. "Free-Living Bacterial Communities Associated with Tubeworm (*Ridgeia piscesae*) Aggregations in Contrasting Diffuse Flow Hydrothermal Vent Habitats at the Main Endeavour Field, Juan de Fuca Ridge." *MicrobiologyOpen* 2(2):259–75.

Fujishima, Masahiro, and Yuuki Kodama. 2012. "Endosymbionts in Paramecium." *European Journal of Protistology* 48(2):124–37.

González-Porras, Miguel-Ángel, Rebecca Ansorge, Harald Gruber-Vodicka, Nicole Dubilier and Nikolaus Leisch. "Conquering the nucleus: Inhibitors of apoptosis are the genomic innovation that originated intranuclear lifestyle in *Hahellaceae*." *In Preparation.*

Grabherr, Manfred G. .., Nir Brian J. Haas, Moran Yassour Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, Lin Fan, Raktima Raychowdhury, Qiandong Zeng, Zehua Chen, Evan Mauceli, Nir Hacohen, Andreas Gnirke, Nicholas Rhind, Federica di Palma, Bruce W., and and Aviv Regev Friedman. 2013. "Trinity: Reconstructing a Full-Length Transcriptome without a Genome from RNA-Seq Data." *Nature Biotechnology* 29(7):644–52.

Gruber-Vodicka, Harald R., Nikolaus Leisch, Manuel Kleiner, Tjorven Hinzke, Manuel Liebeke, Margaret McFall-Ngai, Michael G. Hadfield, and Nicole Dubilier. 2019. "Two Intracellular and Cell Type-Specific Bacterial Symbionts in the Placozoan *Trichoplax* H2." *Nature Microbiology* 4(9):1465–74.

Gruber-Vodicka, Harald, Brandon KB Seah, and Elmar Pruesse. 2019. "PhyloFlash — Rapid SSU rRNA Profiling and Targeted Assembly from Metagenomes: Supplementary Information." *BioRxiv* 521922.

Guo, Shuaiqi, Tyler D. R. Vance, Corey A. Stevens, Ilja Voets, and Peter L. Davies. 2019. "RTX Adhesins Are Key Bacterial Surface Megaproteins in the Formation of Biofilms." *Trends in Microbiology* 27(5):453–67.

Haferkamp, Ilka, Stephan Schmitz-Esser, Michael Wagner, Nadjeschka Neigel, Matthias Horn, and H. Ekkehard Neuhaus. 2006. "Tapping the Nucleotide Pool of the Host: Novel Nucleotide Carrier Proteins of *Protochlamydia amoebophila*." *Molecular Microbiology* 60(6):1534–45.

Hayes, Christopher S., Stephanie K. Aoki, and David A. Low. 2010. "Bacterial Contact-Dependent Delivery Systems." *Annual Review of Genetics* 44(1):71–90.

Henne, W. Mike, Michael L. Reese, and Joel M. Goodman. 2018. "The Assembly of Lipid Droplets and Their Roles in Challenged Cells." *The EMBO Journal* 37(12).

Hinzke, Tjorven, Angela Kouris, Rebecca Ayme Hughes, Marc Strous, and Manuel Kleiner. 2019. "More Is Not Always Better: Evaluation of 1D and 2D-LC-MS/MS Methods for Metaproteomics." *Frontiers in Microbiology* 10(FEB):1–13.

Huson, Daniel H., Alexander F. Auch, Ji Qi, and Stephan C. Schuster. 2007. “MEGAN Analysis of Metagenomic Data.” *Genome Research* 17(3):377–86.

Jemaa, Mohamed, Nathalie Morin, Patricia Cavelier, Julien Cau, Jean Marc Strub, and Claude Delsert. 2014. “Adult Somatic Progenitor Cells and Hematopoiesis in Oysters.” *Journal of Experimental Biology* 217(17):3067–77.

Jensen, Sigmund, Sébastien Duperron, Nils Kåre Birkeland, and Martin Hovland. 2010. “Intracellular *Oceanospirillales* Bacteria Inhabit Gills of *Acesta* Bivalves.” *FEMS Microbiology Ecology* 74(3):523–33.

Katharios, Pantelis, Helena M. B. Seth-Smith, Alexander Fehr, José M. Mateos, Weihong Qi, Denis Richter, Lisbeth Nufer, Maja Ruetten, Maricruz Guevara Soto, Urs Ziegler, Nicholas R. Thomson, Ralph Schlapbach, and Lloyd Vaughan. 2015. “Environmental Marine Pathogen Isolation Using Mesocosm Culture of Sharpsnout Seabream: Striking Genomic and Morphological Features of Novel *Endozoicomonas* Sp.” *Scientific Reports* 5(August):17609.

Katoh, K. 2002. “MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform.” *Nucleic Acids Research* 30(14):3059–66.

Keminer, Oliver, and Reiner Peters. 1999. “Permeability of Single Nuclear Pores.” *Biophysical Journal* 77(1):217–28.

Kleiner, Manuel, Erin Thorson, Christine E. Sharp, Xiaoli Dong, Dan Liu, Carmen Li, and Marc Strous. 2017. “Assessing Species Biomass Contributions in Microbial Communities via Metaproteomics.” *Nature Communications* 8(1).

Knockenhauer, Kevin E., and Thomas U. Schwartz. 2016. “The Nuclear Pore Complex as a Flexible and Dynamic Gate.” *Cell* 164(6):1162–71.

Koonin, Eugene, and L. Aravind. 2002. “Koonin EV, Aravind L.. Origin and Evolution of Eukaryotic Apoptosis: The Bacterial Connection. Cell Death Differ 9: 394-404.” *Cell Death and Differentiation* 9:394–404.

Koren, Sergey, Brian P. Walenz, Konstantin Berlin, Jason R. Miller, Nicholas H. Bergman, and Adam M. Phillippy. 2017. “Canu: Scalable and Accurate Long-Read Assembly via Adaptive κ-Mer Weighting and Repeat Separation.” *Genome Research* 27(5):722–36.

Kräter, Martin, Jiranuwat Sapudom, Nicole Bilz, Tilo Pompe, Jochen Guck, and Claudia Claus. 2018. “Alterations in Cell Mechanics by Actin Cytoskeletal Changes Correlate with Strain-Specific Rubella Virus Phenotypes for Cell Migration and Induction of Apoptosis.” *Cells* 7(9):136.

Le, Bao, and Seung Hwan Yang. 2018. “Characterization of a Chitinase from *Salinivibrio* Sp. BAO-1801 as an Antifungal Activity and a Biocatalyst for Producing Chitobiose.” *Journal of Basic Microbiology* 58(10):848–56.

Lee, D. P., A. S. Deonarine, M. Kienetz, Q. Zhu, M. Skrzypczak, M. Chan, and P. C. Choy. 2001. “A Novel Pathway for Lipid Biosynthesis: The Direct Acylation of Glycerol.”

Journal of Lipid Research 42(12):1979–86.

Liao, Yang, Gordon K. Smyth, and Wei Shi. 2014. “FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features.” *Bioinformatics* 30(7):923–30.

Liem, Ronald K. H. 2016. “Cytoskeletal Integrators: The Spectrin Superfamily.” *Cold Spring Harbor Perspectives in Biology* 8(10).

Ma, Jiong, Alexander Goryaynov, Ashapurna Sarma, and Weidong Yang. 2012. “Self-Regulated Viscous Channel in the Nuclear Pore Complex.” *Proceedings of the National Academy of Sciences of the United States of America* 109(19):7326–31.

Al Makishah, Naief H., and Wilfrid J. Mitchell. 2013. “Dual Substrate Specificity of an N-Acetylglucosamine Phosphotransferase System in *Clostridium beijerinckii*.” *Applied and Environmental Microbiology* 79(21):6712–18.

Marchler-Bauer, Aron, Shennan Lu, John B. Anderson, Farideh Chitsaz, Myra K. Derbyshire, Carol DeWeese-Scott, Jessica H. Fong, Lewis Y. Geer, Renata C. Geer, Noreen R. Gonzales, Marc Gwadz, David I. Hurwitz, John D. Jackson, Zhaoxi Ke, Christopher J. Lanczycki, Fu Lu, Gabriele H. Marchler, Mikhail Mullokandov, Marina V. Omelchenko, Cynthia L. Robertson, James S. Song, Narmada Thanki, Roxanne A. Yamashita, Dachuan Zhang, Naigong Zhang, Chanjuan Zheng, and Stephen H. Bryant. 2011. “CDD: A Conserved Domain Database for the Functional Annotation of Proteins.” *Nucleic Acids Research* 39(SUPPL. 1):225–29.

Martano, Giuseppe, Luca Murru, Edoardo Moretto, Laura Gerosa, Giulia Garrone, Vittorio Krogh, and Maria Passafaro. 2016. “Biosynthesis of Glycerol Phosphate Is Associated with Long-Term Potentiation in Hippocampal Neurons.” *Metabolomics* 12(8):1–8.

McDonald, Kent L. 2014. “Rapid Embedding Methods into Epoxy and LR White Resins for Morphological and Immunological Analysis of Cryofixed Biological Specimens.” *Microscopy and Microanalysis* 20(1):152–63.

Meldrum, Eric, Peter J. Parker, and Amanda Carozzi. 1991. “The PtdIns-PLC Superfamily and Signal Transduction.” *BBA - Molecular Cell Research* 1092(1):49–71.

Melloni, E., and S. Pontremoli. 1989. “The Calpains.” *Trends in Neurosciences* 12(11):438–44.

Merzendorfer, Hans, and Lars Zimoch. 2003. “Chitin Metabolism in Insects: Structure, Function and Regulation of Chitin Synthases and Chitinases.” *Journal of Experimental Biology* 206(24):4393–4412.

Meyer, Heiko, Olga Vitavska, and Helmut Wieczorek. 2011. “Identification of an Animal Sucrose Transporter.” *Journal of Cell Science* 124(12):1984–91.

Mohr, Dagmar, Steffen Frey, Torsten Fischer, Thomas Güttler, and Dirk Görlich. 2009. “Characterisation of the Passive Permeability Barrier of Nuclear Pore Complexes.” *EMBO Journal* 28(17):2541–53.

Montanaro, Jacqueline, Daniela Gruber, and Nikolaus Leisch. 2016. “Improved Ultrastructure of Marine Invertebrates Using Non-Toxic Buffers.” *PeerJ* 2016(3):1–15.

Morrow, Jon S., David L. Rimm, Scott P. Kennedy, Carol D. Cianci, John H. Sinard, and Scott A. Weed. 2011. “Of Membrane Stability and Mosaics: The Spectrin Cytoskeleton.” *Comprehensive Physiology*.

Morrow, Kathleen M., Anthony G. Moss, Nanette E. Chadwick, and Mark R. Liles. 2012. “Bacterial Associates of Two Caribbean Coral Species Reveal Species-Specific Distribution and Geographic Variability.” *Applied and Environmental Microbiology* 78(18):6438–49.

Murphy, Denis J., and Jean Vance. 1999. “Mechanisms of Lipid-Body Formation.” *Trends in Biochemical Sciences* 24(3):109–15.

Muthukrishnan, Subbaratnam, Hans Merzendorfer, Yasuyuki Arakane, and Karl J. Kramer. 2012. *Chitin Metabolism in Insects*.

Muyzer, G., A. Teske, C. O. Wirsen, and H. W. Jannasch. 1995. “Phylogenetic Relationships of Thiomicrospira Species and Their Identification in Deep-Sea Hydrothermal Vent Samples by Denaturing Gradient Gel-Electrophoresis of 16S RDNA Fragments.” *Archives of Microbiology* 164(3):165–72.

Najm, Imad, Peter Vanderklish, Amir Etebari, Gary Lynch, and Michel Baudry. 1991. “Complex Interactions Between Polyamines and Calpain-Mediated Proteolysis in Rat Brain.” *Journal of Neurochemistry* 57(4):1151–58.

Navale, Archana M., and Archana N. Paranjape. 2016. “Glucose Transporters: Physiological and Pathological Roles.” *Biophysical Reviews* 8(1):5–9.

Neave, Matthew J., Craig T. Michell, Amy Apprill, and Christian R. Voolstra. 2017. “Endozoicomonas Genomes Reveal Functional Adaptation and Plasticity in Bacterial Strains Symbiotically Associated with Diverse Marine Hosts.” Scientific Reports 7(January):1–12.

Nguyen, Lam Tung, Heiko A. Schmidt, Arndt Von Haeseler, and Bui Quang Minh. 2015. “IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies.” *Molecular Biology and Evolution* 32(1):268–74.

Niebuhr, Kirsten, Sylvie Giuriato, Thierry Pedron, Dana J. Philpott, Frédérique Gaits, Julia Sable, Michael P. Sheetz, Claude Parsot, Philippe J. Sansonetti, and Bernard Payrastre. 2002. “Conversion of PtdIns(4, 5)P2 into PtdIns(5)P by the *S. flexneri* Effector IpgD Reorganizes Host Cell Morphology.” *EMBO Journal* 21(19):5069–78.

Niebuhr, Kirsten, Noureddine Jouihri, Abdelmounaaim Allaoui, Pierre Gounon, Philippe J. Sansonetti, and Claude Parsot. 2000. “IpgD, a Protein Secreted by the Type III Secretion Machinery of *Shigella flexneri*, Is Chaperoned by IpgE and Implicated in Entry Focus Formation.” *Molecular Microbiology* 38(1):8–19.

Nolan, S. J., Romano, J. D., & Coppens, I. (2017). "Host lipid droplets: An important source of lipids salvaged by the intracellular parasite *Toxoplasma gondii*". PLoS

pathogens, 13(6), e1006362.

Olson, Robert J., Sallie W. Chisholm, Richard Devereux, David A. Stahl, R. I. Amann, and B. J. Binder. 1990. “Combination of 16S rRNA-Targeted Oligonucleotide Probes with Flow Cytometry for Analyzing Mixed Microbial Populations.” *Applied and Environmental Microbiology* 56(6):1919–25.

Olzmann, James A., and Pedro Carvalho. 2019. “Dynamics and Functions of Lipid Droplets.” *Nature Reviews Molecular Cell Biology* 20(3):137–55.

Parks, Donovan H., Michael Imelfort, Connor T. Skennerton, Philip Hugenholtz, and Gene W. Tyson. 2015. “CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes.” *Genome Research* 25(7):1043–55.

Qi, Weihong, Maria Chiara Cascarano, Ralph Schlapbach, Pantelis Katharios, Lloyd Vaughan, and Helena M. B. Seth-Smith. 2018. “*Ca*. Endozoicomonas cretensis: A Novel Fish Pathogen Characterized by Genome Plasticity.” *Genome Biology and Evolution* 10(6):1363–74.

Quast, Christian, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, and Frank Oliver Glöckner. 2013. “The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools.” *Nucleic Acids Research* 41(D1):590–96.

Ray, Katrina, Benoit Marteyn, Philippe J. Sansonetti, and Christoph M. Tang. 2009. “Life on the inside: The Intracellular Lifestyle of Cytosolic Bacteria.” *Nature Reviews Microbiology* 7(5):333–40.

Ribbeck, K., and D. Görlich. 2001. “Kinetic Analysis of Translocation through Nuclear Pore Complexes.” *EMBO Journal* 20(6):1320–30.

Sannasi, A., and H. R. Hermann. 1970. “Chitin in the *Cephalochordata*, *Branchisotoma floridae*.” *Experientia* 26(4):351–52.

Satir, P. 1980. “Structural Basis of Ciliary Movement.” *Environmental Health Perspectives* VOL. 35(April):77–82.

Sato, Tomoyuki, Hirokazu Kuwahara, Kazuma Fujita, Satoko Noda, Kumiko Kihara, Akinori Yamada, Moriya Ohkuma, and Yuichi Hongoh. 2014. “Intranuclear Verrucomicrobial Symbionts and Evidence of Lateral Gene Transfer to the Host Protist in the Termite Gut.” *ISME Journal* 8(5):1008–19.

Schmitz-esser, Stephan, Nicole Linka, Astrid Collingro, L. Beier, H. Ekkehard Neuhaus, Michael Wagner, Matthias Horn, and Cora L. Beier. 2004. “ATP / ADP Translocases : A Common Feature of Obligate Intracellular Amoebal Symbionts Related to Chlamydiae and Rickettsiae.” *Journal of Bacteriology* 186(3):683–91.

Schreiber, Lars, Kasper U. Kjeldsen, Peter Funch, Jeppe Jensen, Matthias Obst, Susanna López-Legentil, and Andreas Schramm. 2016. “*Endozoicomonas* Are Specific, Facultative Symbionts of Sea Squirts.” *Frontiers in Microbiology* 7(JUL):1–15.

Schuett, Christian, Hilke Doepke, Annette Grathoff, and Michael Gedde. 2007. “Bacterial Aggregates in the Tentacles of the Sea Anemone *Metridium senile*.” *Helgoland Marine Research* 61(3):211–16.

Schulz, Frederik, and Matthias Horn. 2015a. “Intranuclear Bacteria: Inside the Cellular Control Center of Eukaryotes.” *Trends in Cell Biology* 25(6):339–46.

Schulz, Frederik, Ilias Lagkouvardos, Florian Wascher, Karin Aistleitner, Rok Kostanjšek, and Matthias Horn. 2014. “Life in an Unusual Intracellular Niche: A Bacterial Symbiont Infecting the Nucleus of Amoebae.” *ISME Journal* 8(8):1634–44.

Schwefel, David, Chris Fröhlich, Jenny Eichhorst, Burkhard Wiesner, Joachim Behlke, L. Aravind, and Oliver Daumke. 2010. “Structural Basis of Oligomerization in Septin-like GTPase of Immunity-Associated Protein 2 (GIMAP2).” *Proceedings of the National Academy of Sciences of the United States of America* 107(47):20299–304.

Seah, Brandon K. B., and Harald R. Gruber-Vodicka. 2015. “Gbtools: Interactive Visualization of Metagenome Bins in R.” *Frontiers in Microbiology* 6(DEC):1–10.

Sedlazeck, Fritz J., Philipp Rescheneder, Moritz Smolka, Han Fang, Maria Nattestad, Arndt Von Haeseler, and Michael C. Schatz. 2018. “Accurate Detection of Complex Structural Variations Using Single-Molecule Sequencing.” *Nature Methods* 15(6):461–68.

Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2015. “BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs.” *Bioinformatics* 31(19):3210–12

Sun, Jin, Yu Zhang, Ting Xu, Yang Zhang, Huawei Mu, Yanjie Zhang, Yi Lan, Christopher J. Fields, Jerome Ho Lam Hui, Weipeng Zhang, Runsheng Li, Wenyan Nong, Fiona Ka Man Cheung, Jian Wen Qiu, and Pei Yuan Qian. 2017. “Adaptation to Deep-Sea Chemosynthetic Environments as Revealed by Mussel Genomes.” *Nature Ecology and Evolution* 1(5):1–7.

Tauchi-Sato, Kumi, Shintaro Ozeki, Toshiaki Houjou, Ryo Taguchi, and Toyoshi Fujimoto. 2002. “The Surface of Lipid Droplets Is a Phospholipid Monolayer with a Unique Fatty Acid Composition.” *Journal of Biological Chemistry* 277(46):44507–12.

Tavares, Sandra, André Filipe Vieira, Anna Verena Taubenberger, Margarida Araújo, Nuno Pimpao Martins, Catarina Brás-Pereira, António Polónia, Maik Herbig, Clara Barreto, Oliver Otto, Joana Cardoso, José B. Pereira-Leal, Jochen Guck, Joana Paredes, and Florence Janody. 2017. “Actin Stress Fiber Organization Promotes Cell Stiffening and Proliferation of Pre-Invasive Breast Cancer Cells.” *Nature Communications* 8(May).

Thorens, Bernard. 1996. “Glucose Transporters in the Regulation of Intestinal, Renal, and Liver Glucose Fluxes.” *American Journal of Physiology - Gastrointestinal and Liver Physiology* 270(4 33-4).

Tietjen, Målin. “Physiology and ecology of deep-sea bathymodiolin symbioses”

Personal communication.

Timney, Benjamin L., Barak Raveh, Roxana Mironska, Jill M. Trivedi, Seung Joong Kim, Daniel Russel, Susan R. Wente, Andrej Sali, and Michael P. Rout. 2016. “Simple Rules for Passive Diffusion through the Nuclear Pore Complex.” *Journal of Cell Biology* 215(1):57–76.

Toker, Alex. 1998. “The Synthesis and Cellular Roles of Phosphatidylinositol 4,5-Bisphosphate.” *Current Opinion in Cell Biology* 10(2):254–61.

Troy, Carol M., and Elena M. Ribe. 2008. “Caspase-2:Vestigial Remnant or Master Regulator?” *Science Signaling* 1(38):1–4.

Turchinovich, Andrey, Harald Surowy, Andrius Serva, Marc Zapatka, Peter Lichter, and Barbara Burwinkel. 2014. “Capture and Amplification by Tailing and Switching (CATS), Anultrasensitive Ligation-Independent Method for Generation of DNA Libraries for Deep Sequencing from Picogram Amounts of DNA and RNA.” *RNA Biology* 11(7):817–28.

Viaud, Julien, Frédéric Boal, Hélène Tronchère, Frédérique Gaits-Iacovoni, and Bernard Payrastre. 2014. “Phosphatidylinositol 5-Phosphate: A Nuclear Stress Lipid and a Tuner of Membranes and Cytoskeleton Dynamics.” *BioEssays* 36(3):260–72.

Watson, M. L. 1955. “The Nuclear Envelope; Its Structure and Relation to Cytoplasmic Membranes.” *The Journal of Biophysical and Biochemical Cytology* 1(3):257–70.

Weiss, Ingrid M. 2012. “Species-Specific Shells: Chitin Synthases and Cell Mechanics in Molluscs.” *Zeitschrift Fur Kristallographie* 227(11):723–38.

Wick, Ryan R., Louise M. Judd, Claire L. Gorrie, and Kathryn E. Holt. 2017. “Unicycler: Resolving Bacterial Genome Assemblies from Short and Long Sequencing Reads.” *PLoS Computational Biology* 13(6):1–22.

Wick, Ryan R., Mark B. Schultz, Justin Zobel, and Kathryn E. Holt. 2015. “Bandage: Interactive Visualization of de Novo Genome Assemblies.” *Bioinformatics* 31(20):3350–52.

Wiśniewski, Jacek R., Alexandre Zougman, Nagarjuna Nagaraj, and Matthias Mann. 2009. “Universal Sample Preparation Method for Proteome Analysis.” *Nature Methods* 6(5):359–62.

Wright, E. M., D. D. F. Loo, M. Panayotova-Heiermann, M. P. Lostao, B. H. Hirayama, B. Mackenzie, K. Boorer, and G. Zampighi. 1994. “‘Active’ Sugar Transport in Eukaryotes.” *Journal of Experimental Biology* 196:197–212.

Zhang, Jianlin, Amanda Felder, Yujie Liu, Ling T. Guo, Stephan Lange, Nancy D. Dalton, Yusu Gu, Kirk L. Peterson, Andrew P. Mizisin, G. Diane Shelton, Richard L. Lieber, and Ju Chen. 2009. “Nesprin 1 Is Critical for Nuclear Positioning and Anchorage.” *Human Molecular Genetics* 19(2):329–41.

Zielinski, Frank U., Annelie Pernthaler, Sebastien Duperron, Luciana Raggi, Olav Giere, Christian Borowski, and Nicole Dubilier. 2009. “Widespread Occurrence of an

Intranuclear Bacterial Parasite in Vent and Seep Bathymodiolin Mussels." *Environmental Microbiology* 11(5):1150–67.

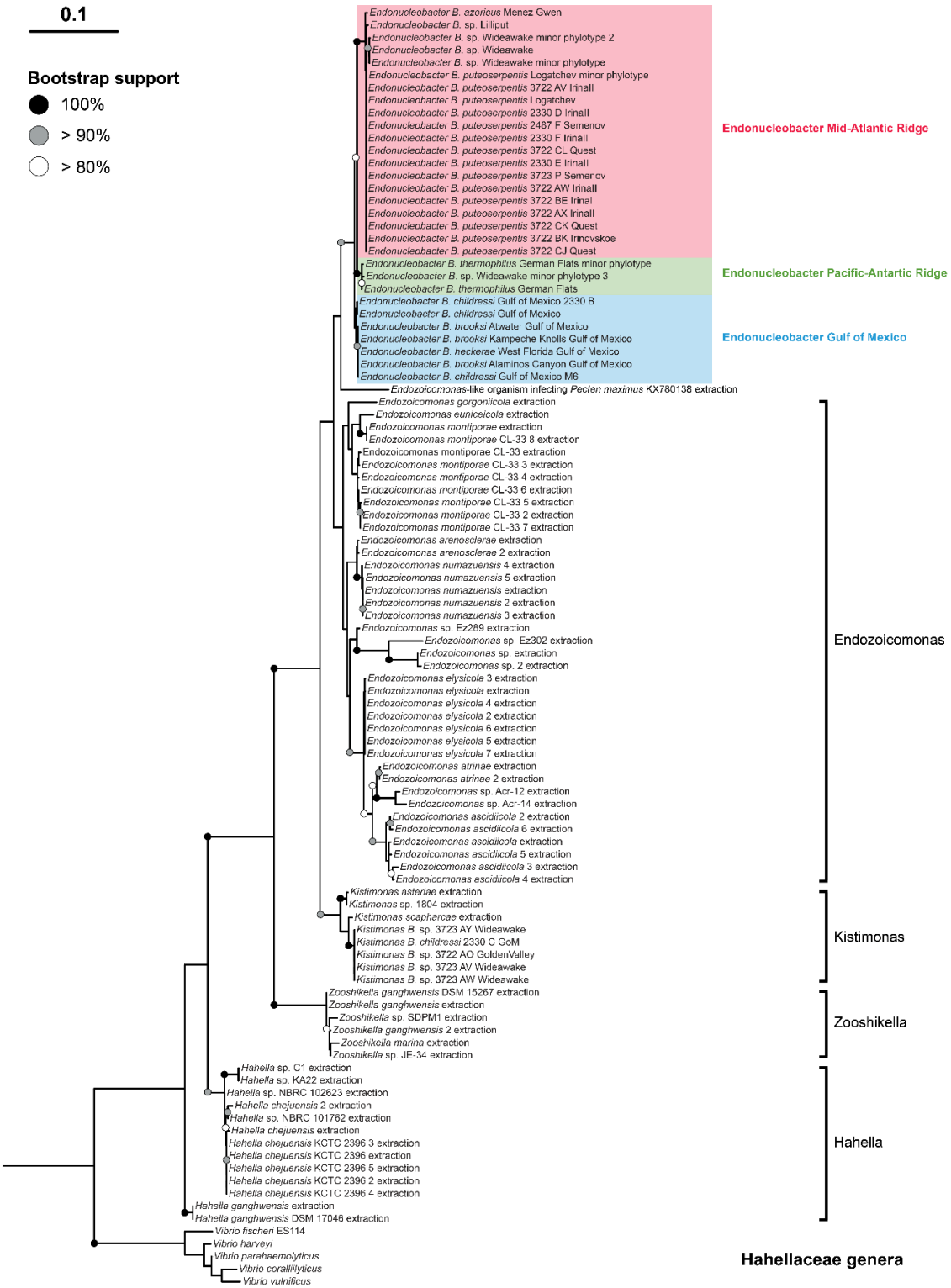
Supplementary Information for chapter II

Phylogenetics

Based on the sequence analysis of 16S ribosomal RNA genes (**Supp. Fig.1**), *Ca.* Endonucleobacter are gammaproteobacterial that belong to the family *Hahellaceae*. They are a monophyletic group (100% support) which is the sister clade of *Endozoicomonas* spp.. The *Ca*. Endonucleobacter clade was formed by three subclades reflecting the mussels' sampling sites as previously reported by Zielinski et al., 2009. Sequences recovered from the hydrothermal vents of the Mid-Atlantic Ridge (MAR) and the hydrothermal vents of the Pacific-Antarctic Ridge (PAR) constituted a monophyletic group (89% support) that clustered separately from the sequences recovered from the cold seeps of the Gulf of Mexico (GOM). A single sequence from an *Endozoicomonas*-like organism infecting *Pecten maximus* (Cano et al., 2018) was placed as the sister clade of *Ca*. Endonucleobacter. Intriguingly, both groups together were not well supported as a monophyletic group (53% support).

Assembly

In this study, we used a hybrid pipeline to assemble the *Ca*. Endonucleobacter genome from gill metagenomes of the deep-sea mussel "*B.*" *childressi*. The draft genome was 3.3 Mb, and estimated to be 96% complete. It was composed of 94 contigs (longer than 1000 bp) with an N50 of 88,503 bp, and had 41% GC content, 3,654 protein-coding genes and 54 tRNA-encoding genes.



Supplementary figure 1. Phylogeny of *Ca.* Endonucleobacter and associated *Oceanospirillales* based on the 16S rRNA gene. The data set included metagenomic 16S rRNA gene sequences (102 sequences total). Bootstrap values below 80% are not shown. 5 *Vibrio* spp. sequences were used to root the tree.

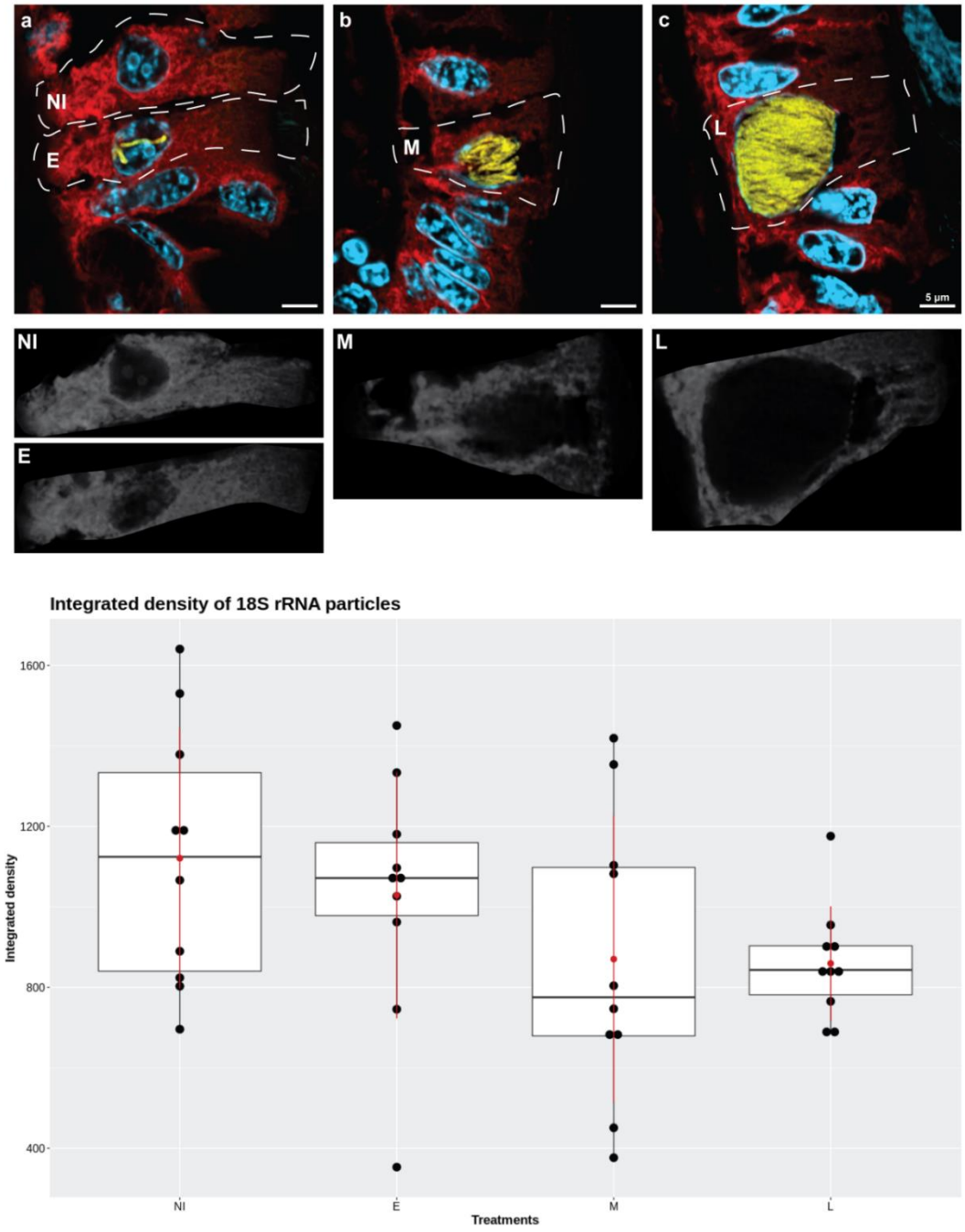
Cytological changes

Ribosomal integrity

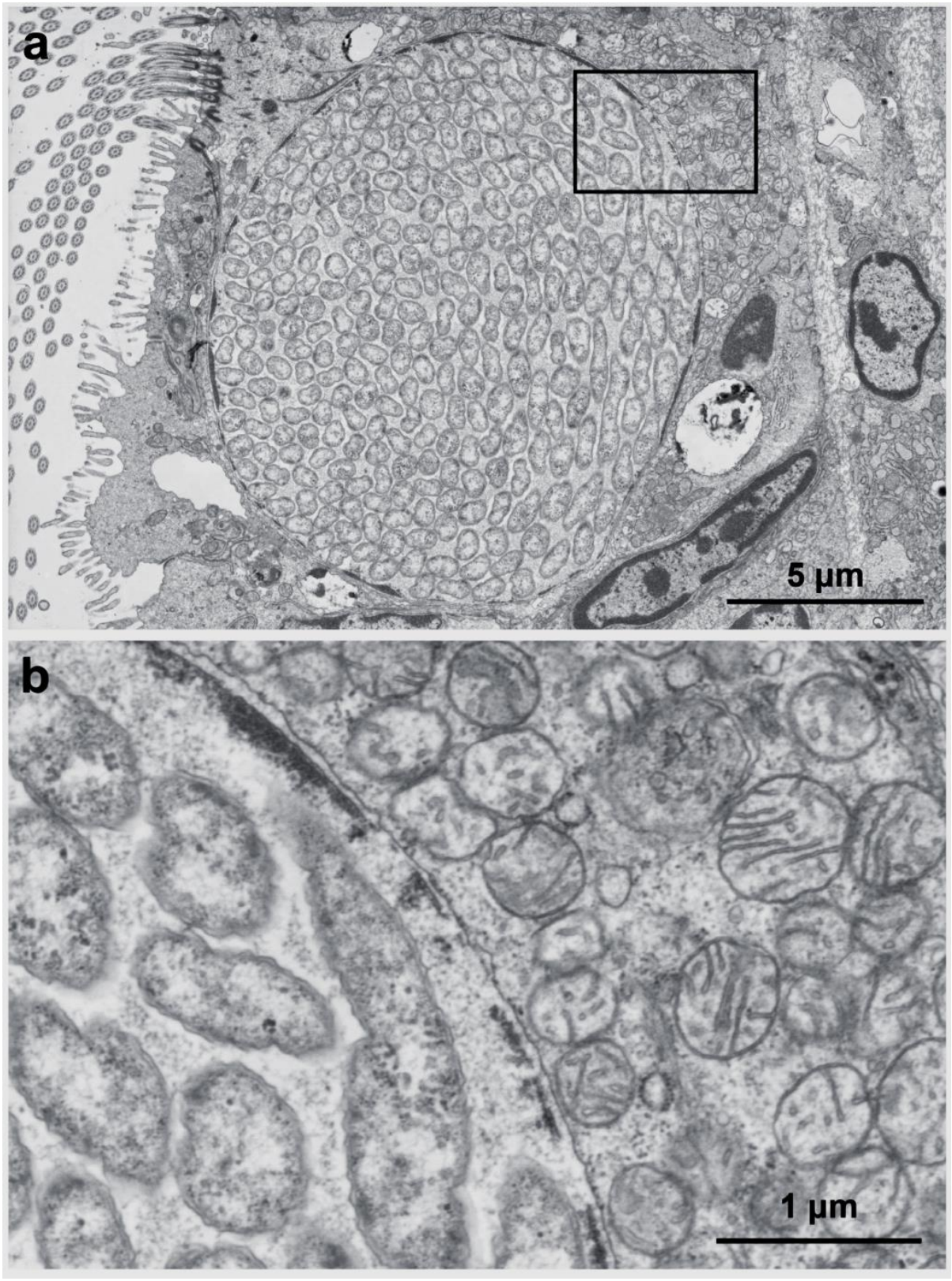
To evaluate the integrity of the host cell ribosomes along the infectious cycle, we co-localized host 18S rRNA and *Ca*. Endonucleobacter 16S rRNA using FISH. **Supp.Fig.2** shows SR-Airyscan snapshots of "*B". childressi* host cells at different infection stages (E, early; M, mid; L, late). The 18S rRNA fluorescence intensity of an uninfected area of tissue was measured as a control (NI, non-infected). The cytoplasmic regions surrounding nuclei at different infection stages showed comparable 18S rRNA fluorescent intensities. Moreover, no differences in 18S rRNA fluorescent intensity were observed between the different infection stages and the control. Our results show that host 18S rRNA fluorescent intensity does not decrease along the infectious cycle, as an indicative of ribosomal and host DNA integrity.

Ultrastructure

To evaluate morphological changes in the host cell along the infectious cycle, we imaged *"B". childressi* infected cells using transmission electron microscopy (TEM). A *"B". childressi* infected cell in late stage (**Supplementary Figure 3, a**) revealed an apparently functional cell. A closer look to the outer nuclear membrane – mitochondria area interface (**Supp. Fig.3, b**) confirmed that the mitochondria are intact in the late stage of infection. Additionally, the host chromatin seemed to be compressed against the inner nuclear membrane. This data suggested that the host cell mitochondria remained functional and that the host chromatin did not disappear at the end of the infectious cycle.



Supplementary figure 2. The integrated density of 18S rRNA particles did not change significantly along the infectious cycle of *Ca.* Endonucleobacter. The fluorescence intensity of 18S rRNA particles was measured as an indicator of transcription activity and DNA integrity of the host cell. Ten equivalent areas from the cell cytoplasm were measured per treatment. **a**, **b**, **c**, Fluorescence *in situ* hybridization on gill cross sections of "*B." childressi*. SR-airyscan snapshots of *"B." childressi* cells at different *Ca*. Endonucleobacter infection stages. Host 18S rRNA (red: Alexa Fluor 647), *Ca.* Endonucleobacter 16S rRNA (yellow: Atto550), DNA (cyan: DAPI). **a**, snapshot of a non-infected (NI) and infected cell at early stage (E). **b**, snapshot of an infected cell at mid stage (M). **c**, snapshot of an infected cell at late stage (L). The grey scale values of the 18S rRNA particles of the four different treatments (**NI**, **E**, **M**, **L**) were measured and normalized to the area of the 10 measured areas per treatment (integrated density, right box plot).



Supplementary figure 3. The mitochondria of an infected cell at late stage seemed intact and completely functional. Ultrastructural analysis of a late stage infected cell by *Ca.* Endonucleobacter. a, TEM overview of infected cell. b, magnification of area of interest. Electrondense chromatin can be observed pushed against the inner nuclear membrane, while mitochondria seem intact and functional.

Samples

Supplementary table 1 contains information about all the specimens used in this study and the analyses done on each of them.

Supplementary table 1. 14 *"B". childressi* gill samples were PCR screened looking for *Ca*. Endonucleobacter infection. The different analyses done to each of the samples included in the study are highlighted with green boxes.

Mussel specimen ID	Ca. Endonucleobacter - *"B". childressi* (PCR screening)	Microscopy	Illumina DNA sequencing	Bulk transcriptomics	PacBio DNA sequencing	LCM transcriptomics	Host de novo transcriptomics	Proteomics
H1423/001-N1	Detected							
H1423/001-N2	Detected							
H1423/001-N4	Detected							
H1423/001-N5	Detected							
H1423/001-N6	Detected							
H1423/002-N9	Detected							
H1423/002-N10	Detected							
H1423/003-N11	Detected							
H1423/003-N15	Detected							
H1423/004-N19	Detected							
H1425/019-N59	Non detected							
H1425/019-N60	Non detected							
H1425/019-N61	Non detected							
H1425/019-N62	Non detected							

Sequencing depth of LCM libraries

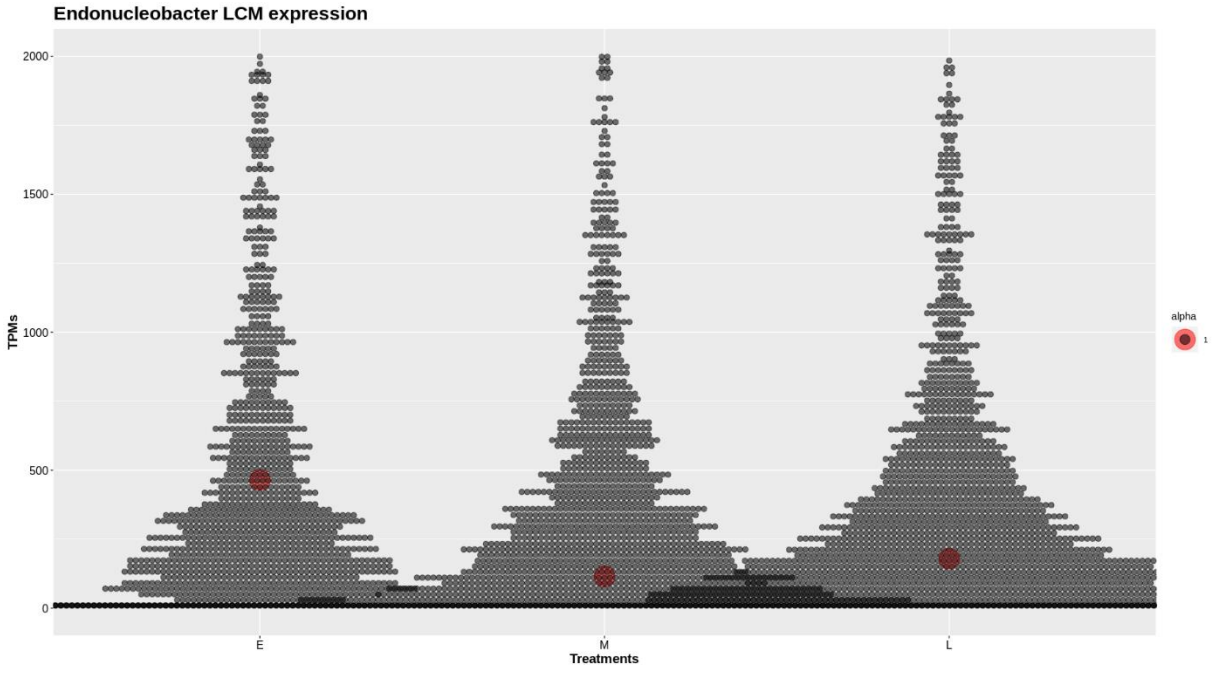
We sequenced preliminarily 3,333,333 reads from each LCM library to survey the mapping rate against the reference genome of *Ca*. Endonucleobacter. Based on this mapping rates, we adjusted the sequencing depth of the LCM libraries to equalize the number of *Ca*. Endonucleobacter mRNA reads mapping to our reference genome. We aimed to get a similar number of reads mapping to the reference genome in all treatments, regardless of the infection stage we were looking at (early, mid or late). Number of reads sequenced at that end are contained in **Supplementary table 2**.

Supplementary table 2. Laser capture microdissection libraries were sequenced at different depths trying to equalize the number of *Ca*. Endonucleobacter reads. Intuitively, the libraries with lower amount of *Ca*. Endonucleobacter biomass (early stage) needed deeper sequencing. Unfortunately, we lost one of the non-infected libraries (negative control for host expression) during the sequencing.

Treatment	Library	Number of reads
Non-infected	NI1	33,333,333
Non-infected	NI2	Lost replicate
Non-infected	NI3	33,333,333
Early	E1	150,000,000
Early	E2	666,666,600
Early	E3	233,333,310
Mid	M1	63,333,327
Mid	M2	53,333,328
Mid	M3	33,333,333
Late	L1	33,333,333
Late	L2	33,333,333
Late	L3	33,333,333

Normalization of Ca. *Endonucleobacter expression to* RecA

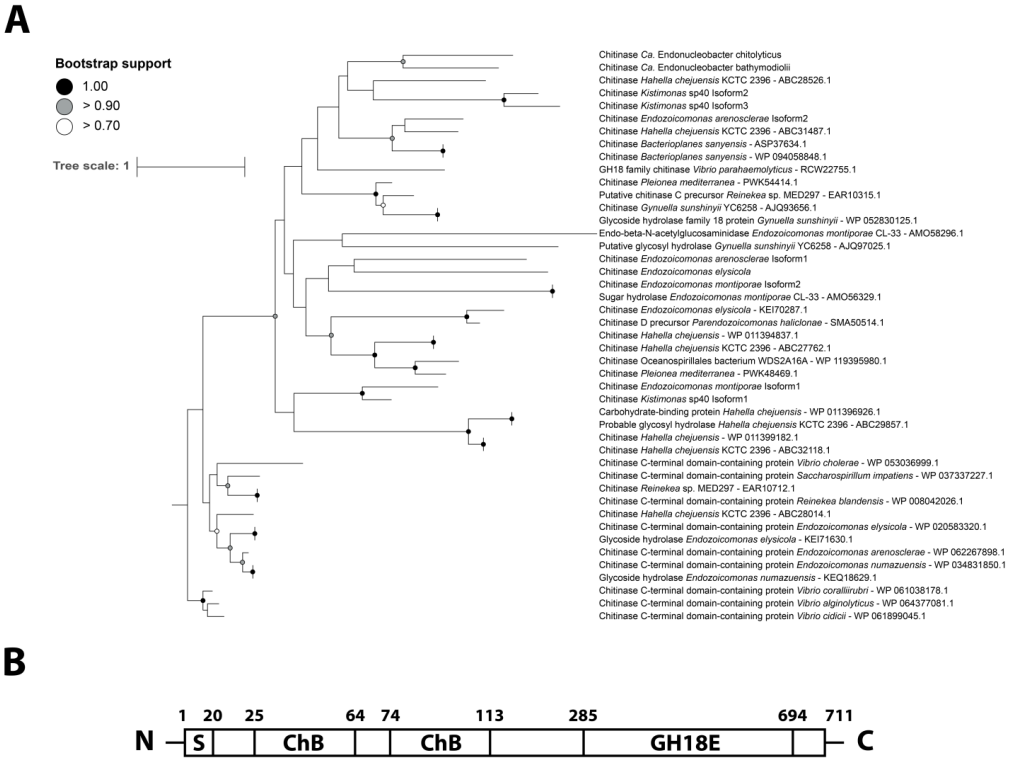
The expression of *Ca.* Endonucleobacter was normalized to the expression of the housekeeping gene *RecA*. For the bulk transcriptomic data represented in **Fig.3**, the expression of every gene was normalized to the expression of *RecA*. For the transcriptomic profiling of the infectious cycle (**Fig.4**), the average expression of every gene (three replicates per treatment) was normalized to the average expression of *RecA*. **Supplementary figure 4** shows the average expression (TPMs) of 3,661 genes annotated for *Ca.* Endonucleobacter (three replicates per treatment). Except *RecA*, which average expression is represented by a red dot, all other genes have been represented with black dots. As the average expression of *RecA* did not change considerably among treatments, we considered it a good candidate to relativize the expression of all other genes.



Supplementary figure 4. Average expression of 3,661 genes annotated for *Ca.* Endonucleobacter, with the expression of *RecA* represented by a red dot. The expression of every gene was calculated averaging the expression of the three libraries per treatment (E, M and L for early, mid and late stage of infection, respectively).

Phylogenetic and protein domain analysis of chitinase

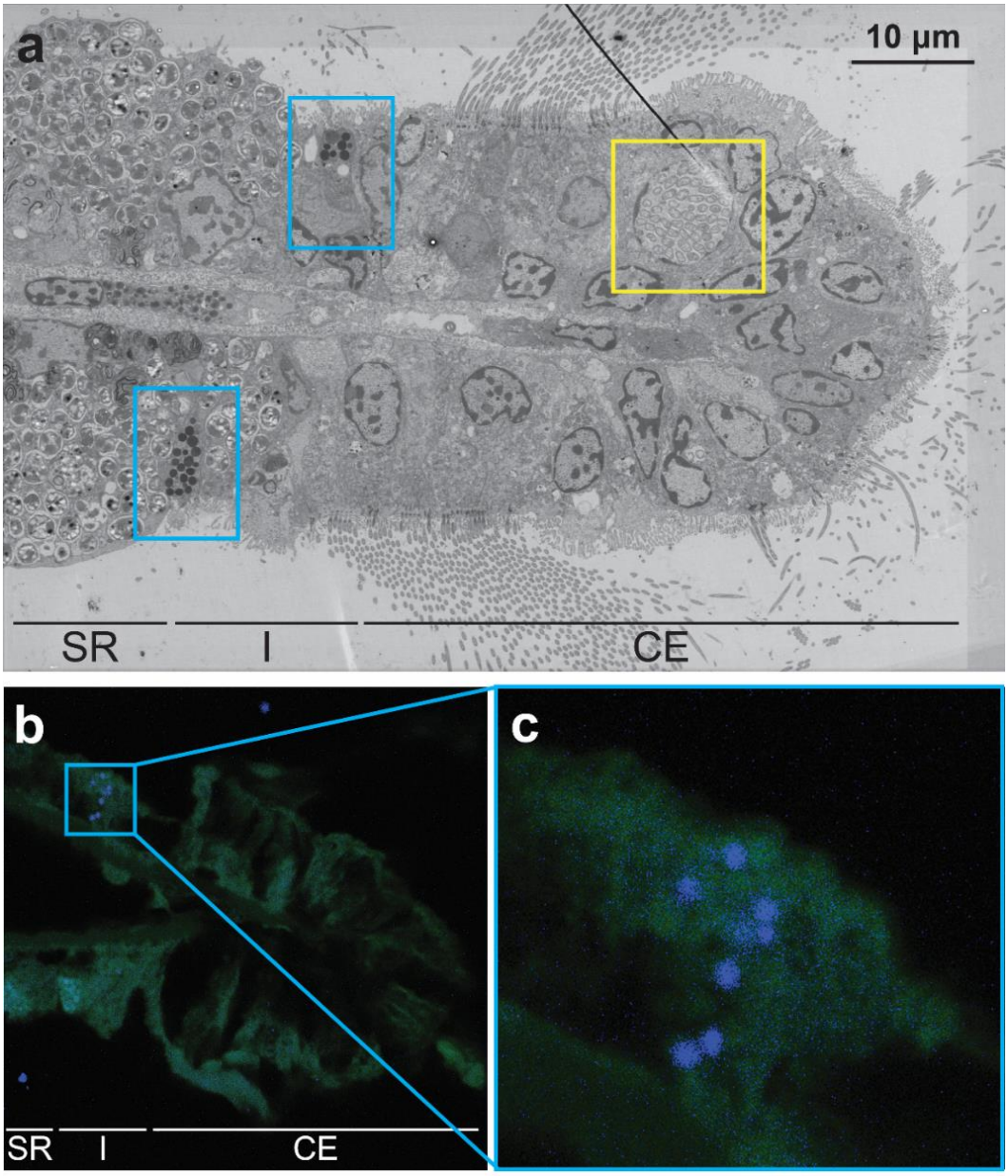
The phylogenetic analysis of 45 chitinases from the GH18 glycosidases family revealed that the chitinase encoded by *Ca*. Endonucleobacter is also a GH18 endochitinase (**Supplementary Figure 5, A**). Chitinases from both *Ca*. Endonucleobacter species clustered together with a bootstrap support >90%. The protein domain analysis of the chitinase encoded by Ca. Endonucleobacter (**Supp. Fig.5, B**) revealed that the enzyme presents a T3SS secretion signal peptide, two chitin-binding domains in tandem and a catalytic domain with a glutamate residue in the active center.



Supplementary figure 5. Endochitinase from *Ca.* Endonucleobacter forms part of the GH18 glycosidases family. A, phylogenetic tree of chitinases of *Oceanospirillales* representatives. 45 chitinase amino acid sequences from representatives of the order Oceanospirillales were included in this study, while 5 chitinase sequences from representatives of the order *Vibrionales* were used to root the tree. **B**, domains representation of chitinase from *Ca.* Endonucleobacter (**S**, secretion signal peptide; **ChB**, chitin-binding domain; **GH18E**, catalytic domain with a glutamate residue in the catalytic center). The numbers indicate the amino acid position in the sequence. Domains' positions are not scaled according to sequence length.

Localization of chitin in the gill filaments of "B." childressi

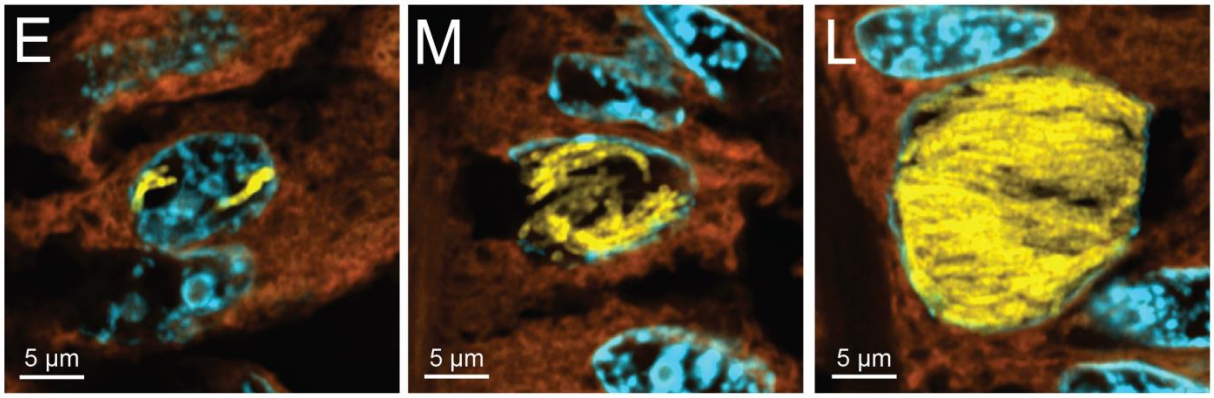
Ultrastructure imaging of the gill filaments of *"B." childressi* revealed the existence of secretory cells in the symbiotic region-ciliated edge interface (**cyan frames in Supplementary Figure 6, a**). The fluorescence localization of chitin (**Supp. Fig.6, b**) revealed chitin-containing vesicles in the same area where secretory cells occur.



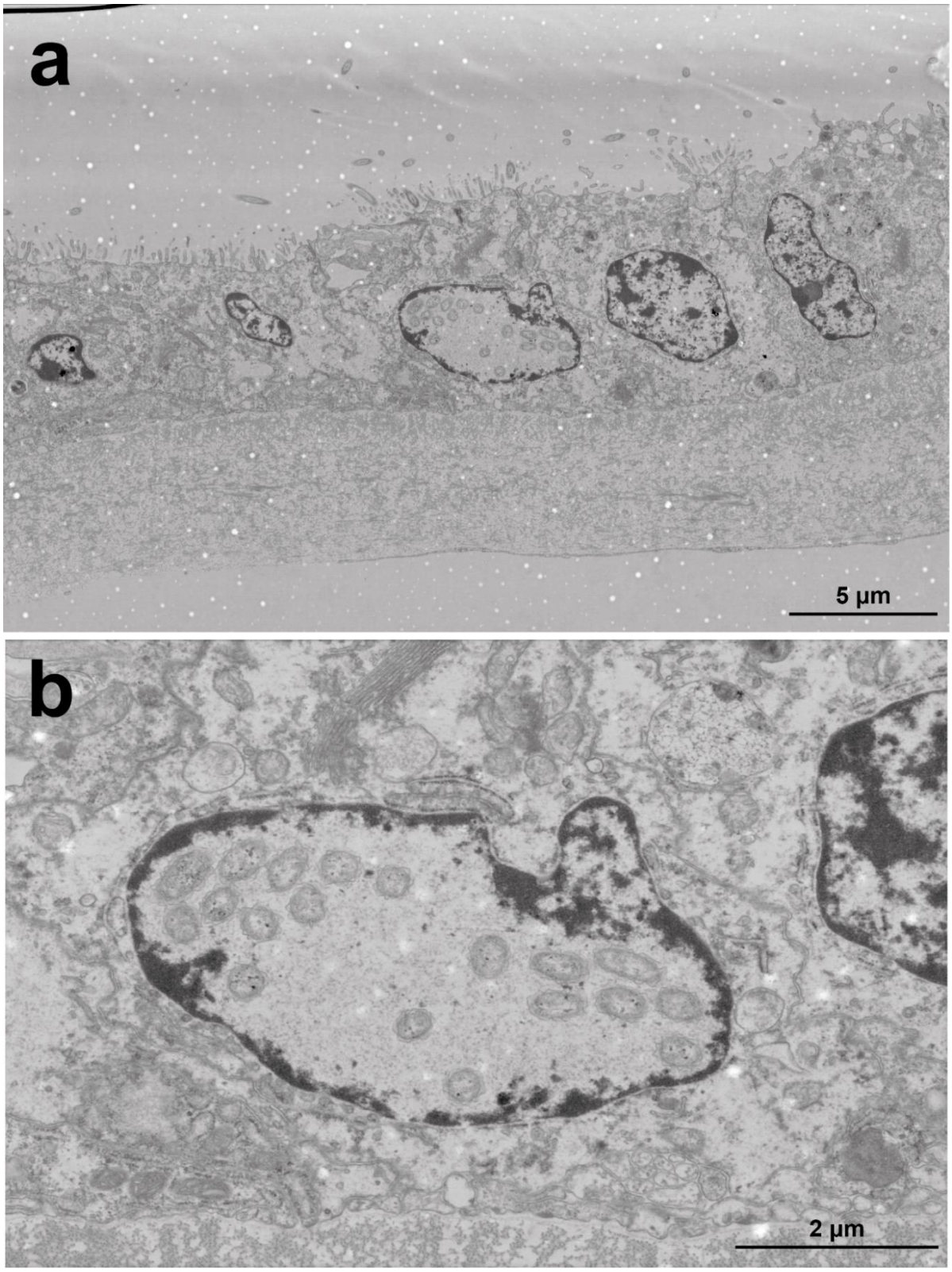
Supplementary figure 6. Chitin-containing vesicles occur in secretory cells located in the symbiont region-ciliated edge interface. a, ultrastructural image (TEM) of *"B." childressi* gill filament cross section. Cyan frames: secretory cells. Yellow frame: nucleus infected by *Ca*. Endonucleobacter. **b**, CLSM image of *"B." childressi* gill filament cross sections. Chitin (blue: Fungi-Fluor Kit), Tissue autofluorescence (green). **c**, amplification of the secretory cell represented in b. **SR:** Symbiotic region. **I:** Symbiotic region - ciliated edge interface. **CE:** Ciliated edge.

***Localization of lipid droplets in the gill filaments of* "B." childressi**

Fluorescence localization of lipids using Nile Red did not reveal any lipid droplet-like structure in cells infected by *Ca*. Endonucleobacter at any point of its life cycle (**Supplementary Figure 7**). All the membrane structures were localized, suggesting that staining with Nile Red functioned optimally.



Supplementary figure 7. Cells infected by *Ca.* Endonucleobacter did not show any lipid droplet-like structure at any point of the infectious life cycle. Airyscan images on *"B". childressi* gill filament cross sections. **E,** early stage of infection. **M,** mid stage of infection. **L,** late stage of infection. Super-resolution Airyscan images of *"B." childressi* gill filaments sections. Lipids (orange: Nile Red), *Ca*. Endonucleobacter 16S rRNA (yellow: Atto550), DNA (cyan: DAPI).



Supplementary figure 8. *Ca.* Endonucleobacter can invade the symbiotic region of *"B." childressi* specimens that have been kept under starvation conditions, which have lost most of their methane-oxidizing symbionts. a, TEM image of symbiotic region of *"B." childressi*. Several bacteriocytes containing few or none methane oxidizing symbionts can be seen. The nucleus infected by *Ca.* Endonucleobacter is in the middle of the image. **b**, amplification of the nucleus infected by *Ca.* Endonucleobacter.

Supplementary table 3. Bulk transcriptomic expression values of *Ca.* Endonucleobacter. ID, gene identifier in *Ca.* Endonucleobacter genome. **Annotation,** RAST annotation. **Symbol in Fig.3,** symbol used to represent each gene in Figure 3 (main paper). **Expression (TPM)**, expression values of each gene in TPMs. **Normalized expression to RecA (TPM)**, expression of each gene normalized to *RecA*.

ID	Annotation	Symbol in Fig.3	Expression (TPM)	Normalized expression to RecA (TPM)
fig|539814.26.peg.2251	RecA protein	NA	81,56	1,00
fig|539814.26.peg.2086	Type III secretion spans bacterial envelope protein (YscO)	*YscO*	314,43	3,86
fig|539814.26.peg.2088	Type III secretion inner membrane protein (YscQ%2Chomologous to flagellar export components)	*YscQ*	376,13	4,61
fig|539814.26.peg.2089	Type III secretion inner membrane protein (YscR%2CSpaR%2CHrcR%2CEscR%2Chomologous to flagellar export components)	*YscR*	180,77	2,22
fig|539814.26.peg.1947	PTS IIA-like nitrogen-regulatory protein PtsN	*PtsN*	26,20	0,32
fig|539814.26.peg.2421	PTS system%2C glucose-specific IIB component (EC 2.7.1.69) / PTS system%2C glucose-specific IIC component	PTSG	18,89	0,23
fig|539814.26.peg.2422	Phosphoenolpyruvate-protein phosphotransferase of PTS system (EC 2.7.3.9)	PTPS	63,01	0,77
fig|539814.26.peg.2889	Aquaporin Z	*aqpZ*	340,12	4,17
fig|539814.26.peg.2862	Porin family	Uporin	10436,50	127,96
fig|539814.26.peg.1576	Type I secretion system%2C outer membrane component LapE	*LapE*	426,53	5,23
fig|539814.26.peg.1578	Type I secretion system%2C membrane fusion protein LapC	*LapC*	100,28	1,23
fig|539814.26.peg.88	Chitinase (EC 3.2.1.14)	*ChiA*	2557,37	31,35
fig|539814.26.peg.203	putative virulence protein ipgD (Shigella flexneri plasmid pINV)	*IpgD*	49,85	0,61
fig|539814.26.peg.1580	Probable RTX	RTX-a	404,35	4,96
fig|539814.26.peg.35	Probable lipase	Lipase	597,92	7,33
fig|539814.26.peg.3371	baculoviral IAP repeat-containing protein 7-like [Betta splendens]"	IAP	40,68	0,50
fig|539814.26.peg.3091	N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25)	*NagA*	122,85	1,51
fig|539814.26.peg.3175	Histidine permease YuiF	*YuiF*	26,19	0,32
fig|539814.26.peg.30	Serine transporter	SERT	27,27	0,33
fig|539814.26.peg.2777	Methionine ABC transporter substrate-binding protein	MSBP	143,47	1,76
fig|539814.26.peg.1167	Putrescine transport ATP-binding protein PotG (TC 3.A.1.11.2)	*PotG*	15,50	0,19
fig|539814.26.peg.1924	Putrescine ABC transporter putrescine-binding protein PotF (TC 3.A.1.11.2)	*PotF*	7,58	0,09
fig|539814.26.peg.235	Oligopeptide transport system permease protein OppB (TC 3.A.1.5.1)	*OppB*	32,73	0,40
fig|539814.26.peg.236	Oligopeptide transport system permease protein OppC (TC 3.A.1.5.1)	*OppC*	52,19	0,64
fig|539814.26.peg.1300	Competence protein F homolog%2C phosphoribosyltransferase domain%3B protein YhgH required for utilization of DNA as sole source of carbon and energy	*ComF*	0,00	0,00
fig|539814.26.peg.305	DNA internalization-related competence protein ComEC/Rec2	*ComEC*	14,18	0,17
fig|539814.26.peg.2218	Single-stranded-DNA-specific exonuclease RecJ (EC 3.1.-.-)	*RecJ*	19,33	0,24
fig|539814.26.peg.567	Amidophosphoribosyltransferase (EC 2.4.2.14)	GPAT	122,74	1,50
fig|539814.26.peg.2962	Carbamoyl-phosphate synthase large chain (EC 6.3.5.5)	CPSase	88,63	1,09
fig|539814.26.peg.577	CTP synthase (EC 6.3.4.2)	CTPS	72,13	0,88
fig|539814.26.peg.3584	Thymidylate synthase (EC 2.1.1.45)	TS	159,71	1,96
fig|539814.26.peg.876	Long-chain fatty acid transport protein	FATP	93,43	1,15
fig|539814.26.peg.1112	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)	LC-FACS	19,83	0,24
fig|539814.26.peg.1308	Acyl-CoA dehydrogenase (EC 1.3.8.7)	ACAD	82,80	1,02
fig|539814.26.peg.2383	ATP synthase alpha chain (EC 3.6.3.14)	ATPα	164,06	2,01
fig|539814.26.peg.2305	Thioredoxin	THIO	758,80	9,30
fig|539814.26.peg.1085	Cobalt-zinc-cadmium resistance protein CzcD	*CzcD*	0,00	0,00
fig|539814.26.peg.2280	Lead%2C cadmium%2C zinc and mercury transporting ATPase (EC 3.6.3.3) (EC 3.6.3.5)%3B Copper-translocating P-type ATPase (EC 3.6.3.4)	CUCDT	19,80	0,24
fig|539814.26.peg.1421	Zinc ABC transporter%2C ATP-binding protein ZnuC	*ZnuC*	0,00	0,00
fig|539814.26.peg.1422	Zinc ABC transporter%2C inner membrane permease protein ZnuB	*ZnuB*	24,62	0,30
fig|539814.26.peg.181	Manganese ABC transporter%2C inner membrane permease protein SitD	*SitD*	0,00	0,00
fig|539814.26.peg.1742	Ferrous iron transport protein B	*FeoB*	75,00	0,92
fig|539814.26.peg.1743	Ferrous iron transport protein A	*FeoA*	491,97	6,03
fig|539814.26.peg.224	Catalase (EC 1.11.1.6)	CAT	81,58	1,00
fig|539814.26.peg.1075	Catalase (EC 1.11.1.6) / Peroxidase (EC 1.11.1.7)	CAT/PER	57,63	0,71
fig|539814.26.peg.3207	Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16)	GFAT	72,26	0,89

Supplementary table 4. Laser-capture microdissection expression values for *Ca.* Endonucleobacter. ID, gene identifier in *Ca.* Endonucleobacter genome. **Annotation,** RAST annotation. **Symbol in Fig.4,** symbol used to represent each gene in Figure 4 (main paper). **E-average, M-average, and L-average**, average expression values (TPMs) for the infection stages treatments early, mid and late, respectively (n = 3 per treatment). **E-average-N, M-average-N, and L-average-N,** average expression values (TPMs) normalized to *RecA* for the infection stages treatments early, mid and late, respectively.

ID	Annotation	Symbol in Fig.4	E-average	M-average	L-average	E-average-N	M-average-N	L-average-N
fig|539814.26.peg.2251	RecA protein	NA	464,77	115,07	179,62	1,00	1,00	1,00
fig|539814.26.peg.2088	Type III secretion inner membrane protein (YscQ%2Chomologous to flagellar export components)	*YscQ*	1346,47	939,96	1270,67	2,90	8,17	7,07
fig|539814.26.peg.2089	Type III secretion inner membrane protein (YscR%2CSpaR%2CHrcR%2CEscR%2Chomologous to flagellar export components)	*YscR*	700,32	643,56	682,40	1,51	5,59	3,80
fig|539814.26.peg.2421	PTS system%2C glucose-specific IIB component (EC 2.7.1.69) / PTS system%2C glucose-specific IIC component	PTS	212,64	216,72	189,40	0,46	1,88	1,05
fig|539814.26.peg.88	Chitinase (EC 3.2.1.14)	*ChiA*	4333,91	5983,56	7674,67	9,32	52,00	42,73
fig|539814.26.peg.35	Probable lipase	Lipase	6285,44	2377,47	3485,50	13,52	20,66	19,41
fig|539814.26.peg.203	putative virulence protein ipgD (Shigella flexneri plasmid pINV)	*IpgD*	256,04	242,01	396,66	0,55	2,10	2,21
fig|539814.26.peg.1580	Probable RTX	RTX-a	941,02	2115,26	1838,87	2,02	18,38	10,24
fig|539814.26.peg.3371	baculoviral IAP repeat-containing protein 7-like [Betta splendens]*	IAP	64,41	120,08	165,82	0,14	1,04	0,92
fig|539814.26.peg.3091	N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25)	*NagA*	458,67	1177,51	595,36	0,99	10,23	3,31
fig|539814.26.peg.3207	Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16)	GFAT	153,60	302,86	297,64	0,33	2,63	1,66
fig|539814.26.peg.876	Long-chain fatty acid transport protein	FATP	173,63	396,27	523,78	0,37	3,44	2,92
fig|539814.26.peg.1112	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)	LC-FACS	99,77	45,72	198,24	0,21	0,40	1,10
fig|539814.26.peg.1308	Acyl-CoA dehydrogenase (EC 1.3.8.7)	ACAD	190,58	378,79	530,92	0,41	3,29	2,96
fig|539814.26.peg.1167	Putrescine transport ATP-binding protein PotG (TC 3.A.1.11.2)	*PotG*	139,32	206,08	146,85	0,30	1,79	0,82
fig|539814.26.peg.1924	Putrescine ABC transporter putrescine-binding protein PotF (TC 3.A.1.11.2)	*PotF*	54,61	193,18	108,57	0,12	1,68	0,60
fig|539814.26.peg.235	Oligopeptide transport system permease protein OppB (TC 3.A.1.5.1)	*OppB*	347,01	152,33	272,00	0,75	1,32	1,51
fig|539814.26.peg.236	Oligopeptide transport system permease protein OppC (TC 3.A.1.5.1)	*OppC*	552,47	608,88	137,13	1,19	5,29	0,76
fig|539814.26.peg.3175	Histidine permease YuiF	*YuiF*	69,64	318,31	148,39	0,15	2,77	0,83
fig|539814.26.peg.30	Serine transporter	SERT	266,78	147,43	376,10	0,57	1,28	2,09
fig|539814.26.peg.2777	Methionine ABC transporter substrate-binding protein	MSBP	268,77	460,23	510,83	0,58	4,00	2,84

Supplementary table 5. Laser-capture microdissection expression values for the host cell. ID, gene identifier in *de novo* host transcriptome. **Annotation,** Trinotate annotation. **Symbol in Fig.4,** symbol used to represent each gene in Figure 4 (main paper). Average (n = 3 per treatment) Log2 fold expression changes when comparing non-infected treatment with early stage treatment (**NI_vs_E**), early stage vs mid stage (**E_vs_M**) and mid stage vs late stage (**M_vs_L**).

ID	Annotation	Symbol in Fig.4	NI_vs_E	E_vs_M	M_vs_L
TRINITY_DN38272_c5_g3_i2	Calpain-5	CAP5	*3,49*	*0,63*	*-0,48*
TRINITY_DN33165_c7_g3_i2	Perilipin-2	PLIN2	*3,75*	-0,37	1,36
TRINITY_DN35012_c2_g1_i2	GTPase IMAP family member 4	GIMAP	*4,18*	*-0,14*	1,29
TRINITY_DN33995_c3_g1_i3	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-1	PLCG	3,37	*1,46*	-0,13
TRINITY_DN33271_c3_g7_i2	Fatty acid synthase	FAS	2,91	*0,03*	*0,68*
TRINITY_DN37954_c5_g2_i1	Glycerol-3-phosphate acyltransferase 3	GPAT	-1,03	1,19	-1,70
TRINITY_DN23084_c0_g1_i1	1-acyl-sn-glycerol-3-phosphate acyltransferase alpha (AGPAT)	AGPAT	*-0,18*	*3,02*	-2,47
TRINITY_DN26157_c0_g1_i2	Phosphatidate phosphatase LPIN2 (PAP)	PAP	*4,21*	*1,23*	-0,09
TRINITY_DN28572_c0_g1_i1	Diacylglycerol O-acyltransferase 1	DGAT	*2,39*	*1,12*	-3,08
TRINITY_DN38344_c6_g6_i1	Nesprin-1	SYNE1	*1,43*	*0,51*	0,04
TRINITY_DN39068_c2_g3_i2	Pyruvate kinase PKM	PKM	*3,70*	-0,16	1,75
TRINITY_DN1665_c0_g1_i1	ATP synthase subunit alpha, mitochondrial	ATPα	-0,19	0,11	0,73
TRINITY_DN36128_c6_g1_i1	ATP synthase subunit beta, mitochondrial	ATPβ	4,91	1,84	*-0,98*
TRINITY_DN39445_c2_g1_i21	Caspase-2	CASP2	*5,37*	1,01	-0,49
TRINITY_DN36387_c2_g1_i1	sodium-coupled neutral amino acid transporter 9	SCNA	*2,46*	0,30	*0,22*
TRINITY_DN32750_c0_g1_i1	Neutral and basic amino acid transport protein rBAT	rBAT	*5,94*	*-0,07*	*0,09*
TRINITY_DN39124_c10_g1_i1	Chitin synthase chs-2 {ECO:0000305}	*Chs2*	*0,64*	*0,82*	*-0,44*
TRINITY_DN39807_c0_g2_i1	Chitobiase	CTBS	*3,83*	*0,28*	*-0,30*
TRINITY_DN39009_c1_g1_i1	N-acetylglucosamine-6-phosphate deacetylase	*NagA*	0,00	1,10	0,96
TRINITY_DN27671_c0_g1_i2	Sugar transporter SWEET1	SWEET	2,43	1,68	-2,00
TRINITY_DN38538_c4_g3_i1	Glutamine--fructose-6-phosphate aminotransferase [isomerizing] 2	GFAT	-0,52	0,26	0,12
TRINITY_DN35359_c2_g2_i2	ATP-dependent 6-phosphofructokinase	PFK-1	-2,03	3,04	1,46
TRINITY_DN35181_c3_g1_i1	Fructose-bisphosphate aldolase	FBPA	2,45	-0,91	0,71
TRINITY_DN32232_c4_g1_i3	Glycerol-3-phosphate dehydrogenase	GPDH	1,91	2,20	-0,96

Supplementary table 6. *Ca.* Endonucleobacter bulk transcriptomic expression of pathways for *de novo* synthesis of nucleotides. ID, gene identifier in *Ca.* Endonucleobacter genome. **Annotation,** RAST annotation. **Pathway,** assigned metabolic route for each gene (Pathway tools v24.0). **Expression (TPM)**, expression values of each gene in TPMs. **Normalized expression to *RecA* (TPM)**, expression of each gene normalized to *RecA*.

ID	Annotation	Pathway	Expression (TPM)	Normalized expression to RecA (TPM)
fig|539814.26.peg.2251	RecA protein	Housekeeping gene	81.5635	*1.00*
fig|539814.26.peg.567	Amidophosphoribosyltransferase (EC 2.4.2.14)	Superpathway of purine nucleotides de novo biosynthesis I	122.74	*1.50*
fig|539814.26.peg.3260	Phosphoribosylamine--glycine ligase (EC 6.3.4.13)	Superpathway of purine nucleotides de novo biosynthesis I	27.0435	*0.33*
fig|539814.26.peg.2102	Phosphoribosylglycinamide formyltransferase (EC 2.1.2.2)	Superpathway of purine nucleotides de novo biosynthesis I	61.3699	*0.75*
fig|539814.26.peg.2709	Phosphoribosylformylglycinamidine synthase%2C synthetase subunit (EC 6.3.5.3) / Phosphoribosylformylglycinamidine synthase%2C glutamine amidotransferase subunit (EC 6.3.5.3)	Superpathway of purine nucleotides de novo biosynthesis I	32.9131	*0.40*
fig|539814.26.peg.2103	Phosphoribosylformylglycinamidine cyclo-ligase (EC 6.3.3.1)	Superpathway of purine nucleotides de novo biosynthesis I	250.541	*3.07*
fig|539814.26.peg.3197	Phosphoribosylaminoimidazole carboxylase catalytic subunit (EC 4.1.1.21)	Superpathway of purine nucleotides de novo biosynthesis I	90.8532	*1.11*
fig|539814.26.peg.3123	Phosphoribosylaminoimidazole-succinocarboxamide synthase (EC 6.3.2.6)	Superpathway of purine nucleotides de novo biosynthesis I	0	*0.00*
fig|539814.26.peg.1929	Phosphoribosylaminoimidazole-succinocarboxamide synthase (EC 6.3.2.6)	Superpathway of purine nucleotides de novo biosynthesis I	64.6157	*0.79*
fig|539814.26.peg.2048	Phosphoribosylaminoimidazole-succinocarboxamide synthase (EC 6.3.2.6)	Superpathway of purine nucleotides de novo biosynthesis I	64.6157	*0.79*
fig|539814.26.peg.2583	Adenylosuccinate lyase (EC 4.3.2.2)	Superpathway of purine nucleotides de novo biosynthesis I	62.6728	*0.77*
fig|539814.26.peg.3578	IMP cyclohydrolase (EC 3.5.4.10) / Phosphoribosylaminoimidazolecarboxamide formyltransferase (EC 2.1.2.3)	Superpathway of purine nucleotides de novo biosynthesis I	42.3546	*0.52*
fig|539814.26.peg.1124	Inosine-5'-monophosphate dehydrogenase (EC 1.1.1.205)	Superpathway of purine nucleotides de novo biosynthesis I towards GTP	80.4748	*0.99*
fig|539814.26.peg.1123	GMP synthase [glutamine-hydrolyzing] (EC 6.3.5.2)	Superpathway of purine nucleotides de novo biosynthesis I towards GTP	85.4359	*1.05*
fig|539814.26.peg.1431	Guanylate kinase (EC 2.7.4.8)	Superpathway of purine nucleotides de novo biosynthesis I towards GTP	34.0342	*0.42*
fig|539814.26.peg.108	Nucleoside diphosphate kinase (EC 2.7.4.6)	Superpathway of purine nucleotides de novo biosynthesis I towards GTP	583.911	*7.16*
fig|539814.26.peg.3299	Ribonucleotide reductase of class Ia (aerobic)%2C beta subunit (EC 1.17.4.1)	Superpathway of purine nucleotides de novo biosynthesis I towards GTP	125.565	*1.54*
fig|539814.26.peg.3300	Ribonucleotide reductase of class Ia (aerobic)%2C alpha subunit (EC 1.17.4.1)	Superpathway of purine nucleotides de novo biosynthesis I towards GTP	107.118	*1.31*
fig|539814.26.peg.108	Nucleoside diphosphate kinase (EC 2.7.4.6)	Superpathway of purine nucleotides de novo biosynthesis I towards GTP	583.911	*7.16*
fig|539814.26.peg.1377	Adenylosuccinate synthetase (EC 6.3.4.4)	Superpathway of purine nucleotides de novo biosynthesis I towards ATP	53.2198	*0.65*
fig|539814.26.peg.2583	Adenylosuccinate lyase (EC 4.3.2.2)	Superpathway of purine nucleotides de novo biosynthesis I towards ATP	62.6728	*0.77*
fig|539814.26.peg.2183	Adenylate kinase (EC 2.7.4.3)	Superpathway of purine nucleotides de novo biosynthesis I towards ATP	125.829	*1.54*
fig|539814.26.peg.3299	Ribonucleotide reductase of class Ia (aerobic)%2C beta subunit (EC 1.17.4.1)	Superpathway of purine nucleotides de novo biosynthesis I towards ATP	125.565	*1.54*
fig|539814.26.peg.3300	Ribonucleotide reductase of class Ia (aerobic)%2C alpha subunit (EC 1.17.4.1)	Superpathway of purine nucleotides de novo biosynthesis I towards ATP	107.118	*1.31*
fig|539814.26.peg.108	Nucleoside diphosphate kinase (EC 2.7.4.6)	Superpathway of purine nucleotides de novo biosynthesis I towards ATP	583.911	*7.16*
fig|539814.26.peg.2963	Carbamoyl-phosphate synthase small chain (EC 6.3.5.5)	UMP biosynthesis II	60.6784	*0.74*
fig|539814.26.peg.2962	Carbamoyl-phosphate synthase large chain (EC 6.3.5.5)	UMP biosynthesis II	88.6279	*1.09*
fig|539814.26.peg.3305	Aspartate carbamoyltransferase (EC 2.1.3.2)	UMP biosynthesis II	0	*0.00*
fig|539814.26.peg.840	Dihydroorotase (EC 3.5.2.3)	UMP biosynthesis II	0	*0.00*
fig|539814.26.peg.841	Dihydroorotase (EC 3.5.2.3)	UMP biosynthesis II	0	*0.00*
fig|539814.26.peg.842	Dihydroorotase (EC 3.5.2.3)	UMP biosynthesis II	0	*0.00*
fig|539814.26.peg.741	Dihydroorotate dehydrogenase (EC 1.3.3.1)	UMP biosynthesis II	17.3956	*0.21*
fig|539814.26.peg.1413	Orotate phosphoribosyltransferase (EC 2.4.2.10)	UMP biosynthesis II	32.2728	*0.40*
fig|539814.26.peg.575	Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23)	UMP biosynthesis II	47.1445	*0.58*
fig|539814.26.peg.3457	Cytidylate kinase (EC 2.7.4.25)	UTP and CTP de novo biosynthesis	85.7158	*1.05*
fig|539814.26.peg.158	Uridine monophosphate kinase (EC 2.7.4.22)	UTP and CTP de novo biosynthesis	423.198	*5.19*
fig|539814.26.peg.108	Nucleoside diphosphate kinase (EC 2.7.4.6)	UTP and CTP de novo biosynthesis	583.911	*7.16*
fig|539814.26.peg.577	CTP synthase (EC 6.3.4.2)	UTP and CTP de novo biosynthesis	72.1264	*0.88*
fig|539814.26.peg.3299	Ribonucleotide reductase of class Ia (aerobic)%2C beta subunit (EC 1.17.4.1)	pyrimidine deoxyribonucleotides de novo biosynthesis I	125.565	*1.54*
fig|539814.26.peg.3300	Ribonucleotide reductase of class Ia (aerobic)%2C alpha subunit (EC 1.17.4.1)	pyrimidine deoxyribonucleotides de novo biosynthesis I	107.118	*1.31*
fig|539814.26.peg.108	Nucleoside diphosphate kinase (EC 2.7.4.6)	pyrimidine deoxyribonucleotides de novo biosynthesis I	583.911	*7.16*
fig|539814.26.peg.1418	Deoxyuridine 5'-triphosphate nucleotidohydrolase (EC 3.6.1.23)	pyrimidine deoxyribonucleotides de novo biosynthesis I	107.031	*1.31*
fig|539814.26.peg.2617	Nucleoside triphosphate pyrophosphohydrolase MazG (EC 3.6.1.8)	pyrimidine deoxyribonucleotides de novo biosynthesis I	0	*0.00*
fig|539814.26.peg.3584	Thymidylate synthase (EC 2.1.1.45)	pyrimidine deoxyribonucleotides de novo biosynthesis I	159.713	*1.96*
fig|539814.26.peg.2286	Permeases of the major facilitator superfamily / Thymidylate kinase (EC 2.7.4.9)	pyrimidine deoxyribonucleotides de novo biosynthesis I	12.5539	*0.15*
fig|539814.26.peg.2677	Thymidylate kinase (EC 2.7.4.9)	pyrimidine deoxyribonucleotides de novo biosynthesis I	202.819	*2.49*
fig|539814.26.peg.108	Nucleoside diphosphate kinase (EC 2.7.4.6)	pyrimidine deoxyribonucleotides de novo biosynthesis I	583.911	*7.16*

References for supplement of chapter II

Cano, I., van Aerle, R., Ross, S., Verner-Jeffreys, D. W., Paley, R. K., Rimmer, G. S. E., Ryder, D., Hooper, P., Stone, D., & Feist, S. W. (2018). Molecular characterization of an *Endozoicomonas*-like organism causing infection in the king scallop (*Pecten maximus* L.). *Applied and Environmental Microbiology*, *84*(3), 1–14.

Zielinski, F. U., Pernthaler, A., Duperron, S., Raggi, L., Giere, O., Borowski, C., & Dubilier, N. (2009). Widespread occurrence of an intranuclear bacterial parasite in vent and seep bathymodiolin mussels. *Environmental Microbiology*, *11*(5), 1150–1167.

Chapter III | Evolution of *Ca*. Endonucleobacter

Conquering the nucleus: Inhibitors of apoptosis are the genomic innovation that originated intranuclear lifestyle in *Hahellaceae*

Miguel-Ángel González-Porras[1], Rebecca Ansorge[2], Harald Gruber-Vodicka[1], Nicole Dubilier[1], Nikolaus Leisch[1]

1 Department of Symbiosis, Max Planck Institute for Marine Microbiology, Bremen, Germany

2 Quadram Institute Bioscience, Norwich Research Park, Norwich, UK

The manuscript is in preparation and it has not been revised by all authors.

Author contributions

MAGP conceived the study, performed the fluorescence microscopy experiments, analyzed the data, prepared the figures/tables and wrote the manuscript. RA helped to conceive the study and contributed to the conceptual design of all genomic analyses. HGV contributed to the conceptual design and data interpretation of the genomic analyses of the study, as well as to the sequencing strategies for data acquisition. ND helped to conceive the study. NL helped to conceive the study and performed the ultrastructure experiments.

Abstract

Most members of the genus Endozoicomonas (*Oceanospirillales, Hahellaceae*) are facultative mutualists of a wide range of marine metazoans. Within the host, *Endozoicomonas* plays a role in carbohydrate transport and cycling and amino acid delivery. *Ca.* Endonucleobacter, a monophyletic clade of intranuclear parasites of bathymodiolin mussels, is the sister clade of *Endozoicomonas*. By comparing *Ca.* Endonucleobacter and *Endozoicomonas* genomes, we studied the factors that might have played a role in the origin of intranuclear lifestyle in *Hahellaceae*. *Ca.* Endonucleobacter genomes were reduced, showed a low GC% content and lacked for more than 10 amino acid synthetic routes. Further, both genera were enriched in mobile elements, suggesting genomic plasticity. In addition, inhibitors of apoptosis were abundant in *Ca.* Endonucleobacter, while virtually absent in *Endozoicomonas*. Last, *Ca.* Endonucleobacter showed functional adaptations to different bathymodiolin hosts. Its reduced genome and loss of metabolic autonomy might indicate that *Ca.* Endonucleobacter has become an obligated intranuclear parasite. Our findings suggest that genomic plasticity might have played a role in the transition from a mutualistic to a parasitic lifestyle. However, we hypothesize that the expansion of inhibitors of apoptosis are the key genomic innovation that allowed certain members of the *Hahellaceae* family to set the nucleus as a permanent residence. By the first time, a genetic comparative approach has been used to elucidate the origin of intranuclear lifestyle in bacteria. This study is a milestone in the understanding of the genomic perquisites for the colonization of nuclei by bacteria.

Keywords: *Ca.* Endonucleobacter; *Endozoicomonas*; *Hahellaceae*; intranuclear parasitism; genome reduction; genome plasticity; inhibitors of apoptosis; bathymodiolin mussels.

Introduction

Endozoicomonas (*Oceanospirillales, Hahellaceae*) is a bacterial genus known to establish symbiotic relationships with a great diversity of marine metazoans, including cnidarians, poriferans, molluscs, annelids, tunicates and fish (Fiore et al. 2015; Forget and Kim Juniper 2013; Jensen et al. 2010; Katharios et al. 2015; Morrow et al. 2012). *Endozoicomonas* has been described as an extremely diverse and flexible group of symbionts occurring worldwide (Neave et al. 2017). Although *Ca.* Endozoicomonas cretensis has been described as pathogen of fish, the other members of this genus are mutualistic symbionts required for the normal functioning of their host (Katharios et al. 2015; Neave et al. 2016; Qi et al. 2018). *Endozoicomonas* genomes are enriched in mobile elements, which have been considered the main drivers of their flexibility in terms of hosts and symbiotic lifestyle that they can adopt (Neave et al. 2017). Mobile elements can provide microorganisms with the genome malleability that is required for rapid adaptation to new niches.

Ca. Endonucleobacter is a monophyletic clade of gammaproteobacterial parasites that infect the nuclei of bathymodiolin mussels (Zielinski et al. 2009, González-Porras et al, *in Prep*). After a single *Ca*. Endonucleobacter cell invades the host nucleus, it proliferates massively (up to 80,000 bacteria per nucleus) and the volume of the nucleus increases up to 50 fold. At the end of the infectious cycle the host cell bursts, releasing a bacterial progeny which is ready to infect new nuclei (González-Porras et

al, *in Prep*; Zielinski et al. 2009). Several phylogenetic studies based on 16S rRNA placed *Ca.* Endonucleobacter as a sister clade of the almost exclusively mutualistic genus *Endozoicomonas* (Pike et al. 2013; Schreiber et al. 2016; Zielinski et al. 2009). Moreover, while *Endozoicomonas* has a wide range of hots, *Ca.* Endonucleobacter is restricted to bathymodiolin mussels. In contrast to *Endozoicomonas*, *Ca.* Endonucleobacter has developed a highly specialized lifestyle adopting the nuclei of bathymodiolin mussels as replication niche (González-Porras et al, *in Prep*; Zielinski et al. 2009).

Intranuclear lifestyle has many potential advantages for bacteria: The nucleus is a nutrient-rich subcellular compartment where they can hide from cytoplasmic defense mechanisms (Schulz and Horn 2015). Yet proliferation within the nucleus is not extent of risks, as eukaryotic cells can sense nuclear deformation and respond to it. The outer nuclear membrane protein nesprin-1 anchors the nucleus to the actin cytoskeleton, acting as a mechanical transmitter during nuclear deformation (Zhang et al. 2009). In metazoans, caspases are the main regulators of apoptosis, a mechanism of programmed cell death that is triggered in compromised cells (Crawford et al. 2012). Because one of the intrinsic triggers of the apoptotic cascade is cytoskeleton deformation (Kräter et al. 2018), proliferation within the nucleus can compromise the long-term viability of the replication niche for intranuclear bacteria. Therefore, intranuclear bacteria infecting metazoans must count on mechanisms that prevent the shutting down of the host cell to complete their life cycle.

In this study, we generated a high-quality genome of the previously described *Ca.* Endonucleobacter bathymodioli (Zielinski et al. 2009). To uncover interspecific

differences within the genus *Ca.* Endonucleobacter, we compared this genome with the genome of *Ca.* Endonucleobacter infecting *"B." childressi* (González-Porras et al, in Prep). Our results suggested that *Ca.* Endonucleobacter showed functional adaptations to different bathymodiolin hosts that were reflected in their differential distribution pattern. We also wanted to identify factors that might have played a role in the origin of intranuclear lifestyle in *Hahellaceae*. To do so, we compared both *Ca.* Endonucleobacter genomes (parasites) with *Endozoicomonas* (mutualists) and *Hahella* (free-living) genomes. *Ca.* Endonucleobacter spp. genomes were reduced, had a low GC% content and lacked for the capability to synthesize more than 10 proteinogenic amino acids, suggesting an obligated parasitic lifestyle. Like *Endozoicomonas*, *Ca.* Endonucleobacter spp. genomes were enriched in mobile elements, which might have conferred the genomic plasticity required to switch between symbiotic lifestyles. In addition, we found that inhibitors of apoptosis were expanded in *Ca.* Endonucleobacter. Our findings suggested that genomic plasticity might have favored the transition between mutualistic and parasitic lifestyles. However, it is ultimately the expansion of apoptosis inhibitors what allowed *Ca.* Endonucleobacter to conquer the nuclei of bathymodiolin mussels. Therefore, we expect this study to shed light over the genetic innovations that might cause the appearance of intranuclear lifestyle in bacteria.

Results

Ca. *Endonucleobacter bathymodioli genome sequencing and assembly*

Metagenomic binning was used to obtain a high-quality genome from *Ca.* Endonucleobacter bathymodioli (98.28% completeness). This genome was formed by 72 contigs, had a 3.8 Mb size, 3,995 protein-coding sequences and a GC% content of 40% (**Supplementary table 1**). *Ca.* E. bathymodioli and *Ca.* Endonucleobacter infecting *"B." childressi* (from now onwards, *Ca.* E. chitolyticus; see Discussion) had an average amino acid identity of 78% (**Supplementary Figure 3**).

Single-copy orthologous genes phylogeny and genomic features

We calculated the pangenome of the 21 genomes included in this study by running an orthology analysis using OrthoFinder (Emms and Kelly 2019). The core genome was formed by 28,298 genes (**Figure 1, A**). From these core genes, 10,794 were assigned to 514 single-copy gene orthogroups. To determine the evolutionary relationship of *Ca.* Endonucleobacter *spp*. and Endozoicomonas *spp*., we analyzed these 514 single-copy orthologous gene sequences using the ete3 pipeline from the ETEToolkit (Huerta-Cepas et al. 2016) (**Fig.1, B**). Our phylogenomic analysis confirmed that *Ca.* Endonucleobacter and *Endozoicomonas* are monophyletic genera (100% support) (**Fig.1, B**). Both genera were sister clades, and they clustered together as a monophyletic group. (100% support). This monophyletic group formed by the *Ca.* Endonucleobacter spp. + *Endozoicomonas* spp. tandem showed a longer branch length than the free-living outgroup (*Hahella* spp.).

The *Hahella* spp. genomes ranged in sizes from 6.5 to 7.2 Mb, while the 16 *Endozoicomonas* spp. genomes ranged in sizes from 4.0 to 6.7 Mb. In contrast, *Ca.* E.

bathymodioli and *Ca.* E. chitolyticus genomes were reduced, with sizes of 3.9 and 3.3 Mb, respectively (**Fig.1, B**). Last, all the genomes from the genera *Hahella* and *Endozoicomonas* showed GC% values close to 50%. Contrarily, *Ca.* Endonucleobacter spp. genomes had GC% values close to 40% (**Fig.1, B**).

Missing routes for amino acid synthesis

Loss of metabolic autonomy is a common feature of obligated symbionts. We wanted to investigate the genomic capabilities for *de novo* synthesis of proteinogenic amino acids of the 21 genomes under study. **Fig.1, B** shows the number of missing routes per genome, but a detailed description of missing amino acids is covered in **Supplementary table 2** and listed next. The members of the genus *Hahella* had the pathways to synthesize all proteinogenic amino acids, except for *Hahella chejuensis*, which lacked the ability to synthesize Phenylalanine (**Suppl. Table 2**). *Endozoicomonas* sp OPT23 and *Ca.* Endozoicomonas cretensis had the pathways to synthesize all proteinogenic amino acids. *Endozoicomonas* sp. AB1 *Bugula neritia* lacked the ability to synthesize Methionine. The other members of the genus *Endozoicomonas* were unable to synthesize Tyrosine. In addition to their inability to synthesize Tyrosine, *Endozoicomonas numazuensis* and *Endozoicomonas arenosclerae* E-MC227 lacked the genomic potential to synthesize Tyrosine (**Suppl. Table 2**). Both *Ca*. Endonucleobacter species lacked the ability to synthesize 10 amino acids (Valine, Leucine, Isoleucine, Threonine, Asparagine, Cysteine, Methionine, Tyrosine, Arginine and Histidine). In addition, *Ca*. E. chitolyticus lacked the ability to synthesize Phenylalanine (**Suppl. Table 2**).

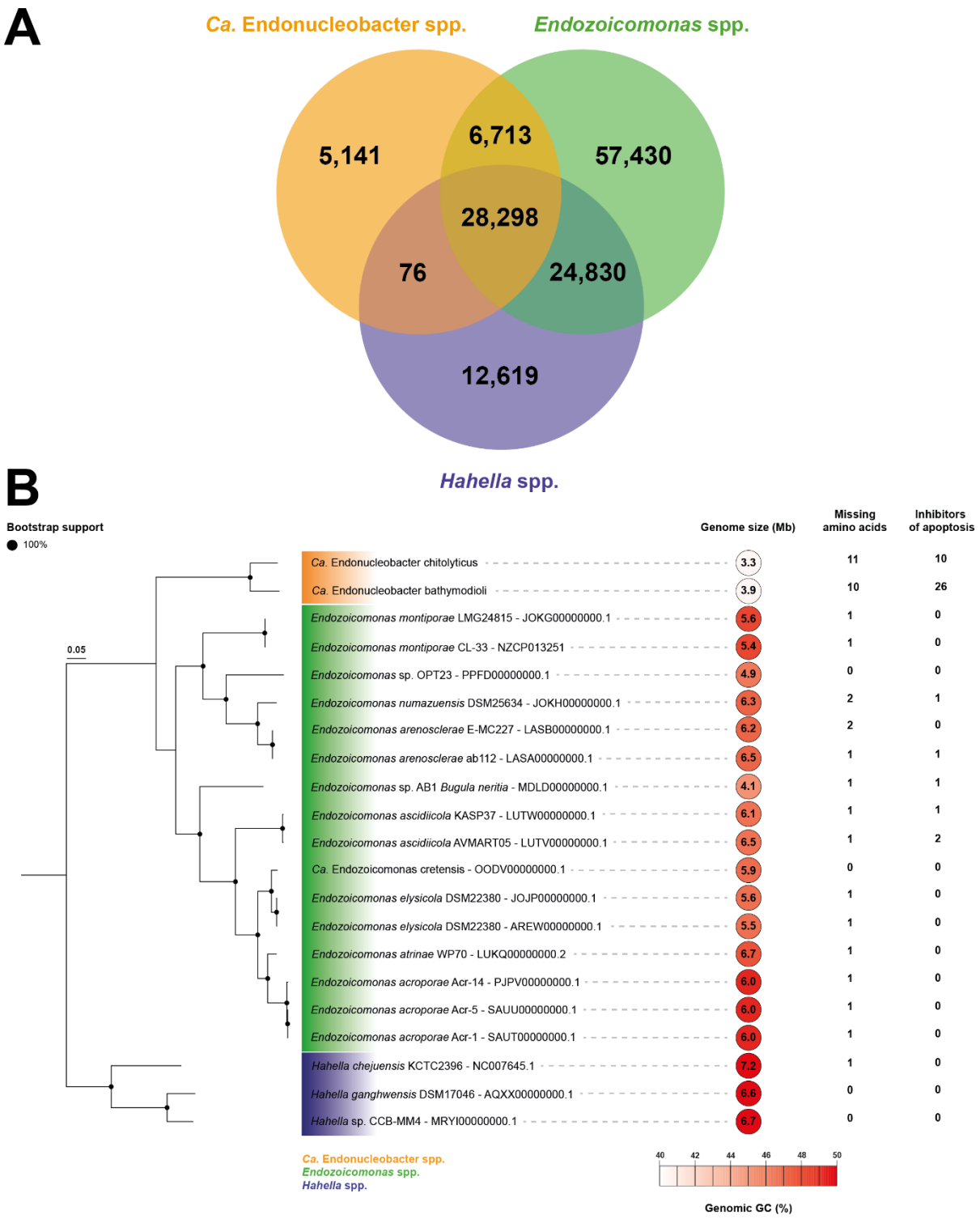


Figure 1. *Ca.* Endonucleobacter spp. and *Endozoicomonas* spp. are monophyletic genera that cluster as sister groups. **A,** Pangenome of the 21 genomes included in this study. The core genome was formed by 28,298 genes. **B,** phylogenetic analysis based on 514 single-copy orthologous genes extracted from the core genome. Genome size, GC% content, number of missing routes for amino acid synthesis and number of inhibitors of apoptosis per genome has been included in the figure.

COG categories

To determine the functionality of the genomes included in the dataset, abundance of genes within clusters of orthologous groups (COG) were compared between *Ca.* Endonucleobacter spp., *Endozoicomonas* spp. and *Hahella* spp. (**Figure 2**). Both *Ca*. *Ca.* Endonucleobacter spp. had less genes (<100) dedicated to amino acid transport and metabolism than *Endozoicomonas* spp. (>200) and *Hahella* spp. (>400). This data supported the lack of amino acid synthetic routes for *Ca.* Endonucleobacter spp (**Supp. Table 2**). *Ca.* Endonucleobacter spp. had less genes (<100) dedicated to lipid and carbohydrate metabolism than *Endozoicomonas* spp. and *Hahella* spp. (>100). *Ca.* Endonucleobacter spp. had less genes (<100) dedicated to transcription than *Endozoicomonas* spp. (>200) and *Hahella* spp. (>100). Remarkably, *Ca.* Endonucleobacter spp had >600 genes for replication, recombination and repair. This was similar for Endozoicomonas spp. (200-700), while lower in the *Hahella* representatives. Due to their reduced genome, 600 genes dedicated to this category accounted for more than 20% of *Ca.* Endonucleobacter spp. total genes. *Ca.* Endonucleobacter spp. had less genes (<100) dedicated to cell motility than *Endozoicomonas* spp. (>100) and *Hahella* spp. (>175). *Ca.* Endonucleobacter had the lowest number of genes dedicated to defense mechanisms (<30) in contrast to *Endozoicomonas* spp and *Hahella* spp. (50-75). Last, Endonucleobacter spp. had less genes (<100) dedicated to post-translational modification, protein turnover and chaperones than *Endozoicomonas* spp. (>125) and *Hahella* spp. (>175). However, genes in this category accounted for more than 4% of the total genes of Ca. Endonucleobacter spp., showing higher relative abundance than *Endozoicomonas* and *Hahella* representatives (<4%).

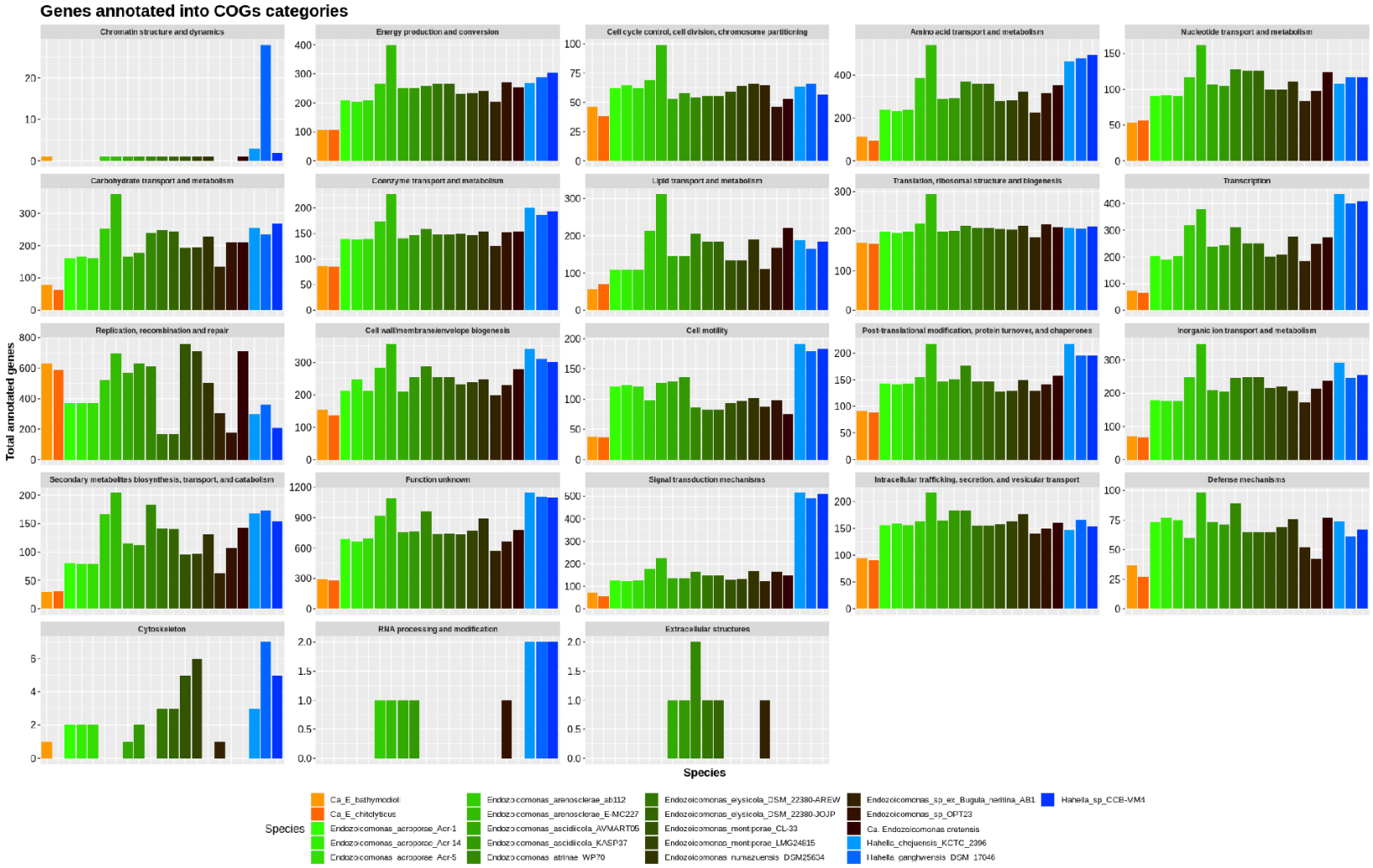


Figure 2. *Ca.* Endonucleobacter spp. encoded for approximately 600 genes involved in recombination and mobile elements, which accounts for more than 20% of their genomic content.

Number of annotated genes assigned to COG categories for *Ca.* Endonucleobacter spp. (orange), *Endozoicomonas* spp. (green) and *Hahella* spp. (blue).

Inhibitors of apoptosis

As highlighted in the previous section, *Ca.* Endonucleobacter spp. showed the highest relative abundance of genes dedicated to post-translational modification, protein turnover and chaperones (>4%), in contrast to *Endozoicomonas* spp. and *Hahella* spp. (<4%). One of the gene families that was assigned to this category were inhibitors of apoptosis (IAPs). *Ca.* E. bathymodioli encoded for 26 IAPs, while *Ca.* E. chitolyticus encoded for 10 IAPs (**Fig.1, B**). In contrast, IAPs were scarcely represented in *Endozoicomonas* representatives. Only three *Endozoicomonas* species encoded for IAPs: *E. ascidiicola* (2), *E. atrinae* (1) and *E. elysicola* (1) (**Fig.1, B**). IAPs were absent in *Hahella* spp. Analysis of orthology grouped all IAPs from the pangenome in 8 orthogroups. The phylogenetic analysis of 6 of the 8 IAPs orthogroups revealed that IAPs from *Endozoicomonas* spp. clustered with *Ca.* Endonucleobacter IAPs (**OG0001941, OG0005541, Supplementary figure 1**).

Interspecific genomic differences of* Ca. *Endonucleobacter spp.

The assembly of a high-quality genome for *Ca.* E. bathymodioli allowed us to investigate differences in genomic potential between *Ca.* E. bathymodioli and *Ca.* E. chitolyticus. To highlight these differences, we adapted the metabolic reconstruction of González-Porras et al*. in Prep* to include those genetic factors that were exclusive for one *Ca.* Endonucleobacter species (**Figure 3**). *Ca.* E. chitolyticus encoded for the Shigella-like factor *IpgD*, involved in cellular colonization. In addition, *Ca.* E. chitolyticus encoded for 9 CRISPR-Cas associated proteins, involved in viral defense. Next, *Ca.* E. chitolyticus also encoded for a fermentation/oxidation switch protein (FRSP) and for an alcohol dehydrogenase (ADH) involved in alcoholic fermentation. Last, *Ca.* E.

bathymodioli codified for 7 subunits of the V-type ATP synthase complex, involved in cellular acidification.

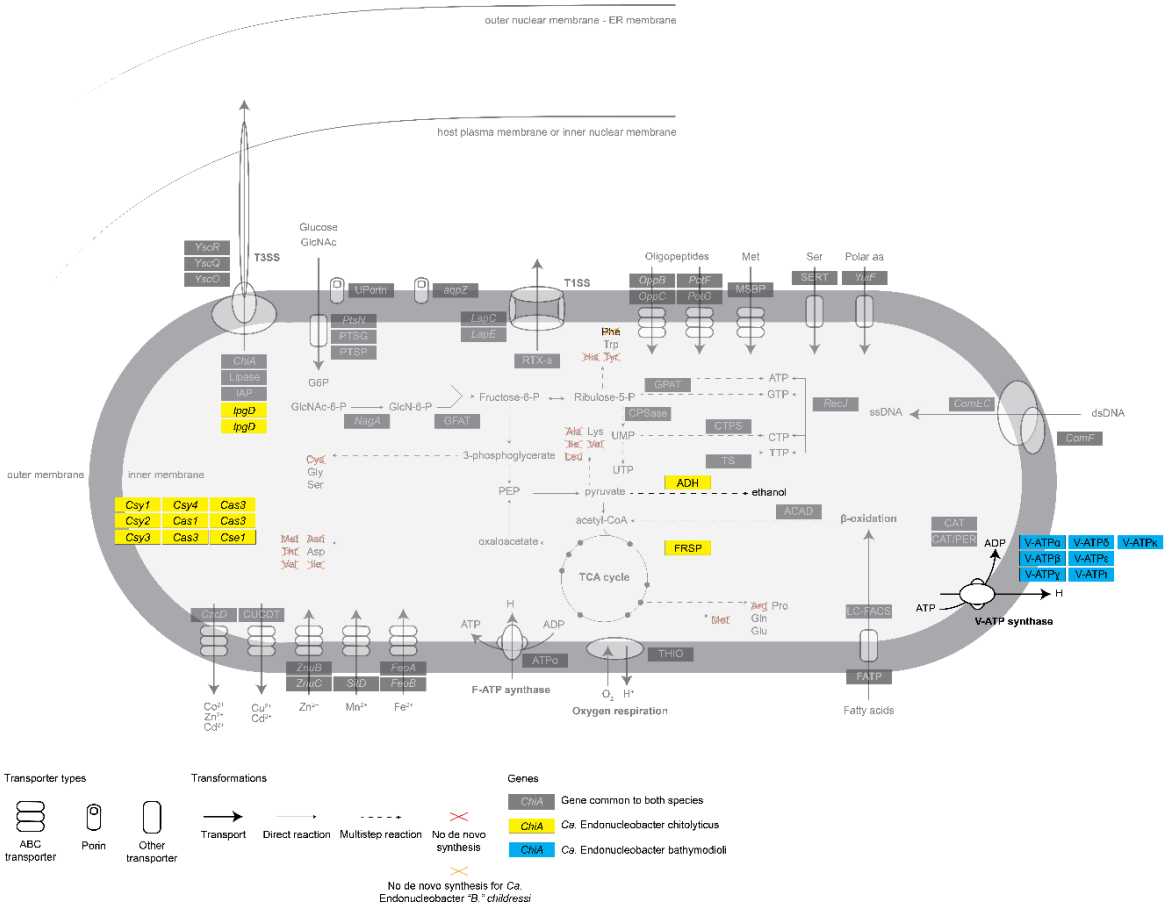


Figure 3. Genomic potential comparison between *Ca.* E. chitolyticus and *Ca.* E. bathymodioli. Genomic potential reconstruction based on RAST annotation and Pathway Tools metabolic modeling. The figure has been adapted from González-Porras et al. *in Prep*. **Yellow:** Genetic factors exclusive of *Ca.* Endonucleobacter chitolyticus. **Blue:** Genetic factors exclusive of *Ca.* Endonucleobacter bathymodioli. **Fade:** Genetic factors and routes present in both microorganisms. ***YscR***, T3SS inner membrane protein *YscR*. ***YscQ***, T3SS inner membrane protein *YscQ*. ***YscO***, T3SS bacterial envelope protein *YscO*. ***ChiA***, chitinase. **Lipase**, lipase. **IAP**, Inhibitor of apoptosis. ***IpgD***, Shigella-like Inositol phosphate phosphatase. **Uporin**, undetermined porin. ***agpZ***, aquaporin. ***LapC***, T1SS membrane fusion protein. ***LapE***, T1SS outer membrane component. **RTX-a**, *Vibrio*-like repeats-in-toxin adhesin. ***NagA***, N-acetylglucosamine-6-phosphate deacetylase. **GFAT**, Glucosamine-fructose-6-phosphate aminotransferase. ***CzcD***, Cobalt-zinc-cadmium resistance protein *CzcD*. **CUCDT**, Copper-translocating P-type ATPase. ***ZnuB***, Zinc ABC transporter inner membrane permease protein *ZnuB*. ***ZnuC***, Zinc ABC transporter ATP-binding protein *ZnuC*. ***SitD***, Manganese ABC transporter inner membrane permease protein *SitD*. ***FeoA***, Ferrous iron transport protein A. ***FeoB***, Ferrous iron transport protein B. **ATPα**, ATP synthase alpha chain. **THIO**, Thioredoxin. **FATP**, Long-chain fatty acid transport protein. **LC-FACS**, Long-chain-fatty-acid-CoA ligase. **ACAD**, Acyl-CoA dehydrogenase. **CAT/PER**, Catalase / Peroxidase. **CAT**, Catalase. ***ComF***, Competence protein F homolog. ***ComEC***, DNA internalization-related competence protein ComEC. ***RecJ***, Single-stranded-DNA-specific exonuclease *RecJ*. ***OppB***, Oligopeptide transport system permease protein *OppB*. ***OppC***, Oligopeptide

transport system permease protein *OppC*. ***PotF***, Putrescine ABC transporter putrescine-binding protein PotF. ***PotG***, Putrescine transport ATP-binding protein *PotG*. **MSBP**, Methionine ABC transporter substrate-binding protein. **SERT**, Serine transporter. ***YuiF***, Histidine permease *YuiF*. **GPAT**, Amidophosphoribosyltransferase. **CPSase**, Carbamoyl-phosphate synthase large chain. **CTPS**, CTP synthase. **TS**, Thymidylate synthase. ***Csy1-4,*** CRISPR-related protein Csy1-4. ***Cas1,*** CRISPR-related protein Cas1. ***Cas3,*** CRISPR-related protein Cas3. ***Cse1,*** CRISPR-related protein Cse1. **FRSP**, Fermentation-respiration switch protein. **ADH**, Alcohol dehydrogenase. **V-ATPα**, V-type ATP synthase subunit alpha. **V-ATPβ**, V-type ATP synthase subunit beta. **V-ATPɣ**, V-type ATP synthase subunit gamma. **V-ATPδ**, V-type ATP synthase subunit delta. **V-ATPε**, V-type ATP synthase subunit epsilon. **ATPι**, V-type ATP synthase subunit iota. **ATPκ**, V-type ATP synthase subunit kappa.

Differential localization pattern of Ca. *Endonucleobacter spp.*

We hypothesized that the differences in genomic potential between *Ca.* E. bathymodioli and *Ca.* E. chitolyticus might be indicative of a differential colonization strategy of their bathymodiolin hosts (*B. puteoserpentis* vs. *"B." childressi*, respectively). To confirm our hypothesis, we determined the localization pattern of *Ca.* E. bathymodioli in *B. puteoserpentis*, and compared it with the localization pattern of *Ca.* E. chitolyticus in *"B." childressi*. *B. puteoserpentis* and *"B." childressi* specimens were sampled in the Mid-Atlantic Ridge and Gulf of México, respectively (**Figure 4, a**). Like other bathymodiolin mussels, *B. puteoserpentis* and *"B". childressi* have huge brown gills (**Fig.4, b**) formed by thousands of gill filaments (**Fig.4, c, f**). Base on the presence or absence of mutualistic symbionts, two anatomically distinct areas can be distinguished in the gill filaments: The symbiotic region, where the mutualistic symbionts occur and the ciliated edge, where there are not mutualistic symbionts (**Fig.4, d, g**). Fluorescence *in situ* hybridization (FISH) targeting *Ca*. Endonucleobacter spp. on whole filaments revealed that *Ca.* E. bathymodioli occurred in the symbiotic region of *B. puteoserpentis* (**yellow scattered signal in Fig.4, c**). In contrast, *Ca*. E. chitolyticus was limited to the ciliated edge in *"B." childressi* (**yellow marginal rim in Fig.4, f**). A closer look confirmed that *Ca.* E. bathymodioli infects symbiont-free cells of the symbiotic region of *B. puteoserpentis* (**Fig.4, d**), while *Ca.* E. chitolyticus never

occurs in the symbiotic region of *"B." childressi* (**Fig.4, g**). Regardless of the host mussel species, *Ca*. Endonucleobacter spp. forms massive bacterial clusters (**Fig.4, e, h**).

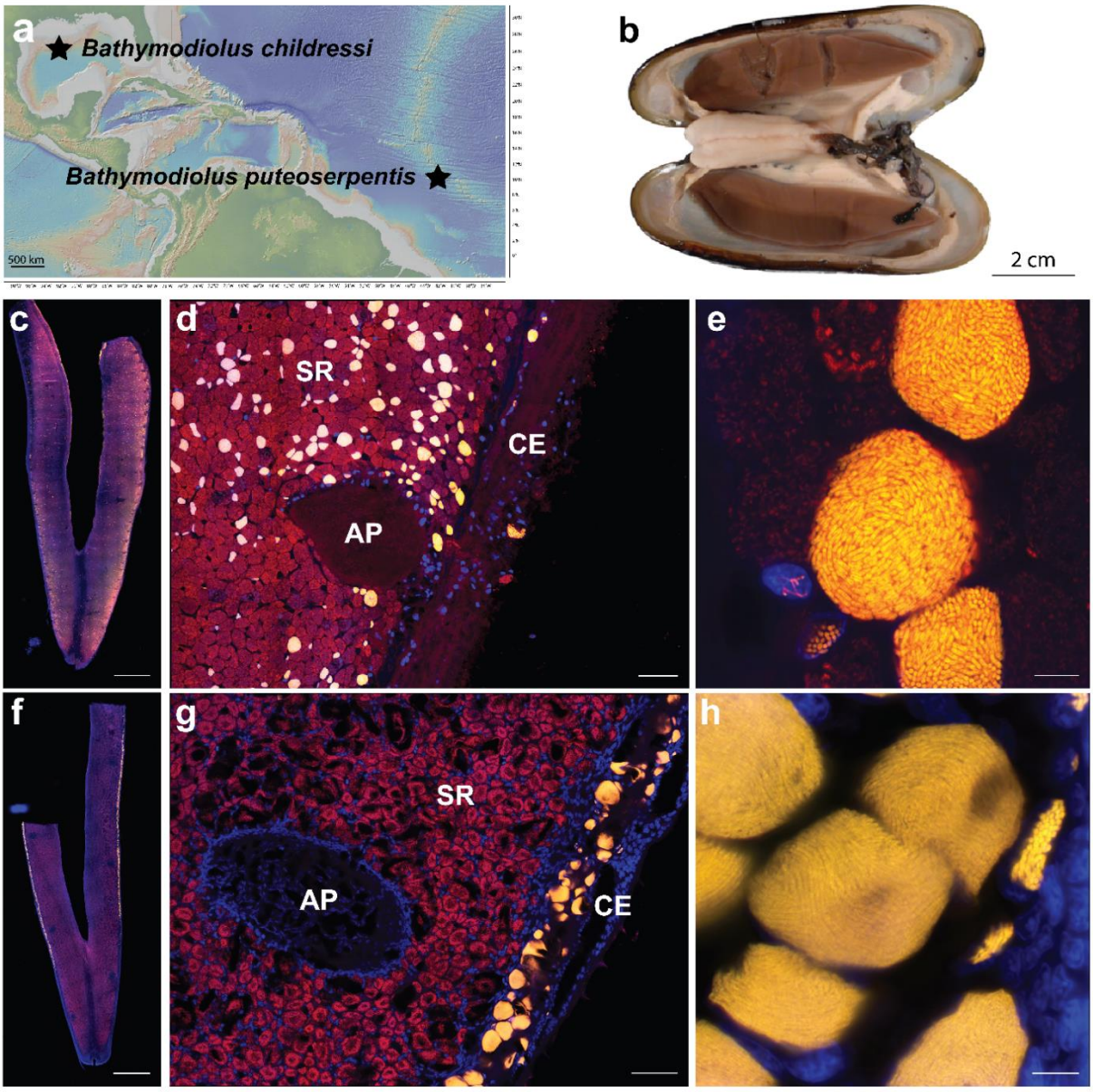


Figure 4. *Ca.* E. bathymodioli infects symbiont-free cells from the symbiotic region in *B. puteoserpentis*, while *Ca.* E. chitolyticus is limited to the ciliated edge in "*B". childressi*. *B. puteoserpentis* was sampled in the Mid-Atlantic Ridge while *"B." childressi* was sampled in the Gulf of México (a). The mussels' gills were dissected (brown structures in b) and single gill filaments (c, f) were subjected to Whole-mount FISH. (**SR**: Symbiotic area, **CE**: Ciliated edge, **AP**: Adhesion patch; d, g). Eubacteria 16S rRNA (red: Alexa Fluor 647), *Ca*. Endonucleobacter 16S rRNA (yellow: Atto550), DNA (blue: DAPI). ***c, d, e,*** *B. puteoserpentis*; ***f, g, h,*** *"B." childressi* . ***c, f,*** epifluorescence overviews of whole filaments, scalebars: 2 mm. ***d, g,*** CLSM images of adhesion patch surroundings, scalebars: 50 µm. ***e, h,*** SR-Airyscan images of infection area, scalebars: 5 µm. *"B." childressi* images have been adapted from González-Porras et al., *in Prep*.

Discussion

Ca. *E. bathymodioli is a separated species from* Ca. *Endonucleobacter infecting "B." childressi*

The assembly of a high-quality genome for *Ca.* E. bathymodioli allowed us to compare its genomic features with the previously published genome for *Ca.* Endonucleobacter infecting *"B." childressi* (González-Porras et al., in Prep). Both genomes had an average amino acid identity of 78% (**Supplementary Figure 3**), below the species cut-off level of 85% (Goris et al. 2007). This data confirmed that *Ca.* E. bathymodioli (which infects *B. puteoserpentis* in the Mid-Atlantic Ridge) is a separated species from *Ca.* Endonucleobacter infecting *"B." childressi*, a deep-sea mussel that occurs in the Gulf of México (Zielinski et al. 2009).

As previously described in González-Porras et al., in Prep, *Ca.* Endonucleobacter infecting *"B." childressi* seems to profit nutritionally in extracellular chitin occuring in the gill filaments of its host. Therefore, we propose the name of *Ca.* E. chitolyticus for *Ca.* Endonucleobacter infecting *"B." childressi*.

Ca. Endonucleobacter spp. are obligated intranuclear parasites

When analyzing the features of the 21 genomes included in this study, we reported that *Endozoicomonas* spp. genomes ranged in sizes between 4.0 to 6.7 Mb, while *Hahella* representatives ranged in sizes between 6.5 to 7.2 Mb. Alternatively, *Ca.* E. bathymodioli and *Ca.* E. chitolyticus genomes were reduced, with sizes of 3.9 and 3.3 Mb, respectively (**Fig.1, B**). The sister clade of *Ca.* Endonucleobacter, *Endozoicomonas*, are thought to have a free-living stage due to their generalized big genome sizes (> 5 Mb) (Neave et al. 2016). The transition from a free-living to an

intracellular lifestyle is commonly associated with the loss of fragments of DNA (Fraser-Liggett, 2005; Gil, et al. 2004). Our results revealed that *Ca.* E. bathymodioli and *Ca.* E. chitolyticus had incomplete *de novo* synthesis routes for 10 and 11 proteinogenic amino acids, respectively (**Fig.1, B; Supp. Table 2**). Loss of metabolic autonomy is another common phenomenon in bacteria that become highly dependent on their hosts (Casadevall 2008). We hypothesized that *Ca.* Endonucleobacter spp. became metabolically dependent on its host cell when acquiring obligated intranuclear lifestyle. The constant supply of nutrients by the host cell might have relaxed the selective pressure over *Ca.* Endonucleobacter spp. to keep functional routes for amino acid synthesis. Our analysis also revealed that the genomes of *Endozoicomonas* and *Hahella* representatives had GC% values closed to 50%. In contrast, *Ca.* Endonucleobacter spp. had GC% values of approximately 40% (**Fig.1, B**). In addition to genome reduction, low genomic GC content has been associated to obligated intracellular lifestyle (Moran 2002). Their reduced genome, loss of metabolic autonomy and low GC% content suggest that *Ca.* Endonucleobacter spp. are obligated intranuclear parasites. This contrasts with the lifestyle of their sister clade, *Endozoicomonas*, which have been described as mutualistic and facultative symbionts in their vast majority (Neave et al. 2016). The opposite life strategies of both sister clades highlights the flexibility of this subgroup of *Hahellaceae* in terms of hosts and symbiotic lifestyles.

***Genomic plasticity and inhibitors of apoptosis might have originated intranuclear lifestyle in* Hahellaceae**

Assignment of annotated genes into COG categories showed that *Ca.* Endonucleobacter spp. had more than 600 genes devoted to replication,

recombination and repair (**Fig.2**), which includes mobile elements. Genes from this COG category accounted for more than 20% of the genomic content of *Ca.* Endonucleobacter spp.. Neave et al. 2017 demonstrated that one of the processes enriched in *Endozoicomonas* genomes was transposition and DNA recombination. The authors hypothesized that this might help *Endozoicomonas* to rapidly adapt to new hosts, or to opportunistically transition between symbiotic lifestyles (mutualistic, commensalistic or parasitic). This would explain the enormous plasticity of *Endozoicomonas* in terms of the range of hosts that they can occupy, but also why some *Endozoicomonas* representatives have switched to a parasitic lifestyle as a secondary adaptation (Katharios et al. 2015; Qi et al. 2018). In an analogous way, mobile elements might have played a role in the evolutionary origin of *Ca.* Endonucleobacter spp. We contemplate the scenario in which the last common antecessor of *Ca.* Endonucleobacter spp. and *Endozoicomonas* spp. capitalized on mobile elements to colonize the nuclei of bathymodiolin mussels, resulting in the divergence of both genera.

Ca. Endonucleobacter spp. showed the highest relative abundance of genes dedicated to post-translational modification, protein turnover and chaperones (>4%), a COG category that includes inhibitors of apoptosis (IAPs). While *Ca.* E. bathymodioli encoded for 26 IAPs and *Ca.* E. chitolyticus encoded for 10 IAPs, IAPs were scarce in *Endozoicomonas* representatives and completely absent in *Hahella* spp. (**Fig.1, B**). Thus, we hypothesize that the expansion of IAPs is a genomic signature of *Ca.* Endonucleobacter spp. IAPs have been described as physiologic caspase inhibitors which are able to arrest the apoptosis cascade (Deveraux et al. 1997; Deveraux and Reed 1999). As introduced previously, nuclear deformation due to bacterial proliferation is transmitted to the host cell cytoskeleton via nesprin-1 (Zhang et al.

2009), which can result in the activation of the caspase-mediated apoptotic cascade (Crawford et al. 2012; Kräter et al. 2018). It is tempting to speculate that Ca. Endonucleobacter spp. is using IAPs to prevent the shutdown of the host cell and to complete its life cycle. More than the ultimate genomic innovation that promoted the appearance of intranuclear lifestyle in *Hahellaceae*, IAPs might be the defining genetic element of *Ca.* Endonucleobacter spp. biology.

Ca. Endonucleobacter spp. shows specific adaptations to the gill filaments of its bathymodiolin host

Ca. Endonucleobacter spp. always occurred in non-symbiotic cells of the gill filaments of bathymodiolin mussels. However, while *Ca.* E. bathymodioli occurred in the symbiotic region of *B. puteoserpentis*, *Ca.* E. chitolyticus was limited to the ciliated edge (**Fig.4**). These differences in distribution pattern might be caused by microscale variations in oxygen concentration in the gills of bathymodiolin mussels. *B. puteoserpentis* inhabits hydrothermal vents from the Mid-Atlantic Ridge, where the water column is enriched in oxygen (van der Heijden et al. 2012). In contrast, *"B". childressi* occurs in hydrocarbon seeps of the Gulf of Mexico, where oxygen is thought to be constantly consumed by methane-oxidizing bacteria due to high influx of environmental methane (Boetius and Wenzhöfer 2013). We hypothesize that the mutualistic symbionts of *B. puteoserpentis* cannot consume all oxygen from the gill filament – water column interface during chemosynthesis, which would allow *Ca.* E. bathymodioli to colonize the symbiotic region of *B. puteoserpentis* (**Fig.4, d**). In contrast, the high rates of oxygen consumption by the methane-oxidizing mutualists of *"B". childressi* might be limiting *Ca.* E. chitolyticus to the ciliated edge (**Fig.4, g**).

Interspecific differences in genomic potential supported this hypothesis. *Ca.* E. chitolyticus encoded for a fermentation/respiration switch protein (FRSP), as well as for alcohol dehydrogenase (ADH), which were absent in *Ca.* E. bathymodioli (**Fig.3**). This suggested that *Ca.* E. chitolyticus might be alternating respiration and fermentation when the oxygen concentration in the gill filament – water column interface fluctuate.

Ca. E. chitolyticus encoded for two copies of *IpgD*, a virulence factor used by Shigella to promote its uptake by non-phagocytic cells (enterocytes) (Niebuhr et al. 2000, 2002). Unlike the cells of the symbiotic region, where *Ca.* E. bathymodioli occurs, the ciliated edge cells of bathymodiolin mussels are non-phagocytic. Limitation to the ciliated edge might have promoted the development of specialized strategies for host cell colonization in *Ca.* E. chitolyticus. We hypothesize that *Ca.* E. chitolyticus is using *IpgD* to force phagocytic behavior in the ciliated edge cells of *"B". childressi*.

The interspecific differences in distribution pattern of *Ca.* Endonucleobacter spp. could also have nutritional implications. Chitinase was one of the highest represented factors in the transcriptome and proteome of *Ca.* E. chitolyticus, suggesting that chitin might be an important nutritional source for the intranuclear parasite of *"B". childressi* (González-Porras et al, *in Prep*). Tietjen, *pers. comm.* found differential expression of chitin synthase 2 (*Chs2*) when comparing ciliated edge (high expression) vs. symbiotic region (low expression) transcriptomes of *"B." childressi* gill filaments. We hypothesize that *Ca.* E. chitolyticus might have specialized in chitin metabolism, as it seems to be an abundant nutritional source in the ciliated edge of *"B". childressi*.

Taken all together, these data suggested that *Ca.* E. chitolyticus has developed a highly specialized lifestyle adapted to the ciliated edge of *"B." childressi*. The

impossibility to colonize the symbiotic region might have promoted the appearance of specialized colonization (*lpgD*) or nutritional (chitinase) strategies.

Interspecific differences in Ca. Endonucleobacter spp. pathogenicity

We wanted to investigate whether differences in genomic potential in *Ca.* Endonucleobacter spp. might reflect differential pathogenicity strategies. *Ca.* E. chitolyticus encoded for nine CRISPR-Cas related proteins, which were absent in *Ca.* E. bathymodioli genome (**Fig.3**). CRISPR-Cas systems have been described as the "adaptive immune system" that protects prokaryotes against bacteriophages (van der Oost et al. 2014). Ultrastructural imaging of *Ca.* E. chitolyticus revealed infection by viral particles (**Supplementary Figure 2**). As hypothesized above, *Ca.* Endonucleobacter spp. are obligated intranuclear parasites that have lost metabolic autonomy. However, the presence of a viral-defense mechanism suggested that *Ca*. Endonucleobacter spp. might have a free-living stage, as they are protected against viral infection within the eukaryotic host cell. We hypothesize that CRISPR-Cas systems might help *Ca.* E. chitolyticus to survive the extracellular phase of its life cycle, increasing its dispersion potential and ability to infect mussels' beds separated in space.

Ca. E. bathymodioli codified for a V-type ATPase complex, which was absent in *Ca.* E. chitolyticus (**Fig.3**). V-type ATP are membrane-embedded protein complexes that operate as ATP hydrolysis-driven proton pumps. V-ATPases complexes are the main drivers of organellar acidification in eukaryotes, occurring rarely in prokaryotes (Vasanthakumar and Rubinstein 2020). The intracellular pathogen *Salmonella*

typhimurium acidifies its own cytoplasm as a response to the acidification of the endophagosome. By doing so, S. *typhimurium* regulates the transcription of several virulence factors through *OmpR,* a transcription regulator sensitive to osmotic changes and acidification (Chakraborty et al. 2015). Although not directly annotated as *OmpR*, our orthology analysis assigned a *Ca.* E. bathymodioli hypothetical protein to the same orthogroup than *OmpR* in Endozoicomonas representatives. One could imagine a hypothetical scenario in which *Ca.* E. bathymodioli is acidifying its own cytoplasm to regulate transcription of virulence factors in response to intracellular pH changes in its host cell.

Conclusion and outlook

This work describes the genomic features that might have originated the genus *Ca.* Endonucleobacter, which is a group of intranuclear parasites within the family *Hahellaceae*. *Ca.* Endonucleobacter spp. are the sister clade of the mutualistic genus *Endozoicomonas*. However, unlike their relatives, *Ca.* Endonucleobacter spp. have reduced genomes, low GC% and are metabolically dependent on their host cells. This might be indicative of an obligated parasitic lifestyle. *Ca.* Endonucleobacter spp. had more than 20% of their genomic content devoted to replication, recombination and repair, a COG category that includes mobile elements. Mobile elements might have given the genomic plasticity to the last common antecessor of *Ca.* Endonucleobacter spp. and *Endozoicomonas* spp. to switch from a mutualistic to a parasitic lifestyle. However, we hypothesize that inhibitors of apoptosis are the ultimate genomic innovation that promoted the appearance of intranuclear lifestyle in *Hahellaceae*. Last, we found interspecific differences in *Ca.* Endonucleobacter spp. when adapting to the gill filaments of their bathymodiolin hosts, as well as interspecific differences in pathogenicity strategies. This study has helped to understand the genomic perquisites for the colonization of nuclei by bacteria. Further analyses would imply the precise identification of the mobile elements that enrich the genomes of *Ca.* Endonucleobacter spp., as well as to track the expansion and/loss of inhibitors apoptosis in the *Ca.* Endonucleobacter and *Endozoicomonas* genera.

Materials and methods

Sample collection

B. puteoserpentis mussels were collected with the remote operated vehicle (ROV) MARUM-QUEST 4000m during the during the Sonne M-126 cruise to the Mid-Atlantic Ridge in April 2016. Mussels were collected during one dive from the Irina II venting site (Irina-II, 14º45' N; -044º59' W) at a water depth of 3,036 m.

DNA extraction and infection screening

To look for *Ca*. E. bathymodioli infection, 5 *B. puteoserpentis* gill samples were PCR screened. DNA was extracted from gill samples using the DNeasy Blood and Tissue Kit (QIAGEN, Germany) and used as template in PCR reactions. The *Ca*. E. bathymodioli 16S rRNA gene was amplified by PCR using the following conditions: Initial denaturation for 3 min at 95 °C, followed by 30 cycles at 95 °C for 30 s, 55 °C for 30 s and 72 °C for 2 min, followed by a final elongation step at 72 °C for 10 min. The *Ca*. Endonucleobacter 16S rRNA gene was amplified using the specific forward primer BNIX64 (AGCGGTAACAGGTCTAGC) (Zielinski et al. 2009) and the general reverse primer GM4 (TACCTTGTTACGACTT) (Muyzer et al. 1995). Taq DNA Polymerase (5 PRIME, Hamburg, Germany) was used in all PCR reactions. PCR products were loaded in agarose gels (2%) and stained with ethidium bromide for 30 min. Band thickness and intensity were considered as indicators of the degree of infection.

DNA library preparation and sequencing

The DNA from the three most infected specimens of the 5 *B. puteoserpentis* gill samples screened was sequenced. DNA quality was assessed with an Agilent 2100 Bioanalyzer. Genomic DNA was sequenced using Illumina HiSeq 2500 technology as follows: 33 million 250 bp paired-end reads were generated. Additionally, the DNA from the most infected specimen out of the previous three was further sequenced using Illumina HiSeq 2500 technology and PacBio technology as follows: 333 million long reads were generated.

Genome analysis

To estimate the abundance of *Ca*. Endonucleobacter within the three DNA libraries, we quantify its 16S rRNA gene sequences using using phyloFlash v3.3 (Gruber-Vodicka et al. 2019). Three individual metagenomes were assembled using Spades v3.7 (Bankevich et al. 2012), after decontamination, quality filtering (trimq = 2) and adapter-trimming of the short illumina reads with the BBDuk tool from the BBMap suite (Bushnell B, https://sourceforge.net/projects/bbmap/). The three *Ca*. E. bathymodioli genomes were binned based on genome coverage, GC content and taxonomic affiliation using Gbtools v2.6.0 (Seah and Gruber-Vodicka 2015). To reassemble the genomes, we re-mapped Illumina short reads to the bins using BBMap with 0.98 minimum identity. The reads were reassembled with Spades v3.7, using a maximum k-mer length of 127. Following manual removal of contigs shorter than 1 kB and contamination screening using Bandage v0.8.1 (Wick et al. 2015), quality metrics were calculated with CheckM v1.0.18. Among the three draft genomes, the one with higher completeness and lower fragmentation (number of contigs) was chosen for additional

Illumina HiSeq 2500 sequencing (see previous section). A new short-reads draft genome was assembled from the additional Illumina sequencing library and chosen to proceed further. The PacBio long reads were mapped against the chosen *Ca*. E. bathymodioli short-reads draft genome using the long read mapper ngmlr v.0.2.7 (Sedlazeck et al. 2018). Mapped long-reads were assembled using the long-read assembler CANU v.0.2.7 (Koren et al. 2017). Following an hybrid assembly strategy, the CANU bin was loaded in Unicycler v0.4.8 (Wick et al. 2017) and supplemented with the illumina short-reads from the short-reads draft genome. The resulting hybrid genome was checked for quality metrics with CheckM v1.0.18 and contigs shorter than 1 kB were manually removed. The hybrid genome was annotated with RAST v2.0 (Aziz et al. 2008). The annotations were manually cross-checked and the annotations for the main genes discussed here were verified using v.2.10.1 NCBI's BLAST.

Dataset

A total of 21 genomes were included in the analyzed dataset. The 2 *Ca.* Endonucleobacter spp. genomes (*Ca.* E. bathymodioli, *Ca.* E. chitolyticus) and the 16 *Endozoicomonas* spp. genomes were considered as symbiotic groups. Additionally, 3 *Hahella* spp. genomes were included as negative control (free-living group). Genomic quality parameters were calculated using CheckM v1.0.18 (Parks et al. 2015). Accession numbers, complete names, and quality parameters of the analyzed genomes can be found in **Supplementary Table 1**.

Orthology analysis

To determine the degree of homology among the genes of the studied microorganisms, we did an orthology analysis using OrthoFinder v.2.4.0 (Emms and Kelly 2019). After classification of the whole genomic dataset into orthogroups, genes shared among genera were extracted and represented using the script "OrthomclToVenn.py" from GitHub repository "https://github.com/philippbayer/orthomclToVenn" (**Fig.1, A**).

Phylogenomic analysis

To resolve the phylogenetic relationship of the microorganisms under study, we analyzed 514 single-copy orthologue genes extracted from OrthoFinder analysis. A total of 10,794 CDS were aligned using Clustal Omega (Sievers et al. 2011) and analyzed using FastTree 2 (Price et al. 2010) as part of the ete3 pipeline from the ETEToolkit (Huerta-Cepas et al. 2016). How the taxa clustered in the final tree was determined based on 100 bootstraps.

De novo *synthesis of amino acids*

The completeness of the routes for de novo synthesis of amino acids was determined using the web server tool from pathway tools software (Karp et al. 2009). Any biosynthesis route from which a key gene was missing was considered as incomplete.

Clusters of Orthologous Groups of Proteins (COG categories)

We did a functional prediction of all genes of the microorganisms under study by analyzing the genomic amino acid sequences with eggNOG 5.0 (Huerta-Cepas et al. 2019). The total amount of genes assigned to clusters of orthologous groups of proteins (COG categories) was extracted and represented using R studio v1.3.959.

Transmission electron microscopy

Upon recovery, *"B". childressi* specimens were fixed in 2.5% glutaraldehyde (GA) in PHEM buffer (piperazine-N, N′-bis , 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, ethylene glycol-bis(β-aminoethyl ether and MgCl2 (Montanaro et al. 2016)). After fixation, samples were stored in buffer for transport. Samples were post fixed with 1% (v/v) osmium tetroxide (OsO4) for 2 h at 4 °C, washed three times with PHEM and dehydrated with an ethanol series (30%, 50%, 70%, 80%, 90% and 100% (v/v)) at -10 °C, each step lasting 10 min. The sample was then transferred into 50:50 ethanol and acetone, followed by 100% acetone and infiltrated with low-viscosity resin (Agar Scientific, UK) using centrifugation embedding (McDonald 2014). Samples were centrifuged for 30s in resin: acetone mixtures of 25%, 50%, 75% and twice in 100%. Finally they were transferred into fresh resin in embedding molds and polymerized at 60–65 °C for 48 h. Ultra-thin (70 nm) sections were cut on a microtome (Ultracut UC7 Leica Microsystem, Austria), mounted on formvar-coated slot grids (Agar Scientific, United Kingdom) and contrasted with 0.5% aqueous uranyl acetate (Science Services, Germany) for 20 min and with 2% Reynold's lead citrate for 6 min. Sections were imaged at 20–30 kV with a Quanta FEG 250 scanning electron microscope (FEI Company, USA) equipped with a STEM detector using the xT microscope control software ver. 6.2.6.3123.

Fluorescence* in situ *hybridization

The distribution pattern of *Ca*. E. bathymodioli in *B. puteoserpentis* was determined using Whole-mount fluorescence *in situ* hybridization (Whole-mount FISH). To that end, a formalin-fixed *B. puteoserpentis* gill sample (ROV 499/1-4) was chosen. Three gill filaments were dissected and hybridized with 16S rRNA-targeting probes. The 16

rRNA-targeting probes were in solution within the hybridization buffer (conforming together the hybridization mixture) containing 35% formamide, 80 mM NaCl, 400 mM Tris HCl, 0.4% blocking reagent for nucleic acids (Roche, Switzerland)), 0.08% SDS (v/v), 0.08 dextran sulfate (w/v), 5 ng·µl^{-1} of the Eubacterial 16S rRNA probe EUB 388 labeled with the dye Alexa Fluor 647 (Amann et al. 1990) and 5 ng·µl^{-1} of the *Ca*. Endonucleobacter 16S rRNA probe BNIX64 labeled with the dye Atto 550. The gill filaments were hybridized in 50 µL of hybridization mixture at 46°C for 3 h. Following hybridization, the gill filaments were washed in pre-warmed 48°C washing buffer (0.07 M NaCl, 0.02 M Tris-HCl pH 7.8, 5 mM EDTA pH 8, and 0.01% SDS (v/v)) for 15 min. After washing, the gill filaments were counterstained with DAPI for 10 min at room temperature, transferred to polylysine glass slides and mounted using the ProLong® Gold antifade mounting media (Thermo Fisher Scientific, MA, USA), cured overnight at room temperature and stored -20°C until visualization. All steps during and after hybridization were done in dark conditions.

Microscopy

Whole-filament overviews were visualized with the epifluorescence microscope Olympus BX53 (Olympus, Germany). The adhesion-patch overviews and the ciliated edge snapshots were visualized with the dual system Zeiss LSM 780 & Airyscan detector (Carl Zeiss Microscopy GmbH, Germany). The samples were continuously illuminated using different excitation sources depending on the fluorophore used. Epifluorescence images were taken with the objective Olympus UCPlanFL 20X/0.70 air transmission using an Orca Flash 4.0 camera (Hamamatsu, Japan). LSM and Airyscan images were taken with two different objectives: A plan-APROCHROMAT 63X/1.4 oil immersion objective and a plan-APROCHROMAT 100X/1.46 DIC M27

Elyra oil immersion objective using an Andor iXon Ultra 897 High Speed EMCCD Camera (Andor, UK). Beam selection and modulation of the laser intensities were controlled in several ways, depending on the laser wavelength and the sample of study. Epifluorescence images were obtained and post-processed using the Olympus cellSens Dimension software v. 1.18 (Olympus, Germany). LSM and Airyscan images were obtained and post-processed using ZEN software (black edition, 64bits, version: 14.0.1.201) (Carl Zeiss Microscopy GmbH, Germany). Prior to image exporting, histograms were slightly modified to increase the contrast between channels. Exported images were brightness and levels-corrected using the software Adobe Photoshop (version: 12.0) (Adobe Systems, CA, USA).

References for chapter III

Aziz, Ramy K., Daniela Bartels, Aaron Best, Matthew DeJongh, Terrence Disz, Robert A. Edwards, Kevin Formsma, Svetlana Gerdes, Elizabeth M. Glass, Michael Kubal, Folker Meyer, Gary J. Olsen, Robert Olson, Andrei L. Osterman, Ross A. Overbeek, Leslie K. McNeil, Daniel Paarmann, Tobias Paczian, Bruce Parrello, Gordon D. Pusch, Claudia Reich, Rick Stevens, Olga Vassieva, Veronika Vonstein, Andreas Wilke, and Olga Zagnitko. 2008. "The RAST Server: Rapid Annotations Using Subsystems Technology." *BMC Genomics* 9:1–15.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, Sergey I. Nikolenko, Son Pham, Andrey D. Prjibelski, Alexey V. Pyshkin, Alexander V. Sirotkin, Nikolay Vyahhi, Glenn Tesler, Max A. Alekseyev, and Pavel A. Pevzner. 2012. "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." *Journal of Computational Biology* 19(5):455–77.

Boetius, Antje, and Frank Wenzhöfer. 2013. "Seafloor Oxygen Consumption Fuelled by Methane from Cold Seeps." *Nature Geoscience* 6(9):725–34.

Casadevall, Arturo. 2008. "Evolution of Intracellular Pathogens." *Annual Review of Microbiology* 62(1):19–33.

Chakraborty, Smarajit, Hideaki Mizusaki, and Linda J. Kenney. 2015. "A FRET-Based DNA Biosensor Tracks OmpR-Dependent Acidification of *Salmonella* during Macrophage Infection." *PLoS Biology* 13(4):1–32

Crawford, E. D., J. E. Seaman, A. E. Barber, D. C. David, P. C. Babbitt, A. L. Burlingame, and J. A. Wells. 2012. "Conservation of Caspase Substrates across Metazoans Suggests Hierarchical Importance of Signaling Pathways over Specific Targets and Cleavage Site Motifs in Apoptosis." *Cell Death and Differentiation* 19(12):2040–48.

Deveraux, Quinn L., and John C. Reed. 1999. "IAP Family Proteins - Suppressors of Apoptosis." *Genes and Development* 13(3):239–52.

Deveraux, Quinn L., Ryosuke Takahashi, Guy S. Salvesen, and John C. Reed. 1997. "X-Linked IAP Is a Direct Inhibitor of Cell-Death Proteases." *Nature* 388(6639):300–304.

Emms, David M., and Steven Kelly. 2019. "OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics." *Genome Biology* 20(1):1–14.

Fiore, Cara L., Micheline Labrie, Jessica K. Jarett, and Michael P. Lesser. 2015. "Transcriptional Activity of the Giant Barrel Sponge, *Xestospongia muta* Holobiont: Molecular Evidence for Metabolic Interchange." *Frontiers in Microbiology* 6(APR).

Forget, Nathalie L., and S. Kim Juniper. 2013. "Free-Living Bacterial Communities Associated with Tubeworm (*Ridgeia piscesae*) Aggregations in Contrasting Diffuse Flow Hydrothermal Vent Habitats at the Main Endeavour Field, Juan de Fuca Ridge." *MicrobiologyOpen* 2(2):259–75.

Fraser-Liggett, Claire M. 2005. “Insights on Biology and Evolution from Microbial Genome Sequencing.” *Genome Research* 15(12):1603–10.

Gil, Rosario, Amparo Latorre, and Andrés Moya. 2004. “Bacterial Endosymbionts of Insects: Insights from Comparative Genomics.” *Environmental Microbiology* 6(11):1109–22.

González-Porras, Miguel-Ángel, Adrien Assié, Målin Tietjen, Marlene Jensen, Manuel Kleiner, Harald Gruber-Vodicka, Nicole Dubilier and Nikolaus Leisch. “The hungry nucleus: The nutritional demands of a chitinolytic intranuclear parasite trigger its host cell to upregulate sugar import.” *In Preparation.*

Goris, Johan, Konstantinos T. Konstantinidis, Joel A. Klappenbach, Tom Coenye, Peter Vandamme, and James M. Tiedje. 2007. “DNA-DNA Hybridization Values and Their Relationship to Whole-Genome Sequence Similarities.” *International Journal of Systematic and Evolutionary Microbiology* 57(1):81–91.

van der Heijden, Karina, Jillian M. Petersen, Nicole Dubilier, and Christian Borowski. 2012. “Genetic Connectivity between North and South Mid-Atlantic Ridge Chemosynthetic Bivalves and Their Symbionts.” *PLoS ONE* 7(7).

Huerta-Cepas, Jaime, François Serra, and Peer Bork. 2016. “ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data.” *Molecular Biology and Evolution* 33(6):1635–38.

Huerta-Cepas, Jaime, Damian Szklarczyk, Davide Heller, Ana Hernández-Plaza, Sofia K. Forslund, Helen Cook, Daniel R. Mende, Ivica Letunic, Thomas Rattei, Lars J. Jensen, Christian Von Mering, and Peer Bork. 2019. “EggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses.” *Nucleic Acids Research* 47(D1):D309–14.

Jensen, Sigmund, Sébastien Duperron, Nils Kåre Birkeland, and Martin Hovland. 2010. “Intracellular *Oceanospirillales* Bacteria Inhabit Gills of *Acesta* Bivalves.” *FEMS Microbiology Ecology* 74(3):523–33.

Karp, Peter D., Suzanne M. Paley, Markus Krummenacker, Mario Latendresse, Joseph M. Dale, Thomas J. Lee, Pallavi Kaipa, Fred Gilham, Aaron Spaulding, Liviu Popescu, Tomer Altman, Ian Paulsen, Ingrid M. Keseler, and Ron Caspi. 2009. “Pathway Tools Version 13.0: Integrated Software for Pathway/Genome Informatics and Systems Biology.” *Briefings in Bioinformatics* 11(1):40–79.

Katharios, Pantelis, Helena M. B. Seth-Smith, Alexander Fehr, José M. Mateos, Weihong Qi, Denis Richter, Lisbeth Nufer, Maja Ruetten, Maricruz Guevara Soto, Urs Ziegler, Nicholas R. Thomson, Ralph Schlapbach, and Lloyd Vaughan. 2015. “Environmental Marine Pathogen Isolation Using Mesocosm Culture of Sharpsnout Seabream: Striking Genomic and Morphological Features of Novel *Endozoicomonas* Sp.” *Scientific Reports* 5(August):17609.

Koren, Sergey, Brian P. Walenz, Konstantin Berlin, Jason R. Miller, Nicholas H. Bergman, and Adam M. Phillippy. 2017. “Canu: Scalable and Accurate Long-Read

Assembly via Adaptive κ-Mer Weighting and Repeat Separation." *Genome Research* 27(5):722–36.

Kräter, Martin, Jiranuwat Sapudom, Nicole Bilz, Tilo Pompe, Jochen Guck, and Claudia Claus. 2018. "Alterations in Cell Mechanics by Actin Cytoskeletal Changes Correlate with Strain-Specific Rubella Virus Phenotypes for Cell Migration and Induction of Apoptosis." *Cells* 7(9):136.

McDonald, Kent L. 2014. "Rapid Embedding Methods into Epoxy and LR White Resins for Morphological and Immunological Analysis of Cryofixed Biological Specimens." *Microscopy and Microanalysis* 20(1):152–63.

Montanaro, Jacqueline, Daniela Gruber, and Nikolaus Leisch. 2016. "Improved Ultrastructure of Marine Invertebrates Using Non-Toxic Buffers." *PeerJ* 2016(3):1–15.

Moran, Nancy A. 2002. "Microbial Minimalism: Genome Reduction in Bacterial Pathogens." *Cell* 108(5):583–86.

Morrow, Kathleen M., Anthony G. Moss, Nanette E. Chadwick, and Mark R. Liles. 2012. "Bacterial Associates of Two Caribbean Coral Species Reveal Species-Specific Distribution and Geographic Variability." *Applied and Environmental Microbiology* 78(18):6438–49.

Muyzer, G., A. Teske, C. O. Wirsen, and H. W. Jannasch. 1995. "Phylogenetic Relationships of *Thiomicrospira* Species and Their Identification in Deep-Sea Hydrothermal Vent Samples by Denaturing Gradient Gel-Electrophoresis of 16S RDNA Fragments." *Archives of Microbiology* 164(3):165–72.

Neave, Matthew J., Amy Apprill, Christine Ferrier-Pages, and Christian R. Voolstra. 2016. "Diversity and Function of Prevalent Symbiotic Marine Bacteria in the Genus *Endozoicomonas*." *Applied Microbiology and Biotechnology* 1–10.

Neave, Matthew J., Craig T. Michell, Amy Apprill, and Christian R. Voolstra. 2017. "Endozoicomonas Genomes Reveal Functional Adaptation and Plasticity in Bacterial Strains Symbiotically Associated with Diverse Marine Hosts." *Scientific Reports* 7(January):1–12.

Niebuhr, Kirsten, Sylvie Giuriato, Thierry Pedron, Dana J. Philpott, Frédérique Gaits, Julia Sable, Michael P. Sheetz, Claude Parsot, Philippe J. Sansonetti, and Bernard Payrastre. 2002. "Conversion of PtdIns(4, 5)P2 into PtdIns(5)P by the *S. flexneri* Effector IpgD Reorganizes Host Cell Morphology." *EMBO Journal* 21(19):5069–78.

Niebuhr, Kirsten, Noureddine Jouihri, Abdelmounaaim Allaoui, Pierre Gounon, Philippe J. Sansonetti, and Claude Parsot. 2000. "IpgD, a Protein Secreted by the Type III Secretion Machinery of *Shigella flexneri*, Is Chaperoned by IpgE and Implicated in Entry Focus Formation." *Molecular Microbiology* 38(1):8–19.

Olson, Robert J., Sallie W. Chisholm, Richard Devereux, David A. Stahl, R. I. Amann, and B. J. Binder. 1990. "Combination of 16S rRNA-Targeted Oligonucleotide Probes with Flow Cytometry for Analyzing Mixed Microbial Populations." *Applied and Environmental Microbiology* 56(6):1919–25.

van der Oost, John, Edze R. Westra, Ryan N. Jackson, and Blake Wiedenheft. 2014. “Unravelling the Structural and Mechanistic Basis of CRISPR–Cas Systems.” *Nature Reviews Microbiology* 12(7):479–92.

Parks, Donovan H., Michael Imelfort, Connor T. Skennerton, Philip Hugenholtz, and Gene W. Tyson. 2015. “CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes.” *Genome Research* 25(7):1043–55.

Pike, Rebecca E., Brad Haltli, and Russell G. Kerr. 2013. “Description of *Endozoicomonas euniceicola* Sp. Nov. and *Endozoicomonas gorgoniicola* Sp. Nov., Bacteria Isolated from the Octocorals *Eunicea fusca* and *Plexaura* Sp., and an Emended Description of the Genus Endozoicomonas.” *International Journal of Systematic and Evolutionary Microbiology* 63(PART 11):4294–4302.

Price, Morgan N., Paramvir S. Dehal, and Adam P. Arkin. 2010. “FastTree 2 - Approximately Maximum-Likelihood Trees for Large Alignments.” *PLoS ONE* 5(3).

Qi, Weihong, Maria Chiara Cascarano, Ralph Schlapbach, Pantelis Katharios, Lloyd Vaughan, and Helena M. B. Seth-Smith. 2018. “*Ca*. Endozoicomonas cretensis: A Novel Fish Pathogen Characterized by Genome Plasticity.” *Genome Biology and Evolution* 10(6):1363–74.

Schreiber, Lars, Kasper Urup Kjeldsen, Matthias Obst, Peter Funch, and Andreas Schramm. 2016. “Description of *Endozoicomonas ascidiicola* Sp. Nov., Isolated from Scandinavian Ascidians.” *Systematic and Applied Microbiology* 39(5):313–18.

Schulz, Frederik, and Matthias Horn. 2015. “Intranuclear Bacteria: Inside the Cellular Control Center of Eukaryotes.” *Trends in Cell Biology* 25(6):339–46.

Seah, Brandon K. B., and Harald R. Gruber-Vodicka. 2015. “Gbtools: Interactive Visualization of Metagenome Bins in R.” *Frontiers in Microbiology* 6(DEC):1–10.

Sedlazeck, Fritz J., Philipp Rescheneder, Moritz Smolka, Han Fang, Maria Nattestad, Arndt Von Haeseler, and Michael C. Schatz. 2018. “Accurate Detection of Complex Structural Variations Using Single-Molecule Sequencing.” *Nature Methods* 15(6):461–68.

Sievers, Fabian, Andreas Wilm, David Dineen, Toby J. Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, Hamish McWilliam, Michael Remmert, Johannes Söding,

Tietjen, Målin. “Physiology and ecology of deep-sea bathymodiolin symbioses” *Personal communication*.

Thompson, Julie D., and Higgins, Desmond G. 2011. “Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega.” *Molecular Systems Biology* 7(539).

Vasanthakumar, Thamiya, and John L. Rubinstein. 2020. “Structure and Roles of V-Type ATPases.” *Trends in Biochemical Sciences* 45(4):295–307.

Wick, Ryan R., Louise M. Judd, Claire L. Gorrie, and Kathryn E. Holt. 2017. “Unicycler:

Resolving Bacterial Genome Assemblies from Short and Long Sequencing Reads." *PLoS Computational Biology* 13(6):1–22.

Wick, Ryan R., Mark B. Schultz, Justin Zobel, and Kathryn E. Holt. 2015. "Bandage: Interactive Visualization of de Novo Genome Assemblies." *Bioinformatics* 31(20):3350–52.

Zhang, Jianlin, Amanda Felder, Yujie Liu, Ling T. Guo, Stephan Lange, Nancy D. Dalton, Yusu Gu, Kirk L. Peterson, Andrew P. Mizisin, G. Diane Shelton, Richard L. Lieber, and Ju Chen. 2009. "Nesprin 1 Is Critical for Nuclear Positioning and Anchorage." *Human Molecular Genetics* 19(2):329–41.

Zielinski, Frank U., Annelie Pernthaler, Sebastien Duperron, Luciana Raggi, Olav Giere, Christian Borowski, and Nicole Dubilier. 2009. "Widespread Occurrence of an Intranuclear Bacterial Parasite in Vent and Seep Bathymodiolin Mussels." *Environmental Microbiology* 11(5):1150–67.

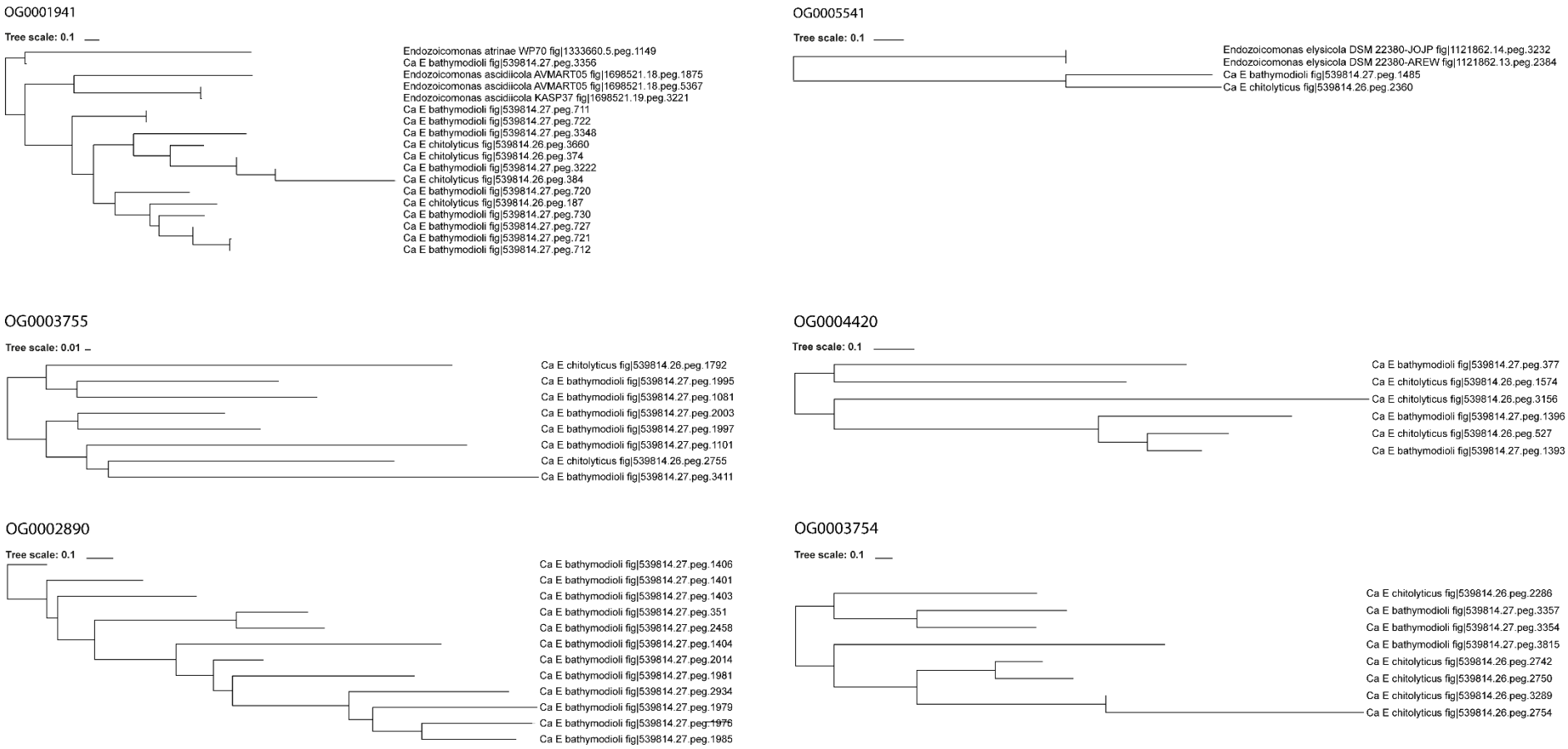
Supplementary Information for chapter III

Supplementary table 1. Assembly quality and RAST v2.0 (Rapid Annotation using Subsystem Technology) annotation results for the 21 genomes included in the study (Aziz et al., 2008).

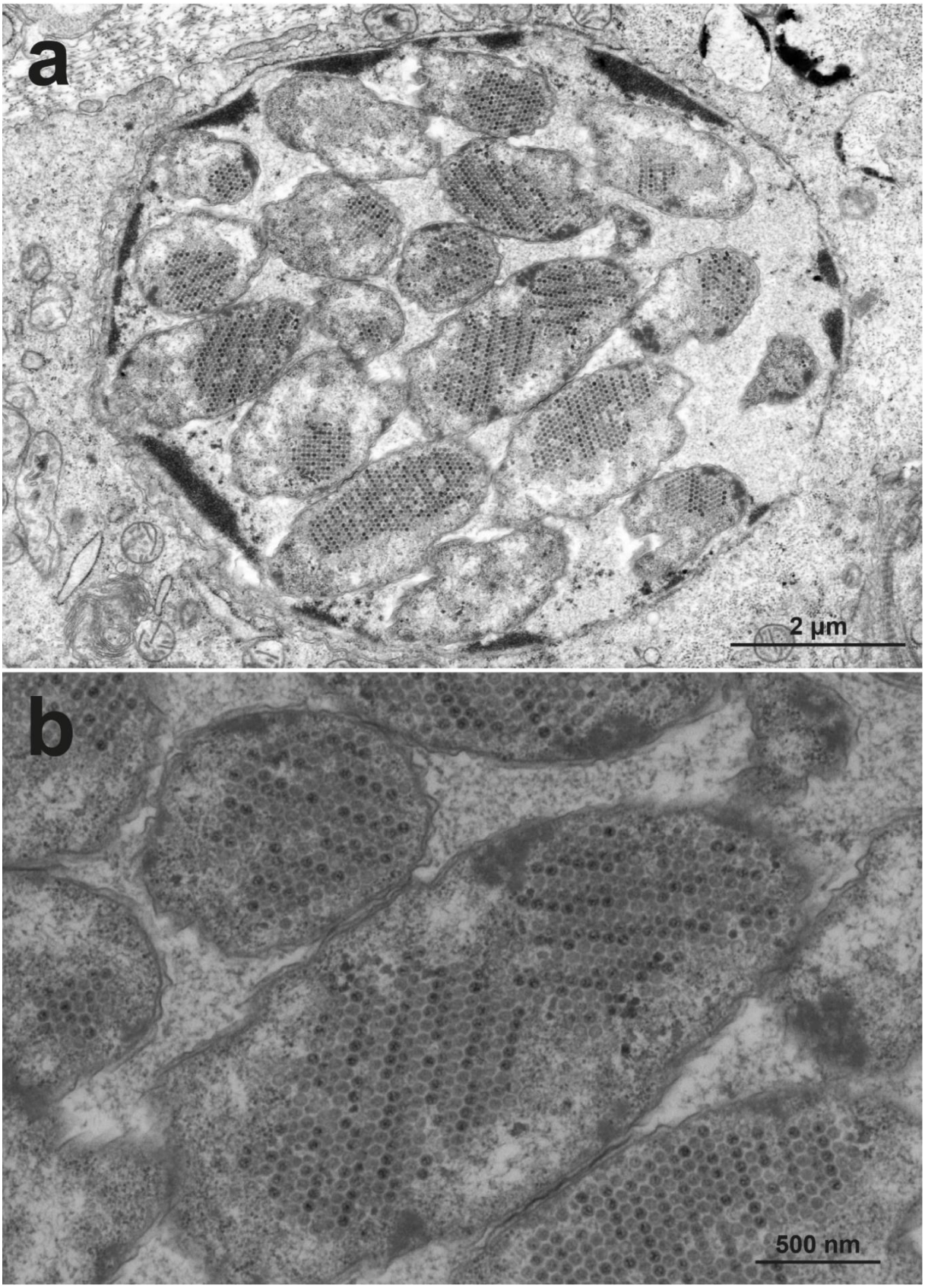
Genome	NCBI ID	RAST ID	Assembly size (bp)	Completeness (%)	Contigs	Scaffolds	Scaffold N50 (bp)	Max scaffold size (bp)	CDS	RNAs	GC%
Ca. E. bathymodioli		539814,27	3.860.465	98,28	72	72	159.565	396.118	3.995	39	40,0
Ca. E. chitolyticus		539814,26	3.326.527	96,55	94	94	88.503	272.057	3.654	54	40,8
Endozoicomonas_elysicola_DSM22380_JOJP	JOJP00000000.1	1121862,1	5.606.375	98,28	21	2	5.569.560	5.569.560	5.256	103	46,8
Endozoicomonas_elysicola_DSM22380_AREW	AREW00000000.1	1121862,1	5.545.378	98,28	33	33	406.323	895.183	5.241	81	46,7
Endozoicomonas_montiporae_LMG24815	JOKG00000000.1	1027273,1	5.602.297	98,28	83	20	1.015.541	1.412.099	5.544	120	48,5
Endozoicomonas_montiporae_CL-33	NZ_CP013251	570277,12	5.430.256	98,28	3	1	5.430.256	5.430.256	5.340	130	48,5
Endozoicomonas_numazuensis_DSM25634	JOKH00000000.1	1137799,2	6.342.227	100,00	131	31	917.146	1.695.894	6.250	97	47,0
Endozoicomonas_arenosclerae_ab112	LASA00000000.1	1633495,6	6.453.554	100,00	328	328	44.889	259.222	6.401	135	47,7
Endozoicomonas_arenosclerae_E-MC227	LASB00000000.1	1633495,7	6.216.773	88,56	2.501	2.501	4.718	34.129	8.613	109	47,2
Endozoicomonas_atrinae_WP70	LUKQ00000000.2	1333660,5	6.687.418	98,28	985	980	21.158	121.488	7.433	89	47,9
Endozoicomonas_ascidiicola_AVMART05	LUTV00000000.1	1698521,2	6.130.497	98,28	287	36	771.815	1.017.847	5.913	100	46,7
Endozoicomonas_ascidiicola_KASP37	LUTW00000000.1	1698521,2	6.512.467	98,28	300	34	675.874	1.128.000	6.264	91	46,6
Endozoicomonas_acroporae_Acr-1	SAUT00000000.1	1701104,1	6.024.033	98,28	299	299	56.565	194.440	5.880	86	49,3
Endozoicomonas_acroporae_Acr-5	SAUU00000000.1	1701104,1	6.034.674	98,28	295	295	58.040	169.237	5.943	91	49,3
Endozoicomonas_acroporae_Acr-14	PJPV00000000.1	1701104,9	6.048.850	98,28	309	309	47.658	161.511	5.836	88	49,2
Endozoicomonas_sp_Bugula_neritia	MDLD00000000.1	1868284,6	4.049.356	92,32	284	272	20.804	84.485	4.118	46	45,3
Uncultured_Endozoicomonas_sp	OODV00000000.1	432652,2	5.876.352	98,28	638	638	19.256	91.550	6.680	87	46,8
Endozoicomonas_sp_OPT23	PPFD00000000.1	2072845,4	4.938.102	100,00	41	30	737.835	1.028.303	4.446	101	46,8
Hahella_ganghwensis-DSM17046	AQXX00000000.1	6666666,6	6.564.965	100,00	144	144	94.036	288.733	6.512	64	49,2
Hahella_sp-CCB-MM4	MRYI00000000.1	1926491,2	6.663.740	100,00	161	161	232.349	805.316	6.278	65	49,8
Hahella_chejuensis_KCTC2396	NC_007645.1	349521,49	7.215.267	98,28	1	1	7.215.267	7.215.267	6.779	82	53,9

Supplementary table 2. Incomplete metabolic routes for de novo synthesis of proteinogenic amino acids. The absence of these metabolic routes has been determined using the comparative analysis function of the pathway tools server mode v24.0 (Karp et al., 2009) in combination with the BioCyc database (Caspi et al., 2016). **Val**, Valine. **Leu**, Leucine. **Ile**, Isoleucine. **Thr**, Threonine. **Asn**, Asparagine. **Cys**, Cysteine. **Met**, Methionine. **Tyr**, Tyrosine. **Arg**, Arginine. **His**, Histidine. **Phe**, Phenylalanine.

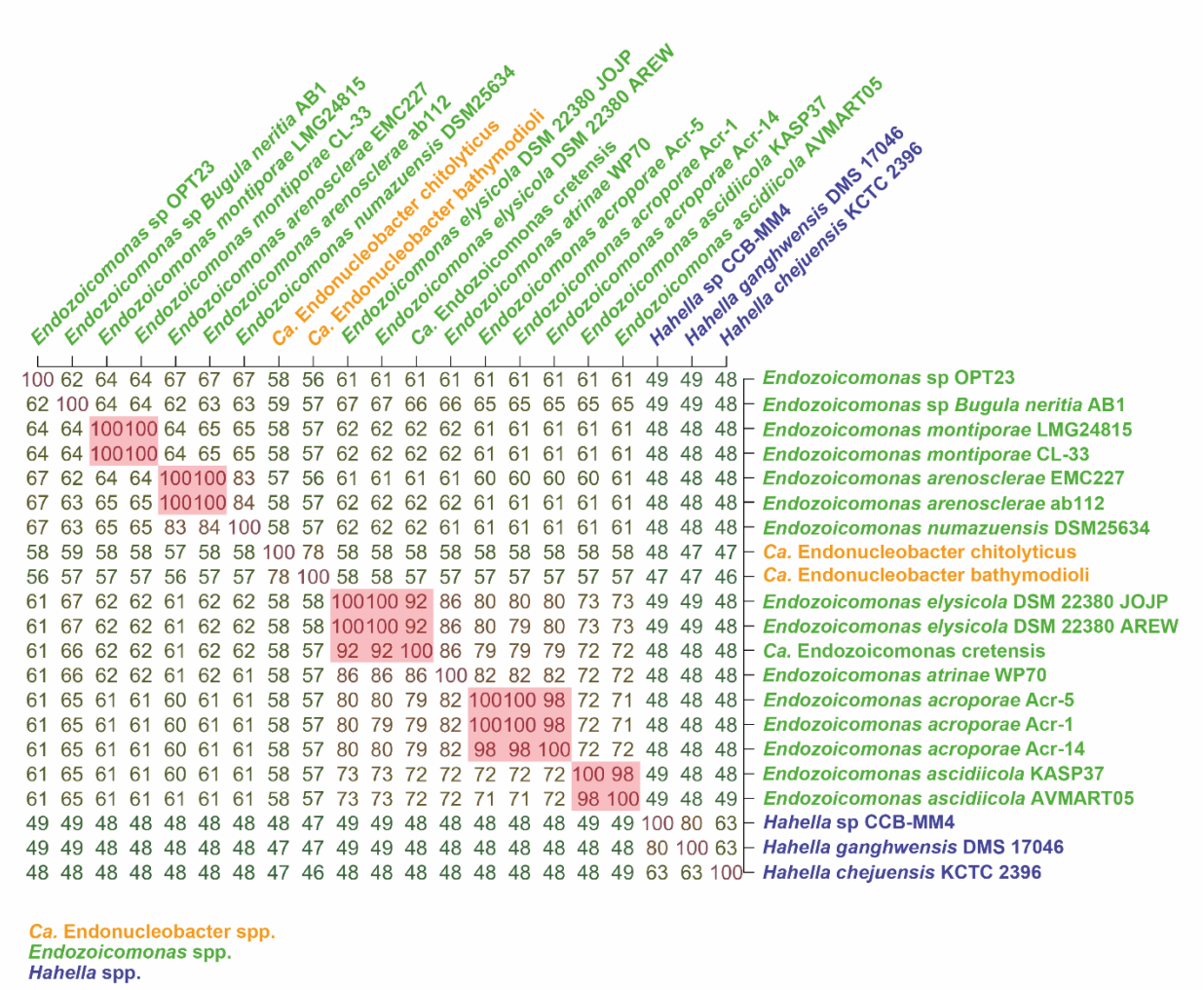
Genome	Incomplete metabolic routes for *de novo* biosynthesis
Ca. E. bathymodioli	Val, Leu, Ile, Thr, Asn, Cys, Met, Tyr, Arg, His
Ca. E. chitolyticus	Val, Leu, Ile, Thr, Asn, Cys, Met, Phe, Tyr, Arg, His
Endozoicomonas_elysicola_DSM22380	Asn
Endozoicomonas_elysicola_DSM22380	Asn
Endozoicomonas_montiporae_LMG24815	Asn
Endozoicomonas_montiporae_CL-33	Asn
Endozoicomonas_numazuensis_DSM25634	Asn, Tyr
Endozoicomonas_arenosclerae_ab112	Asn
Endozoicomonas_arenosclerae_E-MC227	Asn, Tyr
Endozoicomonas_atrinae_WP70	Asn
Endozoicomonas_ascidiicola_AVMART05	Asn
Endozoicomonas_ascidiicola_KASP37	Asn
Endozoicomonas_acroporae_Acr-1	Asn
Endozoicomonas_acroporae_Acr-5	Asn
Endozoicomonas_acroporae_Acr-14	Asn
Endozoicomonas_sp_Bugula_neritia	Met
Uncultured_Endozoicomonas_sp	
Endozoicomonas_sp_OPT23	
Hahella_ganghwensis-DSM17046	
Hahella_sp-CCB-MM4	
Hahella_chejuensis_KCTC2396	Phe



Supplementary figure 1. Inhibitors of apoptosis (IAP) are almost exclusively present in *Ca.* Endonucleobacter. Phylogenetic analysis of 6 of the 8 orthogroups that contain all the IAP from the analyzed dataset (21 genomes). The other two orthogroups (OG0006670, OG0006685) only contained two sequences each, so they could not be phylogenetically analyzed. Trees calculated as part of OrthoFinder pipeline v2.4.0 (Emms & Kelly, 2019).



Supplementary figure 2. Bacteriophages infect *Ca.* E. chitolyticus, which encodes for CRISP-cas related proteins. a, "B." childressi nucleus infected by *Ca.* E. chitolyticus, which is infected by viral particles. b, detail of the bee-hive structure of capsid viral particles within *Ca.* E. chitolyticus.



Supplementary figure 3. *Ca.* E. bathymodioli and *Ca.* E. chitolyticus shared 78% amino acid identity, confirming that they are different species. Average amino acid identity among the 21 genomes included in this study. Legend: *Ca.* Endonucleobacter spp. (orange), *Endozoicomonas* spp. (green) and *Hahella* spp. (blue).

Chapter IV | Spatial segregation of SOX strains

Super-resolution localization of strain-specific markers in sulfur-oxidizing symbionts of *Bathymodiolus azoricus* suggests strain spatial segregation at a single bacteriocyte level

Miguel-Ángel González-Porras[1], Jimena Barrero-Canosa[2], Lizbeth Sayavedra[3], Rebecca Ansorge[3], Nicole Dubilier[1], Nikolaus Leisch[1].

1 Department of Symbiosis, Max Planck Institute for Marine Microbiology, Bremen, Germany

2 Department of Environmental Technology, Technische Universität Berlin, Berlin, Germany

3 Quadram Institute Bioscience, Norwich Research Park, Norwich, UK

The manuscript is in preparation and it has not been revised by all authors.

Author contributions

MAGP conceived the study, performed the microscopy experiments, analyzed the data, prepared the figures/tables and wrote the manuscript. JBC helped to conceive the study and contributed to the conceptual design of the Direct-geneFISH experiments. LS helped to conceive the study. RA helped to conceive the study. ND helped to conceive the study. NL helped to conceive the study and performed the EdU incubations onboard.

Abstract

Bathymodiolin mussels acquire their chemosynthetic symbionts horizontally along their lifetime due to the continuous growth of their gill. These symbionts are harbored in specialized cells of the gills called bacteriocytes. The newly formed bacteriocytes are non-symbiotic, and they must be continuously colonized *de novo*. Metagenomic studies revealed that up to 16 different strains of the sulfur-oxidizing symbiont (SOX) can co-occur in a single mussel. However, if these strains are differentially distributed within their host remains unclear. To answer the question of whether different SOX strains segregate spatially within the gills of *Bathymodiolus azoricus*, we adapted Direct-geneFISH, a technique that co-localizes the 16S rRNA and a gene of interest, to host-microbiome systems. We targeted and localized SOX strain-specific genetic markers methanol dehydrogenase- and hydrogenase-gene clusters with fluorescently labeled probes in *B. azoricus* gill sections. Our results suggested that bacteriocytes are quantitatively dominated by a single SOX strain. Moreover, SOX from the same strain tend to cluster together in neighboring bacteriocytes forming patches. We hypothesize that the SOX pioneers that firstly occupy a newly formed bacteriocyte prevent posterior SOX colonizers from settling in. This spatial exclusion between SOX strains could be explained by ultrastructural modifications of the bacteriocyte cell surface, adaptive selection of specific strains or chemical incompatibility between strains. We also hypothesize that patches are originated in a chain of colonization events between already-colonized bacteriocytes and newly formed, non-symbiotic and adjacent bacteriocytes (“intra-filament” colonization). Our results showed spatial segregation of different SOX strains at a single bacteriocyte scale: A phenomenon of microniche partitioning that might have biological explanations. This study sheds light over the structuring of strain subpopulations in host-microbiome systems at a single

cell scale, revealing a previously hidden level of spatial organization in symbiotic communities within their hosts.

Keywords: Bathymodiolus; sulfur-oxidizing symbiont; bacteriocyte; strain diversity; spatial segregation; patches.

Introduction

Bathymodiolin mussels dominate in terms of biomass the fauna of hydrothermal vents and cold seeps distributed worldwide (Dubilier et al. 2008). These marine invertebrates harbor in their gills chemosynthetic symbionts that allow them to thrive in such ecosystems, where organic carbon is scarce (Fisher, 1990; Cavanaugh et al., 2006). *Bathymodiolus azoricus* is a bathymodiolin mussel that inhabits the hydrothermal vents of the Mid-Atlantic Ridge. *B. azoricus* harbors two types of gammaproteobacterial chemosynthetic symbionts in specialized cells named bacteriocytes (DeChaine and Cavanaugh 2006). One of the two intracellular symbionts harbored by *B. azoricus* is a sulfur-oxidizing symbiont (SOX). *B. azoricus* provides the SOX with sulfur-reduced species such as sulfide, as well as with oxygen and carbon dioxide that the SOX need for their metabolism. In return, the SOX provide nutrition to *B. azoricus* (Duperron et al. 2006). Like the other bathymodiolin mussels, *B. azoricus* acquire their symbionts horizontally along their lifetime along the continuous grow of their gill (DeChaine et al. 2006; Distel et al. 1988; Laming et al. 2014; LePennec, 1988; Wentrup et al. 2014; Won et al. 2003; Won et al. 2008). Up to 16 different strains of the SOX with different metabolic capabilities can co-occur in a single *Bathymodiolus* mussel (Ansorge et al. 2019). This high strain diversity may be explained by multiple symbiont colonization events of newly formed gill tissue. Despite the high strain diversity, not much can be

drawn with regard to the spatial organization of these strains within *B. azoricus* gill filaments based solely on metagenomics analyses. Because of these limitations, microscopy methodologies such as *in situ* hybridization techniques are needed to resolve the spatial organization of different symbiont strains within their host. For instance, Ikuta et al. 2015 demonstrated that heterogeneous symbiont subpopulations that either lacked or encoded for two strain-specific genetic markers clustered separately within the gill filaments of the bathymodiolin mussel *B. septemdierum* using whole-mount *in situ* hybridization. However, it is unknown whether this spatial segregation is originated by host- and/or symbiont-mediated biological mechanisms.

Several phenomena might play a role in spatial segregation between SOX strains in bathymodiolin mussels' gills. The secretion of effectors that induce an ultrastructural transformation of epithelial cells is a common strategy in enteropathogenic intracellular bacteria, such as the enteropathogenic *Escherichia coli* or *Vibrio parahaemolyticus*. These effectors can induce a cytoskeleton rearrangement that culminates in the transformation and/or disappearance of the host cell microvilli, phenomenon known as microvilli effacement (Kaper et al. 2004; Zhou et al. 2014). Microvilli can play a major role during symbiont colonization, and their effacement might impede colonizers to settle ("shut the door mechanism") (Heath-heckman et al. 2013). In bathymodiolin mussels, microvilli effacement takes place when a bacteriocyte is colonized by a pioneer population of symbionts (Franke et al., *in Prep*). Another mechanism that might play a role in rejection between strains is chemical incompatibility. Sayavedra et al. 2015 demonstrated that the SOX symbionts from *B. azoricus* codify for YD repeat proteins, which could participate in competition between closely related bacterial strains (Koskiniemi et al. 2013). SOX strains might also segregate spatially within bathymodiolin mussels' gills when newly formed bacteriocytes get colonized

preferentially by certain strains. Wentrup et al. 2014 demonstrated that newly formed filaments firstly acquire their symbionts in the side that is facing an already colonized and ontogenetically older gill filament. This suggests that the SOX population structure of an ontogenetically older gill filament might determine the SOX population structure of an adjacent and younger gill filament at the strain level.

As introduced above, Ikuta et al. 2015 demonstrated that different strain subpopulations can segregate spatially from each other within the gill filaments of *B. septemdierum*. However, their microscopy approach lacked the resolution to determine whether the spatial segregation of symbiont strain subpopulations also occurs at a single bacteriocyte level. Since the advent of fluorescence in situ hybridization (FISH) in the late 1980s (Giovannoni et al. 1988), several variants of the original protocol have been developed to answer specific questions in different study systems. The development of certain FISH variants during the last decade have allowed the *in situ* localization of strain-specific genetic markers in environmental bacterial samples (Moraru et al. 2010). One of this methods is Direct-geneFISH (Barrero-Canosa et al. 2017), a microscopy technique that allows the simultaneous localization of the 16S rRNA of a free-living microorganism and a specific genetic marker using fluorescently-labeled polynucleotide probes. Direct-geneFISH allows the localization of single-copy marker genes at high efficiencies (> 90% true positives), with the drawback that it was originally developed to work on free-living bacteria. However, its high rate of detection makes it a desirable technique to address the question of how different strains from the same symbiont species organize spatially within their host at a single bacteriocyte scale.

In this study, we hybridized *B. azoricus* gill sections using Direct-geneFISH to answer two questions: 1) whether the spatial segregation of different SOX strains occurs at a single bacteriocyte scale and 2) whether different SOX strains organize into patches in *B. azoricus* gill filaments. To answer these two questions, we used polynucleotide probes to target two strain-specific metabolic markers in SOX symbionts: The methanol dehydrogenase- and the hydrogenase gene clusters (MDH and HYD, respectively). In addition and for the first time, we incubated bathymodiolin mussels with 5-Ethynyl-2´-deoxyuridine (EdU) to localize proliferating cells in gill filaments, allowing us to confirm where the gill growth zones of bathymodiolin mussels occur. The data demonstrated that bacteriocytes that contained any of the two strains localized in this study were fully packaged with a single strain, suggesting biological mechanisms of exclusion between strains at a single bacteriocyte level. Next, SOX symbionts of the same strain cluster together in neighboring bacteriocytes forming uninterrupted patches that cover extended areas of single gill filaments. Finally, the localization of proliferating cells labeled with EdU confirmed that growth zones are deprived of bacterial symbionts, suggesting that newly formed bacteriocytes must be colonized *de novo*. Therefore, one possible explanation for the formation of patches is that newly formed bacteriocytes are colonized by symbionts present in already-colonized and adjacent bacteriocytes from the same gill filament (“intra-filament” colonization). This study reveals a previously hidden level of spatial segregation between strains that occurs at a single cell scale in host-microbiome systems. The results presented here might help to understand the processes responsible for the spatial segregation of symbiont strains within their hosts and the potential repercussions that this phenomenon might have for the stability of the symbiosis.

Results

Direct-geneFISH optimization in host-microbiome systems

To localize the MDH-carrying and the HYD-carrying SOX strains (MDH^+ SOX and HYD^+ SOX, respectively) in *B. azoricus* gill filaments, we optimized the Direct-geneFISH protocol in host-microbiome systems. The complete list of changes to the original protocol can be found in the material and methods section under the paragraph "Direct-geneFISH on *Bathymodiolus azoricus* gill tissue". However, the most noteworthy changes are highlighted next.

We applied Direct-geneFISH in formalin-fixed paraffin-embedded sections of *B. azoricus* gills. To decrease the fluorescent background from the host tissue, we sectioned the samples at 3 µm thickness instead of at the conventional 10 µm. Next, the prolongation of the hybridization time from 2 h to 4 h significantly increased the signal-to-noise ratio when localizing marker genes in *B. azoricus* gill sections. Last, our imaging data suggested that substitution of Alexa 594 by Alexa 647 when targeting one gene marker at a time when using super-resolution microscopy significantly increased the signal-to-noise ratio (**Figure 2, b; Figure 3, b**). When we co-localized both genetic markers MDH and HYD (**Figure 4**), the probes targeting MDH were labeled with Alexa 594 because more probes were designed for this gene (12 probes MDH vs.10 HYD).

Localization of the MDH-gene cluster

For the first time, we localized a single SOX strain (MDH$^+$ SOX) in *B. azoricus* at a whole organ scale (**Figure 1**). To this end, we applied our adapted Direct-geneFISH protocol on whole-gill sections of *B. azoricus* using 12 probes labeled with Alexa 647 and visualize them using CLSM and super-resolution Airyscan microscopy. MDH$^+$ SOX cluster together in neighboring bacteriocytes forming patches (**Fig.1, a; MDH signal in red**). To get an impression of the abundance of the MDH$^+$ SOX patches, we quantified the amount of filaments that had at least one patch. This quantitative analysis revealed that approximately 60% of the filaments had at least 1 MDH$^+$ SOX patch (**Supplementary table 2**). Moreover the spatial proximity between filaments did not determine the presence of facing MDH+ SOX patches between adjacent filaments (**Fig.2, a**). To know whether there is a correlation between the occurrence of patches and their spatial proximity, we also quantified the number of cases in which two adjacent gill filaments had facing MDH$^+$ SOX patches in **Fig.1**. This quantitative analysis revealed that approximately 23% of the MDH$^+$ SOX patches were facing another MDH$^+$ SOX patch in the adjacent filament (**Supplementary table 3**). To assess the abundance of the MDH$^+$ SOX strain in the whole SOX community, we calculated the ratio of SOX 16S rRNA vs MDH (**Supplementary table 4**). We estimated that approximately 6% of SOX 16S rRNA and MDH signals overlapping, indicating that the MDH+ SOX strain is quantitatively low in the global SOX population (less than 10%). Despite the quantitatively low presence MDH$^+$ SOX strain in the global SOX population, it occupied extended uninterrupted areas within individual gill filaments, forming big patches (**red in Fig.2, a**). Moreover, super-resolution Airyscan microscopy revealed that MDH$^+$ SOX was the most abundant strain in the bacteriocytes where it was detected (**Fig.2, b; right side of the filament**). In contrast,

this strain was completely absent in the rest of the bacteriocytes (**Fig.2, b; left side of the filament**). Our result suggests that the MDH$^+$ SOX strain follows a polarized distribution at a single bacteriocyte level: either a bacteriocyte was fully packed with the MDH$^+$ SOX strain or completely deprived of it.

Localization of the HYD-gene cluster

To know whether the spatial segregation is an exclusive phenomenon for the MDH$^+$ SOX strain, we also localized the HYD$^+$ SOX strain in the gills of *B. azoricus*. To that end, we applied our adapted Direct-geneFISH protocol on whole-gill sections of *B. azoricus* using 10 probes labeled with Alexa 647 and visualize them using CLSM and Airyscan microscopy. Localization of the HYD$^+$ SOX strain at a multiple filament scale revealed that the HYD$^+$ SOX strain (HYD signal in blue) displayed a patchy distribution pattern in the gill filaments of *B. azoricus* (**Fig.3, a**), as observed for the MDH$^+$ SOX strain. Furthermore, HYD$^+$ SOX patches extended over large areas of more than 20 neighboring bacteriocytes (**Fig.3, a**). Super-resolution analysis of the HYD$^+$ SOX strain revealed a similar polarized distribution pattern at a single bacteriocyte scale like the MDH$^+$ SOX strain: either a bacteriocyte was fully packed with the HYD$^+$ SOX strain or completely deprived of it (**Fig.3, b**).

Co-localization of MDH- and HYD-gene clusters

To determine whether the MDH$^+$ and the HYD$^+$ SOX strains co-occur in a single bacteriocyte, or whether both strains form segregated patches in the gill filaments of *B. azoricus*, we co-localized both strains using Direct-geneFISH (**Figure 4**). To this end, we applied our adapted Direct-geneFISH protocol on whole-gill sections of *B. azoricus* using 12 probes labeled with Alexa 594 targeting the MDH-gene cluster and

10 probes labeled with Alexa 647 targeting the HYD-gene cluster, and visualize them using super-resolution Airyscan microscopy. Our results indicated that patches from both strains segregated spatially within the gill filaments of *B. azoricus*. Out of 120 bacteriocytes, 8 and 48 were quantitatively dominated by the MDH^+ and the HYD^+ SOX strains, respectively (MDH^+ bacteriocytes and HYD^+ bacteriocytes, respectively). The remaining bacteriocytes (64) were not dominated by any of the two localized strains (referred as MDH^- HYD^- bacteriocytes). Intriguingly, we could co-localize signals of the MDH^+ and the HYD^+ SOX strains in single bacteriocytes, although this was a rare phenomenon. Whenever this happened, either the bacteriocyte was dominated by one of the two strains (either MDH^+ or HYD^+ bacteriocyte), or both strains were scarce in the bacteriocyte (MDH^- HYD^- bacteriocyte).

Extension of single-strain patches

To investigate the extension of single-strain patches in the gill filaments of *B. azoricus*, we localized the MDH^+ SOX strain in whole-gill sagittal sections using 12 probes labeled with Alexa 647 and visualizing them using CLSM microscopy (**Figure 5**). **Fig.5, a** confirmed that MDH^+ SOX patches were also observable in whole-gill sagittal sections. Amplification of the white-framed area in **Fig.5, a** showed that the MDH^+ SOX strain can occupy large uninterrupted areas of a single gill filament (up to 1 mm in length).

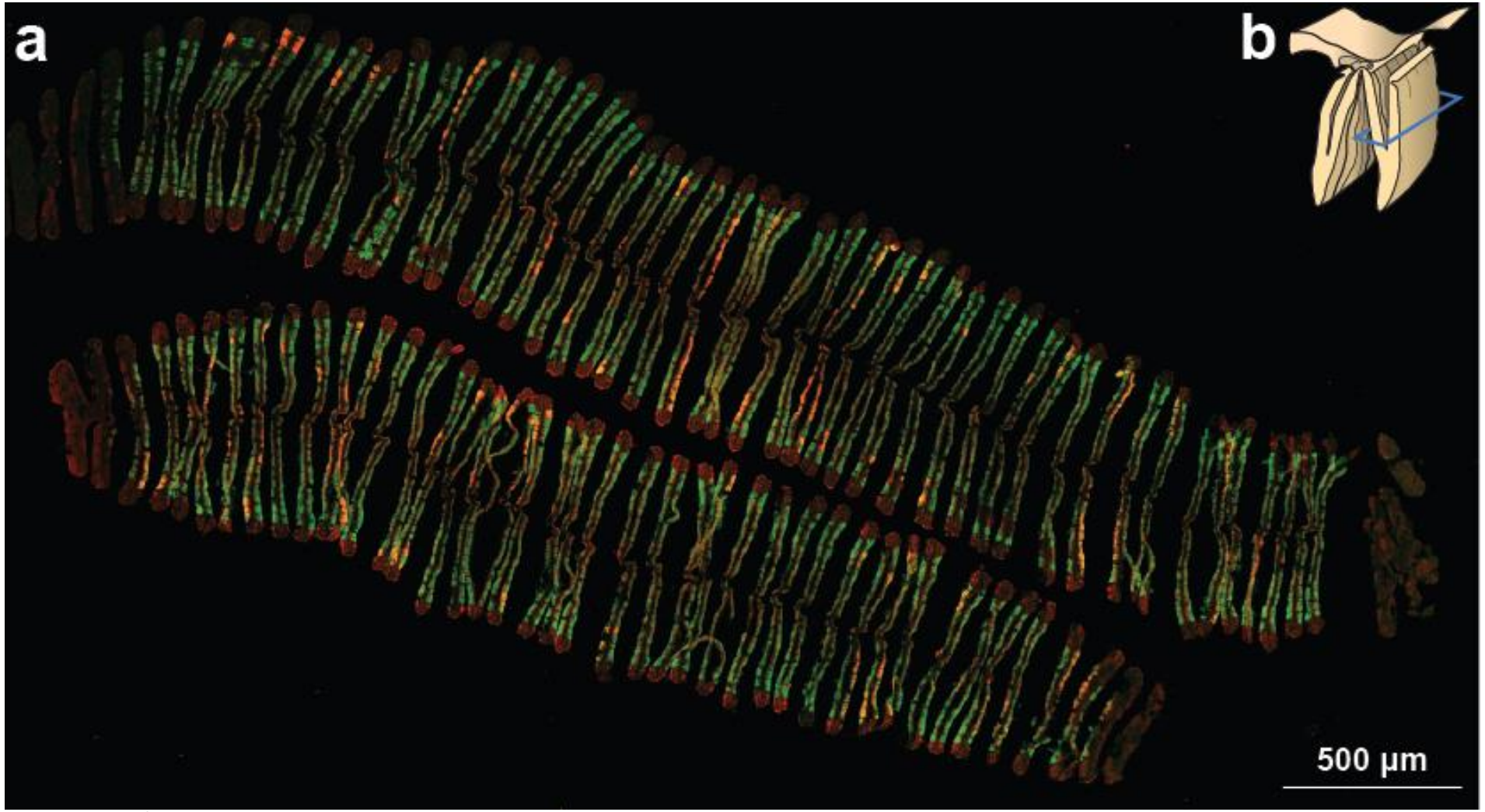


Figure 1. The MDH$^+$ SOX strain clusters together forming patches that create a bacterial metabolic mosaic in the gill filaments of *B. azoricus*. Direct-geneFISH on whole-gill cross sections. The cross sectioning plane (**b**) is represented by the blue frame over the gill diagram (adapted from Pennec 1988). MDH-gene cluster (red: Alexa 647), SOX 16S rRNA (green: atto488). ***a***, CLSM imaging of the MDH+ SOX strain at a whole-gill scale. ***b***, Representation of the cross sectioning plane by the blue frame.

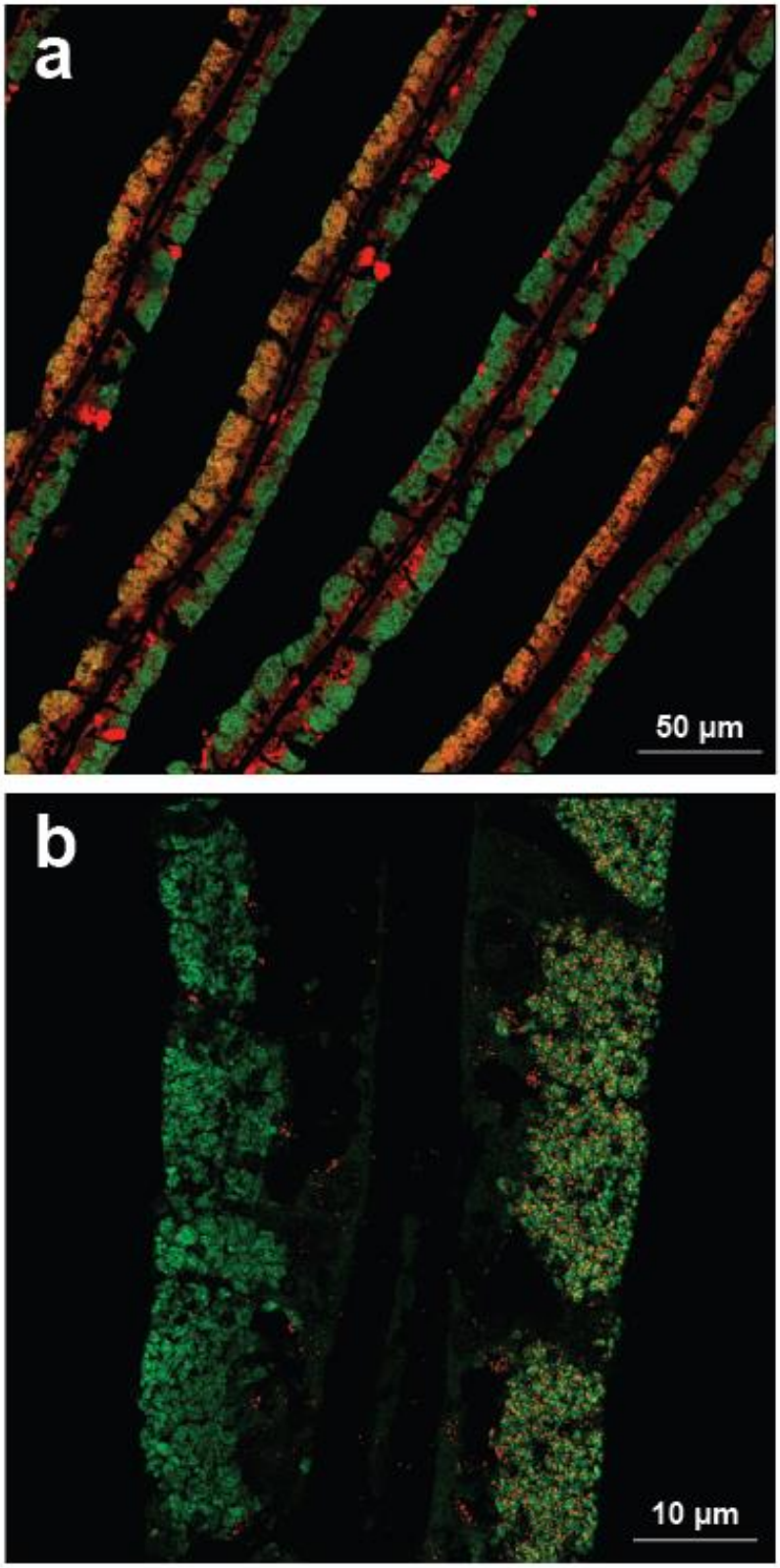


Figure 2. The MDH$^+$ SOX strain dominates quantitatively the bacteriocytes where is present, and these bacteriocytes form patches that do not face each other. Direct-geneFISH on gills' cross sections. MDH-gene cluster (red: Alexa 647), SOX 16S rRNA (green: atto488). ***a***, CLSM imaging of the MDH$^+$ SOX strain at a multiple filaments scale. ***b***, SR-SIM imaging of the MDH$^+$ SOX strain at a single bacteriocyte scale.

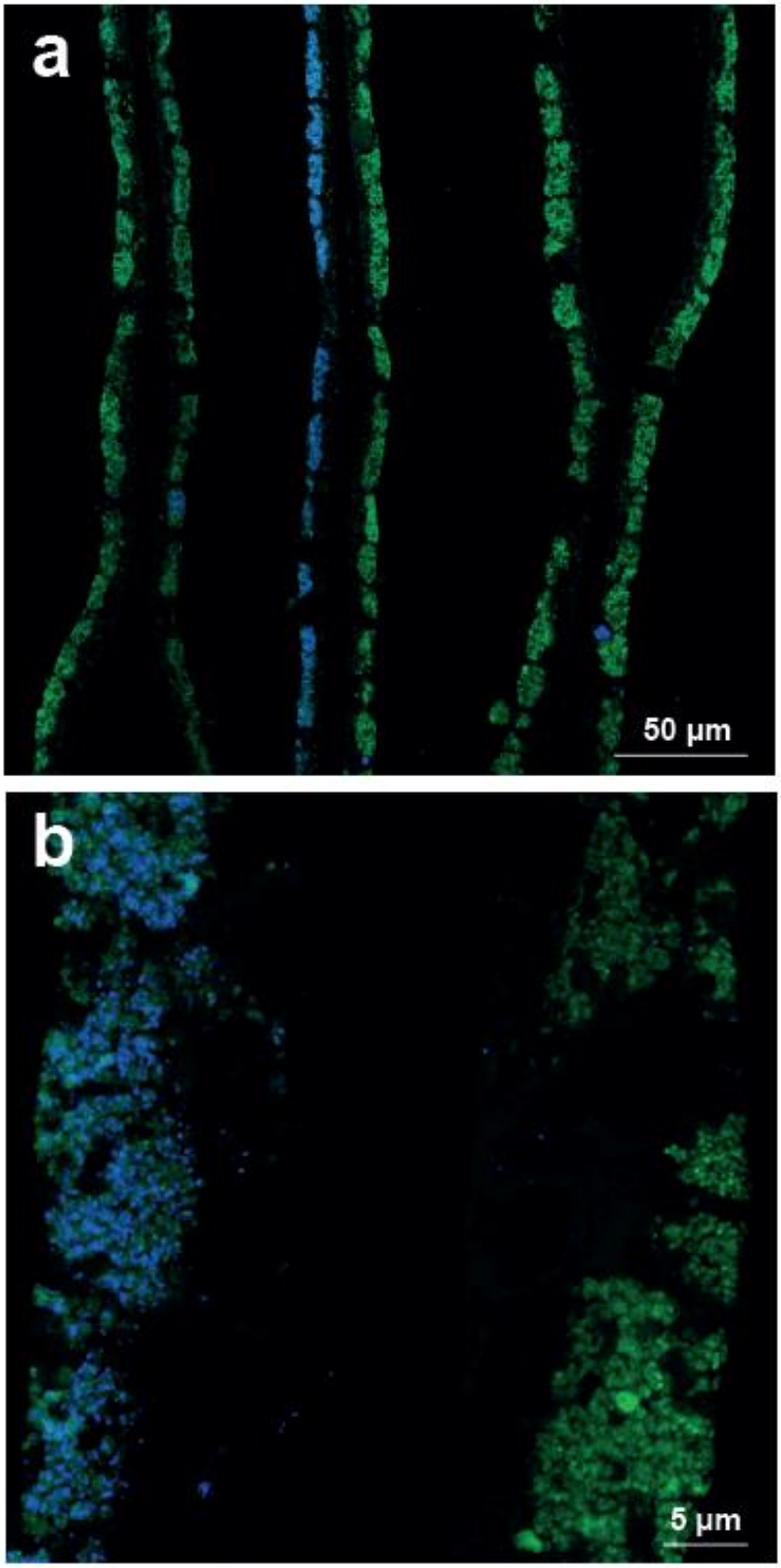


Figure 3. The HYD$^+$ SOX strain dominates quantitatively the bacteriocytes where is present, and these bacteriocytes cluster together forming patches, exactly like the MDH+ SOX strain. Direct-geneFISH on gills' cross sections. HYD-gene cluster (blue: Alexa 647), SOX 16S rRNA (green: atto488). ***a***, CLSM imaging of the HYD$^+$ SOX strain at a multiple filaments scale. ***b***, SR-Airyscan imaging of the HYD$^+$ SOX strain at a single bacteriocyte scale.

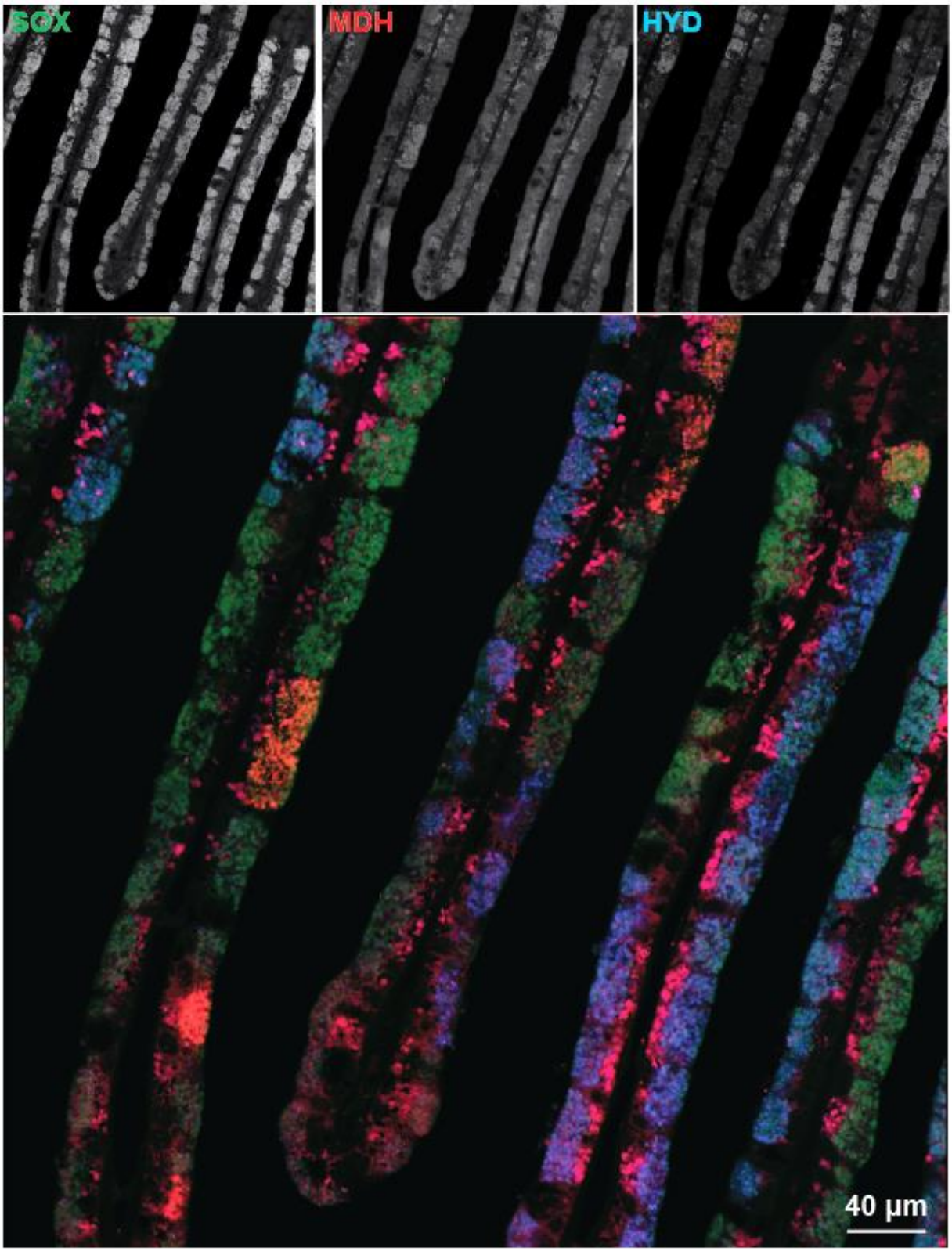


Figure 4. The MDH$^+$ and the HYD$^+$ SOX strains cluster separately in the gill filaments of *B. azoricus*. Double Direct-geneFISH on gills' cross sections. CLSM image of several gill filaments. HYD-gene cluster (blue: Alexa 647), MDH-gene cluster (red: Alexa 594), SOX 16S rRNA (green: atto488). CLSM imaging of HYD$^+$ and the MDH$^+$ SOX strains at a multiple filament scale.

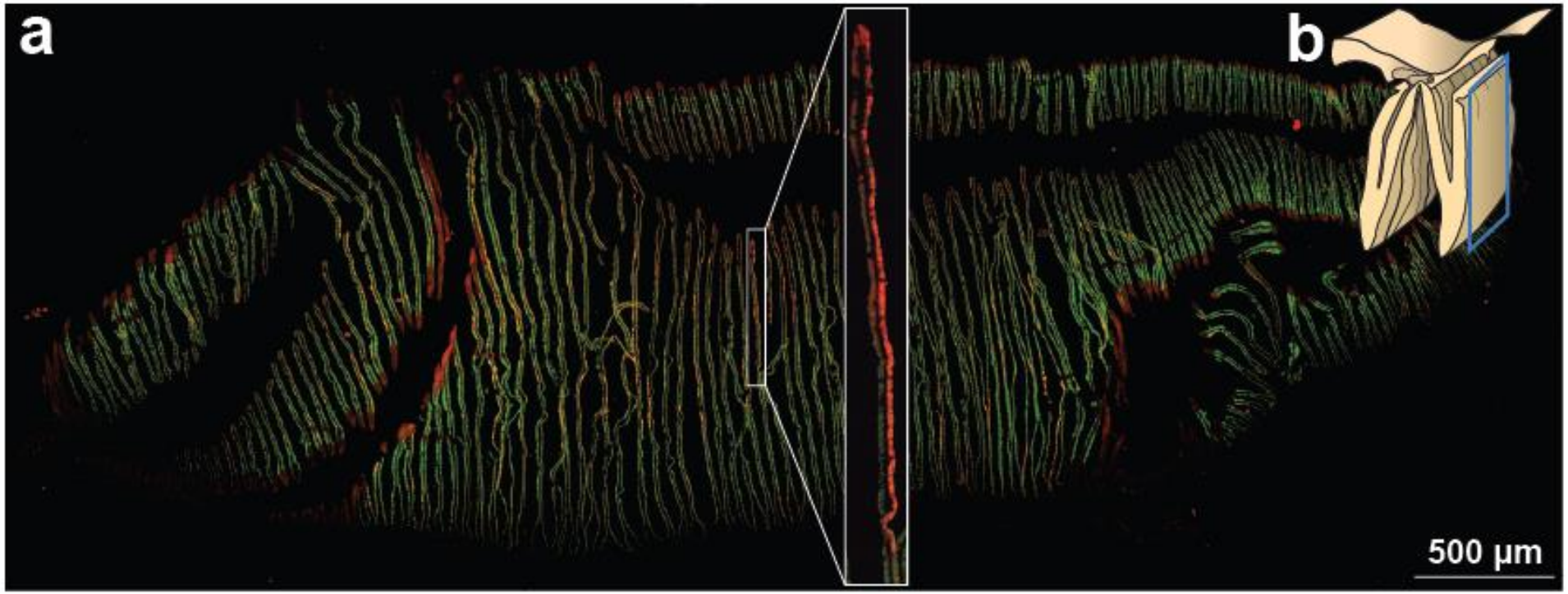


Figure 5. The MDH$^+$ SOX patches can occupy large and uninterrupted areas within single gill filaments. Direct-geneFISH on whole-gills sagittal sections. The sagittal sectioning plane (**b**) is represented by the blue frame over the gill diagram (adapted from Pennec 1988). MDH-gene cluster (red: Alexa 647), SOX 16S rRNA (green: atto488). ***a***, CLSM imaging of MDH-carrying SOX strain at a whole-gill scale. ***b***, Representation of the sagittal sectioning plane by the blue frame.

Cell proliferation

We wanted to investigate where cell proliferation occurs within the gill filaments of bathymodiolin mussels. To do so, we imaged proliferating cells in the gill filaments of EdU-incubated *B. puteoserpentis* specimens. *B. puteoserpentis* is a closely related species to *B. azoricus*, and we expected that both species share the same gill cell proliferation mechanics. The **right side of Figure 6** shows a schematic representation a V-shape gill filament. The **left side of Fig.6** shows six anatomically different areas of a gill filament where proliferating cells were localized. **Images 1 and 2 in Fig.6** correspond to the ciliated edge and the inner edge of the gill filament, respectively. An average of 0.73% of the nuclei in the ciliated edge area were EdU-labeled (**Supplementary table 1**). We did not find any EdU-labeled nuclei in the analyzed inner edge areas (0%). **Images 3 and 4 in Fig.6** show the ventral part of the gill filament. We found that an average of 9.64% nuclei in the ventral area were EdU-labeled. **Image 5 in Fig.6** shows the symbiotic area of the gill filament. In that area we observed that an average of 2.66% of the nuclei were EdU-labeled. **Image 6 in Fig.6** shows the dorsal part of the gill filament, which connects to the mantle tissue. In the dorsal area, 12.69% of the nuclei were EdU-labeled. Evaluation of the screened areas confirmed two proliferation centers (**highlighted in purple in the schematic representation of Fig.6**): The dorsal part (mantle connection) and ventral part (V-tip) of the gill filaments. Remarkably, no symbiotic cells were found in the proliferation centers, suggesting that newly formed bacteriocytes are non-symbiotic. All percentage calculations are included in **Suppl. Table 1**.

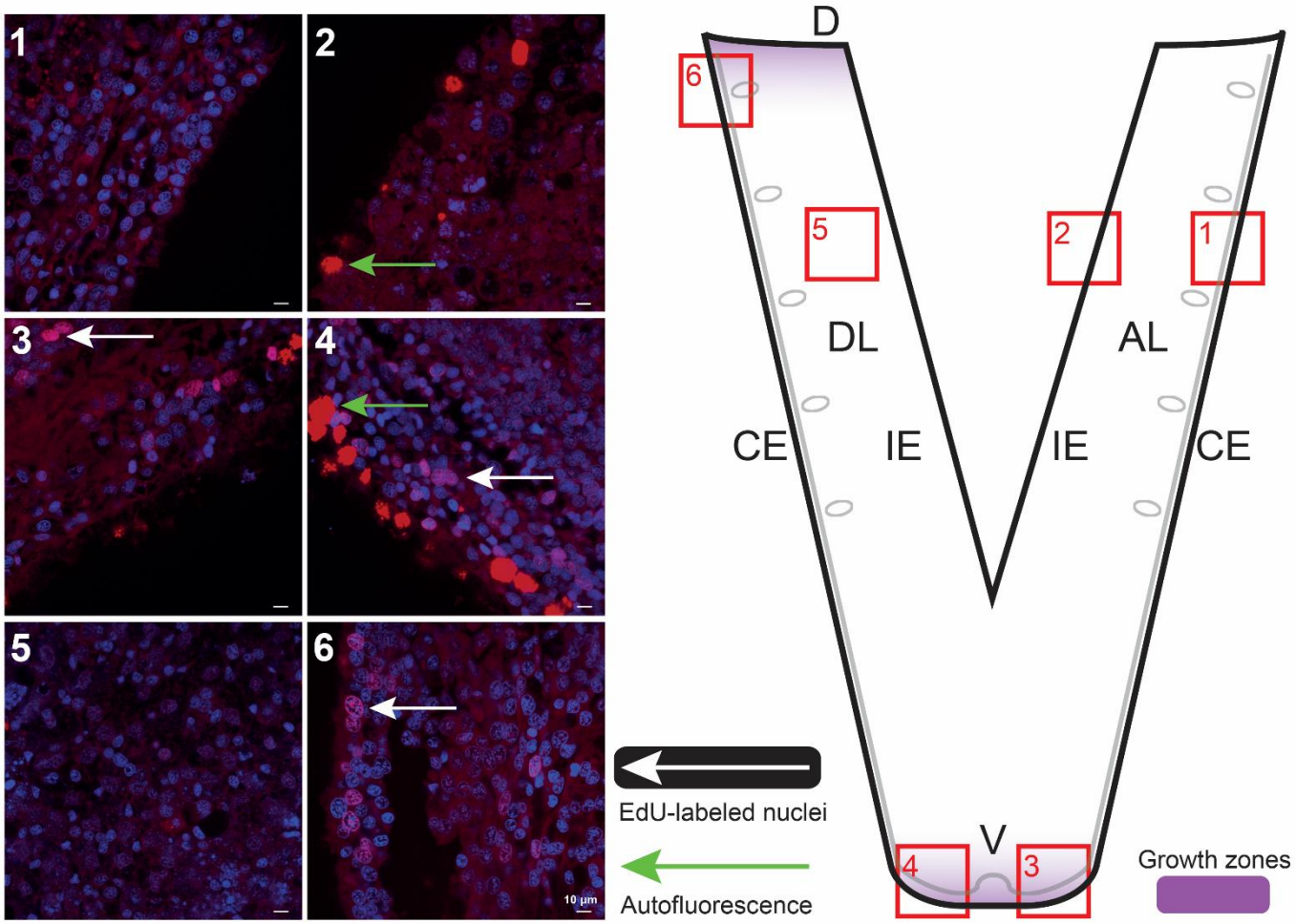


Figure 6. Gill filaments of bathymodiolin mussels grow due to cell proliferation that occurs in the grow zones (dorsal and ventral part of gill filaments). Right-side, schematic representation of gill filament. **D**: dorsal; **DL**: descending lamella; **AL**: ascending lamella; **CE**: ciliated edge; **IE**: inner edge; **V**: ventral. Growth areas have been highlighted in purple. **Left-side**, CLSM images of EdU-labeled cells in whole gill filaments of *B. puteoserpentis*. Proliferating cells labeled with EdU (red: Alexa 594), DNA (blue: DAPI). Red frames in schematic diagram of the gill filament correspond to the images on the left. **1**: ciliated edge area; **2**: inner edge area; **3** & **4**: ventral area; **5**: symbiotic area; **6**: dorsal area. White arrows point proliferating nuclei, while green arrows point autofluorescence (mucus cells).

Discussion

Far-red fluorescent dyes increase the signal-to-noise ratio when resolving strain-specific markers in host-microbiome systems

Our results when resolving strain-specific markers with in *B. azoricus* using Direct-geneFISH confirmed that single-gene localization is possible in host-microbiome systems (**Fig.1, a; Fig.2, b; Fig.3, b**). Super-resolution microscopy techniques require an optimal signal-to-noise ratio to resolve ultrastructures (Schermelleh et al. 2019). One way to increase the signal-to-noise ratio is to decrease the noise, which can be a challenge in host-microbiome systems because symbionts occur in a tissular environment (Chomicki et al. 2020). When localizing a single strain specific marker (**Fig.1, a; Fig.2, b; Fig.3, b**), we obtained the best outcome in terms of image quality by using polynucleotide probes labeled with far-red fluorescent dyes (Alexa 647). Animal tissue shows the highest autofluorescence when excited with light from the middle part of the visible light spectrum (“green” and “yellow” light). In contrast, animal tissue shows the lowest autofluorescence when excited with light from both extremes of the visible light spectrum (“blue” and “far-red” light) (Monici 2005). We also reduced the thickness of the *B. azoricus* gill tissue sections down to 3 µm, reducing the surrounding background as a side-effect of having less tissue volume. We hypothesize that the combination of reducing the sections thickness down to 3 µm together with alexa 647-labeled probes allowed us to optimally resolve strain-specific markers in a host-microbiome system such as *B. azoricus*.

Bacteriocytes are quantitatively dominated by a single SOX strain

The super-resolution localization of the MDH^+ and HYD^+ SOX strains showed that bacteriocytes mainly harbored a single SOX strain (**Fig.2, b; Fig.3, b**). Is this quantitative dominance at a single bacteriocyte level determined biologically and/or environmentally? Intracellular pathogens such as the enteropathogenic *E. coli* or *V. parahaemolyticus* trigger an ultrastructural transformation in their host cell known as microvilli effacement (Kaper et al. 2004; Zhou et al. 2014). Franke et al., *in Prep* showed that the symbiont pioneers that colonize a bacteriocyte induce surface modifications that culminate in microvilli effacement. Are the bacteriocytes' microvilli a requisite for symbionts acquisition? The intracellular enteropathogen *Shigella flexneri* uses its T3SS to deliver into the host enterocyte the virulence factor *IpgD,* which plays a major role during cell invasion (Niebuhr et al. 2000). The delivery of *IpgD* through the T3SS is dependent on the physical interaction between the T3SS and the enterocyte apical microvilli. It is tempting to speculate that the disappearance of microvilli in bacteriocytes might impede symbiont-host plasma membrane recognition and SOX symbiont colonization. It could be that the first SOX symbionts that colonize a newly formed bacteriocyte induce microvilli effacement, "shutting the door" to posterior SOX colonizers. The quantitative dominance of single SOX strains at a single bacteriocyte scale might also be the result of selective advantage. Ansorge et al. 2019 suggested that certain strain-specific genetic markers confer metabolic capabilities to the SOX symbionts that encode them. It is possible that initially, more than one strain colonizes a bacteriocyte, but the strain with selective advantage under certain environmental conditions outnumbers other strains. E.g. higher concentration of hydrogen might promote the proliferation of the HYD^+ SOX strain in detriment of other strains. We contemplate the scenario in which the prevalence of metabolically privileged SOX

strains might confer selective advantage to the bathymodiolin host under certain environmental conditions. An alternative explanation to the quantitative dominance of single SOX strains in individual bacteriocytes could be the result of antagonistic strain-strain interactions. Sayavedra et al. 2015 demonstrated that the SOX symbionts from *B. azoricus* codify for YD repeat proteins. YD repeat protein proteins can participate in competition between closely related bacterial strains (Koskiniemi et al. 2013). It could be that different SOX strains outcompete other strains at a single bacteriocyte level by expressing YD repeat proteins. Our data showed that although bacteriocytes seem to be quantitatively dominated by a single SOX strain, exclusion between strains is not absolute. Although extremely rarely, we co-localized MDH^+ and HYD^+ SOX signals in the same bacteriocyte. Why does the dominant strain not completely outnumber the scarce strains within the same bacteriocyte? A possible explanation to this question is that not all the mechanisms discussed above might be operating in every bacteriocyte. It is possible that selective advantage (environmental conditions) or chemical incompatibility (YD repeat proteins) would lead to a total exclusion between strains at a single bacteriocyte scale. However, it is tempting to speculate that a bacteriocyte that preserves its microvilli can receive more than one SOX strain during the initial phase of colonization.

Intra-filament colonization might explain the formation of patches

Our data showed that SOX symbionts from the MDH^+ or HYD^+ strain cluster together in neighboring bacteriocytes forming patches in the gill filaments of *B. azoricus* (**Fig.1, a; Fig.2, a; Fig.3, a**). Intriguingly, we also observed that MDH^+ patches segregate spatially from HYD^+ patches (**Fig.4**). How are patches formed? Why do patches from different strains segregate spatially?

Wentrup et al. 2014 demonstrated that in bathymodiolin mussels, cell proliferation takes place in the dorsal and ventral parts of the gill filaments (**highlighted in purple in the schematic representation of Fig.6**). Our localization of proliferating cells revealed that 12.69% and .9.64% of the nuclei were undergoing replication in the dorsal and ventral region of gill filaments, respectively (**Suppl. Table 1**). Remarkably, no symbionts were found in these growth zones, suggesting that newly formed bacteriocytes are non-symbiotic. One plausible hypothesis is that newly formed bacteriocytes must be colonized *de novo.* Because new cells form continuously in the gill filaments, symbiont acquisition is a lifetime process for bathymodiolin mussels. Bathymodiolin mussels acquire their symbionts horizontally (DeChaine et al. 2006; Distel et al. 1988; Laming et al. 2014; LePennec, 1988; Wentrup et al. 2014; Won et al. 2003; Won et al. 2008). Newly formed bacteriocytes can get colonized by symbionts from the water column (free-living symbionts or symbionts from co-occurring mussels), from already-colonized bacteriocytes from ontogenetically older and adjacent gill filaments ("inter-filament" colonization) (Wentrup et al. 2014) or from bacteriocytes that have been already colonized within the same gill filament ("intra-filament" colonization).

The newly acquired symbionts might be initially limited in strain diversity. We hypothesize that a batch of newly formed bacteriocytes that get colonized by a pool of symbionts present in the water column with low strain diversity might form a patch. In this hypothetical scenario, the initial limitation in strain diversity of the horizontally acquired symbionts would determine the co-occurrence of the same strain in neighboring bacteriocytes. However, it seems unlikely that such a structured pattern can originate from randomized and independent symbiont colonization events. E.g. the formation of a patch like the one highlighted in **Fig.5** by the white frame would imply that more than 50 neighboring bacteriocytes have uptaken the MDH+ SOX strain from

the water column during their early development during 50 independent colonization events. An alternative explanation to the formation of patches is the prevalence of a certain SOX strain in pre-stablished gill filaments. Wentrup et al. 2014 demonstrated that newly formed filaments firstly acquire their symbionts in the side that is facing an already colonized and pre-stablished gill filament. This suggested that pre-stablished gill filaments might act as symbiont donors to the newly formed and adjacent gill filaments. If a pre-stablished gill filament is mainly occupied by a certain SOX strain, is likely that this SOX strain will preferentially colonize most of the available bacteriocytes from the adjacent and younger gill filament, forming a patch. This would create a "mirrored" distribution of SOX strains between gill filaments. To know if there was a correlation between MDH^+ patches occurrence and the spatial proximity between them, we quantify the number of facing MDH^+ patches in **Fig.1**. Data in **Supp. Table 3** revealed that only 23% of the MDH^+ patches of **Fig.1** were facing another MDH^+ patch in the adjacent gill filament. Our data indicated that the correlation between MDH^+ patches occurrence and the spatial proximity between them was weak, discarding a "mirrored" distribution of SOX strains between gill filaments. We hypothesize that the model of "inter-filament" colonization proposed by Wentrup et al. 2014 (**Fig.7, a**) cannot explain by itself the formation of patches. Which could be the missing biological mechanism responsible for patches formation?

A possible answer to this question is the colonization of newly formed bacteriocytes from symbionts that are present in the same gill filament ("intra-filament" colonization). Filaments grow during their life time due to cell proliferation that occurs in the growth zones (**highlighted in purple in the schematic representation of Fig.6**). Our imaging data in **Fig.6, left side (images 3, 4, 6)** showed that growth zones are deprived of symbionts, suggesting that newly formed bacteriocytes during gill filament

growth must be colonized *de novo*. We hypothesize that non-symbiotic bacteriocytes newly formed in the growth zones get their SOX symbionts from adjacent bacteriocytes that are already colonized. This mechanism of “intra-filament” colonization would act as a “wave of conquest”, in which the SOX strains present in the “old land” (already colonized bacteriocytes) occupy “newly opened terrain” (newly formed bacteriocytes in the growth zones highlighted in purple) (**Figure 7, b**).

Data shown in **Fig.5** highlighted by the white frame demonstrated that MDH^+ patches can occupy large and uninterrupted areas of a single fill filament. As discussed above, it is unlikely that such a structured pattern can be formed by a continuous acquisition of a single strain (MDH^+) from the water column. Also discussed above, the model of “inter-filament” colonization proposed by Wentrup et al. 2014 neither seems to explain the formation of patches. Symbiont acquisition from the water column or from “inter-filament” colonization might operate during the early ontogeny of a gill filament (formation phase) (**Fig.7, a**). However, it is tempting to speculate that “intra-filament” colonization “stretches” the acquired the pioneer SOX strains to newly formed bacteriocytes in the growth zones, forming a patch during its growth phase (**Fig.7, b**). We hypothesize that “intra-filament” colonization favors the formation of patches, but, why do patches from different strains (MDH^+ and HYD^+) segregate spatially (**Fig.4**)? One possible explanation is that patches from different strains do not intermix because each patch originates from an initial bacteriocyte which is quantitatively dominated by a single SOX strain. The population of that bacteriocyte would spread to adjacent newly bacteriocytes that originate in the growth zones in a chain of consecutive colonization events. In the “intra-filament” colonization model proposed here, we contemplate two non-mutually excluding scenarios by which SOX symbionts can be transmitted from stablished bacteriocytes to newly formed bacteriocytes. In the first scenario, symbionts

from stablished bacteriocytes are spelled to the extracellular medium through the apical part of the bacteriocyte and uptaken by adjacent non-symbiotic bacteriocytes. Kádár et al. 2005 demonstrated that *B. azoricus* specimens deprived of dissolved sulfide gradually lost their SOX symbionts. After re-acclimatization in sulfide-supplied seawater, the mussels re-acquired their symbionts. This suggests that bathymodiolin mussels can expel symbionts from their gills to the surrounding medium and re-uptake them. We hypothesize that a similar mechanism might be responsible of the transmission of SOX between stablished bacteriocytes and adjacent and newly formed bacteriocytes. In the second scenario, symbionts from stablished bacteriocytes might be laterally transmitted to adjacent and newly formed bacteriocytes, in a cell-to-cell spreading mechanism. Intracellular pathogens such as *Listeria monocytogenes* or *Shigella flexneri* can spread between enterocytes without transiting through the extracellular medium (Kuehl et al. 2015). After *L. monocytogenes* or *S. flexneri* have primarily infected an enterocyte, they scape the phagocytic vacuole, acquire actin-based mobility and infect an adjacent enterocyte (Weddle and Agaisse 2018). Ultrastructural data demonstrated that SOX symbionts are contained within vacuoles in bacteriocytes (Distel, 1995). We never found ultrastructural evidence that SOX can occur free in the cytoplasm of bacteriocytes, neither evidences for cell-to-cell spreading. Therefore, we hypothesize that SOX colonize newly formed bacteriocytes after being freed in the environment by already colonized and adjacent bacteriocytes (first case scenario).

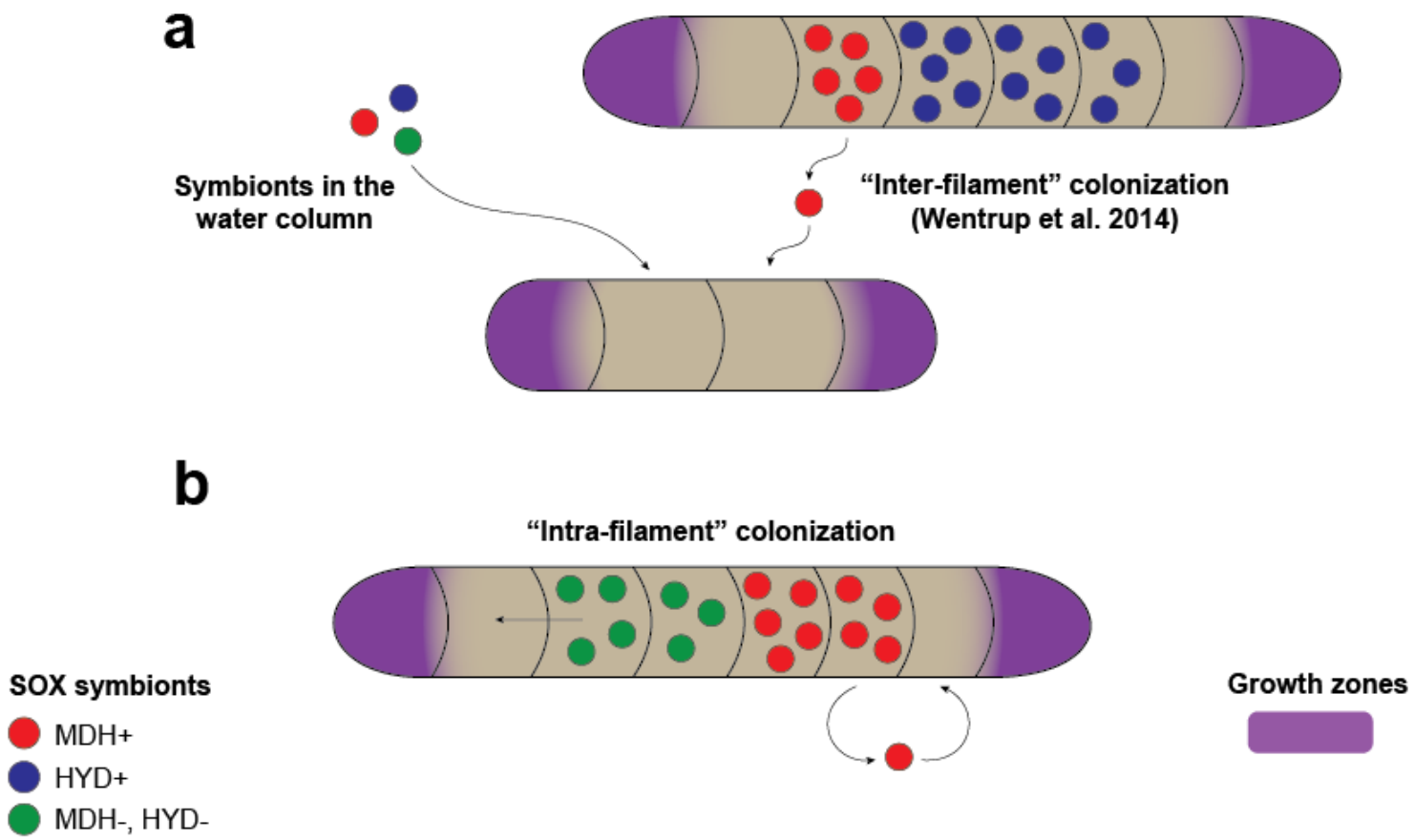


Figure 7. "Intra-filament" colonization might explain the formation of patches in the gill filaments of *B. azoricus*. *a,* schematic representation of the colonization of a newly formed gill filament (formation phase). In the formation phase, we hypothesize that the pioneers symbionts can come from the water column (free-living or from co-occurring mussels) or from adjacent gill filaments ("inter-filament" colonization). *b,* schematic representation of the colonization of a mature gill filament (growth phase). In the growth phase, we hypothesize that newly formed bacteriocytes in the growth zones get colonized by adjacent and pre-colonized bacteriocytes of the same gill filament ("intra-filament" colonization), which would explain the formation of patches. This type of colonization might occur previous release of SOX symbionts to the extracellular medium, or by lateral transference.

Conclusion and outlook

In this study, we aimed to answer two questions: 1) whether the spatial segregation of different SOX strains occurs at a single bacteriocyte scale and 2) whether different SOX strains form patches in the gill filaments of *B. azoricus*. Our super-resolution localization of the MDH^+ and/or the HYD^+ SOX strains revealed that there is quantitative dominance of single SOX strains at a single bacteriocyte scale. We hypothesize that microvilli effacement, selective advantage and/or chemical incompatibility are some of the phenomena that could originate this quantitative dominance. Our simultaneous localization of the MDH^+ and the HYD^+ SOX strains at multiple gill filaments scale revealed that both strain form patches. Moreover, the MDH^+ patches spatially segregated from the HYD^+ patches. Our results also showed that growth zones of gill filaments in *B. puteoserpentis* are deprived of symbionts, suggesting that newly formed bacteriocytes must be colonized *de novo* in bathymodiolin mussels. Newly formed bacteriocytes can be colonized by SOX symbionts present in the water column, present in adjacent gill filaments ("inter-filament" colonization) or present in the same gill filament ("intra-filament" colonization). We hypothesize that SOX symbionts present in the water column or in adjacent gill filaments might be the pioneers when colonizing a young gill filament (formation phase). However, it is tempting to speculate that once the pioneer SOX strains have settle in a young gill filament, they spread to newly formed bacteriocytes of the same filament, creating patches (growth phase). To our understanding, the reason whereby patches from different strains segregate spatially is because they originate from bacteriocytes that are quantitatively dominated by a single SOX strain. The spreading of the pioneer SOX strains to newly formed bacteriocytes during

filament growth (“intra-filament” colonization) would create a bottleneck effect that determines the structure of the SOX population at a single filament scale. To complement this study, we propose the metagenomic analysis of SOX strain diversity at a single bacteriocyte level. This would require the preparation of single-cell DNA libraries from sorted bacteriocytes obtained through the homogenization of *B. azoricus* gill samples. This would allow us to confirm our super-resolution findings: That bacteriocytes might be quantitatively dominated by a single SOX strain. We also propose to resolve SOX strain diversity at a single filament scale. This would require the preparation of DNA libraries from independent gill filaments. This would allow us to corroborate our proposed model for “intra-filament” colonization. Low SOX strain diversity at a single filament scale would imply that there is a bottleneck effect during colonization of newly formed bacteriocytes during the growth phase. To strengthen our model of “intra-filament” colonization, we also propose the 3D reconstruction of SOX strains distribution at a whole-gill scale. This would require the visualization of strain-specific markers in consecutive whole-gill sections using Direct-geneFISH. Last, we localized proliferating cells in the gill filaments of *B. puteoserpentis*. Although the characterization of the growth centers in *B. puteoserpentis* can be extrapolated to *B. azoricus*, we propose the incubation of *B. azoricus* specimens with EdU. The characterization of the growth zones in the same species in which the localization of single SOX strains have been done would allow a better interpretation of the mechanisms that connect cell proliferation and patches formation.

Materials and methods

Sample collection

Mussels were collected with the remote operated vehicle (ROV) MARUM-Quest during two cruises: *B. azoricus* specimens were collected during the BioBaz cruise 2013 in 2013, while *B. puteoserpentis* specimens were collected during the RV Meteor M126 cruise in 2016. *B. azoricus* mussels were collected during one dive from the Eiffel Tower venting site (WGS84, 37º29' N; -032º28' W), while *B. puteoserpentis* mussels were collected during one dive from the Logatchev venting site (Irina II, 14º45' N; -044º58' W) at a water depth of 1,690 and 3,038 m, respectively.

Polynucleotide probes design

A total of 12 and 10 polynucleotide probes were designed using the software Geneious V.11.1.5 (Biomatters, New Zealand) to target the MDH and the HYD-gene clusters, respectively. Hybridization conditions and melting curves of the polynucleotide probes were determined using the Polynucleotide Probe Design Software PolyPro (Moraru et al. 2011). The PCR primers used to amplify the polynucleotide probes, as well as the gene that each polynucleotide probes targets within their respective gene clusters are shown in **Supplementary Table 5**.

Polynucleotide probes synthesis

DNA was extracted from a *B. azoricus* gill homogenate with the DNeasy blood and tissue kit according to the manufacturer instructions (Qiagen, Germany). The extracted DNA was used as a template to amplify the MDH and the HYD-gene

clusters by PCR using Phusion® DNA Polymerase (Thermo Fisher Scientific, MA, USA) with the PCR primers shown in **Supplementary Table 3**. For the amplification of the MDH-gene cluster, the following conditions were used: initial denaturation at 98°C for 30 s, 35 cycles at 98°C for 10 s, 64.6°C for 30 s and 72°C for 4 min, and a final elongation step at 72°C for 10 min. For the amplification of the HYD-gene cluster, the following conditions were used: initial denaturation at 98°C for 30 s, 30 cycles at 98°C for 10 s, 67°C for 30 s and 72°C for 2 min and 30 s, and a final elongation step at 72°C for 10 min. Both PCR products were cloned with the TOPO® XL cloning kit (Invitrogen, CA, USA) according to the manufacturer descriptions. The cloning vectors were purified using the QIAprep Spin Miniprep Kit according to the manufacturer instructions (Qiagen, Germany). The purified cloning vector containing the MDH-gene cluster was used as a template to synthesize the polynucleotide probes by PCR with Taq DNA Polymerase (5 PRIME, Hamburg, Germany) using the following conditions: Initial denaturation at 95°C for 5 min, 35 cycles at 95°C for 1 min, 55°C for 1 min 30 s and 72°C for 2 min, and a final elongation step at 72°C for 10 min. The purified cloning vector containing the HYD-gene cluster was used as a template to synthesize the polynucleotide probes by PCR with Phusion® DNA Polymerase (Thermo Fisher Scientific, MA, USA) using the following conditions: Initial denaturation at 98°C for 30 s, 25 cycles at 98°C for 10 s, 67°C for 30 s and 72°C for 15 s, and a final elongation step at 72°C for 10 min. After amplification, the polynucleotide probes targeting the MDH and the HYD-gene clusters were purified with the GENECLEAN® turbo kit (MP Biomedicals, Germany) and chemically labeled with Alexa Fluor® 647 dyes using the ULYSIS® Nucleic Acid Labeling Kit (Molecular Probes, OR, USA). An additional set of the MDH-targeting

probes was also labeled with Alexa Fluor® 594 dyes using the ULYSIS® Nucleic Acid Labeling Kit (Molecular Probes, OR, USA). After labeling, probes were purified using the gel filtration-based spin columns BioRad Micro Bio-Spin P-30 (BioRad, Hercules, CA). After purification, the labeling efficiency was calculated following manufacturer instructions using a NanoDrop Measurement System ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE). Probes were kept at -20°C until utilization.

MDH$^+$ calculations

To get an impression of the abundance of MDH+ patches in the gill filaments of *B. azoricus*, we analyzed the MDH$^+$ signals in the whole-gill section (**Fig.1, a**). The percentage of gill filaments containing at least one MDH$^+$ patch was calculated as a percentage between the amount of filaments carrying at least one MDH$^+$ patch and the total amount of gill filaments in the section (**Supp. Table 2**). To discuss the hypothesis of inter-filament colonization proposed by Wentrup et al. 2014, we calculated the percentage of facing MDH$^+$ patches. The percentage of gill filaments with at least one MDH$^+$ patch facing another filament with at least one MDH$^+$ patch was calculated as a percentage of the number of pairs of facing filaments with at least one MDH$^+$ patch and the total amount of gill filaments in the section (**Supp. Table 3**). To have an idea about the prevalence of the MDH$^+$ in the global SOX population, we calculated as a ratio the relative abundance of the MDH$^+$ fluorescent signal vs. the total amount of 16S r RNA signals and expressed it as a percentage (**Suppl. Table 4**).

Direct-geneFISH on Bathymodiolus azoricus gill tissue

Upon recovery, *B. azoricus* specimens were sacrificed on board and their gills dissected and fixed in paraformaldehyde (PFA) at 4°C during 8 h. After fixation, samples were kept at -20°C until embedding in the laboratory. Two *B. azoricus* specimens were subjected to a modified version of the original Direct-geneFISH protocol (Barrero-Canosa et al. 2017) as described next. The formalin-fixed paraffin-embedded gills of the mussel specimens were cross sectioned at 3 µm using a conventional microtome and mounted on poly-L-lysine-coated glass slides (Sigma-Aldrich, MO, USA) using a water bath. Sections were left to dry in vertical position at RT for 4 h. Prior to dewaxing, sections were baked at 60°C for 1 h in vertical position for tissue adherence improvement. The gill sections were dewaxed with Roti®-Histol (Carl-Roth, Germany) in three consecutive steps for 10 min each followed by decreasing ethanol series of 96, 80, 70 and 50% (v/v) for 10 min each. Then, tissue sections were washed in milliQ water for 10 min. For permeabilization, the sections were incubated with lysozyme solution (0.5 $mg \cdot ml^{-1}$ lysozyme (Sigma-Aldrich, Germany) in 1X PBS, 0.05 M EDTA and 0.1 M Tris-HCl pH 7.8) for 1 h over ice. After lysozyme incubation, tissue sections were washed with milliQ water for 5 min followed by a quick wash in ethanol 96% (v/v) and dried at 37°C for 30 min. The gene- and rRNA-targeting probes were hybridized simultaneously in a hybridization buffer (conforming together the hybridization mixture) containing 35% formamide (5X SSC buffer, 20% dextran sulfate (v/v), 20 mM EDTA, 35% formamide (v/v), 0.1% SDS (v/v), sheared salmon sperm 0.25 $mg \cdot ml^{-1}$ (Invitrogen, MA, USA), yeast RNA 0.25 $mg \cdot ml^{-1}$ (Invitrogen, MA, USA), 1% blocking reagent for nucleic acids (Roche, Switzerland)), 31 $pg \cdot \mu l^{-1}$ per individual polynucleotide probe (see treatments next)

and 5 ng·µl^{-1} of the SOX 16S rRNA probe BNMARt_193 labeled with the dye MFP-ATTO488 (Duperron et al. 2006). As a negative control, we used 372 pg·µl^{-1} of Alexa Fluor® 647 and/or 594 labeled of the non-sense polynucleotide probe NonPolyPr350. NonPolyPr350 was amplified by PCR from a larger artificial template sequence named NonPolyPr as described in Barrero-Canosa et al. 2017. NonPolyPr was designed not to have significant likeness via BLASTN searches with bacterial/archaeal sequences deposited in the public database GenBank (Johnson et al. 2008). We set three hybridization treatments: $MDH_{(647)}$ + $SOX_{(488)}$, $HYD_{(647)}$ + $SOX_{(488)}$ and $HYD_{(647)}$ + $MDH_{(594)}$ + $SOX_{(488)}$. For each treatment, 500 µl of hybridization mixture was prepared. Each hybridization mixture was equally divided between two *B. azoricus* gill sections (250 µl each). Before pipetting the hybridization mixture over the samples they were surrounded by adhesive silicon isolators (Sigma-Aldrich, MO, USA) to avoid buffer leaking during hybridization. The samples were then placed in a hybridization chamber. To maintain a humid atmosphere within the hybridization chamber during incubation, KIMTECHScience precision wipes (Kimberly-Clark, TX, USA) partially soaked in formamide 35% were located below the samples. The hybridization chamber was then incubated for 40 min at 85°C for the denaturation of the target gene and the polynucleotide probes. After the denaturation step, the samples were hybridized at 46°C for 4 h. Following the hybridization, the samples were rinsed in pre-warmed 48°C washing buffer (0.07 M NaCl, 0.02 M Tris-HCl pH 7.8, 5 mM EDTA pH 8, and 0.01% SDS (v/v)) and transferred to fresh pre-warmed washing buffer for 15 min followed by 20 min in 1X PBS, 1 min in milliQ water, a quick wash in ethanol 96% (v/v) and dried at 37°C for 30 min. After drying, the sections were mounted using the ProLong® Gold antifade mounting

media (Thermo Fisher Scientific, MA, USA), cured overnight at room temperature and stored -20°C until visualization.

Cell proliferation on gill filaments

To visualize proliferative cells on the gill filaments of *B. puteoserpentis*, living mussels were incubated for 46 h on board with 5-ethynyl-2'-deoxyuridine (EdU). After incubation, animals were sacrificed and their gills dissected and fixed in paraformaldehyde (PFA) at 4°C during 8 h. PFA-fixed whole gill filaments were subjected to Click-iT reaction (Click-iT™ EdU Cell Proliferation Kit for Imaging, Alexa Fluor™ 594 dye) (Thermo Fisher Scientific, MA, USA), according to the manufacturer instructions. Next, filaments were washed in PBS 1X for 5 min. After washing, filaments were counterstained with DAPI for 10 min at room temperature, transferred to polylysine glass slides and mounted using the ProLong® Gold antifade mounting media (Thermo Fisher Scientific, MA, USA), cured overnight at room temperature and stored -20°C until visualization. All steps during the Click-iT reaction and DAPI counterstaining were done in dark conditions.

Microscopy

Samples were visualized with the dual system Zeiss LSM 780 & Super Resolution System ELYRA PS.1 (Carl Zeiss Microscopy GmbH, Germany). The samples were continuously illuminated using different excitation sources depending on the fluorophore used. Images were taken with a plan-APROCHROMAT 63X/1.4 oil immersion objective using an Andor iXon Ultra 897 High Speed EMCCD Camera

(Andor, UK). Beam selection and modulation of the laser intensities were controlled in several ways, depending on the laser wavelength, the experimental treatment and the sample of study. Images were obtained and post-processed using ZEN software (black edition, 64bits, version: 14.0.1.201) (Carl Zeiss Microscopy GmbH, Germany). Prior to image exporting, histograms were slightly modified to increase the contrast between channels. Exported images were brightness-corrected using the software Adobe Photoshop (version: 12.0) (Adobe Systems, CA, USA).

Direct-geneFISH super resolution microscopy

Single-field localization of the MDH-gene cluster and SOX symbionts 16S rRNA (**Fig.2, b**) was done using super-resolution structured illumination microscopy (SR-SIM). The MDH-gene cluster was localized exciting the Alexa Fluor® 647 dyes using a 642 nm laser (150 mW) and the beam splitter LP655. The SOX symbiont 16S rRNA was localized exciting the MFP-ATTO488 dyes using a 488 nm laser (100 mW) and the beam splitter BP502-538. Single-field localization of the hydrogenase-gene cluster and SOX symbionts 16S rRNA (**Fig.3, b**) was done using the Airyscan detector. The hydrogenase-gene cluster was localized exciting the Alexa Fluor® 647 dyes using a 633 nm laser (150 mW) and the beam splitter SBS LP 660. The SOX symbiont 16S rRNA was localized exciting the ATTO488 dyes using a 488 nm laser (100 mW) and the beam splitter MBS488.

Direct-geneFISH confocal scanning laser microscopy

Additionally, we aimed to get a wider perspective of the MDH-carrying and/or hydrogenase SOX strain spatial distribution in a multiple-gill filaments scale. To

do so, we localized in image series with neighboring fields of view the MDH-gene cluster (**Fig. 1, a; Fig.2, a; Fig.3, a; Fig.5, a**) or the HYD-gene cluster (**Fig.3, a**) or both gene clusters at the same time (**Fig.4**) together with the SOX 16S rRNA using confocal scanning laser microscopy (CLSM). When localized independently (one gene at a time), The MDH- or the HYD-gene cluster were localized exciting the Alexa Fluor® 647 dyes using a 633 nm laser (5 mW) and the beam splitter MBS488/561/633 while the SOX symbiont 16S rRNA was localized exciting the MFP-ATTO488 dyes using an 488 nm laser (25 mW) and the beam splitter MBS488. When localized simultaneously (**double Direct-geneFISH hybridization, Fig.4**), the hydrogenase-gene cluster was localized exciting the Alexa Fluor® 647 dyes using a 633 nm laser (5 mW) and the beam splitter MBS488/561/633, while the MDH-gene cluster was localized exciting the Alexa Fluor® 594 dyes using a 561 nm laser (5 mW) and the beam splitter MBS488/561/633. Again, the SOX symbiont 16S rRNA was localized exciting the MFP-ATTO488 dyes using a 488 nm laser (25 mW) and the beam splitter MBS488.

Cell proliferation (EdU) confocal scanning laser microscopy

We aimed to localize proliferating cells in the gill filaments of *B. azoricus*. To do so, we imaged samples subjected to the Click-iT reaction and counterstained with DAPI using CLSM. The EdU-enriched nuclei were localized exciting the Alexa Fluor® 594 dyes using a 561 nm laser (5 mW) and the beam splitter MBS488/561/633. All the nuclei were localizing exciting DAPI dyes using a 405 nm laser (5 mW) and the beam splitter MBS-405.

References for chapter IV

Ansorge, Rebecca, Stefano Romano, Lizbeth Sayavedra, Miguel Ángel González Porras, Anne Kupczok, Halina E. Tegetmeyer, Nicole Dubilier, and Jillian Petersen. 2019. “Functional Diversity Enables Multiple Symbiont Strains to Coexist in Deep-Sea Mussels.” *Nature Microbiology* 4(12):2487–97.

Barrero-Canosa, Jimena, Cristina Moraru, Laura Zeugner, Bernhard M. Fuchs, and Rudolf Amann. 2017. “Direct-GeneFISH: A Simplified Protocol for the Simultaneous Detection and Quantification of Genes and rRNA in Microorganisms.” *Environmental Microbiology* 19(1):70–82.

Chomicki, Guillaume, Gijsbert D. A. Werner, Stuart A. West, and E. Toby Kiers. 2020. “Compartmentalization Drives the Evolution of Symbiotic Cooperation.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 375(1808):20190602.

DeChaine, Eric G., Amanda E. Bates, Timothy M. Shank, and Colleen M. Cavanaugh. 2006. “Off-Axis Symbiosis Found: Characterization and Biogeography of Bacterial Symbionts of *Bathymodiolus* Mussels from Lost City Hydrothermal Vents.” *Environmental Microbiology* 8(11):1902–12.

DeChaine, Eric G., and Colleen M. Cavanaugh. 2006. “Symbioses of Methanotrophs and Deep-Sea Mussels (*Mytilidae*: *Bathymodiolinae*).” *Progress in Molecular and Subcellular Biology* 41:227–49.

Distel, D. L., D. J. Lane, G. J. Olsen, S. J. Giovannoni, B. Pace, N. R. Pace, D. A. Stahl, and H. Felbeck. 1988. “Sulfur-Oxidizing Bacterial Endosymbionts: Analysis of Phylogeny and Specificity by 16S rRNA Sequences.” *Journal of Bacteriology* 170(6):2506–10.

Distel, Daniel L., Henri K. W. Lee, and Colleen M. Cavanaugh. 1995. “Intracellular Coexistence of Methano- and Thioautotrophic Bacteria in a Hydrothermal Vent Mussel.” *Proceedings of the National Academy of Sciences of the United States of America* 92(21):9598–9602.

Dubilier, Nicole, Claudia Bergin, and Christian Lott. 2008. “Symbiotic Diversity in Marine Animals: The Art of Harnessing Chemosynthesis.” *Nature Reviews Microbiology* 6(10):725–40.

Duperron, Sébastien, Claudia Bergin, Frank Zielinski, Anna Blazejak, Annelie Pernthaler, Zoe P. McKiness, Eric DeChaine, Colleen M. Cavanaugh, and Nicole Dubilier. 2006. “A Dual Symbiosis Shared by Two Mussel Species, *Bathymodiolus azoricus* and *Bathymodiolus puteoserpentis* (*Bivalvia*: *Mytilidae*), from Hydrothermal Vents along the Northern Mid-Atlantic Ridge.” *Environmental Microbiology* 8(8):1441–47.

Franke, Maximilian, Benedikt Geier, Jörg U. Hammel, Nicole Dubilier and

Nikolaus Leisch. “Symbiont colonization and early development of bathymodiolin mussels.” *In Preparation*.

Giovannoni, S. J., E. F. DeLong, G. J. Olsen, and N. R. Pace. 1988. “Phylogenetic Group-Specific Oligodeoxynucleotide Probes for Identification of Single Microbial Cells.” *Journal of Bacteriology* 170(2):720–26.

Heath-Heckman, E. A., Peyer, S. M., Whistler, C. A., Apicella, M. A., Goldman, W. E., & McFall-Ngai, M. J. (2013). "Bacterial bioluminescence regulates expression of a host cryptochrome gene in the squid-Vibrio symbiosis". mBio, 4(2), e00167-13.

Ikuta, Tetsuro, Yoshihiro Takaki, Yukiko Nagai, Shigeru Shimamura, Miwako Tsuda, Shinsuke Kawagucci, Yui Aoki, Koji Inoue, Morimi Teruya, Kazuhito Satou, Kuniko Teruya, Makiko Shimoji, Hinako Tamotsu, Takashi Hirano, Tadashi Maruyama, and Takao Yoshida. 2015. “Heterogeneous Composition of Key Metabolic Gene Clusters in a Vent Mussel Symbiont Population.” *The ISME Journal* 10(4):1–12.

Johnson, Mark, Irena Zaretskaya, Yan Raytselis, Yuri Merezhuk, Scott McGinnis, and Thomas L. Madden. 2008. “NCBI BLAST: A Better Web Interface.” *Nucleic Acids Research* 36(Web Server issue):5–9.

Kádár, Enikö, Raul Bettencourt, Valentina Costa, Ricardo Serrão Santos, Alexandre Lobo-da-Cunha, and Paul Dando. 2005. “Experimentally Induced Endosymbiont Loss and Re-Acquirement in the Hydrothermal Vent Bivalve *Bathymodiolus azoricus*.” *Journal of Experimental Marine Biology and Ecology* 318(1):99–110.

Kaper, James B., James P. Nataro, and Harry L. T. Mobley. 2004. “Pathogenic Escherichia Coli.” *Nature Reviews Microbiology* 2(2):123–40.

Koskiniemi, Sanna, James G. Lamoureux, Kiel C. Nikolakakis, Claire T. Kin. De Roodenbeke, Michael D. Kaplan, David A. Low, and Christopher S. Hayes. 2013. “Rhs Proteins from Diverse Bacteria Mediate Intercellular Competition.” *Proceedings of the National Academy of Sciences of the United States of America* 110(17):7032–37.

Laming, Sven R., Sébastien Duperron, Marina R. Cunha, and Sylvie M. Gaudron. 2014. “Settled, Symbiotic, Then Sexually Mature: Adaptive Developmental Anatomy in the Deep-Sea, Chemosymbiotic Mussel *Idas modiolaeformis*.” *Marine Biology* 161(6):1319–33.

Monici, Monica B. T. Biotechnology Annual Review. 2005. “Cell and Tissue Autofluorescence Research and Diagnostic Applications.” Pp. 227–56 in Vol. 11. Elsevier.

Moraru, Cristina, Phyllis Lam, Bernhard M. Fuchs, Marcel M. M. Kuypers, and

Rudolf Amann. 2010. “GeneFISH - an in Situ Technique for Linking Gene Presence and Cell Identity in Environmental Microorganisms.” *Environmental Microbiology* 12(11):3057–73.

Moraru, Cristina, Gabriel Moraru, Bernhard M. Fuchs, and Rudolf Amann. 2011. “Concepts and Software for a Rational Design of Polynucleotide Probes.” *Environmental Microbiology Reports* 3(1):69–78.

Niebuhr, Kirsten, Noureddine Jouihri, Abdelmounaaim Allaoui, Pierre Gounon, Philippe J. Sansonetti, and Claude Parsot. 2000. “IpgD, a Protein Secreted by the Type III Secretion Machinery of *Shigella flexneri*, Is Chaperoned by IpgE and Implicated in Entry Focus Formation.” *Molecular Microbiology* 38(1):8–19.

Pennec, M. L. 1988. “Alimentation et Reproduction d’un Mytilidae Des Sources Hydrothermales Profondes Du Pacifique Oriental.”

Petersen, Jillian M., and Nicole Dubilier. 2009. “Methanotrophic Symbioses in Marine Invertebrates.” *Environmental Microbiology Reports* 1(5):319–35.

Sayavedra, Lizbeth, Manuel Kleiner, Ruby Ponnudurai, Silke Wetzel, Eric Pelletier, Valerie Barbe, Nori Satoh, Eiichi Shoguchi, Dennis Fink, Corinna Breusing, Thorsten B. H. Reusch, Philip Rosenstiel, Markus B. Schilhabel, Dorte Becher, Thomas Schweder, Stephanie Markert, Nicole Dubilier, and Jillian M. Petersen. 2015. “Abundant Toxin-Related Genes in the Genomes of Beneficial Symbionts from Deep-Sea Hydrothermal Vent Mussels.” *ELife* 4(September):1–39.

Schermelleh, Lothar, Alexia Ferrand, Thomas Huser, Christian Eggeling, Markus Sauer, Oliver Biehlmaier, and Gregor P. C. Drummen. 2019. “Super-Resolution Microscopy Demystified.” *Nature Cell Biology* 21(1):72–84.

Weddle, Erin, and Hervé Agaisse. 2018. “Principles of Intracellular Bacterial Pathogen Spread from Cell to Cell.” *PLoS Pathogens* 14(12):1–8.

Wentrup, Cecilia, Annelie Wendeberg, Mario Schimak, Christian Borowski, and Nicole Dubilier. 2014. “Forever Competent: Deep-Sea Bivalves Are Colonized by Their Chemosynthetic Symbionts throughout Their Lifetime.” *Environmental Microbiology* 16(12):3699–3713.

Won, Yong-Jin, William J. Jones, and Robert C. Vrijenhoek. 2008. “Absence of Cospeciation Between Deep-Sea Mytilids and Their Thiotrophic Endosymbionts.” *Journal of Shellfish Research* 27(1):129–38.

Won, Yong Jin, Steven J. Hallam, Gregory D. O’Mullan, Irvin L. Pan, Kurt R. Buck, and Robert C. Vrijenhoek. 2003. “Environmental Acquisition of Thiotrophic Endosymbionts by Deep-Sea Mussels of the Genus *Bathymodiolus.*” *Applied and Environmental Microbiology* 69(11):6785–92.

Zhou, Xiaohui, Ramiro H. Massol, Fumihiko Nakamura, Xiang Chen, Benjamin E. Gewurz, Brigid M. Davis, Wayne Lencer, and Matthew K. Waldor. 2014. “Remodeling of the Intestinal Brush Border Underlies Adhesion and Virulence of an Enteric Pathogen.” *MBio* 5(4):1–9.

Supplementary Information for chapter IV

Supplementary table 1. Average percentage of EdU-labeled nuclei (nuclei in proliferation) in the 5 filament areas screened (ciliated edge, inner edge, symbiotic area, ventral part and dorsal part). Area, area of the gill filament screened. **Number of nuclei,** number of DAPI-labeled nuclei in the screened area. **Number of EdU-labeled nuclei,** number of nuclei undergoing proliferation in each of the screened areas. **Percentage (%)**, percentage of nuclei undergoing proliferation in each screened area. **Average**, average percentage of nuclei undergoing proliferation per screened area (n =3).

Area	Number of nuclei	Number of EdU-labeled nuclei	Percentage (%)	Average (%)
Ciliated_edge	191	3	1,57	
Ciliated_edge	159	0	0,00	
Ciliated_edge	160	1	0,63	0,73
Inner_edge	144	0	0,00	
Inner_edge	113	0	0,00	
Inner_edge	160	0	0,00	0,00
Symbiotic_area	67	4	5,97	
Symbiotic_area	165	2	1,21	
Symbiotic_area	126	1	0,79	2,66
Ventral_part	97	10	10,31	
Ventral_part	284	27	9,51	
Ventral_part	242	22	9,09	9,64
Dorsal_part	212	20	9,43	
Dorsal_part	116	8	6,90	
Dorsal_part	69	15	21,74	12,69

Supplementary table 2. Almost 60% of the gill filaments of a whole-gill section had a MDH^+ patch. Image, analyzed image in the main paper. **Number of filaments,** total number of filaments in the whole-gill section. **Number of filaments with patch,** number of filaments with had at least one MDH^+ patch. **Percentage of filament with patch (%)**, percentage of filaments carrying MDH^+ patches.

Image	Number of filaments	Number of filaments with patch	Percentage filaments with patch (%)
Figure 1, a	96	55	57,29

Supplementary table 3. Around 23% of the filaments of a whole-gill section that had a MDH$^+$ patch were facing another filament with at least one MDH$^+$ patche. Image, analyzed image in the main paper. **Number of filaments,** total number of filaments in the whole-gill section. **Number of filaments with facing patches**, number of filaments that carrying at least one MDH$^+$ patch, they were facing another filament with at least one MDH$^+$ patch. **Percentage of filaments with facing patches**, percentage of filaments with MDH$^+$ patches that were facing each other.

Image	Number of filaments	Number of filaments with facing patches	Percentage of filaments with facing patches (%)
Figure 1, a	96	22	22,92

Supplementary table 4. Approximately 6% of the SOX population carried the MDH-gene cluster. Image, analyzed image in the main paper. **Area SOX pixels (µm2),** area of the image occupied by 16S rRNA SOX fluorescent signals. **Area MDH pixels (µm2),** area of the image occupied by MDH-gene cluster fluorescent signals. MDH-SOX-correlated area, ratio of MDH signals co-localizing with SOX 16S rRNA signals expressed as percentage.

Image	Area SOX pixels (µm2)	Area MDH pixels (µm2)	MDH-SOX-correlated area (%)
Figure 1, a	400054,169	23590,963	5,90

Supplementary table 5. Detailed information about primers used to amplify the 12 and 10 polynucleotide probes targeting the MDH- and the HYD-gene clusters, respectively. Primer, name of the primer pair. **Target gene,** target gene within its respective gene cluster. **Forward,** sequence of the forward primer. **Reverse**, sequence of the reverse primer. **Product size**, size of the amplicon (bp). **Temp**, Tm of the primer pair (°C)

Primer	Target gene	Forward (5'-3')	Reverse (5'-3')	Product size (bp)	Temp (°C)
MDH_cluster	MDH gene cluster	TACGCCCCTTTCCAAGTGAC	ATTTTGGTGGCTGTCGTTGC	8334	64.6
PP2	Conserved protein	GCTGCGAGGAAAATCCAGTAC	GCCCTAAATCAACACCACGC	299	55
PP3	S-(hydroxymethyl) glutathione dehydrogenase	GCGGTTGGGAAGCAACTTCT	CCTTCGTGACCTAACACACA	299	55
PP4		TGTGGTCCAGGCGTTAAAGA	ACCACAACCCAGCAAGCATA	299	55
PP5		CGGGTCAACGGTTGCAGTAT	ACTGTAGAAACACCCCAGCC	299	55
PP6	S-formylglutathione hydrolase	GGGCGTACAGAGTTACCAGG	TTCTGTGCTTGAGGTGGTAGG	299	55
PP7		ATGTTAGAACCAAGGCAGGCA	GGCACAAATCAACGCACCAT	298	55
PP8		CTCCGATTTGCAATCCTACACA	CCATATGTTCACCTATAAACGATGCA	299	55
MDHPP1	*xoxF*	TTGTCCAGCAGCGTTAGGTT	ATAGTGTCTTCACCGCCTGC	268	55
MDHPP2		GGTTTTGCCTTGCTACTATTGTTG	AAAGCACCACTCCAAACGGA	257	55
MDHPP3		GGGAATATAAACCAATACAAGACCCG	TGTTAGGGAATGGCGTGTGG	260	55
MDHPP4		ACGGTTACTCACAAGGTCCAG	ACCATACTCACCGCCAGAAAT	268	55
MDHPP5		ATCAAATGACCCCGCACGAT	TCGTGCCCAAATCGTCATAGA	270	55

Primer	Target gene	Forward (5'-3')	Reverse (5'-3')	Product size (bp)	Temp (°C)
hydrogenase-gene cluster	hydrogenase-gene cluster	TGTAATGCGAAGACAGGGGA	TGCGACCATCCACAAAGACA	4667	67.0
HYD-PP1	Uptake hydrogenase small subunit precursor	GACAGGGGATTAGTCGTCGT	ACCATCGTATTTTTCAATAGTCTCATC	305	72.0
HYD-PP2		TACACCTCTTAATCAAGACGGC	GGACGGCCTTGTCTGTCTAA	300	72.0
HYD-PP3		ATCGTCGTCCACATTTTGATGC	CTGTAGCACCAGCAACAGCA	300	72.0
HYD-PP4	Uptake hydrogenase large subunit	TGCCAAGACCAATCAAGCAG	TGCTCTACAAGATGCCAGCG	310	72.0
HYD-PP5		TGCGCATCTTATTCGTGAAATG	CGCCATTAAATCTGCTTCGGG	304	72.0
HYD-PP6		CCTTGGACTTGCAAAAAGAGTT	TAGGATAAGCGCCGTAATCCA	300	72.0
HYD-PP7		CCAGGTGGCGTTATTCTTAATGA	TCCTTGTGCGATTGCTAAAACA	309	72.0
HYD-PP8		GGCACTTTCTGGATTTAACGC	ACACTGGTAATTGTCAATTTTTCCA	307	72.0
HYD-PP9	Hypothetical protein	GCCAATGAGCCGGTTGAAAT	GCGCTTACTATCAACACCGC	304	72.0
HYD-PP12	Hydrogenase 1 maturation protease HyaD	TGAAGTTAAATTAATGGATGGTGGCA	TTCATCACGCAAACTACCGC	301	72.0

Chapter V | Preliminary results, future directions and conclusions

Intranuclear lifestyle is the most sophisticated strategy in bacteria that invade specific organelles within their host cells (Schulz and Horn 2015). Since the discovery of “enigmatic particles” in the nuclei of paramecia in the 19th century, this singular life strategy has captivated the scientific community. Yet, our knowledge about the molecular biology and invasion strategies of intranuclear bacteria is limited. The intranuclear lifestyle has potential advantages: The nucleus has been described as a nutrient-rich organelle, where bacteria has access to nitrogen, carbon and phosphorous in form of chromatin. In addition, intranuclear bacteria are safely hidden from cytoplasmic defense mechanisms (Schulz and Horn 2015), and have the possibility to interfere with nuclear processes by using nucleomodulins (Bierne and Cossart 2012). *Ca.* Endonucleobacter spp. infect the nuclei of bathymodiolin mussels, where they can reach population numbers of up to 80,000 bacteria, increasing the volume of the nucleus up to 50 fold (Zielinski et al. 2009). These authors hypothesized that *Ca.* E. bathymodioli might digest DNA, as they observed disappearance of host heterochromatin towards the onset of bacterial replication. However, the digestion of host genetic material would have dramatic consequences for the viability of the host cell as replication niche (Schulz and Horn 2015). Moreover, the use of the nucleus as replication niche carries risks for bacteria, as nuclear deformation could be reported by eukaryotic cells. Zhang et al. 2009 demonstrated that in mouse cells, nuclear deformation is transmitted to the cell cytoskeleton via nesprin-1. One possibility is that this reporting mechanism of

nuclear deformation is conserved among eukaryotes. So in theory, bathymodiolin mussel cells infected by *Ca.* Endonucleobacter can report nuclear deformation, and react to it activating the caspase-mediated apoptotic cascade (Crawford et al. 2012; Kräter et al. 2018). *Ca.* Endonucleobacter spp. might be accessing alternative host resources for its nutrition rather than chromatin, and it might be using molecular strategies to avoid the shutdown of the host cell. During my studies, I developed a pipeline to microdissect single infected nuclei by *Ca.* E. chitolyticus at different infection stages for downstream ultra-low-input RNAseq. In addition, I produced high-quality genomes for *Ca.* E. chitolyticus and *Ca.* E. bathymodioli using a hybrid assembly strategy using short (Illumina) and long reads (PacBio). This allowed me to characterize the transcriptomic profile of *Ca.* E. chitolyticus and its host cell along the infectious cycle, and to compare the differences in genomic potential of both *Ca.* Endonucleobacter species. My results suggested that *Ca.* E. chitolyticus does not use host chromatin as a nutritional source. Instead, it expresses chitinase and the enzymatic machinery to import and metabolize chitosaccharides. Visualization of chitin in the gill filaments of *"B." childressi* (**Supp. Fig.6 in chapter II**) suggested that this polymer occurs in the extracellular medium. It is interesting to think that *Ca.* E. chitolyticus might be profiting on an extracellular matrix component while sitting in the nucleus of its host cell. The host cell of *Ca.* E. chitolyticus remained transcriptionally active during the whole infectious cycle, and reacted to the infection upregulating sugar transport, glycolysis, lipid droplets synthesis and sensors of nuclear deformation. My results have challenged current hypotheses about intranuclear lifestyle, which have traditionally contemplated the nucleus as a merely nutrient-rich environment from which intranuclear bacteria profit from. The results produced during my

doctoral studies will help to understand how intranuclear bacteria can access alternative nutritional host resources rather than chromatin, even when they occur extracellularly. In addition, my PhD research sheds light on the strategies that intranuclear bacteria might be using to avoid the collapse of their host cell, being interfering with apoptosis or with cytoskeleton components. The strategies displayed by *Ca.* E. chitolyticus to colonize its host cell and to settle and thrive in the nucleus might be part of one of the multiple evolutionary routes that intranuclear bacteria have developed in an event of convergent evolution.

5.1 Molecular biology of *Ca.* Endonucleobacter

Investigating the molecular mechanisms involved in nuclear colonization and exploitation poses an immense challenge when intranuclear bacteria cannot be cultured in the laboratory. This was the first limitation that I had to face when investigating the molecular biology of *Ca.* Endonucleobacter spp. *Ca.* E. chitolyticus is limited to the ciliated edge of gill filaments of *"B." childressi* (**Fig.2, c & d in chapter II**). This allowed me to access to high amounts of bacterial biomass in single sectioning planes, and the posterior application of cultivation-independent technologies to investigate its molecular biology. Laser-capture microdissection (LCM) allows the *in situ* separation of subcellular subsets in environmental samples. I developed a pipeline in which I localized infected nuclei using fluorescence *in situ* hybridization (FISH) targeting *Ca.* Endonucleobacter. Nuclei at different infection stages were microdissected using LCM and sequenced using the CATS ultra-low-input RNA-sequencing protocol (Turchinovich et al. 2014). The workflow that I developed revealed that *Ca.* E. chitolyticus does not thrive on host chromatin, but rather access alternative nutritional sources such as chitin and lipids. Moreover, my pipeline demonstrated

that the host cell was transcriptionally active during the infectious cycle, upregulating sugar import, glycolysis, lipid droplets synthesis and sensors of nuclear deformation. Our FISH-LCM-ultra-low-input RNAseq pipeline allowed us to characterize the transcriptomic profile of an intranuclear parasite-host cell interaction along the infection cycle in environmental samples. This pipeline made possible to identify changes in the expression of *Ca.* Endonucleobacter and its host cell in every stage of the infectious cycle, giving us temporal information. For example, the T3SS and the inositol phosphatase *IpgD* are factors typically expressed by *S. flexneri* during the early stages of cell invasion (Niebuhr et al. 2000, 2002). By applying LCM and studying the transcriptome of *Ca.* E. chitolyticus during mid stage of infection, I could confirm that *Ca.* E. chitolyticus expressed both factors while residing in the nucleus (**Fig.4 in chapter II**). Thus, our combined FISH-LCM-ultra-low-input RNAseq pipeline can resolve the expression profiles of microbiome-host interactions with spatial and temporal resolution. The possible applications of our methodological proof of concept are vast. The FISH-LCM-ultra-low-input RNAseq pipeline could be applied to tease apart the expression profile of spatially organized bacterial populations contained in symbiotic organs of diverse metazoans, from the trophosome of *Riftia pachyptila* to the foregut organ of leaf-eater beetles or even the gut of humans (Hinzke et al. 2019; Salem et al. 2017; Thursby and Juge 2017). Moreover, it could help to deepen the understanding of the molecular interactions between intracellular pathogens such as *Salmonella*, *Shigella*, *Listeria* and *Yersinia* and their host cells, with great importance in medical microbiology. Our FISH-LCM-ultra-low-input RNAseq pipeline could also help to tease apart host-microbiome interactions in plants. To deepen the understanding of the relationship of plants

from the family *Fabaceae* and the nodule-forming nitrogen-fixing bacteria from the genus *Rhizobium* (Liu et al. 2018) or the interactions between plants and phytopathogenic bacteria and fungi would have a great significance in agriculture (Chen and Nan 2015). In addition, our pipeline could also characterize the transcriptomic profile of symbiotic interactions that take place during short windows of time, such as the life cycle progression of *Plasmodium falciparum* or the continuous horizontal symbiont acquisition that takes place in the budding zone of bathymodiolin mussels (Phillips et al. 2017; Wentrup et al. 2014).

5.1.1 "We are in. Now, whatever you do, don´t push the red button."

The colonization of the non-phagocytic cells from the ciliated edge of the gill filaments of *"B." childressi* probably starts with the physical adhesion of *Ca.* E. chitolyticus and its target cell. *Ca.* E. chitolyticus encoded and expressed a RTX-adhesin (**Fig.3, chapter II**), which most likely aids the parasite during adhesion and cohesion to the host cell (Guo et al. 2019). However, the final executioner that promotes the uptake of the parasite by its target cell might be the T3SS-*IpgD* tandem. *Ca.* E. chitolyticus might display an analogous colonization strategy than *S. flexneri*, disentangling the cortical actin cytoskeleton of its target cell and promoting the formation of membrane ruffles and its posterior uptake (Niebuhr et al. 2000, 2002). The mechanisms displayed by *Ca.* E. chitolyticus to avoid lysosomal digestion remain elusive: I did not identify any of the previously described mechanisms for nuclear invasion in *Ca.* E. chitolyticus genome (e.g. genes involved in the formation of invasion tip of *Holospora* spp.). Consequently, I contemplate the possibility that the *Ca.* E. chitolyticus-containing vesicle is being translocated to the nucleus, where it can fuse with the outer nuclear membrane. *Ca.* E. chitolyticus would then be released in the perinuclear space to posteriorly

enter the nucleus through invagination of the inner nuclear membrane and posterior shedding of the host membrane. This mechanism of nuclear invasion has been proposed for intracellular bacteria in *Euglena* spp. (**Fig.2, B in chapter I**) (Shin et al. 2003).

Zielinski et al., 2009 hypothesized that *Ca.* E. bathymodioli could be digesting host chromatin, as the authors observed a decrease of heterochromatin towards the onset of the parasite replication. Once inside the nucleus, one would expect that *Ca.* E. chitolyticus starts digesting the available chromatin, as suggested for *Ca.* E. bathymodioli. However, analysis of ultrastructure images of nuclei infected by *Ca.* E. chitolyticus revealed that heterochromatin did not disappear, but rather was pushed against the inner nuclear membrane (**Suppl. Fig.3 in chapter II**). Moreover, the same ultrastructure data suggested that mitochondria were fully functional towards the onset of parasite replication. In addition, *Ca.* E. chitolyticus did not codify for nucleotide importers. Further, *Ca.* E. chitolyticus expressed all the routes for de novo synthesis of nucleotides (**Supp. table 5 in chapter II**). Next, *Ca.* E. chitolyticus did not expressed proteases neither DNAses, and barely expressed its competence system for DNA acquisition (**Fig.3 in chapter II**). Moreover, progression of the infectious cycle did not have any effect on fluorescence intensity when localizing host cell 18S rRNA using FISH, suggesting that the host cell was transcriptionally active during infection (**Suppl. Fig.2 in chapter II**). This was confirmed by analyzing the transcriptome of the host cell along the life cycle (**Fig.4 in chapter II**). All these results described in **chapter II** suggested that *Ca.* E. chitolyticus does not thrives on host chromatin, but rather access alternative host resources. Why would *Ca.* E. chitolyticus ignore an immediately available source of nitrogen, carbon and phosphorous? The nuclear

genetic material is responsible for directing the functions and physiology of the host cell. In the worst case scenario, chromatin digestion by *Ca.* E. chitolyticus would lead to a collapse of the host cell. In the best case scenario, it would lead to a massive deregulation of the host cell transcription. Zielinski et al. 2009 proposed an alternative theory to the chromatin consumption hypothesis: *Ca.* E. bathymodioli could rely on nutrients that transit between the nucleus and the host cell cytoplasm. This idea has its foundation in the nuclear pore complexes, channels for bidirectional trafficking between the nucleus and the cytoplasm that allow the free transit of molecules under 30-60 kDa (Keminer and Peters 1999; Ma et al. 2012; Mohr et al. 2009; Ribbeck and Görlich 2001). My localization data of *Ca.* E. bathymodioli revealed that chromatin was being pushed against the inner nuclear membrane in late stage of infection (**Fig.1**), as ultrastructure data previously reported for Ca. E. chitolyticus (**Suppl. Fig.3 in chapter II**). This suggested that the chromatin-consumption hypothesis might not apply for any of the species that form the genus *Ca.* Endonucleobacter. It is tempting to speculate that *Ca.* Endonucleobacter spp. have developed a life strategy based on keeping its host cell alive as long as possible to complete its life cycle. While located in the nucleus, *Ca.* Endonucleobacter spp. might avoid chromatin consumption and profit on nutrients present in the host cell cytoplasm. In a metaphorical sense, *Ca.* Endonucleobacter spp. do not “push the red button” of chromatin digestion, which would lead to a collapse of the replication niche. If that holds true, the life strategy of *Ca.* Endonucleobacter spp. might not differ dramatically from the nutrition strategies of other intranuclear bacteria. As commented above, intranuclear rickettsiae, *C. caryophilus* and *Holospora* spp. rely on nucleotide transporters pirate the energy budget of the host cell (Haferkamp et al. 2006; Schmitz-esser

et al. 2004). Although this implies a metabolic burden, it is unlikely intranuclear energy parasites are fatal for the host cell. This brings the interesting possibility that intranuclear lifestyle in bacteria might be linked to nutritional strategies that exploit the nucleus in a non-destructive way. When studying intranuclear life style in bacteria, the nucleus has been perceived merely as a nutrient-rich compartment from which bacteria can nutritionally profit from by digesting chromatin. My studies suggest that the nucleus might be used by intranuclear bacteria as a protective shelter opened to the cytoplasm through the nuclear pore complexes rather than as a hermetic pantry full of nutrients to exploit.

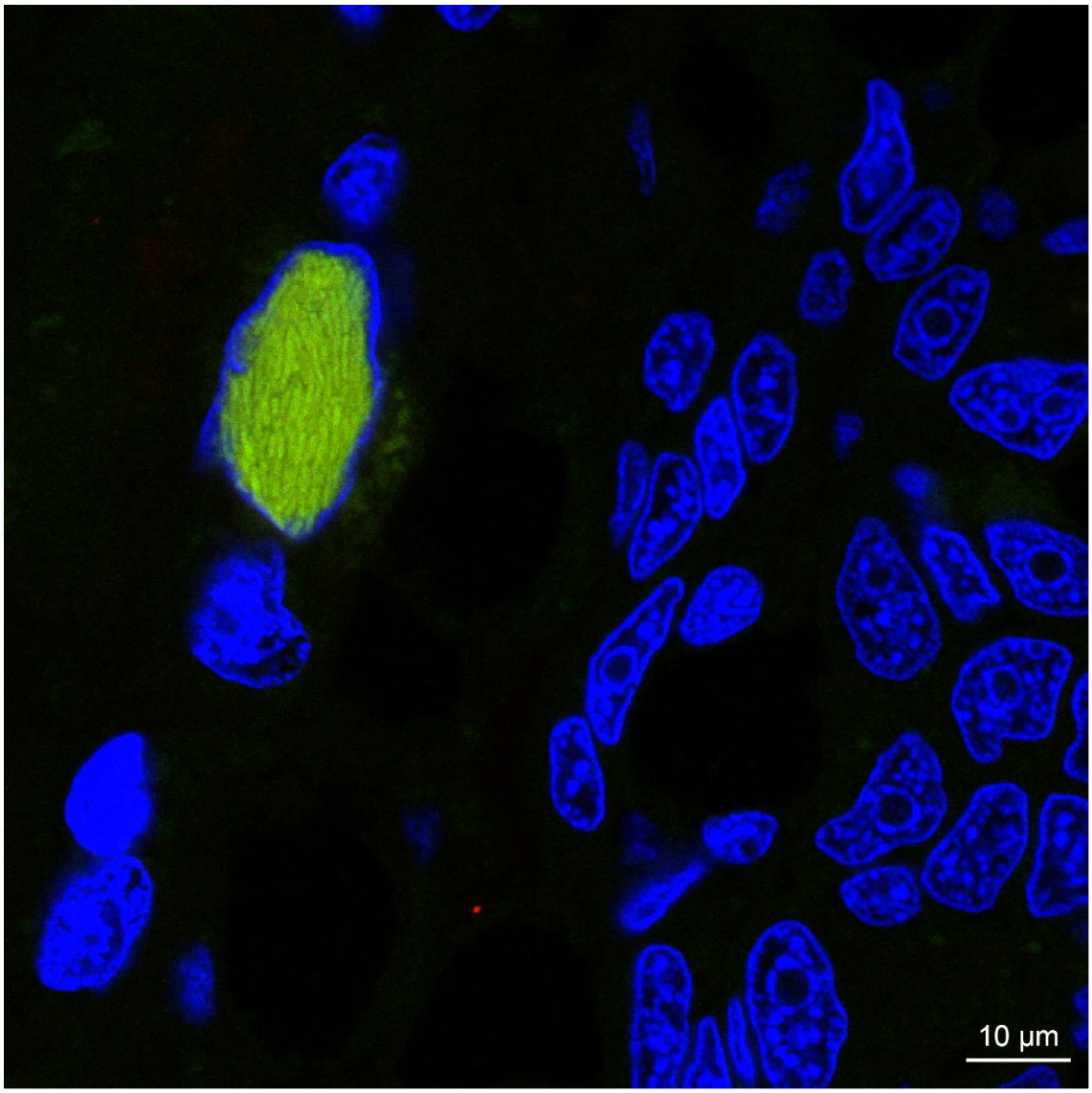


Figure 1. Chromatin was pushed against the inner nuclear membrane in late stage of infection by *Ca.* E. bathymodioli, suggesting that the chromatin-consumption hypothesis is invalided for the whole *Ca.* Endonucleobacter genus. Whole-mount FISH on gill filament of *B. puteoserpentis*. CSLM image colocalizing *Ca.* Endonucleobacter bathymodioli 16S rRNA (yellow: atto550) and DNA (blue: DAPI).

5.1.2 Beyond the nucleus

Ca. E. chitolyticus prevents the disruption of its replication niche by avoiding chromatin consumption. Sitting in the nucleus, *Ca.* E. chitolyticus consumes the host energy budget that freely diffuses from the cytoplasm to a nuclear compartment that acts as a "nutritional black hole". From a parasite's perspective, this seems like the perfect life strategy: *Ca.* E. chitolyticus uses the nucleus as a shelter opened to the influx of nutrients from the cytoplasm of a host cell that is in permanent starvation state. However, the energy budget of the host cell is finite, and intranuclear bacteria might compromise the viability of the replication niche in case of metabolic exhaustion of the host cell. Intranuclear bacteria infecting unicellular eukaryotes take a risk, as the host cell must theoretically cope by itself with their metabolic demands by either increasing import of extracellular nutrients (e.g. *Hartmannella* infected by *Ca.* Nucleicultrix amoebiphila) or increasing carbon fixation through photosynthesis (intranuclear bacteria in *Euglena* spp.) (Schulz et al. 2014; Shin et al. 2003). The inability of the host cell to answer to these metabolic demands would result in a premature collapse of the replication niche, compromising the completion of the intranuclear bacteria life cycle. However, intranuclear bacteria infecting metazoans (such as *Ca*. Endonucleobacter) play safe: They rely on the fact that animal cells occur in a tissular context, and that the host cell would get provided with nutrients present in the extracellular matrix in case of starvation. In this sense, intranuclear bacteria use their host cells as funneling agents that create a constant flow of nutrients between the extracellular matrix and the nucleus. In a metaphorical sense, one could envision the host cell of an intranuclear bacterium as the living funnel that creates a flow of nutrients between the environment and the parasite.

Transcriptomic and proteomic analyses revealed that chitinase was one of the most expressed factors by *Ca.* E. chitolyticus (**Fig.3 & 4 in chapter II**). When localizing chitin in the gill filaments of *"B." childressi*, I could demonstrate that chitin is contained in vesicles within secretory cells that occur in the symbiotic region-ciliated edge interface (**Supp. Fig.6 in chapter II**). The nature and polarity of these secretory cells suggested that chitin is a component of the extracellular matrix of the gill filaments in *"B." childressi*. Thus, *Ca.* E. chitolyticus is doing a considerable metabolic investment in digesting a nutritional source from which it is separated by three cellular membranes. These discoveries imply that *Ca.* E. chitolyticus relies on a nutritional source that is out of its physical grasp. How could an intranuclear parasite digest and access extracellular matrix components? My transcriptomic analyses demonstrated that *Ca.* E. chitolyticus expressed the T3SS while sitting in the nucleus (**Fig.4 in chapter II**). It is tempting to speculate that *Ca.* E. chitolyticus is using the T3SS to pierce the inner nuclear membrane and deliver chitinase to the host secretory pathway. Eventually, chitinase-containing vesicles would reach the host cell plasma membrane, deliver their cargo and digest extracellular chitin. Then, a starving host cell with a growing *Ca.* E. chitolyticus population within its nucleus would eagerly import the chitobiose saccharides resulting from extracellular chitin digestion, trying to satiate the permanent starvation state induced by the parasite. The implication of these hypotheses shake the foundations of research in intranuclear parasitism. In theory, *Ca.* E. chitolyticus is capable to access a nutritional source that occurs not only beyond the nuclear compartment, but outside the host cell. More remarkably, *Ca.* E. chitolyticus could be hijacking the host secretory pathway and altering the host cell metabolic status to nutritionally benefit on extracellular

components. This opens the interesting possibility that bringing the host cell to a starvation state is a common manipulation strategy in intranuclear bacteria. A starving eukaryotic cell would try to compensate the metabolic demands of intranuclear bacteria, increasing the uptake of nutrients from the extracellular medium that would eventually reach the nucleus through the nuclear pore complexes.

5.1.3 Host cell manipulation

Ca. E. chitolyticus hijacking of the host secretory system is one of the strategies to manipulate the host cell, but not the only one. When proliferating, intranuclear bacteria can exert mechanical distress to their host cell by deforming the nucleus. Nuclear deformation is transmitted to the host cell cytoskeleton via nesprin-1 (Zhang et al. 2009), which can trigger the caspase-mediated apoptotic cascade (Crawford et al. 2012; Kräter et al. 2018). Thus, eukaryotic cells are capable of responding to nuclear invasion. Still, intranuclear bacteria are able to complete their life cycles, often with dramatic consequences for the host cell. *Ca.* Endonucleobacter spp. can reach numbers of 80,000 per nuclei, and increase the size of an infected nucleus up to 50 fold (Zielinski et al. 2009). This raised the suspicion that *Ca.* Endonucleobacter spp. might count on molecular mechanisms to prevent the host cell to enter in apoptosis. My comparative genomic analyses revealed that *Ca.* E. bathymodioli and *Ca.* E. chitolyticus encoded for 26 and 10 inhibitors of apoptosis (IAPs), respectively (**Fig.1 in chapter III**). IAPs are physiologic caspase inhibitors that are able to arrest the apoptotic cascade (Deveraux et al. 1997; Deveraux and Reed 1999). Our transcriptomic profiling of *Ca.* E. chitolyticus life cycle demonstrated that IAPs are expressed along the life cycle **(Fig.4 in chapter II)**, suggesting that *Ca.* E. chitolyticus is actively arresting

host cell apoptosis. IAPs were virtually absent in *Endozoicomonas* spp., the sister clade of Ca. Endonucleobacter spp. (**Fig.1 in chapter III**). This led me to the conclusion that IAPs might be one of the genomic innovations that played a role in the evolutionary origin of *Ca.* Endonucleobacter spp. Moreover, IAPs might be the biological signature of the genus. To date, IAPs have not been identified in any other clade of intranuclear bacteria. However, most intranuclear bacteria exert mechanical distress to their host cells while replicating in the nucleus. Therefore, the possibility that IAPs or analogous mechanisms to prevent host cell apoptosis are a common feature of intranuclear bacteria cannot be discarded.

Although *Ca.* Endonucleobacter spp. can arrest host apoptosis by interfering with the caspases cascade and they do not consume host chromatin, it is still surprising that the host cell functioning is not compromised. The mechanical distress that *Ca.* Endonucleobacter spp. exerts to the host cell by increasing its volume up to 50 times fold should compromise its physiology and yet, the host cell remained transcriptionally active during the whole infectious cycle **(Fig.4 in chapter II)**. *Ca.* E. chitolyticus expressed *IpgD* while residing in the nucleus of its host cell, suggesting that *IpgD* might play additional roles in *Ca.* E. chitolyticus biology rather than aiding the parasite during host cell colonization **(Fig.4 in chapter II)**. *IpgD* increases the intracellular levels of phosphatidylinositol 5-phosphate, a molecule that participates in signaling events that control cytoskeleton dynamics. Niebuhr et al. 2002 demonstrated that in HeLa cells transfected with *IpgD*, increasing levels of phosphatidylinositol 5-phosphate lead to disappearance of actin stress fibers, decreasing cell stiffness. It is interesting to think that *Ca.* E. chitolyticus, in an exercise of microniche engineering, is reducing the stiffness of its host cell by expressing *IpgD* while sitting in the

nucleus. The manipulation of the host cell cytoskeleton is an unreported phenomenon in intranuclear bacteria, although is a common strategy during host cell colonization in several intracellular pathogens such as *S. flexneri* or *Listeria monocytogenes* (Dramsi and Cossart 1998; Niebuhr et al. 2000, 2002). To date, *IpgD* has not been found in any other intranuclear parasite. However, the presence of analogous mechanisms that allow intranuclear bacteria to expand the nuclear compartment without prematurely collapsing the host cell through cytoskeleton manipulation should be contemplated.

5.2 How does the host react to the infection?

One of the prevalent questions when studying intranuclear bacteria is how the host cell reacts to the infection. Analysis of the host response does not only allow to investigate the defense strategies displayed by the eukaryotic cell during bacterial invasion and proliferation, but also a better understanding of the molecular arsenal used by intranuclear bacteria. The transcriptomic analysis of nuclei at different stages of infection allowed me to study how the host cell reacted to the infection by *Ca.* E. chitolyticus **(Fig.4 in chapter II)**. One of the metabolic processes that was upregulated after *Ca.* E. chitolyticus infection by the host cell was lipid droplets (LD) synthesis. LD synthesis is a common phenomenon in cells that are compromised by the presence of intracellular parasites. This is because intracellular parasites commonly consume the energy budget of their host cell, bringing it to starvation state (Henne et al. 2018). For example, cells infected by *Toxoplasma gondii* progressively increased LD synthesis until the onset of parasite replication (Nolan et al. 2017). As pointed out above, the nuclear compartment communicates with the cytoplasm through the nuclear pore complexes, which allow free diffusion of molecules under 30-60 kDa.

One could imagine the possibility that in the early stages of infection, *Ca.* E. chitolyticus consumes the host energy budget by expressing amino acids, polypeptides, lipids and sugar transporters **(Fig.4 in chapter II)**. Bringing the host cell to starvation state by consuming its energy budget would have two major consequences. First, the host cell would interpret the lack of nutrients as an environmental cue of nutritional shortage, and it would respond to it by increasing the synthesis of storage structures such as LD (Henne et al. 2018). Second, the host cell would increase the expression of transporters, trying to uptake as many nutrients as possible from the extracellular medium (Efeyan et al. 2015). I observed an upregulation of SWEET sugar transporters by the host cell in mid stage of infection **(Fig.4 in chapter II)**. This suggested that the host cell was trying to compensate the nutritional demands of *Ca.* E. chitolyticus by upregulating sugar uptake. In summary, *Ca.* E. chitolyticus passively instigates the host cell to increase the uptake of nutrients from the extracellular medium. By doing so, *Ca.* E. chitolyticus ensures that the nuclear compartment is continuously provided with nutrients through the nuclear pore complexes. This allows the parasite to growth without the necessity to digest host chromatin and most importantly, without compromising the long-term viability of the replication niche. The capacity to modulate the metabolism of the host cell plays in favor of *Ca.* E. chitolyticus, and it is interesting to think that it might be a common mechanism among intranuclear bacteria. As we observed high expression of lipases during *Ca.* E. chitolyticus life cycle, we also hypothesize that the parasite might be nutritionally profiting on host LD **(Fig.4 in chapter II)**. The fact that we never localized LD might indicate that the parasite is consuming LD along their

formation, which would not allow their accumulation and localization by microscopy techniques (**Supp. Fig.7 in chapter II**).

Another way in which the host cell reacted to *Ca.* E. chitolyticus infection was upregulating the mechanical sensor for nuclear deformation nesprin-1, as well as caspase 2 **(Fig.4 in chapter II)**. As mentioned above, nesprin-1 can transmit nuclear deformation information to the cell cytoskeleton (Zhang et al. 2009), which can ultimately trigger the caspase-mediated apoptotic cascade (Crawford et al. 2012; Kräter et al. 2018). This raised the suspicion that the host cell is aware of *Ca.* E. chitolyticus infection. Moreover, the host cell is trying to respond to the infection by activating apoptosis. One plausible hypothesis is that along evolutionary time, *Ca.* Endonucleobacter spp. and bathymodiolin mussels have established an arms race in which the parasite tries to interfere with apoptosis by expressing IAPs, while the host cell tries to detect and stop infection spreading by activating apoptosis as early as possible. Another possibility is that bathymodiolin mussels are not being positively selected to fight against *Ca.* Endonucleobacter spp. infection. This hypothesis is based on the fact that *Ca.* Endonucleobacter spp. only infects non-symbiotic cells from the gill filaments, which are minority within gill filaments (**Fig.2**). The potential protective role of the mutualistic symbionts of bathymodiolin mussels would avoid *Ca.* Endonucleobacter spp. to be fatal for the animal, as the parasite could only burst a limited fraction of the gills cells.

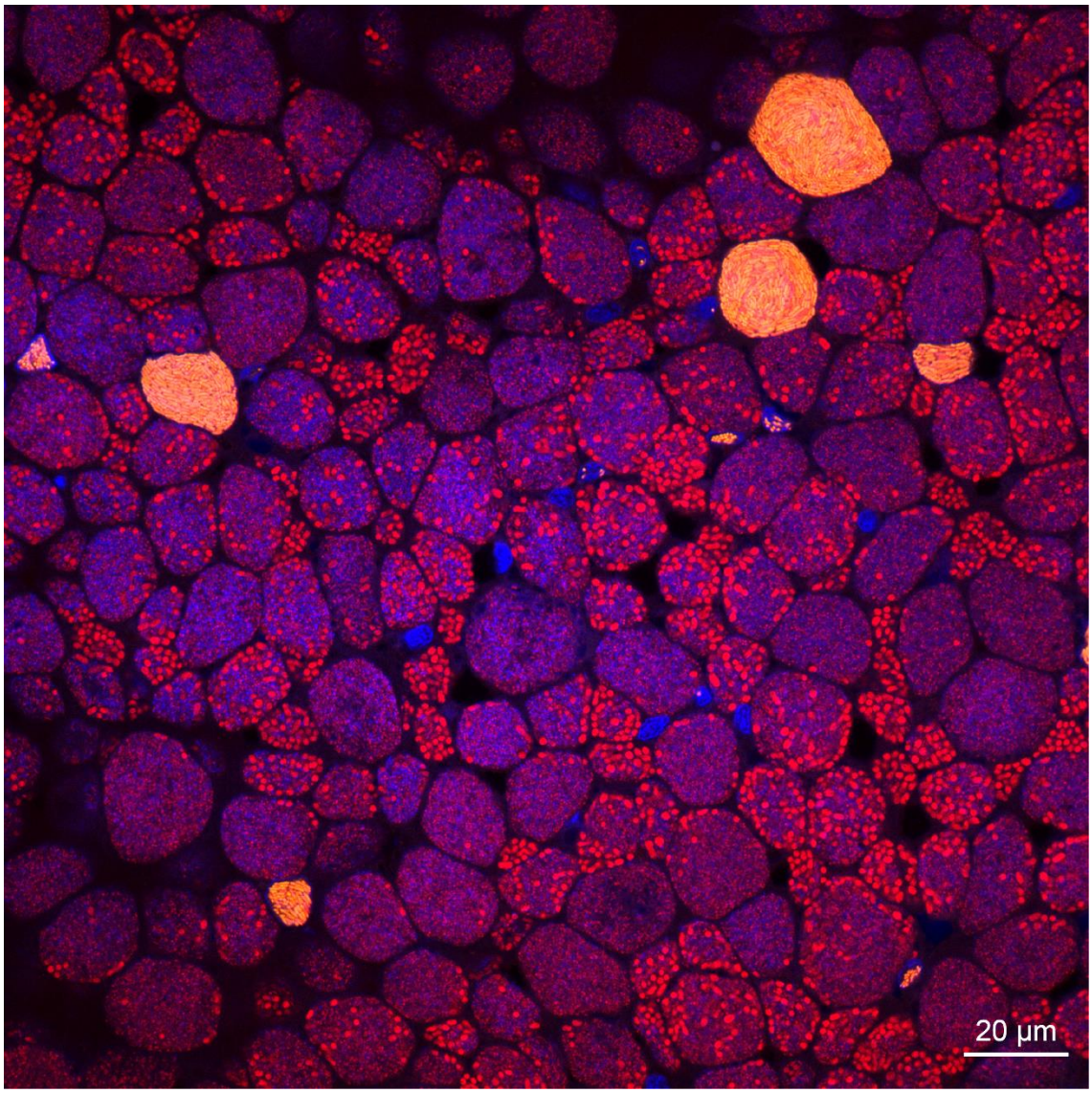


Figure 2. The mutualistic symbionts of bathymodiolin mussels confer protection to the animal, as the intranuclear parasite cannot infect cells already occupied by mutualists. Whole-mount FISH on gill filament of *B. puteoserpentis*. CSLM image colocalizing Eubacteria (red: Alexa 647), *Ca.* Endonucleobacter bathymodioli 16S rRNA (yellow: atto550) and DNA (blue: DAPI).

5.3 The black sheep of the (*Hahellaceae*) family

Ca. Endonucleobacter spp. clustered as the sister clade of *Endozoicomonas* spp. (**Fig.1 in chapter III**), a genus known to establish mutualistic relationships with a great diversity of metazoans (Fiore et al. 2015; Forget and Kim Juniper 2013; Jensen et al. 2010; Katharios et al. 2015; Morrow et al. 2012). Together with the genus *Kistimonas*, *Ca.* Endonucleobacter and *Endozoicomonas* form a monophyletic group within the family *Hahellaceae* that establish symbiotic relationship with marine metazoans (**Fig.8 in chapter I**) (Choi et al. 2010; Schreiber et al. 2016). My studies suggested that *Ca.* Endonucleobacter is a monophyletic group of obligated intranuclear parasites (**chapter III**), while the majority of *Endozoicomonas* spp. have been described to be facultative mutualists with a free-living stage (Neave et al. 2017). The evolutionary events that lead to the divergence of the genera *Ca.* Endonucleobacter (parasites) and *Endozoicomonas* (mutualists) are uncertain. Neave et al., 2017 found that *Endozoicomonas* genomes were enriched in mobile elements, which might have played a role in the adaptation of *Endozoicomonas* spp. to different environments, hosts and symbiotic lifestyles. When analyzing the genomes of *Ca.* E. bathymodioli and *Ca.* E. chitolyticus, I discovered that more than 20% of their genes were assigned to the COG category “replication, recombination and repair”, which includes mobile elements (**Fig.2 in chapter III**). It could be that mobile elements were already present in the last common ancestor of *Ca.* Endonucleobacter and *Endozoicomonas* spp., and that they played a role in the transition from a mutualistic to a parasitic lifestyle. It has been hypothesized that mobile elements might have originated parasitic lifestyle in the only pathogen known within the genus *Endozoicomonas* (*Ca.* Endozoicomonas cretensis),

which infects fish (Katharios et al. 2015; Qi et al. 2018). One possible explanation of *Ca.* Endozoicomonas cretensis being the only pathogen of its genus is that *Endozoicomonas* spp. might be more widespread than initially thought, and the clade might have been undersampled. An alternative hypothesis is that *Ca.* Endozoicomonas cretensis might be a mutualist of a non-yet identified marine invertebrate, and it behaves facultatively as a pathogen when colonizing fish. This would imply that the genus *Endozoicomonas* is exclusively formed by mutualists, and eventually some of the genus representatives can behave as parasites when occurring in hosts with which they do not share a long coevolutionary history (Douglas 2010). The flexibility of *Endozoicomonas* spp. in terms of hosts contrast with what we have hypothesized for *Ca.* Endonucleobacter spp., which might be obligated parasites that exclusively infect bathymodiolin mussels (**chapter III**). Expansion of mobile elements and genome reduction are frequently linked to host restriction and appearance of pathogenicity in symbiotic bacteria (Holden et al. 2009; Parkhill et al. 2003). I discovered that *Ca.* Endonucleobacter spp. had reduced genomes in comparison with their *Endozoicomonas* spp. relatives (**Fig.1 in chapter III**). In addition, *Ca.* Endonucleobacter spp. did not encode for more than 10 routes for *de novo* synthesis of amino acids, suggesting loss of metabolic autonomy (**Fig.1 in chapter III**). The mentioned data suggested that *Ca.* Endonucleobacter spp. might have segregated evolutionarily from *Endozoicomonas* spp. due to mobile elements expansion, and adopted a highly specialized lifestyle in which they rely on their host for amino acid acquisition. *Ca.* Endonucleobacter spp. and *Endozoicomonas* spp. cluster as sister clades within the *Hahellaceae* family and yet, they are in the antipodes of the symbiotic spectrum (parasitism vs. mutualism, respectively). This highlights the flexibility of

the symbiotic representatives of the family *Hahellaceae* when switching between hosts and symbiotic styles. The genomic features that might have originated intranuclear lifestyle in the *Hahellaceae* family could have as well operated in other clades of intranuclear bacteria. The Rickettsia endosymbiont of *Ixodes scapularis* devoted more than 35% of its genome to mobile elements, highlighting its genomic plasticity (Gillespie et al. 2012). All members of *Rickettsiales* are obligated intracellular parasites of a wide range of eukaryotic hosts (Driscoll et al. 2013), but only a minority have conquered the nuclei of their host cells (Schulz and Horn 2015). The cited literature invites to speculate that mobile elements might play a role in the evolutionary origin of intranuclear lifestyle in diverse bacterial phyla. This opens the interesting possibility that genomic enrichment with mobile elements might be the experimentation preamble with which intracellular bacteria acquire intranuclear (or other intraorganellar) lifestyle.

5.4 Symbiont compartmentalization in bathymodiolin mussels

The spatial organization of different SOX functional strains in the gill filaments of bathymodiolin mussels was also a subject of research during my doctoral studies. Using metagenomic approaches, Ansorge et al. 2019 demonstrated that up to 16 SOX functional strains can coexist in a single mussel. Although these numbers shed light over the extend of strain diversity in bathymodiolin symbioses, little was known about the spatial organization of these strains in the symbiotic organ of the host. To fill this gap of knowledge, I optimized Direct-geneFISH (Barrero-Canosa et al. 2017) in bathymodiolin study systems. My results demonstrated that bacteriocytes are quantitatively dominated by a single SOX strain, suggesting mechanisms of spatial segregation between SOX strains at a single bacteriocyte level. Moreover, I could demonstrate that SOX from the same

functional strain co-occur in neighboring bacteriocytes forming patches (**chapter IV**). The fact that bacteriocytes were quantitatively dominated by a single SOX strain (**Fig.2, b & Fig.3, b in chapter IV**) suggested mechanisms of spatial exclusion between SOX strains at single bacteriocyte level. One of the mechanisms proposed for colonization exclusion of symbionts at a single bacteriocyte level is microvilli effacement. Microvilli effacement is an ultrastructural transformation that leads to the disappearance of microvilli in cells that have been colonized by intracellular bacteria. One example are the enteropathogenic *Escherichia coli* or *Vibrio parahaemolyticus*, which trigger microvilli effacement in colonized enterocytes (Kaper et al. 2004; Zhou et al. 2014). Indeed, Franke et al., *in Prep* showed that the mutualistic symbionts pioneers that colonize a bacteriocyte induce surface modifications that culminate in microvilli effacement. Single bacteriocytes could be considered as cellular barriers that aid the host to separate different mutualists strains in an exercise of microscale compartmentalization. In host-microbiome systems, compartmentalization can aid the host to control the symbiotic population in three different ways: First, controlling their reproduction. Second, rewarding cooperative symbionts. Third, avoiding conflict between different strains (Chomicki et al. 2020). I speculated that compartmentalization of different SOX strains can be the result of how newly formed aposymbiotic bacteriocytes are *de novo* colonized by adjacent and symbiotic bacteriocytes (**Fig.7 in chapter IV**). It is tempting to speculate that bathymodiolin mussels might have the capacity to actively compartmentalize their mutualistic symbionts by promoting intra-filament colonization of newly formed bacteriocytes (**chapter IV**). *Ca.* Endonucleobacter spp. cannot colonize symbiotic bacteriocytes in the gill filaments of bathymodiolin

mussels. As previously mentioned, this might have a protective effect for the bathymodiolin host, as cells without symbionts are a minority in the gills. Symbiotic bacteriocytes might avoid *Ca.* Endonucleobacter spp. infection being fatal for the host, as the parasite can only access a small fraction of the total cells in the gills. Competition for space is a commonly described benefit of mutualistic symbionts. Mutualists can occupy the available microniches in the host tissues that would be otherwise occupied by pathogens. One example are *Endozoicomonas* spp. in scleractinian corals, which are thought to have a protective role for the host against opportunistic pathogens (Neave et al. 2016). Bourne et al., 2007 demonstrated that bleached specimens of the scleractinian coral *Acropora millepora* suffered changes in their microbiome composition, switching from an *Endozoicomonas* spp. to a *Vibrio* spp. dominated microbiome. How symbionts organize spatially within their hosts can have a great influence in microbe-microbe interactions, not only among members of the healthy host microbiome, but also against opportunistic pathogens.

5.5 Future directions

During my doctoral studies, I have found an answer for most of the scientific questions that came along the way. Together with my coauthors, I have formulated hypotheses that explain how *Ca*. Endonucleobacter spp. might be thriving nutritionally and manipulating their host cells. However, an increasing in our understanding in *Ca.* Endonucleobacter spp. biology also brought more questions to scene that need further research.

5.5.1 Inhibition of apoptosis might be a perquisite to become an intracellular parasite

IAPs could have played a major role during appearance of intranuclear lifestyle in the family *Hahellaceae*. If mobile elements aided *Ca.* Endonucleobacter spp. to switch from a mutualistic to a parasitic lifestyle, IAPs might have allowed the establishment of the nucleus as a permanent residence. As previously discussed, proliferation in the nucleus would lead to a deformation of this organelle and a subsequent triggering of the apoptotic cascade. I hypothesize that IAPs are a defining element of *Ca.* Endonucleobacter spp. biology, as inflating the nucleus up to 50 times fold must require the secretion of molecular factors that prevent the shutdown of the host cell. Although they might be the key genetic innovation that allowed *Ca.* Endonucleobacter to settle in the nucleus over evolutionary times, IAPs have not been found in other intranuclear bacteria yet. Apoptosis can also be triggered in eukaryotic cells upon infection by intracellular bacteria or viruses (Jorgensen et al. 2018). Not being intranuclear, several intracellular parasites count on molecular mechanisms to impair apoptosis in infected cells (Behar and Briken 2019; Robinson and Aw 2016). For example, the intracellular enteropathogen *S. flexneri* can prevent the apoptotic response of the enterocyte by sequestering caspases with its O-antigen from the lipopolysaccharide (Günther et al. 2019). One could envision the possibility that preventing host cell apoptosis might be a common strategy to all intracellular parasites. IAPs present in *Ca.* Endonucleobacter spp. are only one type of factors that intracellular parasites can use prevent apoptosis. In viruses, the caspases inhibitors p35 and *CmrA* are another examples of factors that can arrest apoptosis (Zhou and Salvesen 2000). Viruses must evade the apoptotic response to infection by the

host cell to take over the cell synthetic machinery and replicate. We have demonstrated that *Ca.* E. chitolyticus encoded for CRISPR-cas related proteins that might defend the intranuclear parasite against bacteriophages while transiting between hosts (van der Oost et al. 2014). Still, *Ca.* E. chitolyticus can get infected by bacteriophages, as demonstrated by our ultrastructure analyses (**Supp. Fig.2 in chapter III**). Transduction is a process of horizontal gene transfer in which a viral particle acts as the molecular vehicle that inserts foreign DNA into the bacterial chromosome (Johnston et al. 2014). It is tempting to speculate that infection by bacteriophages and integration of apoptosis blockers into the recipient bacteria might have favored the appearance of intranuclear lifestyle in different bacteria phyla. Using a comparative genomic approach, I propose the identification of IAPs or analogous mechanisms for apoptosis blockage in publically available genomes of intranuclear bacteria such as *Holospora* spp. or *Ca.* Nucleicultrix amoebiphila.

5.5.2 *Ca.* E. bathymodioli might be thriving on mutualists' chemosynthates

Many of my analyses have been focused on disentangling the molecular biology of *Ca.* E. chitolyticus (**chapter II**). Although I have produced a high-quality genome for *Ca.* E. bathymodioli, many aspects of its molecular biology remain elusive. *Ca.* E. chitolyticus received its name due to its nutritional strategy of digesting and profiting on host chitin. In the deep-sea mussel *"B." childressi*, chitin synthesis seems to occur in the ciliated edge of the gill filaments (Tietjen, *pers. Comm.*). Moreover, I found evidences of chitin synthesis in secretory cells present in the proximity of the ciliated edge of *"B." childressi* (**Supp. Fig.6 in chapter II**). *Ca.* E. bathymodioli infects the symbiont-free cells that occur in the symbiotic region of *B. puteoserpentis* (**Fig.4 in chapter III**). If synthesis of chitin

is limited to the ciliated edge to all bathymodiolin mussels, *Ca.* E. bathymodioli would not be able to use chitin from *B. puteoserpentis* as nutritional source. Thus, the nutritional strategies of *Ca.* E. bathymodioli remain unknown. One possibility is that *Ca.* E. bathymodioli is also bringing its host cell to starvation state. This would stimulate the host cell to upregulate sugar and amino acid transporters to try to satisfy the metabolic demands of *Ca.* E. bathymodioli. One possible hypothesis is that *Ca.* E. bathymodioli is ultimately thriving on the chemosynthates that the mutualistic symbionts of *B. puteoserpentis* are releasing to the extracellular medium, which its host cell is internalizing. The application of my LCM and ultra-low-input RNAseq pipeline would allow the transcriptomic profiling of *Ca.* E. puteoserpentis and its host cell, reveling interspecific differences in expression patterns within the genus *Ca.* Endonucleobacter.

5.6 Conclusions

The data generated during my doctoral studies provide an extensive description of the molecular biology of an intranuclear parasite of deep-sea mussels and how its host cell reacts to the infection. My discoveries about the biology of *Ca.* Endonucleobacter spp. have challenged previous theories about the nutrition and survival strategies of bacteria that inhabit the nuclear compartment. To study intranuclear bacteria has the immense added difficulty of having to maintain host cell lines in laboratory. Therefore, cultivation-independent methods are essential to disentangle the molecular processes that mediate between intranuclear bacteria and their host cell. It is necessary to continue with single-cell techniques such as laser-capture microdissection that allow to study *in situ* the transcriptome of environmental samples. Many analyses of this thesis were focused on the molecular biology of *Ca.* E. chitolyticus. In future studies these should be

extended to *Ca.* E. bathymodioli. This will help us to understand interspecific differences in nutritional and pathogenic strategies that might reflect the different host species they occupy. In addition, a deeper study of the evolutionary origin of the genus *Ca.* Endonucleobacter will help us to understand the genomic innovations that might originate intranuclear lifestyle, which might be common to all intranuclear bacteria. I expect that the development of single-cell technologies and further comparative genomic analyses will allow us to understand not only the genus *Ca.* Endonucleobacter, but the molecular and evolutionary principles of intranuclear lifestyle in bacteria.

References for chapter V

Ansorge, Rebecca, Stefano Romano, Lizbeth Sayavedra, Miguel Ángel González Porras, Anne Kupczok, Halina E. Tegetmeyer, Nicole Dubilier, and Jillian Petersen. 2019. “Functional Diversity Enables Multiple Symbiont Strains to Coexist in Deep-Sea Mussels.” *Nature Microbiology* 4(12):2487–97.

Barrero-Canosa, Jimena, Cristina Moraru, Laura Zeugner, Bernhard M. Fuchs, and Rudolf Amann. 2017. “Direct-GeneFISH: A Simplified Protocol for the Simultaneous Detection and Quantification of Genes and rRNA in Microorganisms.” *Environmental Microbiology* 19(1):70–82.

Behar, Samuel M., and Volker Briken. 2019. “Apoptosis Inhibition by Intracellular Bacteria and Its Consequence on Host Immunity.” *Current Opinion in Immunology* 60:103–10.

Bierne, Hélène, and Pascale Cossart. 2012. “When Bacteria Target the Nucleus: The Emerging Family of Nucleomodulins.” *Cellular Microbiology* 14(5):622–33.

Bourne, David, Y. Iida, Sven Uthicke, and Carolyn Smith-Keune. 2007. “Changes in Coral-Associated Microbial Communities during a Bleaching Event.” *ISME J* 2:350–63.

Chen, Tao, and Zhibiao Nan. 2015. “Effects of Phytopathogens on Plant Community Dynamics: A Review.” *Acta Ecologica Sinica* 35(6):177–83.

Choi, Eun Ju, Hak Cheol Kwon, Young Chang Sohn, and Hyun Ok Yang. 2010. “*Kistimonas asteriae* Gen. Nov., Sp. Nov., a Gammaproteobacterium Isolated from *Asterias amurensis*.” *International Journal of Systematic and Evolutionary Microbiology* 60(4):938–43.

Chomicki, Guillaume, Gijsbert D. A. Werner, Stuart A. West, and E. Toby Kiers. 2020. “Compartmentalization Drives the Evolution of Symbiotic Cooperation.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 375(1808):20190602.

Crawford, E. D., J. E. Seaman, A. E. Barber, D. C. David, P. C. Babbitt, A. L. Burlingame, and J. A. Wells. 2012. “Conservation of Caspase Substrates across Metazoans Suggests Hierarchical Importance of Signaling Pathways over Specific Targets and Cleavage Site Motifs in Apoptosis.” *Cell Death and Differentiation* 19(12):2040–48.

Deveraux, Quinn L., and John C. Reed. 1999. “IAP Family Proteins - Suppressors of Apoptosis.” *Genes and Development* 13(3):239–52.

Deveraux, Quinn L., Ryosuke Takahashi, Guy S. Salvesen, and John C. Reed. 1997. “X-Linked IAP Is a Direct Inhibitor of Cell-Death Proteases.” *Nature* 388(6639):300–304.

Dramsi, S., and P. Cossart. 1998. “Intracellular Pathogens and the Actin Cytoskeleton.” *Annual Review of Cell and Developmental Biology* 14(1):137–66.

Driscoll, Timothy, Joseph J. Gillespie, Eric K. Nordberg, Abdu F. Azad, and Bruno W. Sobral. 2013. “Bacterial DNA Sifted from the *Trichoplax adhaerens* (*Animalia*: *Placozoa*) Genome Project Reveals a Putative Rickettsial Endosymbiont.” *Genome Biology and Evolution* 5(4):621–45.

Efeyan, Alejo, William C. Comb, and David M. Sabatini. 2015. “Nutrient-Sensing Mechanisms and Pathways.” *Nature* 517(7534):302–10.

Fiore, Cara L., Micheline Labrie, Jessica K. Jarett, and Michael P. Lesser. 2015. “Transcriptional Activity of the Giant Barrel Sponge, *Xestospongia muta* Holobiont: Molecular Evidence for Metabolic Interchange.” *Frontiers in Microbiology* 6(APR).

Forget, Nathalie L., and S. Kim Juniper. 2013. “Free-Living Bacterial Communities Associated with Tubeworm (*Ridgeia piscesae*) Aggregations in Contrasting Diffuse Flow Hydrothermal Vent Habitats at the Main Endeavour Field, Juan de Fuca Ridge.” *MicrobiologyOpen* 2(2):259–75.

Gillespie, Joseph J., Vinita Joardar, Kelly P. Williams, Timothy Driscoll, Jessica B. Hostetler, Eric Nordberg, Maulik Shukla, Brian Walenz, Catherine A. Hill, Vishvanath M. Nene, Abdu F. Azad, Bruno W. Sobral, and Elisabet Caler. 2012. “A Rickettsia Genome Overrun by Mobile Genetic Elements Provides Insight into the Acquisition of Genes Characteristic of an Obligate Intracellular Lifestyle.” *Journal of Bacteriology* 194(2):376–94.

Günther, Saskia D., Melanie Fritsch, Jens M. Seeger, Lars M. Schiffmann, Scott J. Snipas, Maria Coutelle, Thomas A. Kufer, Paul G. Higgins, Veit Hornung, Maria L. Bernardini, Stefan Höning, Martin Krönke, Guy S. Salvesen, and Hamid Kashkar. 2019. “Cytosolic Gram-Negative Bacteria Prevent Apoptosis by Inhibition of Effector Caspases through Lipopolysaccharide.” *Nature Microbiology*.

Haferkamp, Ilka, Stephan Schmitz-Esser, Michael Wagner, Nadjeschka Neigel, Matthias Horn, and H. Ekkehard Neuhaus. 2006. “Tapping the Nucleotide Pool of the Host: Novel Nucleotide Carrier Proteins of *Protochlamydia amoebophila*.” *Molecular Microbiology* 60(6):1534–45.

Henne, W. Mike, Michael L. Reese, and Joel M. Goodman. 2018. “The Assembly

of Lipid Droplets and Their Roles in Challenged Cells." *The EMBO Journal* 37(12).

Hinzke, Tjorven, Manuel Kleiner, Corinna Breusing, Horst Felbeck, Robert Häsler, Stefan M. Sievert, Rabea Schlüter, Philip Rosenstiel, Thorsten B. H. Reusch, Thomas Schweder, and Stephanie Markert. 2019. "Host-Microbe Interactions in the Chemosynthetic *Riftia pachyptila* Symbiosis." *MBio* 10(6):1–46.

Holden, Matthew T. G., Helena M. B. Seth-Smith, Lisa C. Crossman, Mohammed Sebaihia, Stephen D. Bentley, Ana M. Cerdeño-Tárraga, Nicholas R. Thomson, Nathalie Bason, Michael A. Quail, Sarah Sharp, Inna Cherevach, Carol Churcher, Ian Goodhead, Heidi Hauser, Nancy Holroyd, Karen Mungall, Paul Scott, Danielle Walker, Brian White, Helen Rose, Pernille Iversen, Dalila Mil-Homens, Eduardo P. C. Rocha, Arsenio M. Fialho, Adam Baldwin, Christopher Dowson, Bart G. Barrell, John R. Govan, Peter Vandamme, C. Anthony Hart, Eshwar Mahenthiralingam, and Julian Parkhill. 2009. "The Genome of *Burkholderia cenocepacia* J2315, an Epidemic Pathogen of Cystic Fibrosis Patients." *Journal of Bacteriology* 91(1):261–77.

Jensen, Sigmund, Sébastien Duperron, Nils Kåre Birkeland, and Martin Hovland. 2010. "Intracellular *Oceanospirillales* Bacteria Inhabit Gills of *Acesta* Bivalves." *FEMS Microbiology Ecology* 74(3):523–33.

Johnston, Calum, Bernard Martin, Gwennaele Fichant, Patrice Polard, and Jean-Pierre Claverys. 2014. "Bacterial Transformation: Distribution, Shared Mechanisms and Divergent Control." *Nature Reviews Microbiology* 12(3):181–96.

Jorgensen, Ine, Manira Rayamajhi, and Edward A. Miao. 2018. "Programmed Cell Death as a Defence against Infection GSDMD Forms a Pore in Membranes." 17(3):151–64.

Kaper, James B., James P. Nataro, and Harry L. T. Mobley. 2004. "Pathogenic *Escherichia coli*." *Nature Reviews Microbiology* 2(2):123–40.

Katharios, Pantelis, Helena M. B. Seth-Smith, Alexander Fehr, José M. Mateos, Weihong Qi, Denis Richter, Lisbeth Nufer, Maja Ruetten, Maricruz Guevara Soto, Urs Ziegler, Nicholas R. Thomson, Ralph Schlapbach, and Lloyd Vaughan. 2015. "Environmental Marine Pathogen Isolation Using Mesocosm Culture of Sharpsnout Seabream: Striking Genomic and Morphological Features of Novel *Endozoicomonas* Sp." *Scientific Reports* 5(August):17609.

Keminer, Oliver, and Reiner Peters. 1999. "Permeability of Single Nuclear Pores." *Biophysical Journal* 77(1):217–28.

Kräter, Martin, Jiranuwat Sapudom, Nicole Bilz, Tilo Pompe, Jochen Guck, and Claudia Claus. 2018. “Alterations in Cell Mechanics by Actin Cytoskeletal Changes Correlate with Strain-Specific Rubella Virus Phenotypes for Cell Migration and Induction of Apoptosis.” *Cells* 7(9):136.

Liu, Ailin, Carolina A. Contador, Kejing Fan, and Hon-Ming Lam. 2018. “Interaction and Regulation of Carbon, Nitrogen, and Phosphorus Metabolisms in Root Nodules of Legumes.” *Frontiers in Plant Science* 9:1860.

Ma, Jiong, Alexander Goryaynov, Ashapurna Sarma, and Weidong Yang. 2012. “Self-Regulated Viscous Channel in the Nuclear Pore Complex.” *Proceedings of the National Academy of Sciences of the United States of America* 109(19):7326–31.

Mohr, Dagmar, Steffen Frey, Torsten Fischer, Thomas Güttler, and Dirk Görlich. 2009. “Characterisation of the Passive Permeability Barrier of Nuclear Pore Complexes.” *EMBO Journal* 28(17):2541–53.

Morrow, Kathleen M., Anthony G. Moss, Nanette E. Chadwick, and Mark R. Liles. 2012. “Bacterial Associates of Two Caribbean Coral Species Reveal Species-Specific Distribution and Geographic Variability.” *Applied and Environmental Microbiology* 78(18):6438–49.

Neave, Matthew J., Amy Apprill, Christine Ferrier-Pages, and Christian R. Voolstra. 2016. “Diversity and Function of Prevalent Symbiotic Marine Bacteria in the Genus *Endozoicomonas*.” *Applied Microbiology and Biotechnology* 1–10.

Neave, Matthew J., Craig T. Michell, Amy Apprill, and Christian R. Voolstra. 2017. “*Endozoicomonas* Genomes Reveal Functional Adaptation and Plasticity in Bacterial Strains Symbiotically Associated with Diverse Marine Hosts.” *Scientific Reports* 7(January):1–12.

Niebuhr, Kirsten, Sylvie Giuriato, Thierry Pedron, Dana J. Philpott, Frédérique Gaits, Julia Sable, Michael P. Sheetz, Claude Parsot, Philippe J. Sansonetti, and Bernard Payrastre. 2002. “Conversion of PtdIns(4, 5)P2 into PtdIns(5)P by the *S. flexneri* Effector IpgD Reorganizes Host Cell Morphology.” *EMBO Journal* 21(19):5069–78.

Niebuhr, Kirsten, Noureddine Jouihri, Abdelmounaaim Allaoui, Pierre Gounon, Philippe J. Sansonetti, and Claude Parsot. 2000. “IpgD, a Protein Secreted by the Type III Secretion Machinery of Shigella Flexneri, Is Chaperoned by IpgE and Implicated in Entry Focus Formation.” *Molecular Microbiology* 38(1):8–19.

Nolan, S. J., Romano, J. D., & Coppens, I. (2017). "Host lipid droplets: An important source of lipids salvaged by the intracellular parasite *Toxoplasma*

gondii". PLoS pathogens, 13(6), e1006362.

van der Oost, John, Edze R. Westra, Ryan N. Jackson, and Blake Wiedenheft. 2014. “Unravelling the Structural and Mechanistic Basis of CRISPR–Cas Systems.” *Nature Reviews Microbiology* 12(7):479–92.

Parkhill, Julian, Mohammed Sebaihia, Andrew Preston, Lee D. Murphy, Nicholas Thomson, David E. Harris, Matthew T. G. Holden, Carol M. Churcher, Stephen D. Bentley, Karen L. Mungall, Ana M. Cerdeño-Tárraga, Louise Temple, Keith James, Barbara Harris, Michal A. Quail, Mark Achtman, Rebecca Atkin, Steven Baker, David Basham, Nathalie Bason, Inna Cherevach, Tracey Chillingworth, Matthew Collins, Anne Cronin, Paul Davis, Jonathan Doggett, Theresa Feltwell, Arlette Goble, Nancy Hamlin, Heidi Hauser, Simon Holroyd, Kay Jagels, Sampsa Leather, Sharon Moule, Halina Norberczak, Susan O’Neil, Doug Ormond, Claire Price, Ester Rabbinowitsch, Simon Rutter, Mandy Sanders, David Saunders, Katherine Seeger, Sarah Sharp, Mark Simmonds, Jason Skelton, Robert Squares, Steven Squares, Kim Stevens, Louise Unwin, Sally Whitehead, Bart G. Barrell, and Duncan J. Maskell. 2003. “Comparative Analysis of the Genome Sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica.*” *Nature Genetics* 35(1):32–40.

Phillips, Margaret A., Jeremy N. Burrows, Christine Manyando, Rob Hooft Van Huijsduijnen, Wesley C. Van Voorhis, and Timothy N. C. Wells. 2017. “Malaria.” *Nature Reviews Disease Primers* 3.

Qi, Weihong, Maria Chiara Cascarano, Ralph Schlapbach, Pantelis Katharios, Lloyd Vaughan, and Helena M. B. Seth-Smith. 2018. “*Ca.* Endozoicomonas cretensis: A Novel Fish Pathogen Characterized by Genome Plasticity.” *Genome Biology and Evolution* 10(6):1363–74.

Ribbeck, K., and D. Görlich. 2001. “Kinetic Analysis of Translocation through Nuclear Pore Complexes.” *EMBO Journal* 20(6):1320–30.

Douglas, E. A. 2010. “The Symbiotic Habit.” *Symbiosis* 51(2):197–98.

Robinson, Keith S., and Rochelle Aw. 2016. “The Commonalities in Bacterial Effector Inhibition of Apoptosis.” *Trends in Microbiology* 24(8):665–80.

Salem, Hassan, Eugen Bauer, Roy Kirsch, Aileen Berasategui, Michael Cripps, Benjamin Weiss, Ryuichi Koga, Kayoko Fukumori, Heiko Vogel, Takema Fukatsu, and Martin Kaltenpoth. 2017. “Drastic Genome Reduction in an Herbivore’s Pectinolytic Symbiont.” *Cell* 171(7):1520-1531.e13.

Schmitz-esser, Stephan, Nicole Linka, Astrid Collingro, L. Beier, H. Ekkehard Neuhaus, Michael Wagner, Matthias Horn, and Cora L. Beier. 2004. “ATP / ADP

Translocases : A Common Feature of Obligate Intracellular Amoebal Symbionts Related to Chlamydiae and Rickettsiae." *Journal of Bacteriology* 186(3):683–91.

Schreiber, Lars, Kasper Urup Kjeldsen, Matthias Obst, Peter Funch, and Andreas Schramm. 2016. "Description of *Endozoicomonas ascidiicola* Sp. Nov., Isolated from Scandinavian Ascidians." *Systematic and Applied Microbiology* 39(5):313–18.

Schulz, Frederik, and Matthias Horn. 2015. "Intranuclear Bacteria: Inside the Cellular Control Center of Eukaryotes." *Trends in Cell Biology* 25(6):339–46.

Schulz, Frederik, Ilias Lagkouvardos, Florian Wascher, Karin Aistleitner, Rok Kostanjšek, and Matthias Horn. 2014. "Life in an Unusual Intracellular Niche: A Bacterial Symbiont Infecting the Nucleus of Amoebae." *ISME Journal* 8(8):1634–44.

Shin, Woongghi, Sung Min Boo, and Lawrence Fritz. 2003. "Endonuclear Bacteria in Euglena Hemichromata (Euglenophyceae): A Proposed Pathway to Endonucleobiosis." *Phycologia* 42(2):198–203.

Thursby, Elizabeth, and Nathalie Juge. 2017. "Introduction to the Human Gut Microbiota." *Biochemical Journal* 474(11):1823–36.

Tietjen, Målin. "Physiology and ecology of deep-sea bathymodiolin symbioses" *Personal communication*.

Turchinovich, Andrey, Harald Surowy, Andrius Serva, Marc Zapatka, Peter Lichter, and Barbara Burwinkel. 2014. "Capture and Amplification by Tailing and Switching (CATS), Anultrasensitive Ligation-Independent Method for Generation of DNA Libraries for Deep Sequencing from Picogram Amounts of DNA and RNA." *RNA Biology* 11(7):817–28.

Wentrup, Cecilia, Annelie Wendeberg, Mario Schimak, Christian Borowski, and Nicole Dubilier. 2014. "Forever Competent: Deep-Sea Bivalves Are Colonized by Their Chemosynthetic Symbionts throughout Their Lifetime." *Environmental Microbiology* 16(12):3699–3713.

Zhang, Jianlin, Amanda Felder, Yujie Liu, Ling T. Guo, Stephan Lange, Nancy D. Dalton, Yusu Gu, Kirk L. Peterson, Andrew P. Mizisin, G. Diane Shelton, Richard L. Lieber, and Ju Chen. 2009. "Nesprin 1 Is Critical for Nuclear Positioning and Anchorage." *Human Molecular Genetics* 19(2):329–41.

Zhou, Qiao, and Guy S. Salvesen. 2000. "[12] - Viral Caspase Inhibitors CrmA and P35." Pp. 143–54 in *Apoptosis*. Vol. 322, edited by J. C. B. T.-M. in E. Reed. Academic Press.

Zhou, Xiaohui, Ramiro H. Massol, Fumihiko Nakamura, Xiang Chen, Benjamin E. Gewurz, Brigid M. Davis, Wayne Lencer, and Matthew K. Waldor. 2014. “Remodeling of the Intestinal Brush Border Underlies Adhesion and Virulence of an Enteric Pathogen.” *MBio* 5(4):1–9.

Zielinski, Frank U., Annelie Pernthaler, Sebastien Duperron, Luciana Raggi, Olav Giere, Christian Borowski, and Nicole Dubilier. 2009. “Widespread Occurrence of an Intranuclear Bacterial Parasite in Vent and Seep Bathymodiolin Mussels.” *Environmental Microbiology* 11(5):1150–67.

Acknowledgements

This thesis would not have been possible without the professional and emotional support of all the people that have accompanied me either physically or from the distance (my beloved Málaga).

Thank you **Prof. Dr. Nicole Dubilier** for allowing me to conduct my doctoral studies in the Symbiosis Department. I will always remember the interview in which together with Jill you offered me the PhD position ("We love your sense of humor and your skills in the lab!"). Also thank you for the encouragement all along the way, for allowing me to work from Cologne in these last difficult times and for reviewing this thesis. Many thanks also to **Prof. Dr. Matthias Horn** for reviewing this work and taking the time to join my examination committee. I would also like to take the opportunity to thank **Dr. Bernhard Fuchs** and **Prof. Dr. Michael W. Friedrich** for joining the examination committee.

I could dedicate two pages to thank **Dr. Nikolaus Leisch** for his constant and outstanding support along these 5 years and I would not be even close to express myself properly. Niko, you being the sun and I being the earth, I say we are in the habitable zone: Far away enough to see things with perspective and close enough to have made this experience a joy beyond professional achievements. Thank you for being a tutor, a mentor, a friend and an all-in-expert. I truly appreciate everything you have taught me, especially about being an independent researcher. I also thank you for being part of the examination committee and of course, for proof-reading part of this thesis.

Thank you **Dennis Jakob** for being part of the examination committee, and good luck with Endonucleobacter. It is a pretty cool bug!

The **Symbiosis Department** was a great place to grow, not only professionally, but also personally. Many thanks to all the people that made the working atmosphere a joy. A special consideration goes to **Martina Meyer**, with whom I have shared really delightful moments, both working and laughing. Thank you for bringing such a warm spirit to the group Martina.

Everybody is coming to this part to know what I have written to you, so let´s please the plebs. Thank you **Benedikt**, for being a friend, a brother and the rock in the middle of the sea that kept my sanity when everything was melting down around me. Your friendship and camaraderie are the best memories I take from Germany. Keep everything from yourself intact and don´t change the minimal thing. You got a passionate friend for the rest of your life. Also thank you for proof-reading part of this thesis. P.S. I promise you I will keep more attention to fascinating things like floating microbes or particle accelerators that I am sure they will help me a lot in my future research.

Thank you **Miriam**, from your early support when settling in Germany and your day-by-day assistance during my hardest moments. Thank you also for these nice dinners and wine evenings and all the music we have enjoyed together. A warm place in my hearth has your name. Also thank you for helping me with the German parts of this thesis!

Max, young brother! Thank you for being such an outstanding colleague and friend. I am really happy that you took your imaging skills so far.

Adrien, you are my reference for many things. Not only for being an outstanding scientist, but also for being able to keep your hobbies and personal interests while developing your scientific career. Thank you for the inspiration, all the good advices and for giving me Endonucleobacter, the coolest bug ever.

Liz, thank you for being such a wonderful master thesis supervisor and having such a warm character. It was a pleasure to work with you.

Thank you **Jimena** for your patience and kindness when teaching me the awesome technique that gave me the entry ticket to the MPI. Also, thank you for reminding me more than one time that I should not underestimate myself.

Juliane, my spirit postdoc. I have missed you since the day you stepped out of the institute! The rest of the things I want to tell you are quite inappropriate, so better in person.

Rahel, why did you have to go before me? Still laughing so much about the endless jokes we made together. I am glad that somebody else is enjoying your awesome personality. Keep it up!

Harald, your help with the sequencing services, bioinformatics guidance and shaping of endless projects in the Symbiosis Department deserves a huge recognition. My part is here: Thank you for your guidance and assistance with all the computer work, and also for your emotional support when there was nobody else there.

Laetitia, thank you for the spark of joy you bring to the group. It is a shame we have not been able to be expend more time talking, this year was a weird one! May the socks always be with you. Thank you for proof-reading part of this thesis.

Thank you **Manuel** for letting metabolomics to go in the Endonucleobacter project. Also your expertise and tips during my PhD were mostly appreciated, most of them related to how-not-to burn out!

Thank you **Målin** for sharing with me all these support talks in the tea kitchen. It was not easy, but it was worth. Also a thousand thanks for the great transcriptomic expertise you brought to the Endonucleobacter project. Thank you also for your help finding me a super nice flat. It´s half yours!

Janine, thank you for your effort bringing metabolomics to the Endonucleobacter project, it was a great pleasure to work with you!

Rebecca, I thank you for all the “positive vibes” you brought to the last stage of my PhD. I wish I could have enjoyed your presence a bit longer. I was a pleasure to work with you but overall, to enjoy your warm soul.

Thank you **Dolma** for your smart sense of humor, your polite manners, your encouragement, your support in difficult times, your scientific input and your good aura. You are this kind of glue that keeps people together, and such an elegant glue! P.S. Thank you also for your gastronomic standards. Glad I could criticize with you the generalized feeding behavior. Also thank you for proof-reading part of this thesis!

Brandon, thank you for being such an awesome flat mate and your outstanding helps in bioinformatics. Also thank you for all the gastronomic talks we shared together!

Anna, a thousand thanks for your support with bioinformatics. I would have killed myself without you, binary girl. Please, think about me every Easter, and do not forget about the reptilians.

Merle, thank you for being such an awesome cruise partner. I surely enjoyed our honey moon in Matamanoa. And thank you also for being such an organized person, the second place reminds me that I still have hope to be considered a human!

Tina, thanks a thousand for your bioinformatics and phylogeny input. Good luck with FISH, I am sure you are going to rock it!

Oliver, thank you for reminding me in every retreat that my figures have improved since I stopped using paint! Jokes aside, it was great to grow next to you. Thank you for your scientific criticism and constant input, and for making me see what is in my science blind spot. Thank you also for all the help finding and setting my flat. P.S. I cannot give you the other half, because I already gave a half to Målin and I need a place to live.

I miss you **Mario**. Thank you for all the awesome laughs we had at the beginning of my PhD. Taking your dark crown was painful, but I sadistically enjoyed it!

Burak, it was great to have you around during my PhD. The MPI and I surely miss you (ah-ah-ah).

Rafa, thank you for the continuous checkup *"¿Cómo estás Miguel?".* Since the very first moment you have demonstrated to be an outstanding friend. Keep up the good work and the name of our old nation quite high.

Antonio and **Marina**, thank you for all the awesome chats and meals in your great flat. I enjoyed so much your company and caring.

Silvia, thank you for your patience and your kindness with me. I know I was quite annoying always mentioning “that topic”. You are a cheerful and delightful person to be around, thank you for making my stay in Germany a bit brighter.

Mónica, thank you for all the times you have sweeten my day. I wish you all the best for the future and it was a pleasure to have you around.

Thank you **Andreas Ellrott** for your technical assistance with the CLSM.

I would like to thank **Daniela Tienken** for your introduction to the LMD.

Thank you **Dr. Christiane Glöckner**, for all the administrative (and emotional) support you have given me during my PhD. I will surely not forget your constant smile.

Thank you **Mrs. Anita Tingberg** for helping me with a lot of administration at the beginning of my adventure in Germany. I will always remember the day you picked me up from the train station in Bremen!

Thank you for nothing **COVID-19**.

Germany, thank you for your awesome weather, your easy language, your outstanding social life, your friendly neighbors and your straight forward administrative procedures. And now the truth: Thank you Germany, for giving me the dignity as a scientist that my own country has denied me.

Muchas gracias a **mi familia** por su apoyo incondicional durante tantos años, por el cariño y la comprensión derramados y por las ganas con las que me recogían

en el aeropuerto todas las vacaciones. **Mamá**, **Vitorilla**, os quiero mucho, y lo que más deseo en el mundo es que volvamos a estar juntos en Málaga, de la que ya no me sacarán ni con agua hirviendo.

Muchas gracias **José Manuel Morales García**, por tu compañía y tu ayuda durante los momentos más difíciles de mi vida. Gracias también por tu confianza incondicional en mi persona. Una parte de esta tesis es mérito tuyo.

Gracias al **Consejo de Sabios**, que me ha apoyado desde la distancia. **Pedro (hiii-hiii-hiii)**, **Eva**, **Salvador (Cabra)**, **Jean**, **Juan**, **Sonia**, **Lorenzo** y **Charo**, os echo mucho de menos a todos. Espero volver a Málaga pronto y que volvamos a estar juntos.

A mis compañeros de la Universidad de Málaga y grandísimos amigos, **Laura (termita reina)**, **Álvaro (Jodico)** y **Rodri**. Gracias por recibirme siempre con los brazos abiertos en Málaga y aguantar mi relación amor-odio con la investigación.

Gracias a mi gran compañera de videojuegos **Chiti**. Ojalá una décima parte de las personas de este mundo tuvieran tu gran corazón y bondad. Te quiero más de lo que tú quieres a los gatos.

A mis compañeros de casas rurales y videojuegos: ¡¡Gracias!! No sabéis lo duro que es veros 2 o 3 veces con suerte al año. Cada vez que compartimos momentos juntos se me ilumina el corazón. **Elena**, muchas gracias por haberme visitado tantas veces en Bremen. Tu compañía me lo hizo todo más fácil. **Pedro**, gracias por recordarme de vez en cuando lo imbécil que soy. **Jose** y **Dani**, tenemos un par de cosas pendientes y ninguna es decente. **Borja**, que me queo sin comé. **Flamenquita**, yo sé que quieres tenerme allí para ti sola, ya mismo llegará el momento. **Palomo**, sigue siendo tan auténtica y no cambies nada,

siempre me acordaré de lo que me dijiste una vez: “El hombre de tu vida eres tú”. Y por último a mi gran amigo, hermano y confidente **Chechusín**: gracias por saber mejor que yo que es lo que quiero en la vida (y lo que no).

Ana, **Carlos**, **Rafa**, **Susi**, **Sonia**, **Joselu**, **Proyecto Cárnico**, **Pablo**, **Patri**, **Raúl**, **Migue**l, **Carmela y la Sierva de la Naturaleza**, muchísimas gracias por el apoyo incondicional desde España. Espero que sigamos viéndonos muchos años, celebrando bodas en fechas que jodan bastante y muchísimos baby showers.

A mis filetitos, **Edu**, **Arancha**, **Dani Kekito**, **Mari**, **Marta, Santi, Rocío** y por supuesto, la única decente de ese grupo, **Manuela**. Muchísimas gracias por darme tantos buenos ratos cada vez que bajo a Málaga.

A mis amigos de la privada: Muchas gracias por el apoyo desde la distancia. Espero dejar al 29006 a la altura que se merece. Especial mención a **Antonio (Negro)**, **Jose (Tote)** y **Juan (Wanillo)**, mis amigos de la infancia.

Y por último y no por ello menos importante, muchísimas gracias a la **Dr. María Altamirano Jeschke**. María, has sido tutora, mentora y amiga todos estos años, incluso cuando no estaba allí contigo en la Universidad de Málaga. Espero poder incorporarme un día a trabajar en la UMA y que podamos colaborar en proyectos interesantes juntos. El visitarte es una de las grandes experiencias que me llevo cada vez que bajo a mi tierra.

Contribution to each manuscript included in this thesis

Manuscript 1 (chapter II)

Conceptual design: 60%

Data acquisition and experiments: 80%

Analysis and interpretation of results: 80%

Preparation of figures and tables: 95%

Writing the manuscript: 90%

Manuscript 2 (chapter III)

Conceptual design: 70%

Data acquisition and experiments: 30%

Analysis and interpretation of results: 70%

Preparation of figures and tables: 95%

Writing the manuscript: 90%

Manuscript 3 (chapter IV)

Conceptual design: 70%

Data acquisition and experiments: 90%

Analysis and interpretation of results: 90%

Preparation of figures and tables: 100%

Writing the manuscript: 100%

Universität Bremen

Bremen, 17. September. 2020

Versicherung an Eides Statt

Ich, Miguel Ángel González Porras
Friedrich-Ebert-Strasse 170, 28201 Bremen, Matrikelnummer: 3066043

versichere an Eides Statt durch meine Unterschrift, dass

1. Ich die vorstehende Arbeit mit dem titel „Molecular biology and evolution of the intranuclear bacterial intranuclear parasite *Ca.* Endonucleobacter“ selbständig und ohne fremde Hilfe angefertigt,

2. Ich alle Stellen, die ich wörtlich dem Sinne nach aus Veröffentlichungen entnommen habe, als solche kenntlich gemacht habe, mich auch keiner anderen als der angegebenen Literatur oder sonstiger Hilfsmittel bedient habe.

3. Die elektronische Version der Dissertation identisch mit der abgegebenen gedrukten Version ist.

Ich versichere an Eides Statt, dass ich die vorgenannten Angaben nach bestem Wissen und Gewissen gemacht habe und dass die Angaben der Wahrheit entsprechen und ich nichts verschwiegen habe.

Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

_____

Bremen, 17.09.2020, Miguel Ángel González Porras