Genetic tools for the trypanosomatid

Angomonas deanei

help to dissect host-endosymbiont interactions

Inaugural-Dissertation

Zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Sofia Kokkori

aus Athen, Griechenland

Düsseldorf, Juli 2018

aus der Emmy Noether-Gruppe für Mikrobielle Symbiose und Organellenevolution

der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der

Mathematisch-Naturwissenschaftlichen Fakultät der

Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Dr. Eva C. M. Nowack

2. Prof. Dr. Johannes Hegemann

Tag der mündlichen Prüfung: 27.09.2018

Summary

Parasitic and endosymbiontic bacteria can be considered as two different products of a process that would begin by infecting the host cells, avoiding the host defense response and multiplying within the host. Nevertheless, in endosymbiotic associations the host evolves mechanisms to manipulate its bacterial endosymbionts according to its needs. Understanding the molecular basis of those mechanisms will help to better comprehend the fascinating biology of endosymbiosis and might assist in developing strategies to control parasitic bacteria. Despite the prevalence of bacterial endosymbionts across eukaryotic phyla the molecular mechanisms that intervene in host-endosymbiont associations are poorly understood. This gap of knowledge reflects the lack of effective molecular tools that enable functional studies of endosymbiotic systems. This thesis focused on developing molecular biological tools for *A. deanei*, a promising model system for studying endosymbiosis. *A. deanei* is a fast growing flagellated trypanosomatid that contains a single proteobacterial endosymbiont, which provides the host with various amino acids and co-factors. The availability of full genome and transcriptome sequences, high frequencies of homologous recombination (HR), ease of transfection and rapid growth have allowed the generation of a toolbox for genetic manipulation of *A. deanei*. Thus, an electroporation protocol, hygromycin, neomycin and phleomycin resistances as markers for selection of transfectants and *γ-* and *δ-amastin* loci as suitable genome sites for insertion and high-level expression of transgenes were established. The molecular toolbox developed allows for the first time to generate null mutants of specific genes by HR and stable expression of transgenes in a system with vertically transmitted bacterial endosymbionts. This system could be used to explore the trafficking of proteins between the two partners and to investigate the role of host-encoded proteins that are targeted to the endosymbiont. Using this newly established genetic tool, Endosymbiont-targeted Protein 1 (ETP1) was the first host-encoded protein proved to target the endosymbiont of *A. deanei*. ETP1 has no homologs in other organisms and several attempts to generate null mutants of ETP1 were unsuccessful, suggesting that it is essential for the survival of the cells. Co-Immunoprecipitation assay coupled to mass spectrometry analysis and 2-dimensional Clear native/SDS-PAGE identified porin, Kinetoplastid membrane protein-11 and Endosymbiont-targeted Protein 2 as possible interaction partners of ETP1 and first evidence suggest that these four proteins localize to the periphery of the endosymbiont. It is still early to propose that ETP1 has any role in the transport of different

metabolites and/or cell division. Nevertheless, these first findings give some evidence on how host-encoded proteins might shape the symbiosis with bacterial endosymbionts in *A. deanei*.

Zusammenfassung

Parasitische und endosymbiontische Bakterien können als zwei verschiedene Produkte eines Prozesses betrachtet werden, der beginnt, indem Bakterien eine Wirtszellen infizieren, die Abwehrreaktion der Wirtszellen vermeiden und sich in der Wirtszelle multiplizieren. Im Gegensatz zu parasitischen Interaktionen entwickelt die Wirtszelle in endosymbiotischen Assoziationen Mechanismen, um ihre bakteriellen Endosymbionten entsprechend ihren Bedürfnissen zu manipulieren. Das Verständnis der molekularen Basis dieser Mechanismen wird helfen, die faszinierende Biologie der Endosymbiose besser zu verstehen und könnte in der Zukunft dazu beitragen Strategien gegen Infektionen mit parasitischen Bakterien zu entwickeln. Trotz der Häufigkeit bakterieller Endosymbionten in allen eukaryotischen Superphyla sind die molekulare Mechanismen, die die Interaktion zwischen Wirtzelle und Endosymbiont vermitteln, weitgehen unverstanden. Dieser Wissensmangel spiegelt den Mangel an effektiven molekularen Werkzeugen wider, die funktionelle Studien von endosymbiotischen Systemen ermöglichen. Diese Doktorarbeit konzentrierte sich auf die Entwicklung molekularbiologischer Werkzeuge für *Angomonas deanei*, ein vielversprechendes Modellsystem zur Untersuchung der Endosymbiose. *A. deanei* ist ein schnell wachsender Trypanosomatid, der einen proteobakteriellen Endosymbionten enthält, der die Wirtszelle mit verschiedenen Aminosäuren und Cofaktoren versorgt. Die Verfügbarkeit von langen (kern) genomischen Contigs und einem großen Transkriptomdatensatz, die hohe Rate homologer Rekombination (HR), die Leichtigkeit der Transfektion und schnelles Wachstum ermöglichten die Entwicklung eines Protokolls zur genetischen Manipulation von *A. deanei*. So wurden ein Elektroporationsprotokoll, Hygromycin-, Neomycin- und Phleomycinresistenzen als Marker für die Selektion von Transfektanten etabliert und die γ- und δ-Amastin Loci als geeignete Genomstellen für die Insertion und die Expression von Transgenen auf hohem Niveau charakterisiert. Die entwickelte molekulare Toolbox ermöglicht erstmalig die Generierung von Null-Mutanten spezifischer Gene durch HR und die stabile Expression von Transgenen in einem System mit vertikal vererbten bakteriellen Endosymbionten. Dieses System steht nun zur Verfügung, um die Funktion von Wirtsproteinen, die eine Rolle in der Wirts/Symbionteninteraktion spielen, zu charakterisieren. Unter Verwendung dieses neu etablierten genetischen Werkzeugs konnte für Endosymbiont-targeted Protein 1 (ETP1), als erstem kernkodierten Protein, eindeutig eine Endosymbiontenlokalization nachgewiesen werden. ETP1 hat keine Homologie zu Proteinen in anderen Organismen und mehrere Versuche Null-Mutanten von ETP1 zu generieren waren nicht erfolgreich, was darauf hinweist, dass dieses Protein für das Überleben der Zellen

essentiell ist. Co-Immunopräzipitations Assays gekoppelt mit massenspektrometrischer Analyse und 2-dimensionale Clear native/SDS-PAGE identifiziert Porin, Kinetoplastid Membran Protein-11 und Endosymbiont-targeted Protein 2 als mögliche Interaktionspartner von ETP1. Erste Ergebnisse deuten darauf hin, dass diese vier Proteine in der Peripherie des Endosymbionten lokalisieren. Es ist noch früh, definitive Aussagen über die zelluläre Funktion von ETP1 zu treffen. Trotzdem geben diese ersten Befunde wichtige Hinweise für die funktionelle Charakterisierung von ETP1 and können dazu beitragen zu verstehen, welche Rolle kernkodierte Proteine bei der Bildung einer synergistischen und homöostatischen Interaktion zwischen einem bakteriellen Endosymbionten und einer eukaryotischen Wirtszelle spielen.

List of Abbrevations

%	Percentage	h	Hour
Amp	Ampicillin	*hyg*	Resistance marker against hygromycin
AMP	antimicrobial peptide	HR	Homologous recombination
aa	Amino acid	*in silico*	By using computational simulation
BSA	Bovine serum albumin	*in situ*	On site the phenomenon is occurring
CAM	Chloramphenicol	*in vitro*	Independent of a living organism
Co-IP	Co-immunoprecipitation	kb	Kilobases
crTP	Chromatophore transit peptide	kDa	Kilodalton
C-terminal	Carboxyterminal	KMP-11	Kinetoplastid membrane protein-11
dH$_2$O	Distilled water	mRNA	Messenger RNA
DIG	Digoxigenin	*neo*	Resistance marker against neomycin
DNA	Deoxyribonucleic acid	N-terminal	Aminoterminal
dNTP	Deoxynucleoside triphosphate	nt	Nucleotide
EtOH	Ethanol	ORF	Open reading frame
ETP1	Endosymbiont-targeted Protein 1	*phleo*	Resistance marker against phleomycin
ETP2	Endosymbiont-targeted Protein 2	RNA	Ribonucleic acid
et al.	Et alii (And others)	RT	Room temperature
EGT	Endosymbiotic gene transfer	v/v	Volume per volume
FR	Flanking region	w/v	Weight per volume
gDNA	Genomic DNA	WT	Wild-type
GFP	Green fluorescent protein	x g	x fold gravitational force

Table of Contents

1 Introduction

1.1 Bacterial Symbiosis and Pathogenesis

All forms of life on Earth interact with the environment that surrounds them. Within the enormous diversity of organisms that live on Earth, prokaryotes are characterized by a broad spectrum of metabolic capabilities that has allowed them to colonize the entire planet and inhabit every imaginable ecological niche, from the most favorable to the most extreme. In general, prokaryotes have developed tight metabolic associations with other organisms around them, such as other bacteria, protozoa, fungi, plants or animals.

Three types of bacteria-host interactions are generally recognized. These are symbiotic, commensal, and pathogenic interactions depending on whether the influence of one species on another is positive, neutral, or negative respectively (Steinert *et al.*, 2000). The term symbiosis originates from the Greek word "Symbioun" meaning "to live together" and was defined by Anton deBary in 1879 as "the living together of two dissimilar organisms, usually in intimate association, and usually to the benefit of at least one partner"(De Bary, 1879). The partners that establish a symbiotic association are called host and symbiont. The host organism is defined as the provider of resources or the resource base, while the symbionts are the consumers of such resources, and may or may not provide services in return (Ferrière *et al.*, 2007). When the symbiont lives intracellularly within its host, the integration is called endosymbiosis and the symbiont is called an endosymbiont. Bacterial parasites, on the other hand, range from agents of infectious diseases that exploit hosts before infecting new individuals, to bacteria that are transmitted vertically and manipulate host reproduction to favor their own spread (Stouthamer *et al.*, 1999). Over the last century, parasitic infectious bacteria have received intense attention from researchers because of their clinical and economic importance (Sachs *et al.*, 2011). But it was not until relatively recent times that the research interest on beneficial microbial infections, where hosts gain new capabilities from their endosymbionts, have come into the spotlight (Webster, 2014).

Both parasitic and endosymbiotic bacteria need initially a way to contact and enter the host cell, overcome digestion and interact with the host in a way to prevent and/or escape the cellular or immunological attack (Nowack and Grossman, 2015; Gil *et al.*, 2004). However, in the case of endosymbiotic associations, the eukaryotic host additionally evolved mechanisms to control and manipulate its bacterial endosymbionts according to its needs. Unvailing the molecular mechanisms underlying host-symbiont communication and host control over an

endosymbiotic bacterium is exciting in its own right. Additionally, defining those mechanisms might help researchers to come up with new strategies to gain control over parasitic bacteria that cause infectious diseases and to gain fundamental insights into the poorly studied immune system of unicellular eukaryotes (protists).

1.2 Roles of symbiosis in different systems

Endosymbiosis has acted in the past as a major catalyst of eukaryotic evolution giving rise to mitochondria and plastids (McCutcheon and Moran, 2012). The conversion of an endosymbiont into an organelle (mitochondrion or plastid) depends on the transfer of endosymbiont genes into the nucleus of the host and the evolution of a protein import apparatus to transport proteins encoded by such transferred genes back into the organelle (Cavalier-Smith and Lee, 1985). However, there are many known examples of more recently established endosymbiotic interactions that are spread across the eukaryotic phyla where the host is provided with new physiological capabilities that allow it to thrive in formerly inaccessible niches.

In general, the endosymbionts are provided with relatively stable and nutrient rich conditions in the intracellular environment of the host. As a result, vertically transmitted endosymbionts that are permanently confined to the host intracellular environment tend to have reduced genomes when compared to their "free-living" counterparts. The genome reduction in these intracellular bacteria is characterized by the loss of genes that become superfluous or redundant in the intracellular environment, while retaining and sometimes amplifying those genes needed for functions that are beneficial for the interaction with its host (McCutcheon and Moran, 2012). In this sense, genome reduction in endosymbionts reflects functional integration within the host (Wernegreen, 2016), where the gene repertoire of the host and endosymbiont can complement each other filling gaps in metabolic pathways between the partner organisms (Alves *et al.*, 2013; Singer *et al.*, 2017). An exponential growth of genome studies has offered new opportunities to analyze the associations between endosymbionts and their hosts, introducing a rich informational framework and allowing a significant progress in the field of endosymbiosis. However, only few systems are studied at protein biochemical level. Three well studied examples of symbiotic associations are presented below.

1.2.1 *Paulinella chromatophora*

P. chromatophora is a thecate freshwater amoebae that belongs to the Cercozoa (Rhizaria). It harbors two photosynthetic organelles termed "chromatophores" that originated only 90-140 million years ago (MYA) from an α-cyanobacterium (Delaye *et al.*, 2016), through an

endosymbiotic event that is independent of the one that gave rise to plastids in the Archaeplastida (Marin *et al.*, 2005). *P. ovalis*, one of the closest non-photosynthetic relatives of photosynthetic *Paulinella* species, is preying on various bacteria including single-celled cyanobacteria, as was shown using single-cell genomics (Bhattacharya *et al.*, 2012). This finding suggests that chromatophores likely evolved from a preyed cyanobacterium that overcame digestion and was subsequently established in the cytoplasm of its former predator.

Despite the fact that the genome of the chromatophore is reduced to approximately one-third compared to the size of the genome of its free-living ancestor (Nowack *et al.*, 2008) and approximately 2,500 genes were lost, only a few more than 70 genes seem to have been transferred to the nucleus of the host cell, by a process known as endosymbiotic gene transfer (EGT) (Nowack *et al.*, 2011; Nowack *et al.*, 2016). Among the EGT-derived genes, two genes encoding low molecular weight subunits of photosystem I, PsaE and PsaK, were shown to be translated on cytoplasmic ribosomes and targeted to the chromatophore where they assembled with the chromatophore-encoded subunits of photosystem I, although their protein import pathway still remains unclear (Nowack and Grossman, 2012).

A recent study that analyzed isolated chromatophores by protein mass spectrometry identified over 200 nucleus-encoded proteins which appear to be targeted to the chromatophore (Singer *et al.*, 2017). Those chromatophore-targeted proteins were divided into two classes, the "short" and "long" import candidates. Long imported proteins (>268 amino acids) carry a conserved N-terminal sequence of approximately 200 amino acids, that the authors interpreted as a putative chromatophore transit peptide (crTP). Interestingly, when a recombinant protein carrying the *P. chromatophora* crTP at the N-terminus of the yellow fluorescent protein (crTP::YFP) was heterologously expressed in leaves of the tobacco plant *Nicotiana benthamiana*, YFP fluorescence was observed over the chloroplast suggesting that conserved features or principles are present in the protein import pathways of chromatophores and chloroplasts. In the same study, *in silico* predictions of imported proteins using conserved motives or physicochemical properties found in the crTP sequence extended the catalog of nucleus-encoded chromatophore-targeted proteins to almost 450 proteins. In contrast, short imported proteins (<90 amino acids) lack recognizable N-terminal targeting signals. Many of those short import candidates (39 proteins) contain evident cysteine motifs (CxxC or CxxxxC) and/or stretches of positively charged amino acids that are characteristic of antimicrobial peptides (Singer *et al.*, 2017).

Physicochemical properties of the crTP sequence as well as symbiotic AMPs might hold the key to understand the evolution of a protein import system that allows a eukaryotic host cell to target cytoplasmically synthesized proteins into a bacterial endosymbiont. However, barriers like the slow growth of *P. chromatophora*, its extremely large nuclear genome (approximately 9.6 Gbp) which is not fully assembled and the absence of established genetic tools make it difficult to use these cells as biological model for endosymbiosis research.

1.2.2 *Sitophilus* genus

Also in other endosymbiotic associations the host seems to have the capacity to target at least few proteins into the bacterial endosymbiont. There are three species of the weevil *Sitophilus* (*S. oryzae, S. zeamais, and S. granarius*), all of which harbor an endosymbiont of γ-proteobacterial origin, which is called *Sitophilus* primary endosymbiont (SPE). SPEs are housed within specific host cells, the bacteriocytes, which group together to form the bacteriome organ. These endosymbionts are vertically transmitted through female germ cells to the offspring and they provide at least amino acids and vitamins to the insect (Lefèvre *et al.*, 2004; Heddi, 2003). Considering its recent association as an endosymbiont (around 20 MYA), it is suggested that SPE could be a replacement of an ancestral endosymbiont in the family Dryophtoridae (Lefèvre *et al.*, 2004). In contrast to other more ancient bacterial endosymbionts, the SPE genome is only slightly reduced (3.0 Mb) and its GC content is similar to that found in the closely related free-living bacterium *Escherichia coli* (Charles *et al.*, 1997; Heddi *et al.*, 1998).

Previous studies on SPEs demonstrated that attachment and invasion of the host by the endosymbiont rely on functional genes of the type III secretion system (T3SS), which are closely related to those of the pathogenic bacteria *Salmonella* and *Shigella* (Dale *et al.*, 2002; Oakeson *et al.*, 2014). T3SS is required for secretion of proteins that help pathogenic bacteria to infect their host. Moreover, in other studies it was shown that permanent infection of bacteriocytes with SPE caused a systemic induction of coleoptericin-A (*colA*), a gene that encodes an antimicrobial peptide (AMP) (Anselme *et al.*, 2008). Immunohistochemistry experiments and immunogold electron microscopy (EM) using antibodies against ColA showed that this AMP is expressed in all tissues which are housing SPE, co-localizing with the endosymbionts in the bacteriocytes (Login *et al.*, 2011). Furthermore, the same study showed bactericidal activity of ColA against the Gram-positive *Micrococcus luteus* and the Gram-negative *Escherichia coli* strains, causing bacterial gigantism in the latter. The function of ColA in host-endosymbiont interactions needs to be studied further.

The limited understanding of the role of ColA and of the mechanisms that regulate initiation of symbiosis is mainly due to the difficulties of working with multicellular organisms, isolating insect endosymbionts from their insect hosts and cultivating them *in vitro*, and the lack of efficient molecular tools.

1.2.3 Acyrthosiphon pisum

The pea aphid *A. pisum* has a mutualistic association with the γ-proteobacterial endosymbiont *Buchnera aphidicola,* which has a greatly reduced genome. A total of 12 prokaryotic genes have been found in the *A. pisum* genome, most of which were likely acquired through horizontal gene transfers from a close relative of *Wolbachia* (α-proteobacteria, Rickettsiales) (Nikoh *et al.*, 2010) and only two truncated pseudogenes were likely acquired through endosymbiotic gene transfers from *B. aphidicola*. Among these 12 genes, five copies of the gene encoding RlpA (rare lipoprotein A), for which source and function are still unclear, were identified. Immunofluorescence microscopy and immunogold EM using anti-RlpA4 antibody verified that RlpA4 is targeted to *B. aphidicola* cells (Nakabachi *et al.*, 2014).

For all the symbiotic systems described above, the molecular mechanisms that regulate the metabolic cooperation, the protein targeting and import, and the signaling between the two partners are poorly understood. The reasons for that are the limitations derived from the absence of simple and efficient molecular tools that would enable functional studies in these endosymbiotic systems and the complexity of the multicellular organisms. In that sense, it is necessary to establish a new biological model that would provide a flexible system to develop molecular tools allowing its genetic manipulation to address critical questions regarding the evolution, mechanisms and physiological consequences of the host-endosymbiont interactions.

1.3 Trypanosomatids

Trypanosomatids are a group of protozoan parasites that belong to the family Kinetoplastidae. This group includes the species *Trypanosoma cruzi*, *Trypanosoma brucei* and *Leishmania* spp. that cause the Chagas disease, sleeping sickness and leishmaniases, respectively, infecting around 25 million people and with 400 million more at risk of infection, in tropical and subtropical areas of the planet where they cause significant economic loses and for which no available treatments exist to date (Barrett *et al.*, 2003; Guerin *et al.*, 2002). Whereas these pathogenic trypanosomatids show a dixenic life cycle characterized by several

morphologically and biologically distinct life stages alternating between invertebrate vectors and vertebrate or plant hosts, most trypanosomatid species are monoxenic (i.e. inhabit a single invertebrate host) (Wheeler *et al.*, 2013).

In general, trypanosomatids are characterized by several unique biological features such as the presence of highly compacted mitochondrial DNA arranged in concatenated maxicircles and minicircles forming a structure known as the kinetoplast (Sela *et al.*, 2008), a glycolysis compartmentalized into specialized peroxisomes called glycosomes (Haanstra *et al.*, 2016), the apparent lack of introns and promoter regions for RNA polymerase II in protein-coding genes and the transcription of genes in large polycistronic units that resemble that of prokaryotes (Araújo and Teixeira, 2011).

Due to their gene organization in polycistronic gene clusters (PGCs), mRNA processing in trypanosomatids differs remarkably from most other eukaryotes. PGCs can be transcribed either divergently (towards the telomeres) or convergently (away from the telomeres) (Martínez-Calvillo *et al.*, 2010). Polycistronic pre-mRNAs generated by RNA polymerase II transcription are processed into monocistronic mature mRNAs by cleavage into single open reading frames followed by the addition of a 39 nt trans-spliced leader sequence (SL) at the 5´-end and polyadenylation at the 3´-end of the mature transcripts (Michaeli, 2011) (Figure 1.3-1). Specific signals within the intergenic regions are necessary for trans-splicing and enhance stability of the mature transcripts (Martínez-Calvillo *et al.*, 2010). As a consequence of polycistronic transcription, regulation of gene expression occurs mostly posttranscriptionally (Haile and Papadopoulou, 2007). Especially the 3´-untranslated region has a major influence on mRNA stability and thus levels of gene expression (Nozaki and Cross, 1995).

The relatively simple arrangement of the protein coding genes in PGCs, the fact that the stability of the mature mRNAs depends on the upstream and downstream regions, the availability of fully sequenced genomes for many species and the high levels of homologous recombination have allowed the development of a plethora of genetic manipulation tools, including gene knock-out and knockdown, heterologous gene expression and CRISPR/Cas9 system, to study trypanosomatid biology (Burle-Caldas *et al.*, 2015).
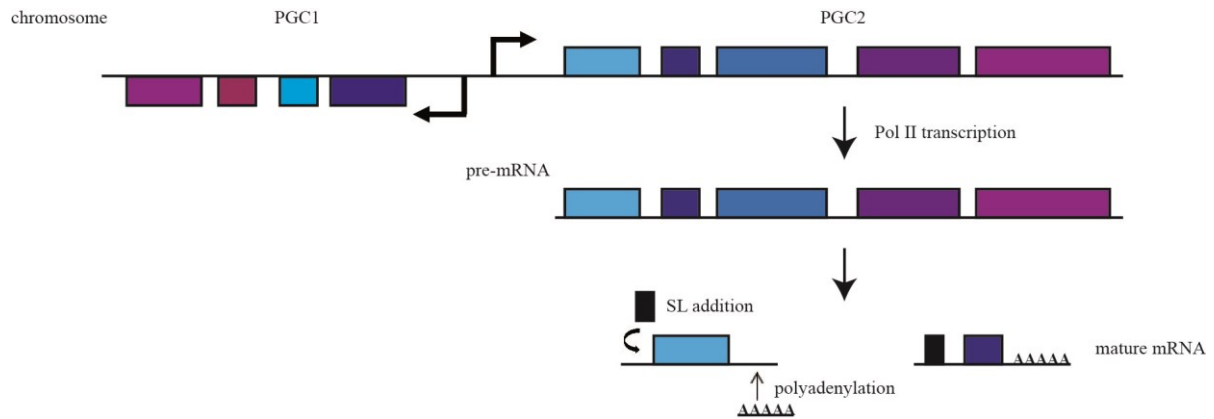
Figure 1.3-1: Gene expression in trypanosomatids. Polycistronic gene clusters (PGC) of functionally unrelated genes (arrow boxes) are transcribed as polycistronic pre-mRNAs. The polycistronic pre-mRNAs are cleaved into monocistronic mRNAs followed by the addition of a splice leader sequence (SL) through a trans-splicing reaction coupled to polyadenylation. These processing reactions are guided by polypyrimidine tracts that are present in every intergenic region. Mature mRNAs are exported to the cytoplasm where their stability depends on elements present in their untranslated regions.

1.3.1 Symbiont-harboring trypanosomatids

Although trypanosomatids are mostly known for the species that are responsible for human, animal and plant infections, there is a large diversity of non-pathogenic, monoxenic trypanosomatid species (Penha *et al.*, 2016). Among them four genera have been recognized to harbor bacterial endosymbionts: a) *Angomonas*, b) *Strigomonas*, c) *Kentomonas* and d) *Novymonas* (Teixeira *et al.*, 2011; Votýpka *et al.*, 2014; Kostygov *et al.*, 2016). The first three genera are members of the subfamily Strigomonadinae. The members of this subfamily share a common ancestor which acquired a β-proteobacterium of the family of *Alcaligenaceae*. This means that they almost certainly evolved from a single endosymbiotic event (Votýpka *et al.*, 2014; Borghesan *et al.*, 2018) (Figure 1.3-2). Their endosymbionts belong to the genus *Candidatus* Kinetoplastibacterium and are closely related to the species of *Taylorella equigenitalis* which is an intracellular pathogenic bacterium (Alves *et al.*, 2013). These β-proteobacterial endosymbionts divide in synchrony with the host cell and each daughter cell contains strictly one single endosymbiont. Whereas most trypanosomatids are nutritionally fastidious, members of the Strigomonadinae can be cultivated in defined media lacking tryptophan, phenylalanine, isoleucine, leucine, valine, heme or a purine source because many metabolites can be synthesized by the endosymbiont and delivered to the host (Noronha *et al*., 1991; Alves *et al.*, 2013).
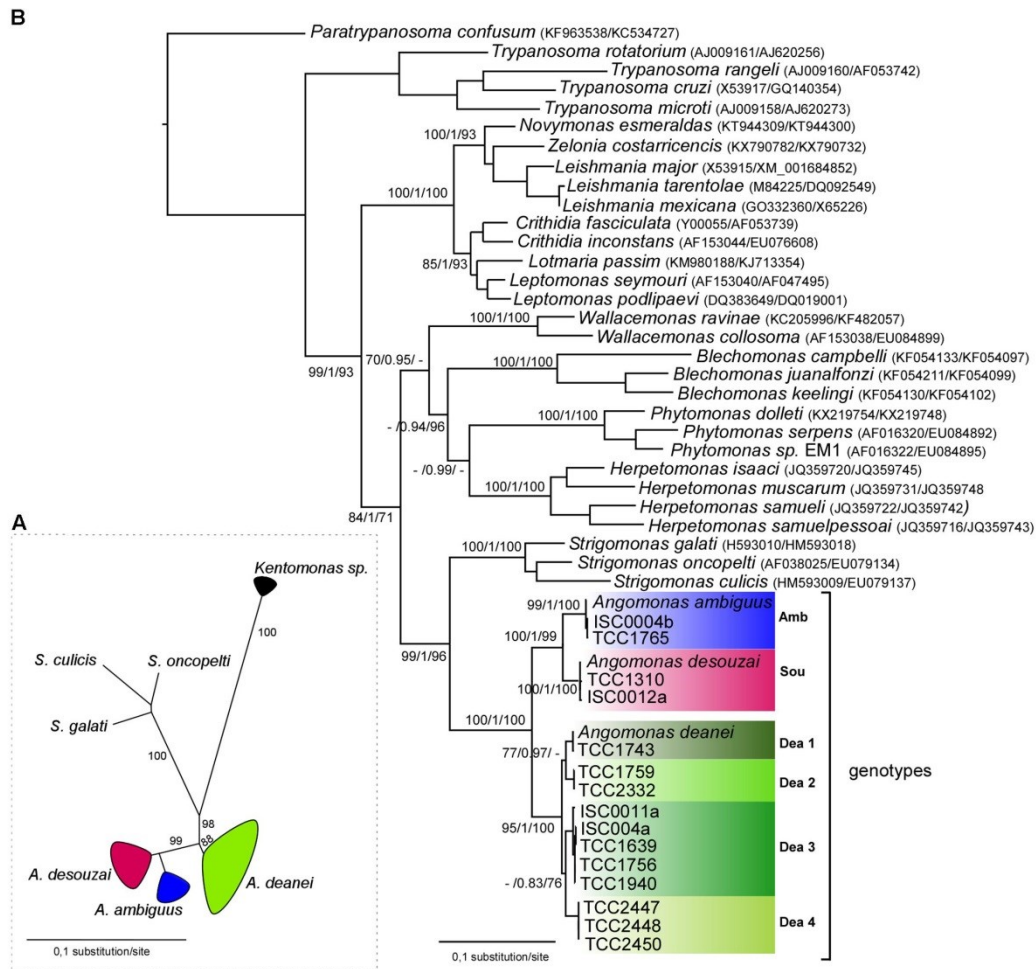
Figure 1.3-2: Phylogenetic analysis of *Angomonas* species and genotypes. (A) Dendrogram analyses performed by maximum likelihood (ML) of V7V8 SSU rRNA sequences from 175 cultures of *A. deanei*, *A. desouzai*, and *A. ambiguous*, and species of the other genera of symbiont-harboring trypanosomatids: *Strigomonas culicis, Strigomonas oncopelti, Strigomonas galati*, and *Kentomonas sorsogonicus*. Numbers at the nodes are ML support values (>50%) from 500 replicates; (B) Positioning of representatives of each the three *Angomonas* species and four genotypes of *A. deanei* (Dea1–Dea4) shown in the phylogenetic tree of Trypanosomatidae based on concatenated gGAPDH and V7V8 SSU rRNA gene sequences, inferred by ML (– Ln = –16564.401081) and Bayesian analyses. Numbers at the nodes correspond, respectively, to ML/Bayesian inference/P support values (>50%) from 500 replicates. Dea, *A. deanei*; Sou, *A. desouzai*; Amb, *A. ambiguous*. (Borghesan *et al*., 2018) http://creativecommons.org/licenses/by/4.0/

The endosymbiotic acquisition event leading to the Stringomonadinae subfamily was considered to have happened only once in the evolutionary history of trypanosomatids. However, recently a novel endosymbiotic association between the trypanosomatid *Novymonas esmeraldas* and the bacterium *Candidatus* Pandoraea novymonadis was discovered (Kostygov *et al.*, 2016). This symbiotic relationship seems to be relatively recent, since the host is not able to control the number of bacteria harbored in its cytoplasm. Like *Angomonas*, *Novymonas* is closely related to the genus *Leishmania* but does not group with the subfamily Strigomonadinae (Figure 1.3-2). Its endosymbiotic bacterium belongs to the

genus *Pandoraea* within the family *Burkholderiaceae* which is distantly related to the genus *Candidatus* Kinetoplastibacterium present in the other known endosymbiont-harboring trypanosomatids (Kostygov *et al.*, 2016). All these facts confirm an independent origin of this novel association. Considering the relatively similar microenvironment of the symbiont-harboring members of Strigomonadinae and *Novymonas* within insect's midgut it is possible that they share common biological features that allowed them to acquire an endosymbiont.

1.3.2 Angomonas deanei

The genus of *Angomonas* consists of three species*: A. deanei*, *A. desouzai* and *A. ambiguous* (Teixeira *et al.*, 2011). *A. deanei* is one of the most studied symbiont-harboring trypanosomatids. It was originally isolated from a predator reduviid, *Zelus leucogrammus* (Carvalho and Deane, 1974). Nevertheless, it is highly prevalent not only in hemiptera but also in diptera insects, prevailing in the Calliphoridae family of blowflies (Borghesan *et al.*, 2018). Moreover, *A. deanei* is considered to be a cosmopolitan species widely distributed in the Neotropical, Afrotropical, European and Papua New Guinea habitats (Borghesan *et al.*, 2018).

The body of *A. deanei* is elliptical in shape, with a prominent tail-like flagellum at its posterior end that emerges from a flagellar pocket. Its endosymbiont, *Candidatus* Kinetoplastibacterium crithidii*,* is in close proximity to the host nucleus and is surrounded by glycosomes (Figure 1.3-3) (Cristina Motta, 2010). It is enclosed by two membranes but unlike other gram-negative bacteria its periplasmic space is occupied by a reduced peptidoglycan layer, which is essential for cell division (Motta *et al.*, 1997).
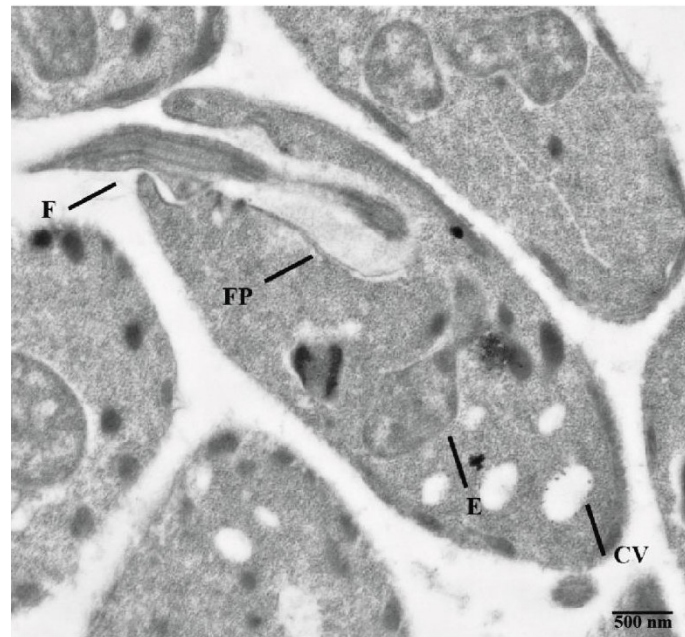
Figure 1.3-3: Transmission electron microscopy of *A. deanei*. A dividing endosymbiont (E) enveloped by two membranes, a flagellum (F) that emerges from a flagellar pocket (FP) and contractile vacuoles (CV) observed by TEM. Scale bar: 500 nm.

The genome sequence of endosymbiont and host are available (Alves *et al.*, 2013; Motta *et al.*, 2013). Comparative genomic analyses revealed that the genome of *Candidatus* Kinetoplastibacterium crithidii was reduced to half and bears half as many protein-coding genes, when compared to the genome of *T. equigenitalis*, its closest known relative. The difference in genome size is more noticeable in comparison with the closely related free-living bacterium *Achromobacter xylosoxidans*, which has 9 times larger genome compared to the genome of the endosymbiont. Among the genes that were lost are those related to transcriptional regulation, DNA repair, cell division and cell wall synthesis (Alves *et al.*, 2013). Moreover, in the closely related pathogen *T. equigenitalis* genes of the Sec-SRP, Tat, type IV and type VI secretion systems are present, whereas in the endosymbiont genome only genes of the Sec-SRP secretion system were preserved. Interestingly, genes that complement host biosynthetic pathways were retained. The bacterium supplies the host with enzymes for completing the metabolic pathways for synthesis of heme, vitamins, coenzymes, lipids and amino acids. Figure 1.3-4 shows as an example the biosynthetic pathways for the amino acids isoleucine and valine. There is an almost complete set of enzymes for isoleucine and valine biosynthesis encoded in the bacterial genome, however, the branched-chain amino acid transaminase which catalyses the last step in both pathways is missing in the endosymbiont,

but is encoded in the nucleus of the host, which highlights the level of genomic/metabolic complementation between *A. deanei* and its endosymbiont (Alves *et al.*, 2013).
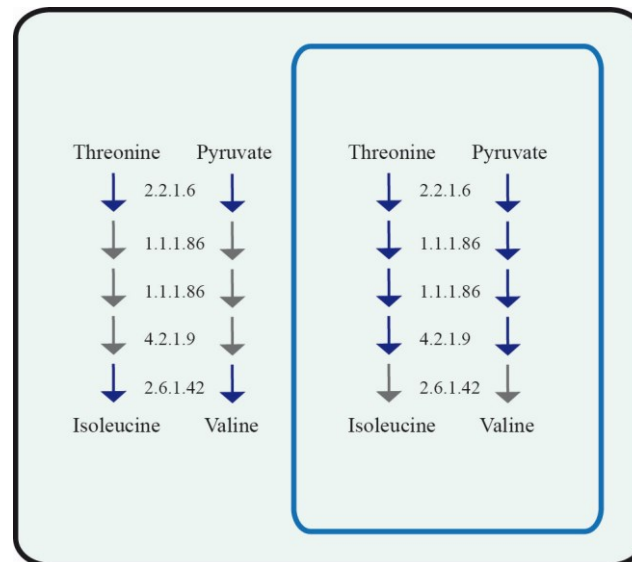
Figure 1.3-4: Isoleucine and valine biosynthetic pathway. The scheme shows the host (black square) and endosymbiont-encoded (blue square) biosynthetic pathways of isoleucine and valine. Blue arrows indicate presence of the gene in the respective compartment, while grey arrows represent their absence. 2.2.1.6, acetolactate synthase small and large subunits; 1.1.1.86, ketol-acid reductoisomerase; 4.2.1.9, dihydroxy-acid dehydratase; 2.6.1.42, branched-chain amino acid transaminase. The picture was made according to Alves *et al.*, 2013.

The endosymbiont also seems to provide the host with purines and improves the production of polyamines (Motta *et al.*, 2013), which in return complements the bacterium with ATP (Motta *et al.*, 1997) and phosphatidylcholine, an essential component of the endosymbiont membrane (de Azevedo-Martins *et al.*, 2007). Another result of the coevolution of symbiont-harboring trypanosomatids of the Strigomonadinae and their bacterial endosymbionts is the coordinated division of the bacterium with the host cell nucleus, with each daughter cell having only one symbiont (Motta *et al.*, 2010). For *A. deanei* it was shown that, inhibiting protozoan protein synthesis with cycloheximide prevented symbiont cytokinesis, indicating that components produced by the host are essential for symbiont division and may control the bacterial cell number (Catta-Preta *et al.*, 2015). Additionally, when compared to trypanosomatids that do not harbor an endosymbiont there are visible changes in the cell structure of the host which accommodates the symbiont such as the reduction of paraflagellar structure (Gadelha *et al.*, 2005), the large and loosely packed kinetoplast (Votýpka *et al.*, 2014), the change on the

surface charge of the host cell (Oda *et al.*, 1984) and the modified expression of gp63-like protease (d'Avila-Levy *et al.*, 2008).

The tightly associated pairs of host-symbiont genotypes suggest coevolution, consistent with interdependence between host and symbiont (Borghesan *et al.*, 2018). The two partners complement each other in a way that the endosymbiont is unable to survive once isolated from the host trypanosomatid. In contrast, aposymbiotic *A. deanei* cells (host cells that lack the endosymbiotic bacterium) were reported to have been obtained after treatment with Chloramphenicol and external addition of nutrients that were previously provided by the endosymbiont (Mundim and Roitman, 1977).

1.4   Aim of this thesis

The application of recent technological developments (next-generation sequencing, proteomic analysis) has allowed a significant progress in the field of endosymbiosis research and offered new opportunities to scrutinize the associations between endosymbiotic bacteria and their hosts. Despite this progress, the molecular mechanisms that regulate the metabolic cooperation, the protein targeting and import between the two partners, and synchronization of host and endosymbiont cell cycles are poorly understood. This lack of understanding is mainly caused by the absence of molecular tools enabling functional studies in current endosymbiotic models and the complexity of the multicellular systems that are being used. Therefore, this project focuses on developing a molecular toolbox for the trypanosomatid *A. deanei*. It was hypothesized that, like other trypanosomatids, *A. deanei* would be amenable to genetic manipulation using homologous recombination and transgene expression would occur by read through without the need of a dedicated promoter. Moreover, the relatively short doubling time (approximately 6 h), the availability of the genome and the metabolic interdependence of the two partners (Motta *et al.*, 2013; Alves *et al.*, 2013), in combination with the tight regulation of the division of the bacterium by the host, renders *A. deanei* an excellent model to study host-symbiont interactions.

The second part of this project focuses on the identification of host genes that might be responsible for the control of endosymbiosis. Establishing an aposymbiotic strain and comparing the transriptomes of symbiont-containing and aposymbiotic *A. deanei* cultures would be a valuable tool to identify those host genes. Additionally, previous proteomic analysis of whole cell lysates versus isolated endosymbionts that was done by Dr. Jorge

Morales revealed high enrichment of the host-encoded Endosymbiont-targeted Protein 1 (ETP1; UniProt accession S9VAC8) in the endosymbiont fraction. Assignment of the sequence to nuclear contig number 14406 (Motta *et al.*, 2013) and the presence of a typical SL at the 5´-end of the ETP1 mRNA clearly demonstrated its nuclear origin. These findings suggest that ETP1 represents a host-encoded protein that plays a role in the host-symbiont interactions. Verification of the subcellular localization of ETP1 at the endosymbiont and an initial biological characterization of this protein is the final aim of this project.

2 Results

2.1 Establishment of genetic tools for *A. deanei*

To develop a molecular toolbox that would allow the genetic manipulation of *A. deanei* the antibiotic resistances that could be used as selection markers, the loci that would be suitable for the expression of heterologous genes in *A. deanei*, the vectors which carry the resistance markers and target the loci of interest and a transfection protocol needed to be established.

2.1.1 Identification of suitable selection markers

Firstly, drugs that would be suitable selection markers able to kill wild-type *A. deanei* cells and select the cells that were successfully transfected needed to be identified. Cultures of wild-type *A. deanei* cells (1 x 10^4 cells/ml) were incubated at 28°C in the presence of different concentrations of G418 or hygromycin. It was determined that 300 µg/ml of either G418 or hygromycin were sufficient to kill wild-type *A. deanei* cells following exposure of less than 48 h (Figure 2.1-1). In both cases, the concentration 500 µg/ml was chosen for the selection of *A. deanei* transfectants to ensure that no wild-type cells would survive the treatment. These concentrations are comparable to those used in other trypanosomatids. In *T. cruzi* for example, concentrations of 100-500 µg/ml G418 or 100-1000 µg/ml hygromycin are used for the selection of transfectants (Burle-Caldas *et al.*, 2015). Additionally, it was demonstrated in our lab by an intern student (Pia Saake) that 50 µg/ml of phleomycin were sufficient to kill wild-type *A. deanei* cells and the concentration of 100 µg/ml was chosen for the selection of the cells that were successfully transfected. On the contrary, concentrations of puromycin up to 2 mg/ml were not adequate to kill wild-type cells, while concentrations up to 50 µg/ml are used for selection in other trypanosomatids (Castanys-Muñoz *et al.*, 2012). For this reason, puromycin resistance cannot be used as a selection marker in *A. deanei*.
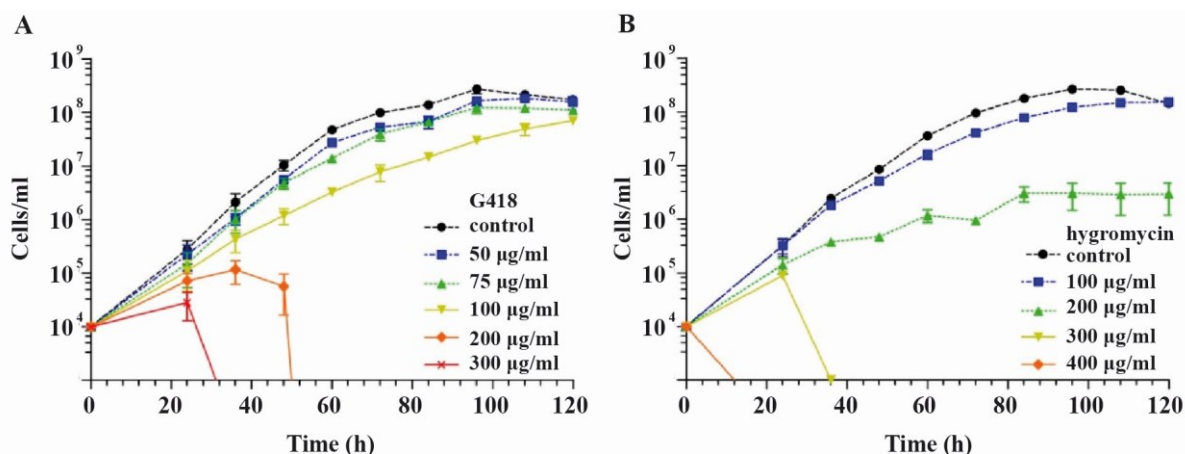
Figure 2.1-1: (A) G418 and (B) hygromycin sensitivity of *A. deanei*. Cultures of wild-type cells grown to mid-log phase were diluted in 10 ml of fresh medium to 1×10^4 cells/ml supplemented with the indicated concentration of drug and incubated at 28°C without agitation. Cells were counted every 12 h and the average and standard deviation from three independent cultures was plotted against time.

2.1.2 Selection of loci for transgene expression

Amastins are surface glycoproteins of unknown function that can be grouped into four subfamilies (termed α to δ) and occur in all trypanosomatids studied so far as tandemly arrayed multicopy genes (de Paiva *et al.*, 2015). Several members of each subfamily are encoded on the *A. deanei* genome (Motta *et al.*, 2013), but sequences of the various members of the subfamilies differ at the nucleotide and amino acid sequences and their intergenic regions show no homology. *Amastin* loci have been previously used to enhance expression of heterologous genes in other trypanosomatids (Araújo *et al.*, 2011; Boucher *et al.*, 2002) and deletion of single *amastin* genes by knock-out does not seem to affect viability of insect stages in several trypanosomatid species (de Paiva *et al.*, 2015). Specifically, in *A. deanei*, the loci for γ- and δ-*amastin* are among the most highly expressed genes as shown by RNAseq data (Morales *et al.*, 2016). For these reasons, γ- or δ-*amastin* genes were selected to initially develop a protocol for targeted gene knock-outs via homologous recombination (HR) in *A. deanei* through replacement of these genes with antibiotic resistance genes. Then, the γ- or δ-*amastin* loci were used for expression of heterologous genes in *A. deanei*.

2.1.3 Vector design

To test for targeted insertion by HR and expression of heterologous genes in *A. deanei*, plasmids containing cassettes consisting of the amino 3'-glycosyl phosphotransferase (*neo*) or hygromycin B phosphotransferase (*hyg*), to provide resistance markers against G-418 and

hygromycin B, respectively, flanked by 1000 bp of the 5′ and 3′ flanking regions of the γ- or δ-amastin genes were constructed. In total, four plasmids (pAdea-γ-AMA/HYG (pAdea004), pAdea-γ-AMA/NEO (pAdea003), pAdea-δ-AMA/HYG (pAdea002), and pAdea-δ-AMA/NEO (pAdea001)) were obtained carrying the cassettes which were excised by restriction reactions before the transfection into *A. deanei* (as shown in Figure 2.1-2).
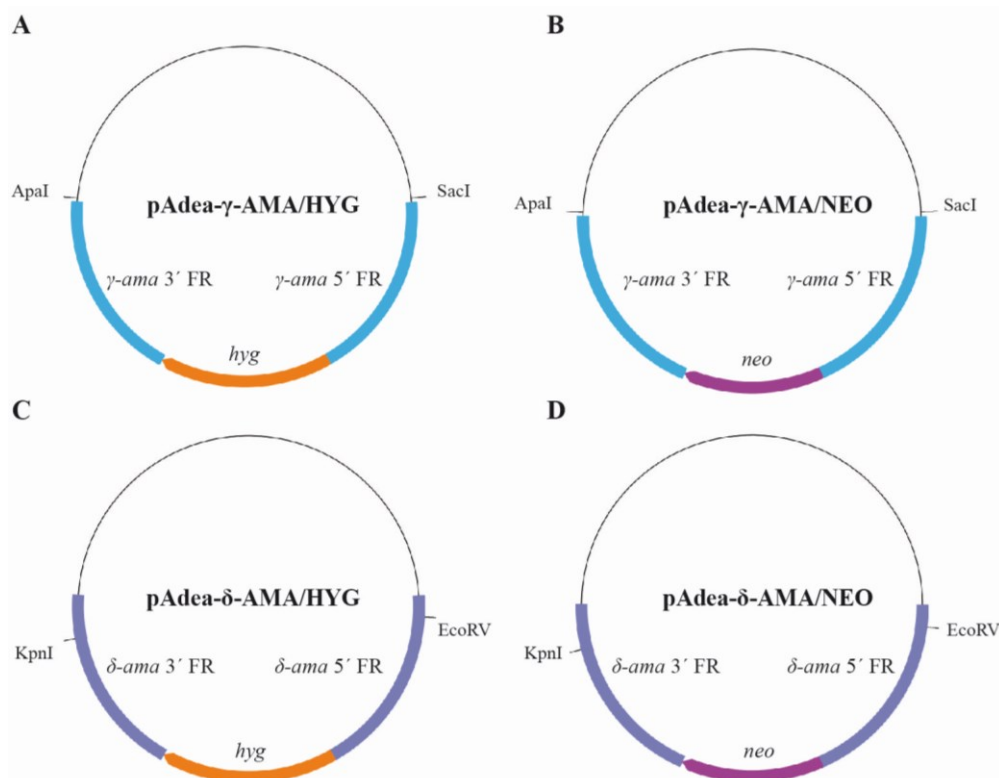
Figure 2.1-2: pAdea series of plasmids containing the *neo* and *hyg* replacement cassettes. Fragments of 1 kb of the 5′- and 3′-flanking regions (FRs) of the *γ-amastin* gene (A and B) or *δ-amastin* gene (C and D) were amplified from *A. deanei* gDNA and the selectable marker genes (*hyg* or *neo*) were inserted between the FRs. The entire *γ-* or *δ-amastin* coding sequences were eliminated and replaced with the drug resistance gene which maintained the same orientation as the original *amastin* target genes. The restriction sites indicate the excision product used for the transfections described in this work. Vector backbone: pGEM-Teasy (Invitrogen); pAdea-γ-AMA/HYG, pAdea004; pAdea-γ-AMA/NEO, pAdea003; pAdea-δ-AMA/HYG, pAdea002; pAdea-δ-AMA/NEO, pAdea001.

2.1.4 Transfection and efficient homologous recombination in *A. deanei*

Transfection of *A. deanei* was performed with the drug resistant cassettes by electroporation (as described in 4.2.2.3). Transfectants were selected in 500 µg/ml of G418 or hygromycin for approximately 2 weeks, and clonal cultures were obtained by limiting dilutions.

Cells transfected with the pAdea-γ-AMA/HYG (pAdea004) cassette were analyzed by PCR using genomic DNA (gDNA) from transformants (compared to the wild-type gDNA) as a template and primers that anneal to gDNA just outside of the flanking region (FR) sequences

used for targeting. This analysis showed a product at 3.2 kb corresponding to the *γ-amastin* locus and an additional PCR product at 3.6 kb corresponding to the *hyg* insertion (Figure 2.1-3A).
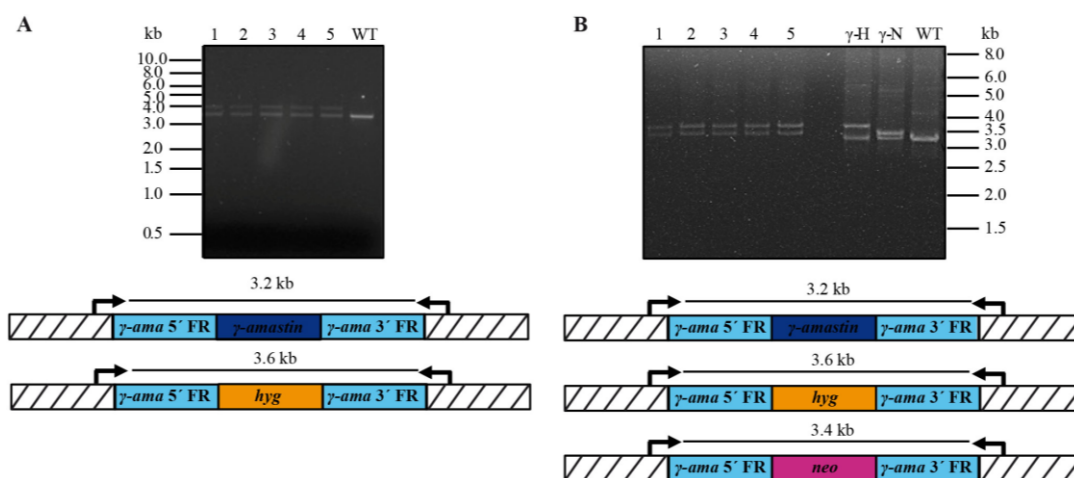
Figure 2.1-3: High efficiency of homologous recombination in *A. deanei*. Around 2 weeks after transfection, clonal cultures were generated by limiting dilution. A week later, around 20 clones were picked, the gDNA extracted, and the insertion analyzed by PCR. For the sake of clarity the results from 5 clones are shown, but in all clones the results were the same. (A) A single band at 3.2 kb corresponding to the *γ-amastin* locus can be seen in the wild-type (WT), while an additional band at 3.6 kb corresponding to the insertion of *hyg* appears in all Δ-γ-AMA[HYG] (Adea0004) SKO clones examined; (B) in all Δ-γ-AMA[HYG]/Δ-γ-AMA[NEO] (Adea0009) DKO clones examined, the expected bands for the insertion of *hyg* and *neo* appear at 3.6 and 3.4 kb, respectively, while the band at 3.2 kb corresponding to the WT *γ-amastin* locus disappears. For comparison also gDNA from WT and the SKO lines Δ-γ-AMA[HYG] (γ-H) and Δ-γ-AMA[NEO] (γ-N) was tested. The **striped rectangles** represent genomic regions up and downstream of the insertion sites for the cassettes; **light blue rectangles** represent the 5´- and 3´-flanking regions (FRs) of the *γ-amastin* gene, that are present in the cassettes; **blue rectangles** represent the *γ-amastin* ORFs; *hyg* and *neo* are represented by **orange** and **pink rectangles**, respectively.

Sequence analysis of PCR products using primers binding to the gDNA outside of the cassette and inside the resistance markers confirmed the insertion of the γ-AMA/HYG cassette (Figure 2.1-4).
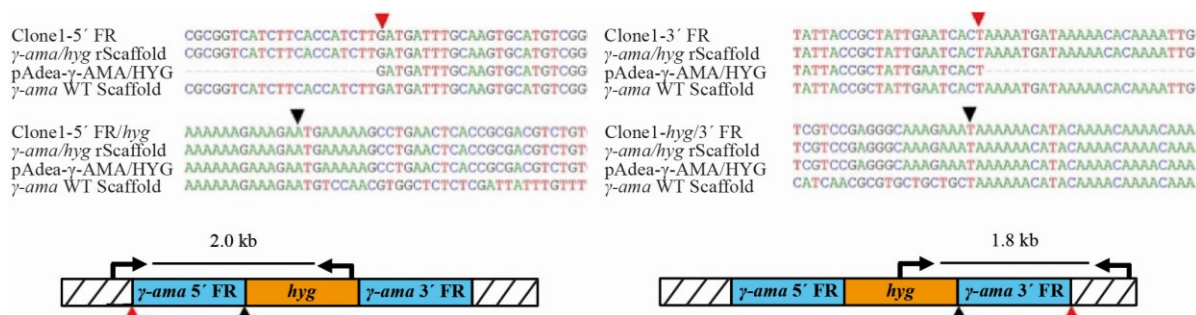
Figure 2.1-4: Sequence analysis confirms insertion of resistance cassette in the *γ-amastin* locus. Genomic DNA (gDNA) extracted from Δ-γ-AMA^{HYG} (Adea0004) SKO clone 1 was used as a template for PCR with primers binding to the gDNA outside of the cassette and inside the resistance marker. The PCR products obtained were sequenced. The alignment shows: **Clone1-5′ or 3′ FR**, which are sections of the genomic sequence of the transformant resulting from HR between the gDNA and the resistance cassette (**pAdea-γ-AMA/HYG**); **clone1-5′ FR/*hyg* and clone1-*hyg*/3′ FR,** which are sections of the genomic sequence of the transformant surrounding the start and stop codon of the *hyg*, respectively; ***γ-ama/hyg* rScaffold**, which are the expected genomic sequence organization of the recombinant *γ-ama/hyg* locus; ***γ-ama* WT Scaffold**, which is the wild-type sequence at the *γ-amastin* locus. The map underneath the alignment indicates primer binding sites (**arrows**) and sizes of resulting PCR products. **Red arrowheads** indicate the position at which the cassette inserted by HR into the genome and **black arrowheads** denote the start and stop codon of the *hyg* ORF. The **striped rectangles** represent genomic regions upstream and downstream of the insertion sites for the cassettes. **Light blue rectangles** represent the 5′- and 3′-flanking regions (FR) of the *γ-amastin* gene that are present in the cassette. **Orange rectangles** represent *hyg*.

In addition, Southern blot analysis using gDNA of the transformants restricted with *Apa1* and *SgrA1* revealed a single band at 4.7 kb hybridizing to the hygromycin probe and a single band at 2.7 kb hybridizing to the γ-amastin probe, with reduced hybridization signals in the transformants Δ-γ-AMA^{HYG} (Adea004) compared to wild-type (Figure 2.1-5A). Taken together these results indicate that HR occurred specifically at the targeted locus, resulting in Δ-γ-AMA^{HYG} single knock-out (SKO) cell lines. Also for the other three plasmids pAdea-γ-AMA/NEO (pAdea003), pAdea-δ-AMA/NEO (pAdea001) and pAdea-δ-AMA/HYG (pAdea002) (Figure 2.1-2) the same strategy was followed in order to generate SKO cell lines with similar results.

Since trypanosomatids are asexual diploids, the generation of null mutants requires disruption of both alleles of any specific gene (i.e. generation of a double knock-out (DKO)). The DKO Δ-γ-AMA^{HYG}/Δ-γ-AMA^{NEO} (Adea0009) was readily obtained by transfection of the Δ-γ-AMA^{HYG} (Adea0004) SKO cell line with the *Apa1*/*Sac1* restriction fragment from pAdea-γ-AMA/NEO (pAdea003), which includes the *neo* resistance marker flanked by *γ-amastin* FRs (Figure 2.1-2). PCR analysis using the Δ-γ-AMA^{HYG}/Δ-γ-AMA^{NEO} (Adea0009) DKO cell line showed a product at 3.6 kb corresponding to the *hyg* insertion and an additional PCR product at 3.4 kb corresponding to the *neo* insertion (Figure 2.1-3B). In addition, Southern blot

analysis using the DKO gDNA restricted with *ApaI* and *SgrAI* showed bands of expected sizes for both *hyg* and *neo* insertion while no signal was detected when the γ-amastin probe designed against the native gene was used, indicating the complete removal of this gene from the genome (Figure 2.1-5B).
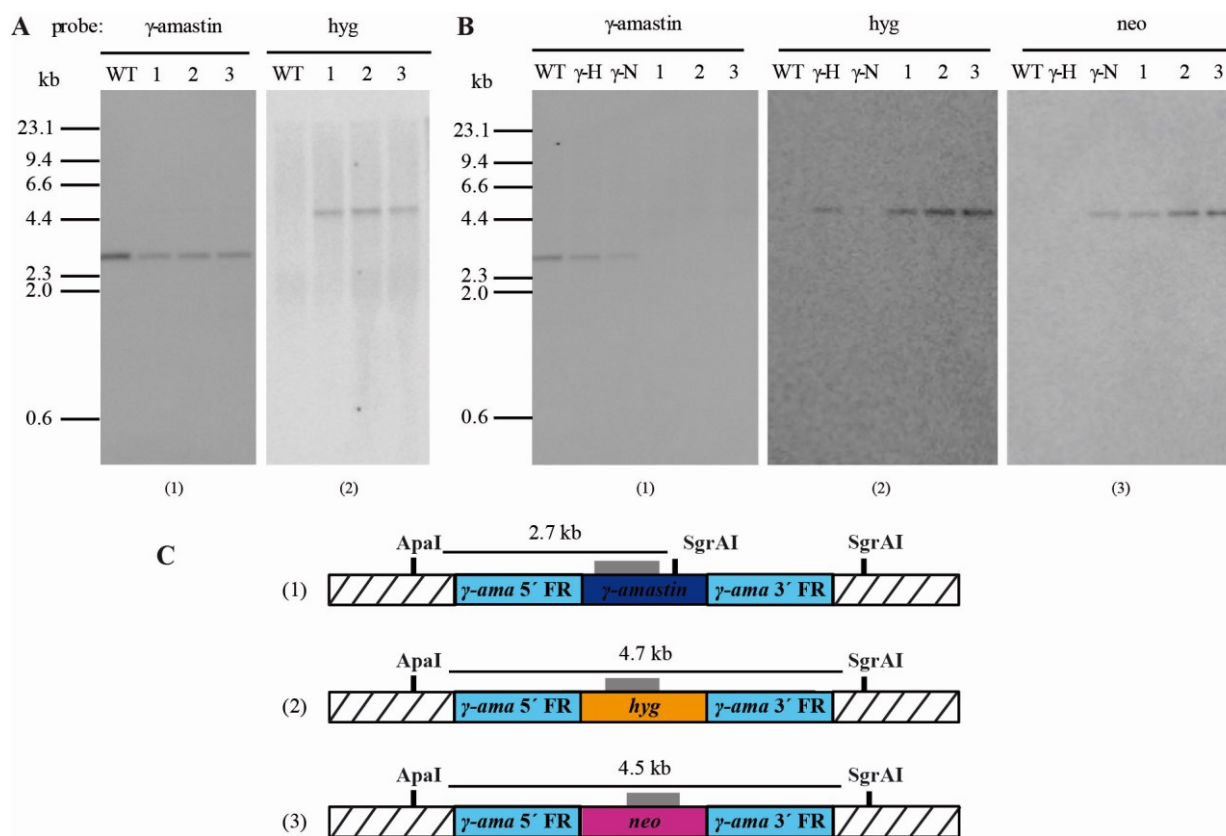
Figure 2.1-5: Homologous recombination is highly specific in *A. deanei*. Genomic DNA was restricted with the specified enzymes and analyzed by Southern blots using probes against *γ-amastin*, *hyg*, or *neo* as indicated above each blot. The wild-type (WT) and three clones were analyzed for Δ-γ-AMA^HYG (Adea0004) SKO (A) and Δ-γ-AMA^HYG/Δ-γ-AMA^NEO (Adea0009) DKO (B). (A) In the Δ-γ-AMA^HYG SKO cell lines, a band shift of 2 kb between the *γ-amastin* WT locus and the recombinant locus containing *hyg* could be readily observed in all three clones analyzed, indicating replacement of the first *γ-amastin* allele by *hyg*; (B) a second transfection with a cassette encoding *neo* targeting the *γ-amastin* locus, completely abolished the *γ-amastin* WT locus in all three Δ-γ-AMA^HYG/Δ-γ-AMA^NEO DKO clones analyzed; for comparison also gDNA from the SKO lines Δ-γ-AMA^HYG (**γ-H**) and Δ-γ-AMA^NEO (**γ-N**) was tested; (C) Restriction maps are provided showing the expected band size depending on the DNA probe that was used in each case (**1**, **2** and **3**). The **striped rectangles** represent genomic regions up and downstream of the insertion sites for the cassettes; **light blue rectangles** represent the 5′- and 3′-flanking regions (FR) of the *γ-amastin* gene, that are present in the cassettes; **blue rectangle** represents the *γ-amastin* ORF; *hyg* and *neo* are represented by **orange** and **pink rectangles**, respectively; **gray bar** highlights the binding region for the DNA probe used for hybridization.

There were no apparent differences in cell growth between wild-type cells and any of the recombinant cell lines generated, suggesting that a SKO or DKO of the *γ-amastin* gene, or a SKO of *δ-amastin* are not detrimental to the cells, at least under the conditions that were used for growth (Figure 2.1-6).
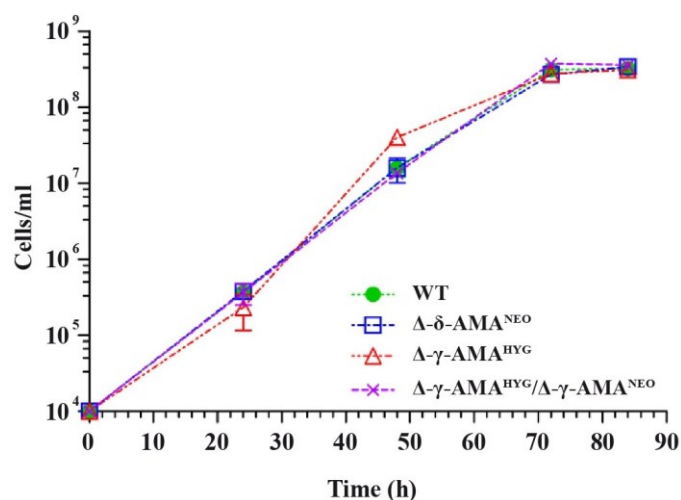
Figure 2.1-6: Knock-out of the *amastin* genes does not affect the growth of *A. deanei*. Cells from mid-log phase cultures of the Δ-δ-AMA^{NEO} (Adea0001) and Δ-γ-AMA^{HYG} (Adea0004) SKOs and the Δ-γ-AMA^{HYG}/Δ-γ-AMA^{NEO} (Adea0009) DKO cell lines were diluted to 1 x 10^4 cells/ml in fresh growth medium supplemented with 500 μg/ml of the corresponding drug(s) and cells were counted every day. Values represent average and standard deviation of three independent experiments.

Additionally, in order to test the stability of genomic insertions in the absence of selection markers, ten clones of the Δ-δ-AMA^{NEO} (Adea0001) SKO were grown in the absence of G418 for a period of more than two months. As demonstrated by the PCR banding patterns using primers that anneal to gDNA just outside of the FR sequences, integrations of the inserted resistance cassettes were stable even after removing the selection pressure for the tested time period (Figure 2.1-7).
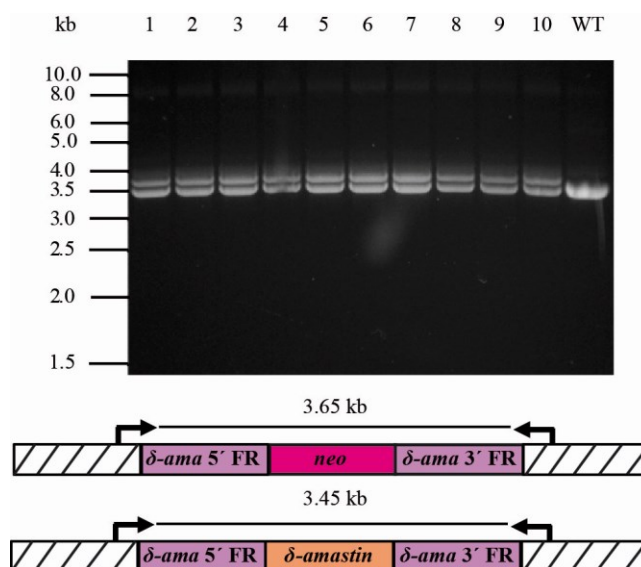
Figure 2.1-7: Stability of genomic insertion in *A. deanei*. Ten clones of the Δ-δ-AMA^NEO (Adea0001) SKO were grown in the absence of G418 for a period of two months. Then, gDNA was extracted and presence of *neo* tested by PCR. A single band at 3.45 kb corresponding to the *δ-amastin* gene can be seen in the wild-type (WT), while an additional band at 3.65 kb corresponding to the insertion of *neo* appears in all 10 Δ-δ-AMA^NEO clones examined (lanes 1-10). The **striped rectangles** represent genomic regions up and downstream of the insertion sites for the cassettes; **purple rectangles** represent the 5′- and 3′-flanking regions (FRs) of the *δ-amastin* gene, that are present in the cassettes; **orange rectangle** represents the *δ-amastin* ORF; *neo* is represented by **pink rectangle**.

2.2 Establishment of an aposymbiotic culture

In the 1970ies it has been reported that the host trypanosomatid *A. deanei* loses its bacterial endosymbiont when treated with chloramphenicol (CAM) while metabolites that are normally provided by the endosymbiont are externally provided, thus generating a cured or aposymbiotic strain (Mundim and Roitman, 1977). Establishing an aposymbiotic strain would be a valuable tool to identify host genes that are involved in control of endosymbiosis by characterizing and comparing the transcriptomes and proteomes of symbiont-containing and aposymbiotic *A. deanei* cultures.

In order to generate an aposymbiotic culture of *A. deanei* ATCC PRA-265 (i.e. the strain used in this thesis), initially, *A. deanei* wild-type cells were treated with different concentrations of CAM (10-800 µg/ml). CAM inhibits bacterial protein synthesis by inhibiting the peptidyl transferase activity of the bacterial ribosome (Kostopoulou *et al.*, 2011). Cells were grown at 28°C in the same medium as *A. deanei* wild-type cells (as described in 4.1.2.1) supplemented with 10 mg/ml folic acid, since the aposymbiotic strain has additional growth requirements (Mundim and Roitman, 1977). The presence or absence of the symbiont was determined by fluorescence in situ hybridization (FISH) using a 5′-Cy3 labeled probe against the bacterial

16S rRNA (Eub33) (Amann *et al.*, 1996). At low concentrations of CAM (10-400 µg/ml) cells retained the endosymbiont. Interestingly, at higher concentrations (500-800 µg/ml) cells that lost their endosymbiont were initially obtained. However, attempts to generate clonal cultures of aposymbiotic cells failed (i.e. all cell lines recovering from limiting dilutions of CAM-treated *A. deanei* cultures contained an endosymbiont and showed resistance against high concentrations (500-800 µg/ml) of CAM) (Figure 2.2-1A). In similar experiments, various concentrations of ciprofloxacin (inhibition of DNA gyrase), and a combination of streptomycin (inhibition of protein synthesis) and penicillin (inhibition of peptidoglycan synthesis) were also tested (Table 2.2-1). Clonal cultures were obtained by limiting dilutions in each case of antibiotic treatment. Nevertheless, in all cases *A. deanei* failed to lose their endosymbiont (Figure 2.2-1). These findings suggest that in *A. deanei* strain ATCC PRA-265 presence of the endosymbiont became obligate for host survival.

Table 2.2-1: Treatment of *A. deanei* cells with various concentrations of different antibiotics failed to generate an aposymbiotic strain. The table provides the concentrations of the antibiotics that were used, the effect that these antibiotics had on the endosymbiont or *A. deanei* cells and what was observed after FISH analysis of the clonal cultures

Antibiotic concentrations	Effect on endosymbionts	Effect on *A. deanei* cells	FISH analysis of clones recovered from limiting dilution
10-400 µg/ml Chloramphenicol	No effect	No effect	Endosymbionts identified in all clones
500-800 µg/ml Chloramphenicol	Negative	Negative	Endosymbionts identified in all clones
10-100 µg/ml Ciprofloxacin	Negative	No effect	Endosymbionts identified in all clones
250 µg/ml Streptomycin and 100 U/ml Penicillin	No effect	No effect	Endosymbionts identified in all clones
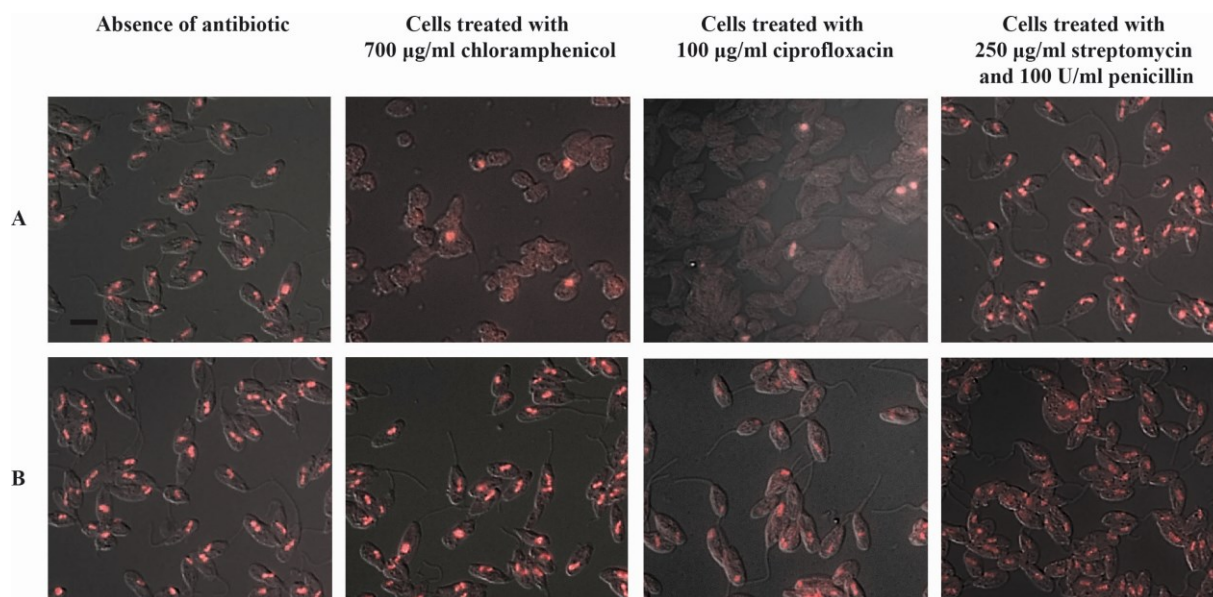
Figure 2.2-1: Treatment of *A. deanei* ATCC PRA-265 cells with various concentrations of different antibiotics failed to generate an aposymbiotic strain. (A) Cultures of wild-type *A. deanei* ATCC PRA-265 cells grown to mid-log phase were diluted in 10 ml of fresh medium to 1 x 10^5 cells/ml supplemented with 10 mg/ml folic acid and 700 μg/ml chloramphenicol or 100 μg/ml ciprofloxacin or 250 μg/ml streptomycin and 100 U/ml penicillin and incubated at 28°C without agitation. Five days after the treatment of the cells with the indicated concentration of drug the presence of the symbiont was checked by fluorescence in situ hybridization (FISH) using a 5′-Cy3 labeled probe against the bacterial 16S rRNA (Eub33). (B) Then, clonal cell lines were generated by limiting dilution of these antibiotic-treated *A. deanei* cells. All cell lines recovering from limiting dilution in the presence of the indicated concentration of drugs were checked by FISH and contained an endosymbiont. Scale bar: 2 μm.

2.3 Biological characterization of ETP1

2.3.1 Sequence analysis of ETP1

Since no aposymbiotic strain of *A. deanei* ATCC PRA-265 could be obtained, an alternative approach was chosen to identify symbiosis-related host genes. As described above, proteomic analysis of the whole cell lysate versus the purified endosymbiont fraction, that was previously conducted in our lab by Dr. Jorge Morales, revealed enrichment of the host-encoded Endosymbiont-targeted Protein 1 (ETP1) in the endosymbiont fraction. Using *A. deanei* genome assembly released in 2013 (Motta *et al.*, 2013), ETP1 was assigned to the nuclear contig number 14406. However, this gene-oriented genome assembly is still very fragmented (contig N50=2.5 kb, GenBank: ATMG00000000, Motta *et al.*, 2013). Blast analysis using a new *A. deanei* genome assembly with more than 10-fold improved linkage, that was obtained by our lab in collaboration with Daniel Rokhsar's lab at the DOE, Joint Genome Institute in Walnut Creek, California (contig N50=28.8 kb, GenBank: LXWQ00000000, Chapman *et al.*, 2016), revealed that there are two isoforms of ETP1.

"ETP1 Q-less", encoded by scaffold 1854, consists of 381 amino acids and "ETP1 Q-rich", encoded by scaffold 3029, contains a glutamine-rich (Q-rich) stretch and a couple of mutations at amino acid positions 141 and 142 (both E to Q), 263 (R to S) and 264 (I to L) as compared the Q-less version (Figure 2.3-1). Scaffold 1854 is 110 kb and scaffold 3029 is 2.2 kb long. Both scaffolds do not cover the full *etp1* ORF as shown in figure 2.3-1. More specifically, scaffold 3029 ends 50 nucleotides before the end of *etp1* ORF, whereas scaffold 1854 is fragmented into several contigs and misses also the last 50 nucleotides of *etp1* ORF and a part of its downstream region (*etp1* 3′ FR), but provides downstream -after a gap of around 1300 nucleotides- more sequence information that is linked to the *etp1* locus. In order to obtain the full sequence of *etp1* ORF and the *etp1* 3′ FR, primers that bind to the end of the known sequence of *etp1* ORF and the beginning of the next contig of scaffold 1854 (Figure 2.3-2) were used to amplify and sequence the missing part. The sequence result of *etp1* 3′ FR that was acquired by this procedure is shown in Appendix (Figure A1). Considering that *A. deanei* is a diploid organism, the upstream intergenic regions of the two scaffolds show only minor differences, and in both scaffolds the same gene (*AGDE_03718*) sits around 0.5 kb upstream of *etp1* gene, scaffold 1854 and scaffold 3029 likely can be considered as two slightly divergent alleles. Reanalysis of the proteomic data using a database containing amino acid sequence information from both gene models showed evidence for both ETP1 forms being enriched in the endosymbiont fraction. Unless otherwise stated, the term ETP1 in this thesis refers to the "ETP1 Q-less" form, for the sake of simplicity.
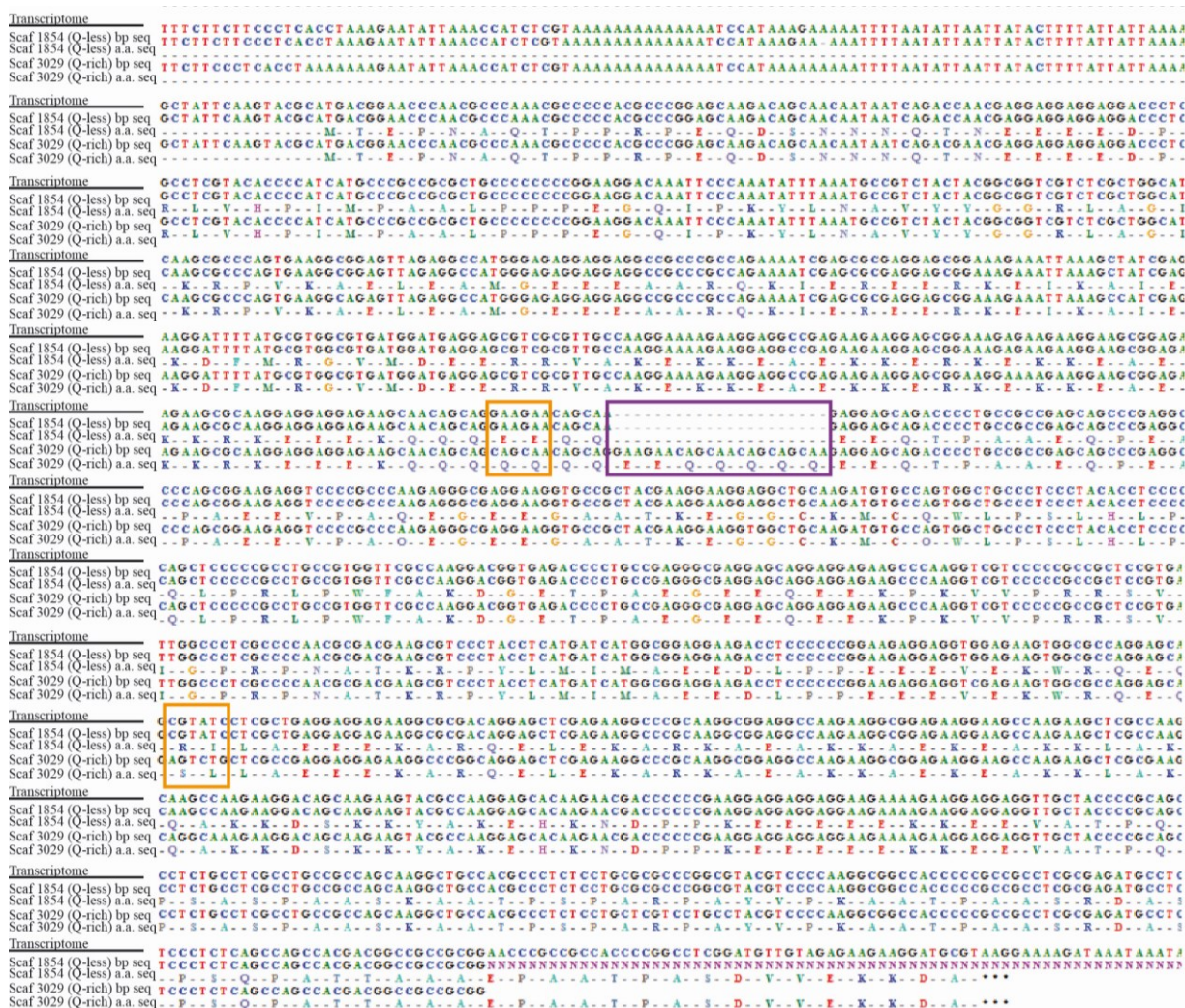
Figure 2.3-1: Alignment of the ORFs corresponding to gDNA scaffold 1854 (Q-less) and gDNA scaffold 3029 (Q-rich) forms of *etp1* in *A. deanei*. Only the Q-less isoform of *etp1* was identified in the *A. deanei* transcriptome; the "ETP1 Q-less" isoform is missing a seven amino acid sequence highlighted by the **purple rectangle**; Q-less and Q-rich isoform sequences differ in four amino acids highlighted by the **orange rectangles**; **bp** and **a.a. seq** refers to the nucleotide and amino acid sequence of the two isoforms, respectively.

Figure 2.3-2: Analysis of the 3′ end of *etp1* ORF and *etp1* 3′ flanking region corresponding to gDNA scaffold 1854 (Q-less). Genomic DNA extracted from *A. deanei* WT cells was used as a template for PCR with primers (ETP1 seq pr fw and ETP1 seq pr rev) binding to the gDNA close to end of *etp1* ORF and in the beginning of the next contig of scaffold 1854. The PCR product obtained was sequenced. The alignment shows: **scaffold 1854 (Q-less) bp and a.a. sequence** which is the nucleotide and amino acid sequence of scaffold 1854 close to the 3′ end of *etp1* ORF and its 3′ flanking region; **black arrows** and **boxes** represent the primer binding sites.

ETP1 has a molecular weight of 43.5 kDa and its theoric isoelectric point was estimated to be 5.2 using the Isoilectric Point Calculator (IPC) (Kozlowski, 2016). Blastp (Altschul *et al.*, 1997) analysis (against database Reference proteins) identified no homologs in other species and DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST) (Boratyn *et al.*, 2012) showed that ETP1 has a possible glycosyltransferase domain. Using PRED-TMBB (Bagos *et al.*, 2004) it was predicted that ETP1 is not a β-barrel outer membrane protein and no transmembrane domains were predicted using TMHMM (transmembrane hidden Markov model) (Sonnhammer *et al.*, 1998). Interestingly, two potential N-linked glysosylated sites (amino acid position 19 and 234 in ETP1) were predicted by GlycoEP (Prediction of glycosites in eukaryotic glycoproteins) (Chauhan *et al.*, 2013) and two predicted sites for palmitoylation (amino acid position 176 and 179 in ETP1) were observed using GPS (Group-based Prediction System)-Lipid predictor (Xie *et al.*, 2016).

2.3.2   Localization of ETP1 in *A. deanei* cells

The newly established genetic tools were used to confirm the specific targeting of ETP1 to the endosymbiont. Two expression vectors were constructed based on the previous pAdea-δ-AMA/NEO (pAdea001) by adding between the *neo* resistance gene and the 3′-FR of the δ-*amastin* gene a nucleotide sequence containing the intergenic region between the *A. deanei* glyceraldehyde 3-phosphate dehydrogenase I and II genes (GAPDH IR) and the *etp1*

26

sequence fused to the N- or C-terminus of enhanced green fluorescent protein (EGFP) resulting in the plasmids pAEX-nGFP-ETP1 (pAdea039) and pAEX-ETP1-cGFP (pAdea040) (Figure 2.3-3A). The GAPDH IR contains all the signals necessary for transcript maturation and was used before to increase the stability of the mRNAs of heterologously expressed genes in *T. brucei* and *T. cruzi* (Michels *et al.*, 1986; Kelly *et al.*, 1992). Furthermore, suitability of this sequence for the separation of two heterologously expressed transgenes in *A. deanei* had been demonstrated before in our lab by Dr. Jorge Morales (Morales *et al.*, 2016). For transfection, the expression cassettes were excised from the plasmids pAEX-nGFP-ETP1 and pAEX-ETP1-cGFP at the *Eco*RV restriction sites, as shown in Figure 2.3-3A. Transfectants were selected in 500 µg/ml G418 and clonal cultures obtained by limiting dilution were verified for the correct insertion at the *δ-amastin* locus by PCR analysis using primers against the regions next to the FRs, as described before. Δ-δ-AMA$^{\text{nGFP-ETP1, NEO}}$ (Adea0035) and Δ-δ-AMA$^{\text{ETP1-cGFP, NEO}}$ (Adea0036) clones were analyzed by FISH using a 5′-Cy3 labeled probe against the bacterial 16S rRNA (Eub33) in combination with an immunofluorescence assay (IFA) using a GFP-specific primary antibody. Epifluorescence microscopy showed the specific targeting of both, the N- and C-terminal fusions of ETP1 to EGFP to the bacterial endosymbiont, indicating that the targeting function of the ETP1 sorting signal was maintained in both of the fusion proteins (Figure 2.3-3B). The same fluorescence pattern was observed by direct analysis of the EGFP signal (Figure 2.3-4). However, this type of analysis cannot be combined with FISH since the treatment of the sample during FISH damages GFP signal.
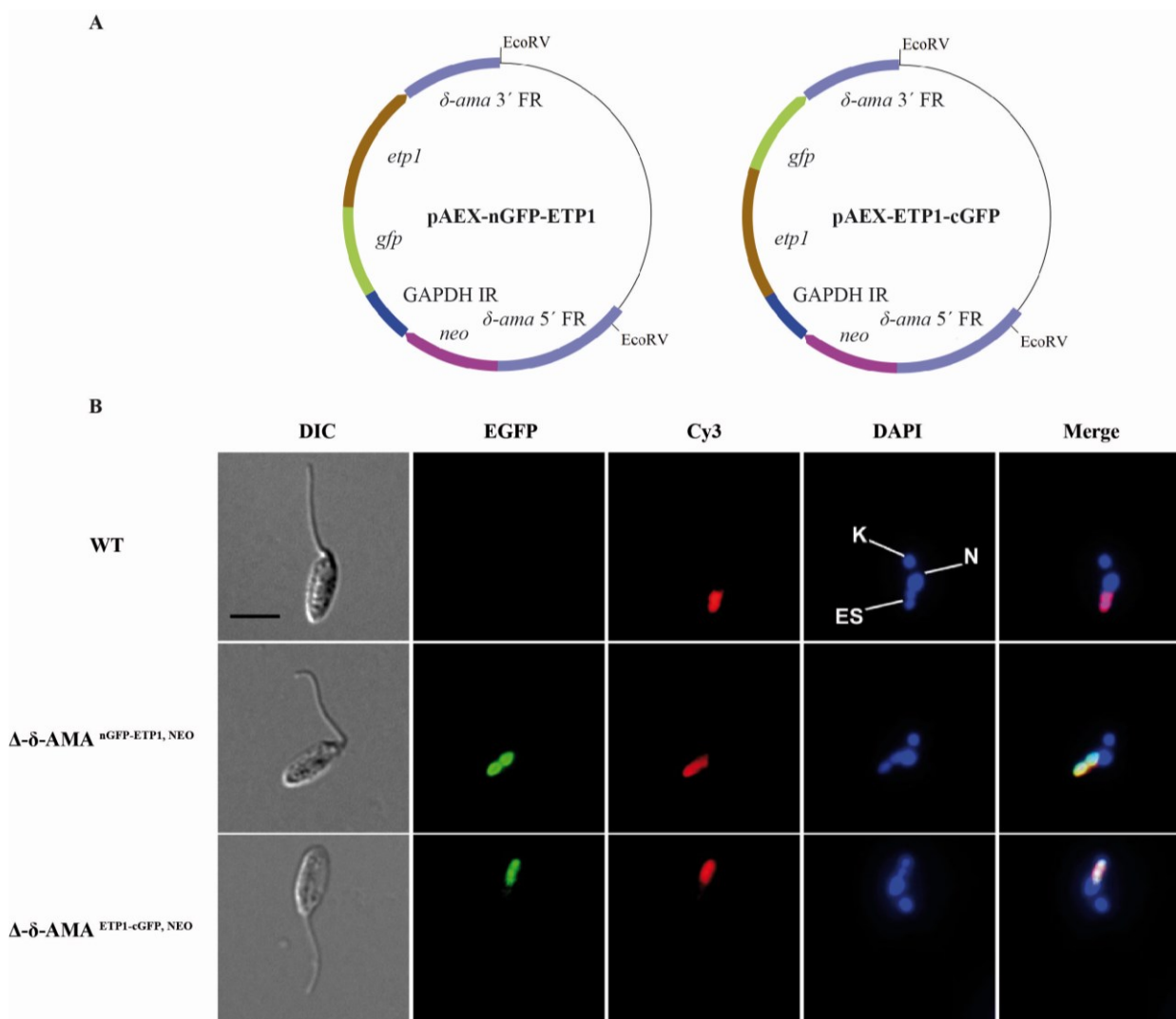
Figure 2.3-3: The host-encoded ETP1 is specifically targeted to the β-proteobacterial endosymbiont in A. deanei. (A) pAEX-nGFP-ETP1 (pAdea039) and pAEX-ETP1-cGFP (pAdea040) plasmids for expression of ETP1 fused to the C- or N-terminus of EGFP, respectively. **Purple rectangles**, *δ-amastin* 5′- and 3′-flanking regions (*δ-ama* FRs); **blue rectangle**, glyceraldehyde 3-phosphate dehydrogenase I-II intergenic region (GAPDH IR); **pink arrow**, *neo*; **green rectangle or arrow**, *enhanced green fluorescence protein* (*gfp*) gene; **brown rectangle or arrow**, *etp1* gene. The restriction sites indicate the excision product used for the transfections described in this work. (B) Cells expressing the N- or C-terminal fusions of ETP1 to EGFP were analyzed by FISH-IFA using a 5′-Cy3 labeled probe against the bacterial 16S rRNA and an α-GFP IgG as described in 4.2.2.5. **Δ-δ-AMA**[nGFP-ETP1, NEO] (Adea0035) and **Δ-δ-AMA**[ETP1-cGFP, NEO] (Adea0036), cell lines expressing in *δ-amastin* locus nGFP-ETP1 or ETP1-cGFP fusions, respectively; **ES**, eight-shaped endosymbiont; **K**, kinetoplast; **N**, nucleus. Scale bar: 2 µm.
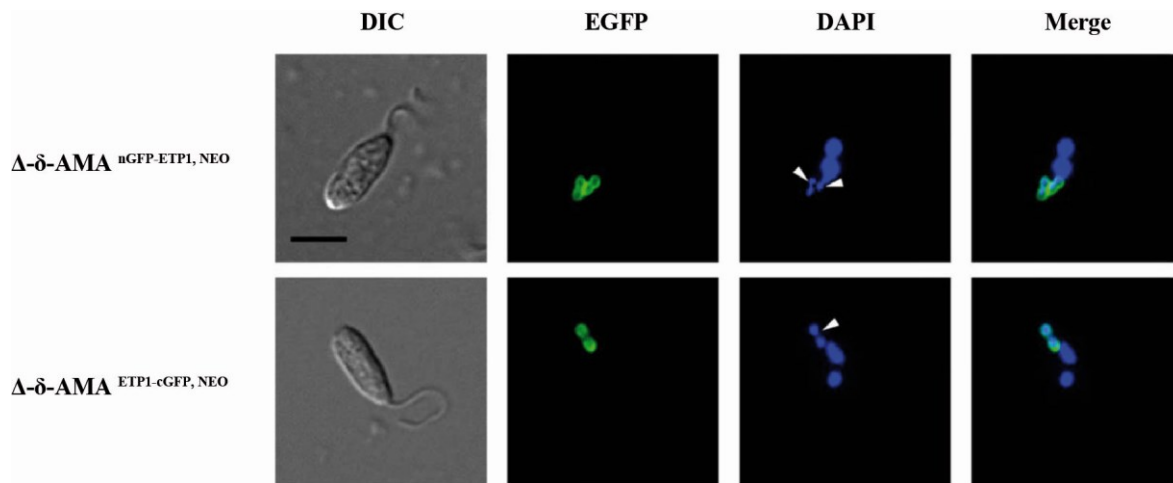
Figure 2.3-4: EGFP fluorescence pattern confirms targeting of host-encoded ETP1 to the endosymbiont in _A. deanei_. Cells expressing nGFP-ETP1 or ETP1-cGFP fusion proteins were fixed in PFA, DAPI stained, and analyzed by epifluorescence microscopy. Δ-δ-AMA^nGFP-ETP1, NEO (Adea0035) and Δ-δ-AMA^ETP1-cGFP, NEO (Adea0036), cell lines expressing in _δ-amastin_ locus nGFP-ETP1 or ETP1-cGFP fusions, respectively; **white arrow heads** highlight the eight-shaped endosymbiont in the DAPI panel. Scale bar: 2 μm.

ETP1 localization was further analyzed using confocal microscopy and stimulated emission depletion (STED) microscopy, considering that these techniques enable increased resolution imaging. Signals from both analyses were detected only in the periphery of the endosymbiont (Appendix, Figure A2) (Master thesis, Georg Ehret).

### 2.3.3 SDS-PAGE analysis of nGFP-ETP1 expressing cells

To check further the enrichment of the nGFP-ETP1 fusion protein in the endosymbionts compared to the cells, equal protein amounts of Δ-δ-AMA^nGFP-ETP1, NEO (Adea0035) cells and endosymbionts purified from this cell line (as described in 4.2.2.7) were analyzed by SDS-PAGE and Western blot alalysis using anti-GFP antibody. A band of 84 kDa was identified in both, cell and endosymbiont samples, showing higher intensity in the latter sample (Figure 2.3-5 A). Interestingly, this band does not correspond to the expected size of the nGFP-ETP1 fusion protein which is 70 kDa. To check whether this discrepancy of 14 kDa between the expected and the observed size of the fusion protein was a result of disulfide bond formation, Δ-δ-AMA^nGFP-ETP1, NEO cells were denatured in the absence of any or in the presence of high concentrations of reducing agents (dithiothreitol (DTT) or β-mercaptoethanol) and analyzed by SDS-PAGE and Western blot analysis. Dithiothreitol (DTT) and β-mercaptoethanol are responsible for the reduction of typical disulfide bonds. In the absence of reducing agent a band of ~145 kDa and an additional band of 84 kDa were identified, while after treatment with up to 600 mM dithiothreitol (DTT) or up to 10% β-mercaptoethanol a 84 kDa band and an additional less intense band of 63 kDa were observed (Figure 2.3-5B). The disappearance of

the ~145 kDa band after the treatment with reducing agents suggests the presence of disulfide bonds of ETP1 with other peptides in the sample. On the other hand, in the 84 kDa band only a fraction of the intensity is reduced when the sample is treated with DTT or β-mercaptoethanol, but still most of the signal is retained at this molecular weight. Together, these results suggest that the band observed at 63 kDa is probably the result of degradation, and that the 84 kDa band corresponds to the stable form of ETP1 inside the cells. Some post-translational modification(s) or the conformation of the fusion protein could drive the irregular migration of this protein under denaturing conditions.
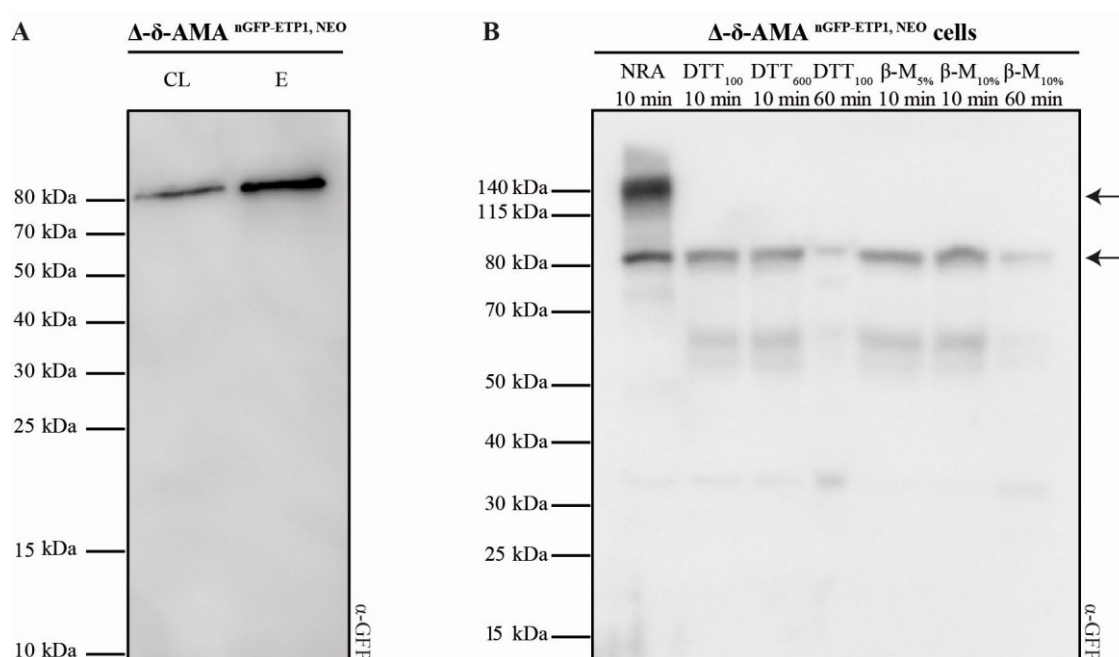
Figure 2.3-5: Analysis of ETP1 molecular weight under reducing conditions. (A) To check the enrichment of nGFP-ETP1 fusion protein in the endosymbiont fraction compared to the cell lysate, 50 µg of Δ-δ-AMA$^{nGFP-ETP1, NEO}$ cell lysate (**CL**) and endosymbionts (**E**) purified from this cell line were analyzed by Western blot analysis using anti-GFP (**α-GFP**). This analysis revealed a band of 84 kDa. (B) Cells expressing nGFP-ETP1 fusion (200µl from a culture in mid-log phase) were spun down (at 5000 g for 5 min) and after the removal of the medium were incubated at 95°C in the absence or in the presence of high concentrations of dithiothreitol (DTT) or β-mercaptoethanol (β-M) for 10 or 60 min and subjected to Western blot analysis. The fusion protein nGFP-ETP1 was detected, using α-GFP antibody, at ~145 kDa and 84 kDa in the absence of reducing agents and at 84 kDa and 63kDa in the presence of high concentrations of reducing agents. **NRA**, no reducing agent; **dTT$_{100}$**, 100 mM dTT; **dTT$_{600}$**, 600 mM dTT; **β-M$_{5\%}$**, 5% β-M; **β-M$_{10\%}$**, 10% β-M.

2.3.4 Generation of ETP1 SKO and DKO cell lines

An important aim for the functional characterization of ETP1 was the generation of ETP1 SKO and DKO cell lines. The first step was the construction of two vectors in which the 5´-FR of the *etp1* gene was fused to the 5´-end of the *neo* or *hyg* marker gene followed by the 3´-FR of the *etp1* gene (pAdea-ETP1/NEO (pAdea042), pAdea-ETP1/HYG (pAdea063); Figure 2.3-6). For transfection, the expression cassettes were excised from the plasmids pAdea-ETP1/NEO and pAdea-ETP1/HYG at the restriction sites, as shown in Figure 2.3-6.
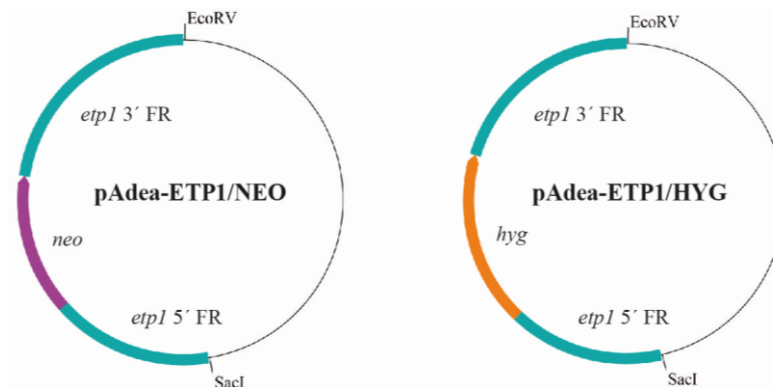
Figure 2.3-6: Plasmids containing the *neo* and *hyg* resistance marker genes surrounded by the *etp1* 5´ and 3´ flanking regions (FRs). Fragments of 1 kb of the 5´- and 3´-FRs of the *etp1* gene were amplified from *A. deanei* gDNA and the selectable marker genes (*neo* or *hyg*) were inserted between the FRs. **Malachite green rectangles**, *etp1* 5´- and 3′-flanking regions (*etp1* FRs); **pink arrow**, *neo*; **orange arrow**, *hyg*. The restriction sites indicate the excision product used for the transfections to generate ETP1 SKOs or DKOs, as described in this work. Vector backbone: pUMA 1467.

Transfectants were selected in 500 µg/ml G418 or 500 µg/ml hygromycin and clonal cultures were generated by limiting dilution. The PCR analysis of both, Δ-ETP1[NEO] (Adea0038) and Δ-ETP1[HYG] (Adea0071) SKO transfectants using primers binding to the gDNA outside of the cassette and inside the *neo* or *hyg* resistance marker showed the expected band pattern and confirmed the insertion of the ETP1/NEO or ETP1/HYG cassette, respectively (Figure 2.3-7). Δ-ETP1[NEO] and Δ-ETP1[HYG] *A. deanei* cell lines showed normal growth characteristics under standard growth conditions and comparison of wild-type and Δ-ETP1[HYG] SKO cells, using TEM revealed no obvious phenotypic difference between the two cell lines (Figure 2.3-8).
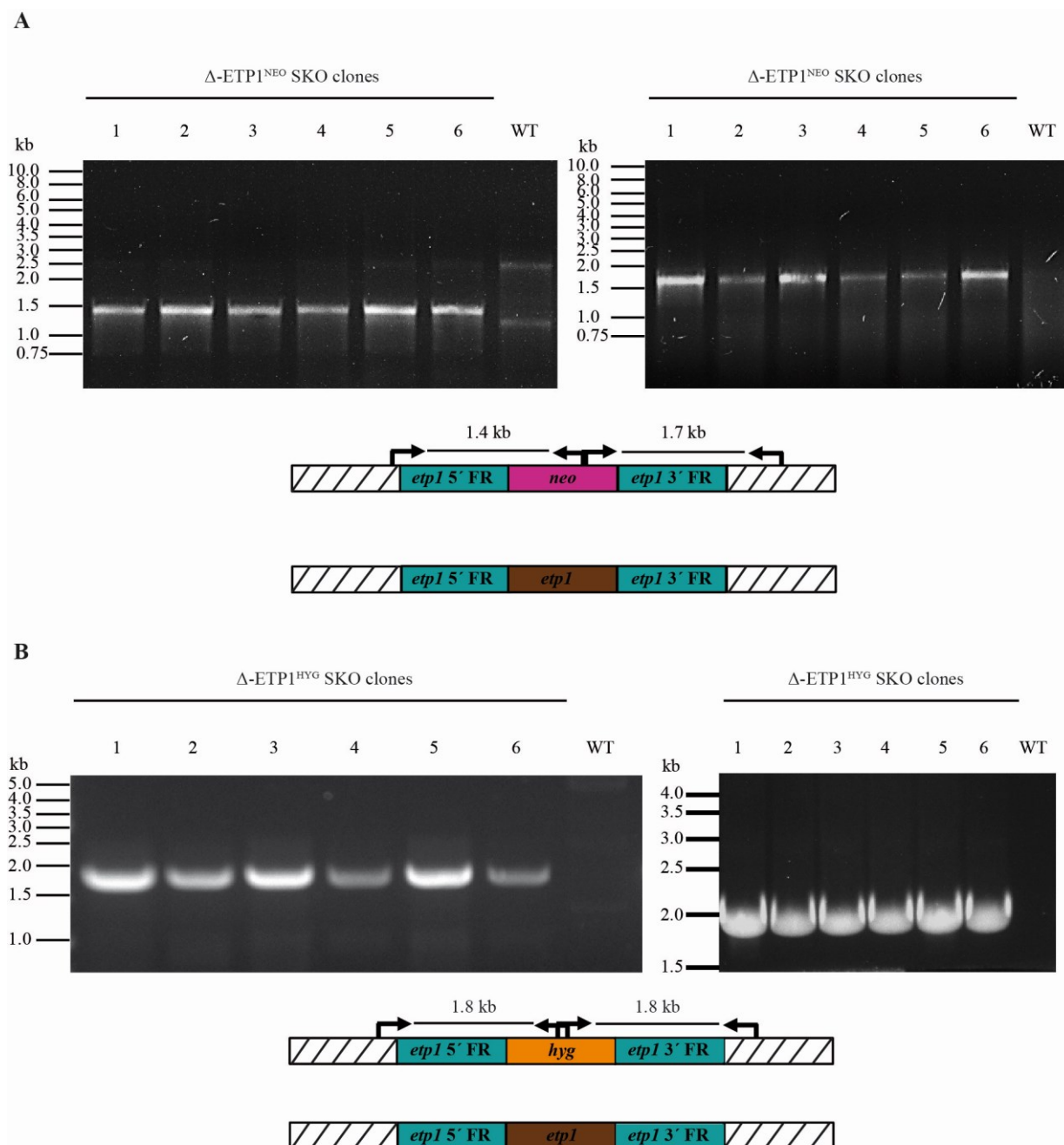
Figure 2.3-7: Generation of Δ-ETP1^NEO and Δ-ETP1^HYG SKOs. Around 2 weeks after transfection with restricted pAdea-ETP1/NEO (pAdea042) (A) or pAdea-ETP1/HYG (pAdea063) (B), clonal cultures were generated by limiting dilution. One week later, clones were selected and the insertion analyzed by PCR. (A) A band at 1.4 kb or at 1.7 kb corresponding to the insertion of *neo* appeared in all SKO clones examined when primers that bind to the gDNA outside of the cassette and inside the *neo* resistance gene were used. (B) A band at 1.8 kb corresponding to *hyg* insertion was observed when primers that bind to the gDNA outside of the cassette and inside the *hyg* resistance were used. As expected no band was identified in the case of the WT. **Striped rectangles**, genomic regions up and downstream of the insertion sites for the cassettes; **malachite green rectangle**, 5′- and 3′-flanking regions (FR) of *etp1* gene, that are present in the cassettes; **brown rectangle**, *etp1* ORF; **pink and orange rectangles**, *neo* and *hyg*, respectively.
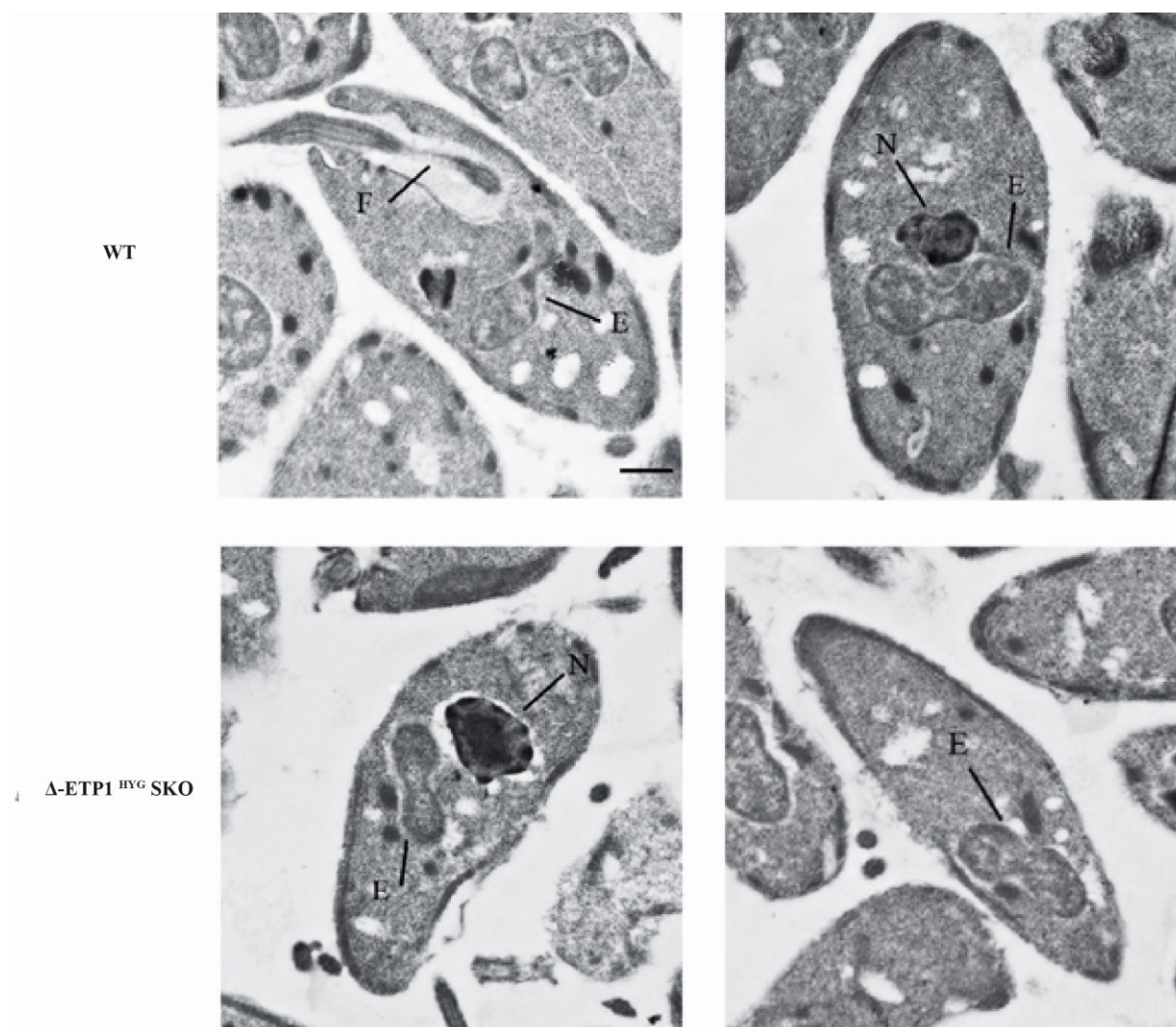
Figure 2.3-8: Comparison of WT and Δ-ETP1^{HYG} (Adea0071) SKO thin sectioned *A. deanei* cells by transmission electron microscopy revealed no phenotypic difference. (E) endosymbiont; **(F)** flagellum; **(N)** nucleus. Scale bar: 500 nm.

Interestingly, PCR analysis of Δ-ETP1^{NEO} (Adea0038) and Δ-ETP1^{HYG} (Adea0071) SKO clonal cultures with primers that anneal to the "Q-rich" sequence of *etp1* gene and to the end of *etp1* sequence showed that in all 16 SKO cell lines analyzed specifically the "*etp1* Q-rich" isoform was removed (Figure 2.3-9). Additionally, the specific removal of "*etp1* Q-rich" form was confirmed using a different primer pair (data not shown).
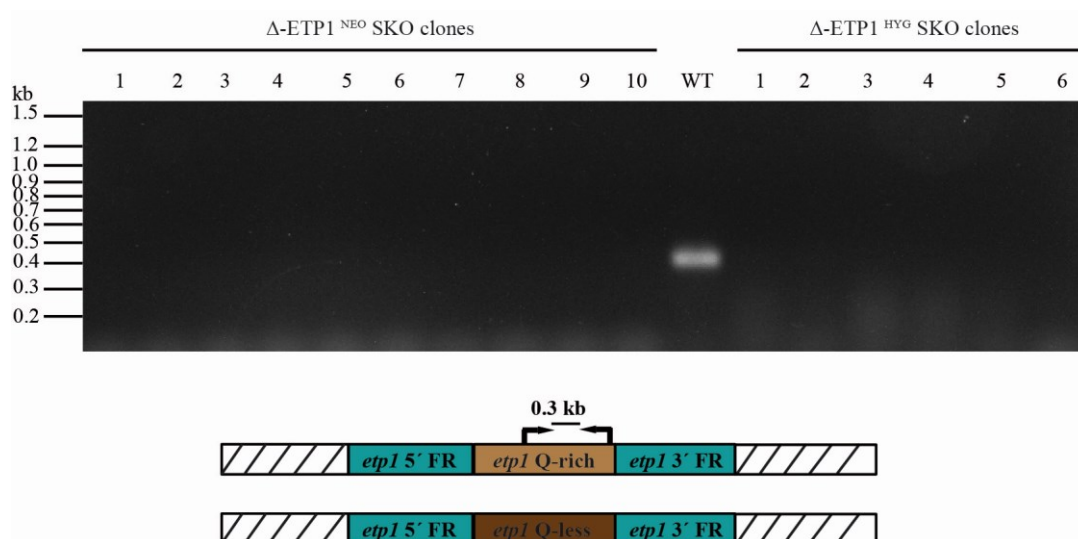
Figure 2.3-9: ETP1 encoded by Q-rich isoform (scaf 3029) is not essential for the survival of the cells. Genomic DNA extracted from Δ-ETP1^NEO (Adea0038) and Δ-ETP1^HYG (Adea0071) SKO clonal cultures was used as a template for PCR with primers binding specifically to the Q-rich version of *etp1* revealed that in all SKO clones examined, specifically the "*etp1* Q-rich" isoform was removed. **Malachite green rectangle**, 5′- and 3′-flanking regions (FR) of *etp1* gene; **light brown rectangle**, ORF of "*etp1* Q-rich" form (*etp1* scaffold 3029); **dark brown rectangle**, ORF of "*etp1* Q-less" form (*etp1* scaffold 1854).

As mentioned before, *A. deanei* is an asexual diploid, and thus, the generation of an ETP1-DKO is needed to create null mutants of ETP1. For this reason Δ-ETP1^NEO (Adea0038) SKO was transfected with the expression cassette that was excised from pAdea-ETP1/HYG (pAdea063) plasmid (Figure 2.3-5). This transfection was repeated three times and the transfectants were selected in 500 µg/ml G418 and 500 µg/ml hygromycin. After the first and second transfection, cells did not survive the treatment with antibiotics. Only after a third time transformants were obtained and clonal cultures resistant to both neomycin and hygromycin were acquired. However, PCR analysis of the presumptive Δ-ETP1^NEO/Δ-ETP1^HYG DKO clonal cultures, using primers that anneal to gDNA just outside of the FR sequences, revealed that the ETP1/HYG cassette was not inserted in the desired *etp1* locus. This was indicated by the absence of a 3.3 kb product corresponding to the *hyg* insertion and the presence of a 3.5 kb product corresponding to the WT *etp1* locus (Figure 2.3-10). Taken together these results indicate that generation of an ETP1 DKO is probably not possible and that ETP1 might be essential for the survival of the cells.
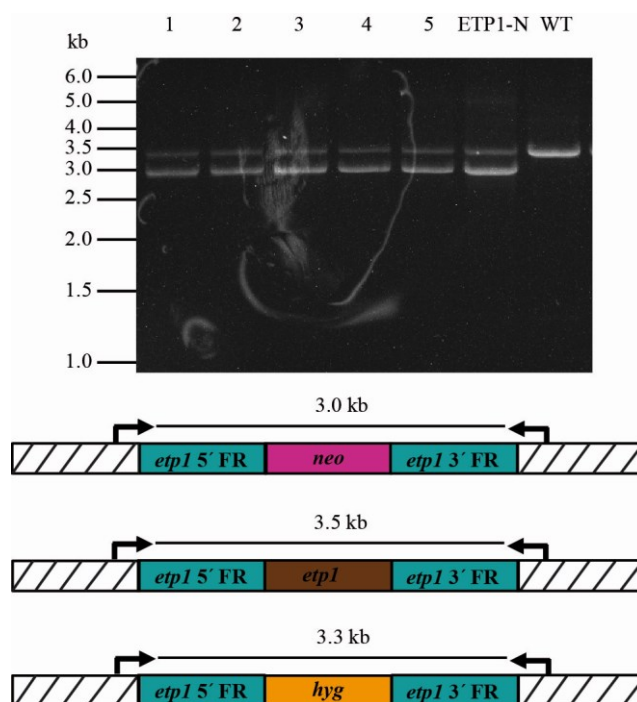
Figure 2.3-10: An ETP1 DKO could not be obtained. Δ-ETP1^NEO clone was transfected with restricted pAdea-ETP1/HYG (pAdea063) and after limiting dilution 5 clones were selected and analyzed by PCR using primers that bind to gDNA outside the inserted cassete. In all clones examined a band at 3.0 kb and a band at 3.5 kb were observed corresponding to the *neo* insertion and the *etp1* WT locus, respectively. No band at 3.3 kb corresponding to *hyg* insertion was identified in any case. **Malachite green rectangles**, 5´- and 3´-flanking regions (FRs) of *etp1* gene, that are present in the cassettes; **brown rectangle**, *etp1* ORF; **pink and orange rectangles**, *neo* and *hyg*, respectively.

2.3.5 Identification of interacting partners of ETP1

2.3.5.1 Co-Immunoprecipitation (Co-IP) assay

To identify interaction partners of ETP1, the Δ-δ-AMA^nGFP-ETP1, NEO (Adea0035) cell line (expressing nGFP-ETP1 from the *δ-amastin* locus) was used in a Co-IP assay coupled to mass spectrometry (MS) analysis. Initially, Δ-δ-AMA^nGFP-ETP1, NEO cells were incubated in lysis buffer (10 mM Tris/HCl pH 7.5; 150 mM NaCl; 0.5 mM EDTA; 2% n-Dodecyl-beta-Maltoside (DDM), Glycon Biochemicals GmbH; and 1 mg/ml lysozyme, Sigma Aldrich) for 30 min with extensively pipetting every 10 min. Then, the insoluble material was spun down at 20000 g for 10 min and the supernatant was loaded on GFP-Trap magnetic beads. After 1h incubation the flow through was withdrawn, the beads were washed 6 times and the proteins that were bound on the beads were eluted. Despite the presence of protease inhibitors (1x cOmplete™ Protease Inhibitor Cocktail, 1 mM PMSF and 1 μM Pepstatin A) and strict work at 4°C intense degradation of the nGFP-ETP1 fusion protein was reproduceably observed in the elution fraction (Figure 2.3-11A). Therefore, endosymbionts isolated from the Δ-δ-

AMA$^{nGFP-ETP1, NEO}$ cell line (nGFP-ETP1 endosymbionts) (as described in 4.2.2.7) were used as starting material. In this case, following the same procedure described above no degradation of nGFP-ETP1 in the elution fraction was observed, as shown in Figure 2.3-10B. Nevertheless, it was noticed that after the lysis of the endosymbionts and the centrifugation of the endosymbiont lysate most of the nGFP-ETP1 remained in the pellet (Figure 2.3-11B). To overcome this problem freshly purified nGFP-ETP1 endosymbionts were treated with lysozyme and subjected to osmotic shocks, in order to form spheroplasts (as described in 4.2.2.8), and then the effects of incubation in a buffer containing 10 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA and different concentrations of detergents or sonication on cell lysis were tested. After centrifugation of the endosymbiont lysate at 20000 g for 10 min, resulting supernatants and pellets were analyzed by SDS-PAGE and Western blot. As shown in Figure 2.3-12, addition of 1% Triton X-100, 0.1% SDS and 1% deoxycholate (RIPA) was the best condition to lyse the endosymbionts and solubilize ETP1. Despite the fact that ETP1 does not contain any predicted transmembrane domains, milder solubilization conditions were insufficient to solubilize ETP1, as most of ETP1 was identified in the pellet.
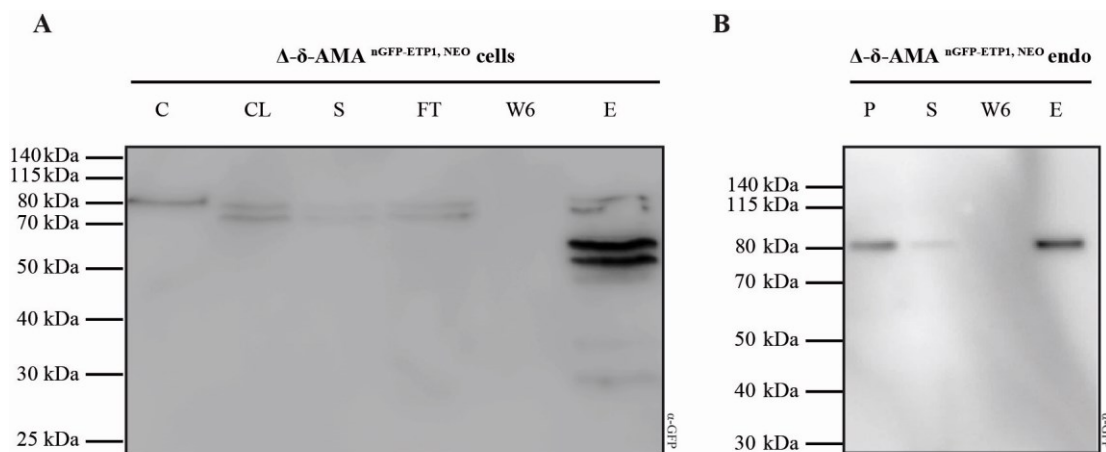
Figure 2.3-11: Co-Immunoprecipitation (Co-IP) assay using Δ-δ-AMA^{nGFP-ETP1, NEO} cells or endosymbionts. Δ-δ-AMA^{nGFP-ETP1, NEO} (Adea0035) cells (4 mg of total protein) (A) or endosymbionts purified from this cell line (100 μg of total protein) (B) were lyzed in 500 μl buffer (10 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA) supplemented with 2% n-Dodecyl-beta-Maltoside (DDM) and 1 mg/ml lysozyme for 30 min at 4°C with extensively pipetting every 10 min. Cell lysis was followed by centrifugation at 20000 g for 10 min. The pellet was resuspended in the same volume of buffer and stored for later analysis. The resulting supernatant was loaded on GFP-Trap magnetic beads and after incubation for 1 h the flow through was withdrawn. Then, the beads were washed 6 times using the same buffer (excluding DDM and lysozyme) mentioned above and the proteins that were bound on the beads were eluted by incubation in 2x SDS sample buffer at 95°C for 10 min (elution fraction). Finally, 5 μl of cell and cell lysate samples and 25 μl of all the other samples in 1x SDS sample buffer were resolved on a 4-12% polyacrylamide gel by SDS-PAGE and subjected to Western blot analysis using anti-GFP antibody (**α-GFP**). Protease inhibitors were added in each buffer of the procedure (1x cOmplete™ Protease Inhibitor Cocktail, 1 mM PMSF and 1 μM Pepstatin A). This analysis revealed that the use of the endosymbionts as starting material of the Co-IP assay reduced degradation of nGFP-ETP1 in the elution fraction. **Δ-δ-AMA^{nGFP-ETP1, NEO} endo**, endosymbionts purified from Δ-δ-AMA^{nGFP-ETP1, NEO} cell line; **C**, Δ-δ-AMA^{nGFP-ETP1, NEO} cells; **CL**, Δ-δ-AMA^{nGFP-ETP1, NEO} cell lysate; **S**, supernatant; **FT**, flow through fraction; **W6**, wash 6 fraction; **E**, elution fraction; **P**, pellet.
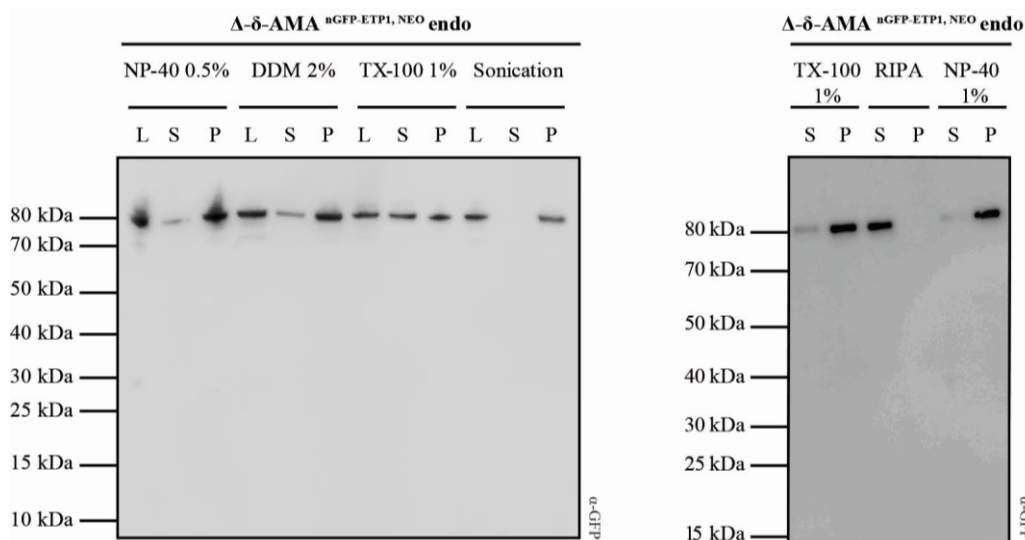
Figure 2.3-12: Optimization of condition for solubilization of endosymbiont fractions. To determine the best condition for endosymbiont solubilization, freshly purified endosymbionts from *A. deanei* cells expressing nGFP-ETP1 from the *δ-amastin* locus (**Δ-δ-AMA**[nGFP-ETP1, NEO] **endo**) were treated with lysozyme, in order to form spheroplasts and then incubated (overnight at -20°C) in buffer (10 mM Tris/HCl pH 7.5; 150 mM NaCl; 0.5 mM EDTA) containing different concentrations of detergents or sonicated. The total amount of protein was 120 μg in each case. After centrifugation (10 min at 20000 g) of the endosymbiont lysate (**L**) the supernatant (**S**) was withdrawn and pellet (**P**) was resuspended in the same volume of buffer as the supernatant. All samples were boiled for 5 min in 1x SDS sample buffer, spun down for 2 min at 1000 g, resolved on a 4-12% polyacrylamide gel by SDS-PAGE and subjected to Western blot analysis using anti-GFP antibody (**α-GFP**). Incubation of endosymbionts with RIPA buffer proved to be the best condition for lysis and solubilization of endosymbionts. Protease inhibitors were added in each buffer of the procedure (1x cOmplete™ Protease Inhibitor Cocktail, 1 mM PMSF, 1 μM Pepstatin A, 10 μM E64, 100 μM TLCK). **NP-40**, Tergitol-type NP-40; **DDM**, n-Dodecyl β-D-maltoside; **TX-100**, Triton X-100; **RIPA**, 1% Trition X-100, 0,1% SDS and 1% deoxycholate.

According to these results, RIPA buffer (10 mM Tris/HCl pH 7.5; 150 mM NaCl; 0.5 mM EDTA, 1% Trition X-100, 0.1% SDS and 1% deoxycholate) was used to solubilize nGFP-ETP1 endosymbionts in following Co-IP experiments. To ensure that interaction partners identified in the GFP-trap approach interact directly with nGFP-ETP1 and are not part of membrane vesicles that ETP1 might interact with, any remaining membrane fragments in the RIPA buffer-solubilized sample were spun down by ultracentrifugation (1 h at 150000 g). The resulting supernatant (500 μg of total protein) was subjected to Co-IP assay coupled to MS analysis, as described in 4.2.3.8. As a control, to test if any endosymbiont-localized protein interacts directly with the beads, the same experiment was performed using endosymbionts isolated from wild-type *A. deanei* cells (WT control) (500 μg of total protein). A further control to identify endosymbiont-localized proteins that interact with GFP and not with ETP1 (GFP control), was needed. A problem was that until now the genetic manipulation of the endosymbiont has not been established. Thus, it was not possible to generate a cell line in which the endosymbiont expresses GFP. For this reason, the cell lysate from a cell line of

A. deanei expressing cytosolic GFP (Adea0010) was prepared and incubated with the GFP-trap beads, as described above. The GFP pre-loaded beads were then incubated with a lysate obtained from wild-type endosymbionts (500 μg of total protein). The amount of the GFP that was added to the beads in the GFP control was calculated to be the same as in the nGFP-ETP1 sample (Figure 2.3-13). A total of three biological replicates were repeated for the Co-IP assay of each cell line.
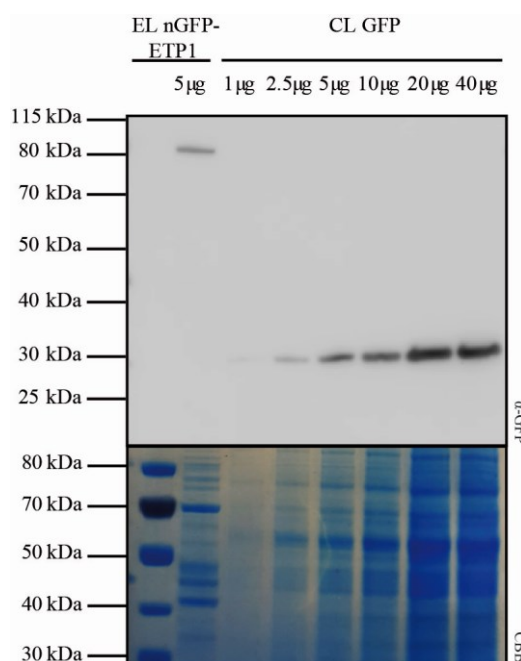
Figure 2.3-13: Comparison of GFP expression levels between lysate of endosymbionts purified from nGFP-ETP1 expressing cell line and whole cell lysate of a cell line that expresses cytosolic GFP (Adea0010). 5 μg of endosymbiont lysate (**EL**) from Δ-δ-AMA$^{nGFP-ETP1, NEO}$ (Adea0035) cell line and different protein amounts of whole cell lysate from the cell line that expresses cytosolic GFP (**CL GFP**) were subjected to SDS-PAGE and Western blot analysis using anti-GFP (**α-GFP**) and the proteins in the gel were visualized using Coomassie Brilliant Blue G-250 (**CBB**) staining. It was identified that 5 μg of nGFP-ETP1 EL contain comparable levels of GFP as 3 μg of CL GFP.

The flow through, wash 6 and elution fractions from the WT control, GFP control and nGFP-ETP1 sample were subjected to SDS-PAGE and Western blot analysis. There were no bands observed in the elution fraction in silver or coomassie brilliant blue (CBB) stained gels in the WT control, whereas there was only one band detected in the case of the GFP control corresponding to the GFP. Interestingly, one band corresponding to nGFP-ETP1 and one additional band corresponding to the size of 45 kDa were identified by CBB staining in the case of the nGFP-ETP1 sample (Fig. 2.3-14A shows Western blot and CBB-stained gel for one representative experiment; results in the remaining two replicates were essentially the same). More bands were detected when silver stain instead CBB was used to stain the gel.

The results were comparable in all replicates (Figure 2.3-14B). Western blot analysis using anti-GFP identified no bands in the WT control, one band in the elution fraction of the GFP control sample corresponding to the GFP and one band in the flow through and wash 6 fractions of the nGFP-ETP1sample corresponding to nGFP-ETP1 fusion. In the elution fraction of nGFP-ETP1 sample a main band corresponding to nGFP-ETP1 and a certain amount of degradation was observed (Figure 2.3-14A).
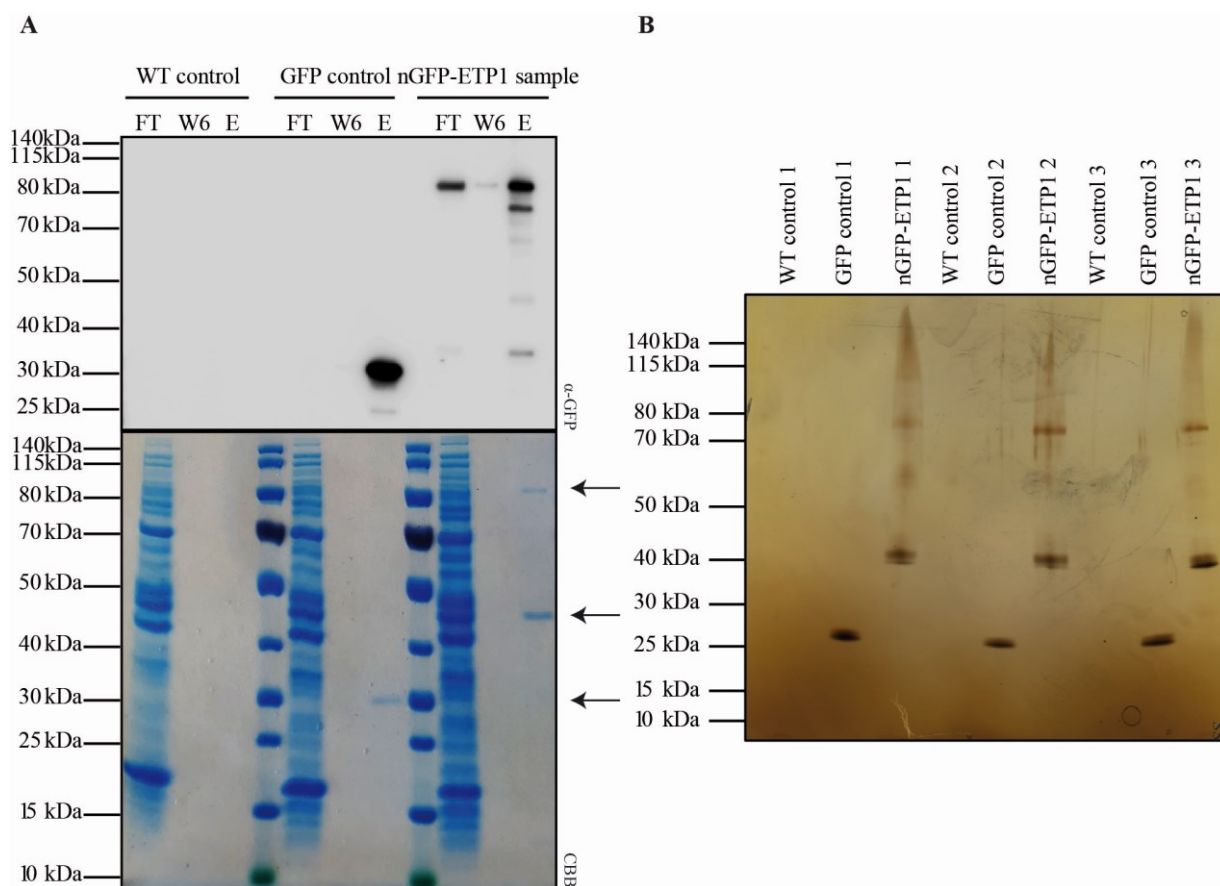
Figure 2.3-14: Co-Immunoprecipitation of nGFP-ETP1 from purified endosymbionts. Freshly purified endosymbionts from wild-type (WT) or nGFP-ETP1 expressing cell line (Δ-δ-AMA$^{nGFP-ETP1, NEO}$) of *A. deanei* were used to generate spheroplasts and then solubilized in RIPA buffer overnight at -20°C. Solubilization was followed by centrifugation at 150000 g for 1 h and loading of the supernatant on the beads. After 6 washes, proteins that interacted with the beads were eluted by incubation in 2x SDS sample buffer at 95°C for 10 min (elution fraction). In the case of the GFP control, beads were incubated initially with lysate from a cell line that express cytosolic GFP, washed 6 times and then incubated with lysate of WT endosymbionts and treated as the other samples. Protease inhibitors were added in each buffer of the procedure (1x cOmplete™ Protease Inhibitor Cocktail, 1 mM PMSF, 1 μM Pepstatin A, 10 μM E64, 100 μM TLCK). (A) 20 μl of flow through (**FT**) and 20 μl of wash fraction 6 (**W6**) in 1x SDS sample buffer and 12 μl of elution fraction (**E**) from the **WT**, **GFP** and **nGFP-ETP1** samples were subjected to SDS-PAGE and Western blot analysis. Detection was performed using anti-GFP (**α-GFP**) in the Western blot analysis and the proteins in the gel were visualized using Coomassie Brilliant Blue G-250 (**CBB**) staining. Results from one biological replicate are presented. The results in the other two replicates were essentially the same. (B) 2 μl of elution fraction from the **WT**, **GFP** and **nGFP-ETP1** samples were analyzed by SDS-PAGE and visualized using silver staining (**SS**). Results from the three biological replicates are presented.

To identify the proteins that precipitated with ETP1 half of the elution fraction was subjected to tandem MS analysis and the results of this analysis are shown in Appendix Table A1. These MS results were visualized in a heat map, which was generated by Dr. Gereon Poschmann using Perseus (version 1.6.1.1) (Figure 2.3-15). Unfortunately, the results of the MS analysis were not satisfying and the intensities and spectral counts obtained did not fit with the amount and distribution of proteins present in the samples as judged by the bands observed in the CBB- and silver-stained gels. Nevertheless, proteins that were identified in the nGFP-ETP1 samples with average ion intensities higher than the 10% of the average ion intensity of ETP1 (average intensities >763927) and not in the WT and GFP controls are considered as possible interaction partners of ETP1. Applying these criteria, the nucleus-encoded kinetoplastid membrane protein-11, uncharacterized protein S9U3N9, protein FAM184A, Endosymbiont-targeted Protein 2 (ETP2) and, the endosymbiont-encoded 30S ribosomal protein S19 are potential interaction partners of ETP1. On the other hand, the endosymbiont-encoded porin is detected in the WT and GFP controls with average ion intensities of 1012400 and 41252567, respectively. However, it is also detected in the nGFP-ETP1 sample with an average ion intensity of 1390886667, approximately 34 times higher intensity than in the GFP control and 1300 times higher intensity than in the WT control. Moreover, as observed in Figure 2.3-14 there is an intense band of 45 kDa corresponding to porin in all the replicates of nGFP-ETP1 sample detected in the CBB- and silver-stained gels, whereas there is no band detected in WT and GFP control samples at this size. For these reasons porin is also considered as a potential interaction partner of ETP1.
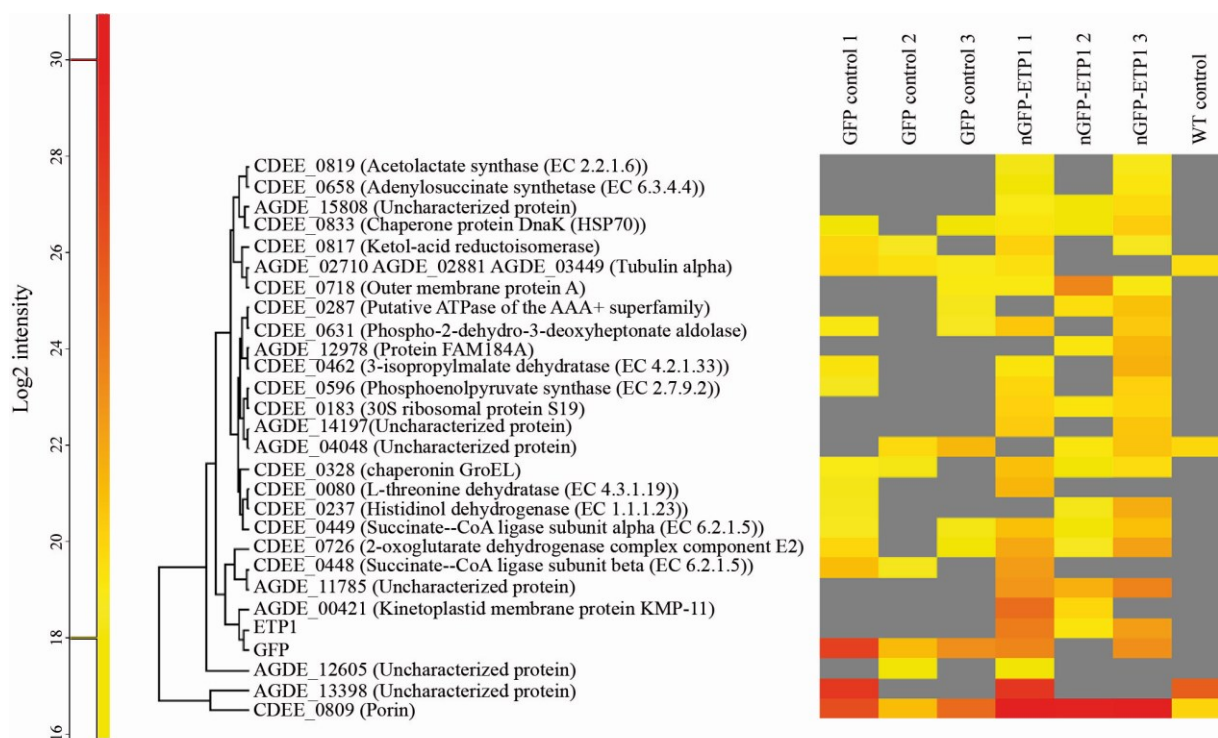
Figure 2.3-15: Heat map visualization of mass-spectrometry analysis of the elution fraction from nGFP-ETP1 Co-IP assay. The heat map was generated with Perseus (version 1.6.1.1 Max Planck Institute of Biochemistry, Planegg, Germany) on log2 label-free quantification intensities. Row trees were generated by a hierarchical cluster analysis using a Euclidean distance matrix with average linkage. Generated by Dr. Gereon Poschmann.

2.3.5.2 Two-Dimensional Clear-native/SDS-polyacrylamide-gel-electrophoresis (2-D CN/SDS-PAGE)

Two-Dimensional CN/SDS-PAGE was used as a second method to identify possible interaction partners of ETP1. *A. deanei* endosymbionts, purified from a cell line that expresses ETP1 fused to the N-terminus of the V5 tag (Δ-δ-AMA$^{ETP1-cV5, NEO}$; Adea0040), were solubilized using 1% DDM (n-Dodecyl-beta-Maltoside, Glycon Biochemicals GmbH), insoluble material was removed by centrifugation (10 min at 20000 g) and CNE was used for separation of DDM solubilized proteins under native conditions, as described in 4.2.3.2. Then, ETP1-cV5 was detected using anti-V5 antidody by Western blot analysis and the band was extracted from the CNE gel (Figure 2.3-16A). Part of this band was then equilibrated in a SDS-containing buffer to promote denaturation of the proteins and later loaded on top of a 4-12% SDS-PAGE gel, where peptides were separated according to their molecular masses (Figure 2.3-16B). Another part of this ETP1-cV5 band was analyzed further by tandem MS analysis, as shown in Appendix Table A2. In total 77 proteins were detected by this MS analysis. Among those only porin, Kinetoplastid membrane protein-11, and ETP2 were

identified also by the Co-IP assay. Considering that, those proteins were identified as interacting partners of ETP1 by two independent methods (Co-IP and 2-D CN/SDS-PAGE) and using different ETP1 tags (GFP and V5, respectively) they are regarded as putative interaction partners of ETP1.
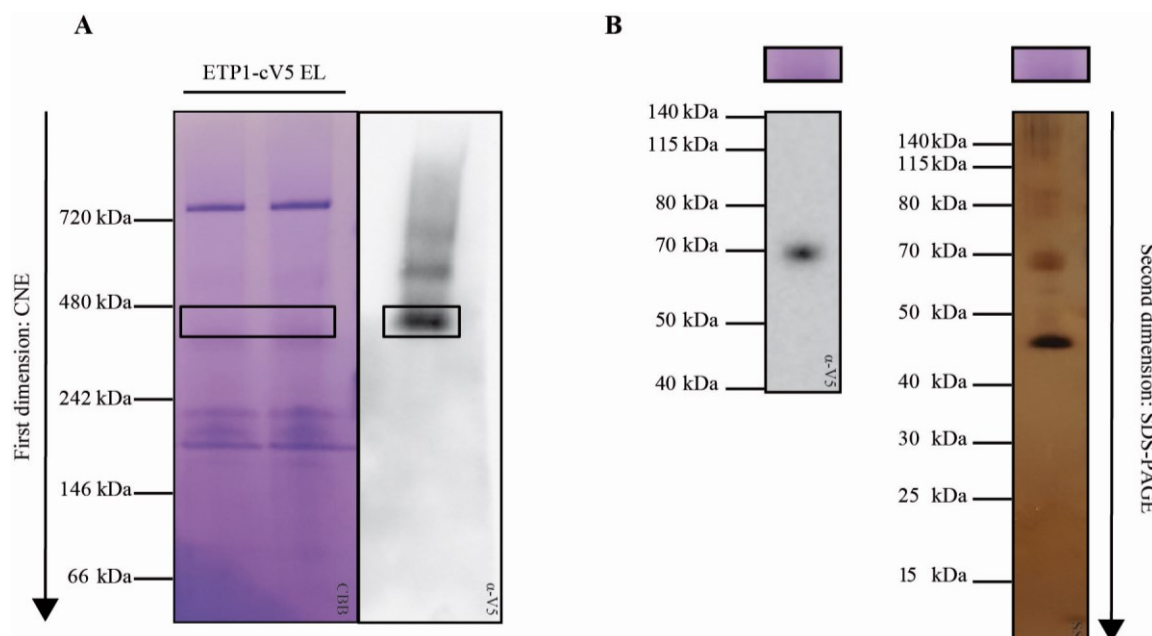
Figure 2.3-16: Two-dimensional Clear-Native/SDS-PAGE analysis of endosymbiont fractions purified from ETP1-cV5 expressing cell line (Δ-δ-AMA$^{\text{ETP1-cV5, NEO}}$; Adea0040). (A) *A. deanei* ETP1-cV5 endosymbionts were solubilized in buffer (10 mM Tris/HCl pH 7.5; 150 mM NaCl; 0.5 mM EDTA) using 1% n-Dodecyl β-D-maltoside (DDM) in the presence of protease inhibitors and after centrifugation (10 min at 20000 g) the supernatant was subjected to Clear-Native electrophoresis (**CNE**). Proteins were stained by Coomassie Brilliant Blue R-250 (**CBB**) (left panel; the same sample was loaded into two lanes because of its volume) and ETP1-cV5 was visualized by Western blot analysis (right panel). Then, the band of the CNE gel that contained ETP1-cV5 and putative interaction partners was extracted. (B) The extracted band was analyzed on a second dimension on a denaturing SDS-PAGE and visualized by Western blot analysis (left panel) and silver staining (**SS**) (right panel). For Western detection anti-V5 (**α-V5**) was used. **ETP1-cV5 EL**, solubilized endosymbionts purified from ETP1-cV5 expressing cell line.

2.3.5.3 Functionality of the nGFP-ETP1 fusion protein

Since the nGFP-ETP1 fusion was used for the identification of proteins that interact with ETP1 in the Co-IP assay, the functionality of this fusion protein needed to be tested. Previously in this work, it was shown that the generation of an ETP1 DKO cell line was not possible and it was suggested that ETP1 is essential for the survival of the cells. Taking that into consideration, the expression of nGFP-ETP1 fusion from the *δ-amastin* locus and the successful substitution of both *etp1* ORFs by resistance genes would prove that the nGFP-ETP1 fusion protein is able to complement the function of ETP1. For this reason, the Δ-δ-AMA$^{\text{nGFP-ETP1, NEO}}$ (Adea0035) cell line was transfected with the expression cassette of

pAdea-ETP1/HYG (pAdea063). Insertion of the *hyg* resistance gene in the transformants was confirmed by PCR analysis using primers that bind to the gDNA outside the cassette and inside the resistance gene (Figure 2.3-17). At this point a third antibiotic that could be used as a resistance marker was needed in order to knock-out the second *etp1* allele. It was shown previously in our lab that 50 µg/ml of phleomycin were sufficient to kill *A. deanei* WT cells and a concentration of 100 µg/ml could be used for the selection of transfectants. Thus, a vector in which the 5′-FR of the *etp1* gene was fused to the 5′-end of the gene that encodes bleomycin resistance protein (*ble)*, that provides resistance against phleomycin, followed by the 3′-FR of the *etp1* gene (pAdea-ETP1/BLE (pAdea083)) was constructed (Figure 2.3-17A). For transfection of Δ-δ-AMA$^{nGFP-ETP1, NEO}$/Δ-ETP1^{HYG} (Adea0052), the expression cassette was excised from the plasmid, the transfectants were selected in 100 µg/ml phleomycin, 500 µg/ml G418 and 500 µg/ml hygromycin, and the clonal cultures were generated by limiting dilution. The PCR analysis of these clonal cultures, using primers that anneal to gDNA outside of the cassette and inside the *ble*, showed that *ble* was not inserted in the *etp1* locus. Additionally, PCR analysis using primers that bind to gDNA outside of the cassette and inside *etp1* proved the retainment of *etp1* WT allele (Figure 2.3-17). Taken together these data suggest that *ble* was inserted in another locus in the genome. To test whether other tagged versions of ETP1 (i.e. nV5-ETP1, ETP1-cV5) could complement ETP1 function, cell lines expressing the V5-tagged versions of ETP1 from the *δ-amastin* locus (as shown in table 2.3.5-1) were transfected with the cassette that was excised from plasmid pAdea-ETP1/BLE (pAdea083) in order to generate ETP1 SKOs in the recombinant ETP1 overexpression background. Unexpectedly, in all clones that were tested by PCR *ble* was not inserted in the *etp1* locus (data not shown), indicating that targeted insertion of *ble* in the *etp1* locus by HR using plasmid ETP1/BLE (pAdea083) was not efficient (Table 2.3.5-1). For this reason an alternative knock-out strategy was chosen. The plasmid pAdea089 was constructed based on plasmid ETP1/BLE (pAdea083) with the difference that instead of the *etp1* 3′ FR it contains the end of the *etp1* coding sequence (700 bp) as a region of homology. Then, Δ-δ-AMA$^{nGFP-ETP1, NEO}$/Δ-ETP1^{HYG} (Adea0052) was transfected with the expression cassette of plasmid pAdea089 but PCR analysis of the clones showed again that *ble* was not inserted in the *etp1* locus as observed for the other clones (Table 2.3-1; gels not shown).
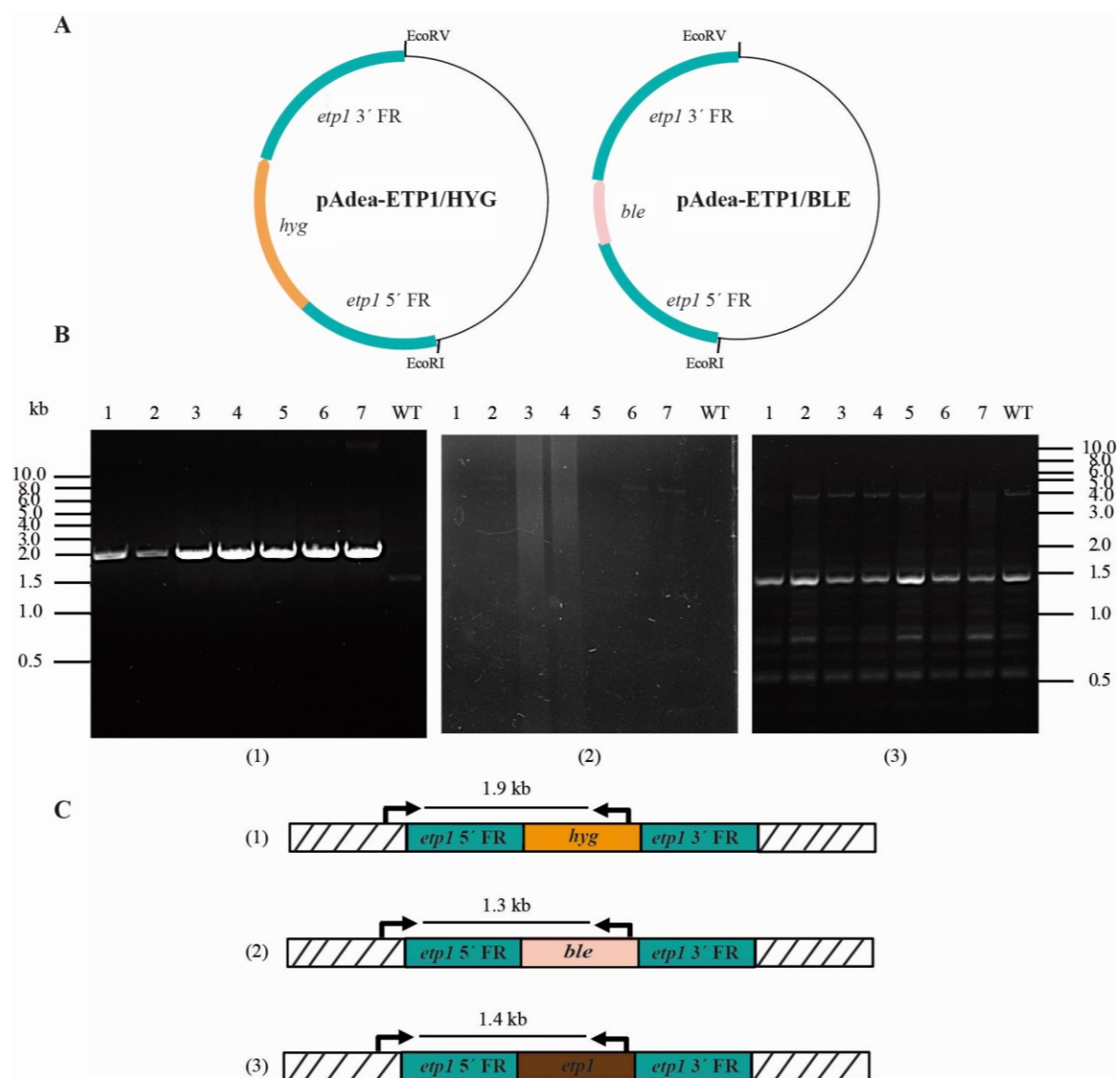
Figure 2.3-17: Testing the functionality of nGFP-ETP1 fusion protein by expressing nGFP-ETP1 fusion from the *δ-amastin* locus and the successful substitution of both *etp1* ORFs by resistance genes. (A) Fragments of 1 kb of the 5′- and 3′-FRs of the *etp1* gene were amplified from *A. deanei* gDNA and the selectable marker genes (*hyg* or *ble*) were inserted between the FRs. The restriction sites indicate the excision product used for the transfections described in this work. Vector backbone: pUMA 1467. (B) Δ-δ-AMA[nGFP-ETP1, NEO] (Adea0035) cell line was transfected with the expression cassette of pAdea-ETP1/HYG (pAdea063). Clonal cultures were selected and the insertion analyzed by PCR using primers that bind in the gDNA outside the cassette and inside the *hyg* gene. A band at 1.9 kb corresponding to the *hyg* insertion was identified in all clonal cultures tested **(left panel)**. Then, Δ-δ-AMA[nGFP-ETP1, NEO]/Δ-ETP1[HYG] (Adea0052) clone 1 was transfected with the cassette of pAdea-ETP1/BLE (pAdea083). Surprisingly, PCR analysis of the clones using primers that bind in the gDNA outside the cassette and inside the *ble* revealed no band corresponding to *ble* insertion **(middle pannel)**. Additionaly PCR using primers that bind in the gDNA outside the cassette and inside ETP1 showed a band at 1.4 kb proving the retainment of *etp1* WT allele **(right panel)**. **Striped rectangles**, genomic regions up and downstream of the insertion sites for the cassettes; **malachite green rectangle**, 5′- and 3′-flanking regions of *etp1* gene (*etp1* FRs), that are present in the cassettes; **brown rectangle**, *etp1* ORF; **orange and pink rectangles**, *hyg* and *ble*, respectively.

Table 2.3-1: Substitution of the function of *etp1* WT gene by recombinant ETP1 proteins was not succesful. *A. deanei* cell lines that express tagged verions of ETP1 (i.e. nGFP-ETP1, nV5-ETP1 and ETP1-cV5) from the *δ-amastin* locus were transfected with different vectors in order to knock-out endogenous *etp1* genes and test for the functionality of the recombinant ETP1 proteins.

Cell line transfected	plasmid	Cell line generated and used for second transfection	plasmid	Cell line generated
Δ-δ-AMA$^{\text{nGFP-ETP1, NEO}}$ (Adea0035)	pAdea-ETP1/HYG (pAdea063)	Δ-δ-AMA$^{\text{nGFP-ETP1, NEO}}$/Δ-ETP1$^{\text{HYG}}$ (Adea0052)	pAdea-ETP1/BLE (pAdea083)	Adea0072 *(ble* was not inserted successfully in *etp1* locus)
Δ-δ-AMA$^{\text{nGFP-ETP1, NEO}}$ (Adea0035)	pAdea-ETP1/HYG (pAdea063)	Δ-δ-AMA$^{\text{nGFP-ETP1, NEO}}$/Δ-ETP1$^{\text{HYG}}$ (Adea0052)	pAdea-ETP1/BLE^2 (pAdea089)	Adea0075 *(ble* was not inserted successfully in *etp1* locus)
Δ-δ-AMA$^{\text{nV5-ETP1, NEO}}$ (Adea0041)	pAdea-ETP1/BLE (pAdea083)	Adea0076 *(ble* was not inserted successfully in *etp1* locus)		
Δ-δ-AMA$^{\text{ETP1cV5, NEO}}$ (Adea0040)	pAdea-ETP1/BLE (pAdea083)	Adea0077 *(ble* was not inserted successfully in *etp1* locus)		

Since *A. deanei* is a newly established model system, there is no experience with triple selection using a combination of G-418, hygromycin, and phleomycin in *A. deanei* yet. Thus, it cannot be excluded that the failure to obtain a rescued ETP1 DKO cell line was caused by not yet identified challenges of the triple selection rather than a lack of functionality of the nGFP-ETP1 fusion protein. As an alternative strategy, that does not rely on triple selection was tried to replace one of the endogenous *etp1* alleles by the nGFP-ETP1 fusion protein. To this end, a vector was constructed which would allow for the replacement of the remaining *etp1* allele in the Δ-ETP1$^{\text{HYG}}$ SKO cell line with an nGFP-ETP1 fusion protein. This vector included the 5′-FR of the *etp1* gene fused to the 5′-end of the *neo* followed by the GAPDH IR and then the *etp1* ORF fused to the C-terminus of EGFP followed by the 3′-FR of *etp1* gene (pAEX-ETP1/nGFP-ETP1 (pAdea048); Figure 2.3-18A). The Δ-ETP1$^{\text{HYG}}$ SKO (Adea0071) cell line was transfected with the expression cassette that was excised from the pAEX-ETP1/nGFP-ETP1 plasmid. Transfectants were selected in 500 µg/ml G418 and 500 µg/ml hygromycin and clonal cultures obtained through limiting dilution. After PCR analysis of the

resulting clonal cultures, using primers that anneal to gDNA just outside of the FR sequences and inside *hyg* the expected band corresponding to *hyg* insertion was observed in all clones tested. Surprisingly, PCR using primers that anneal to gDNA outside of the FR sequences and inside *etp1* revealed two bands, one band corresponding to the insertion of *neo*-GAPDH IR-*gfp-etp1* and one band corresponding to the WT allele of *etp1* (Figure 2.3-18B). In addition, Southern blot analysis using gDNA of three Δ-ETP1$^{\text{HYG}}$/Δ-ETP1$^{\text{nGFP-ETP1, NEO}}$ transformants restricted with *EcoRV* and *SacII* revealed one band at 4.2 kb and one band at 2.3 kb hybridizing to the etp1 probe corresponding to the *neo*-GAPDH IR-*gfp-etp1* insertion and the *etp1* WT allele, respectively. Only one band at 2.3 kb corresponding to the *etp1* WT allele was detected in the WT and the Δ-ETP1$^{\text{HYG}}$ SKO. Restriction with *EcoRV* and *BamHI* revealed a single band at 2.3 kb hybridizing to the hygromycin probe in the transformants Δ-ETP1$^{\text{HYG}}$/Δ-ETP1$^{\text{nGFP-ETP1, NEO}}$ and in Δ-ETP1$^{\text{HYG}}$ SKO. Thus, Southern blot confirmed the results of the PCR analysis (Figure 2.3-19). Moreover, epifluorescence microscopy showed that in Δ-ETP1$^{\text{HYG}}$/Δ-ETP1$^{\text{nGFP-ETP1, NEO}}$ clones nGFP-ETP1 fusion targeted the endosymbiont in all cells that were checked, as previously reported in this study.
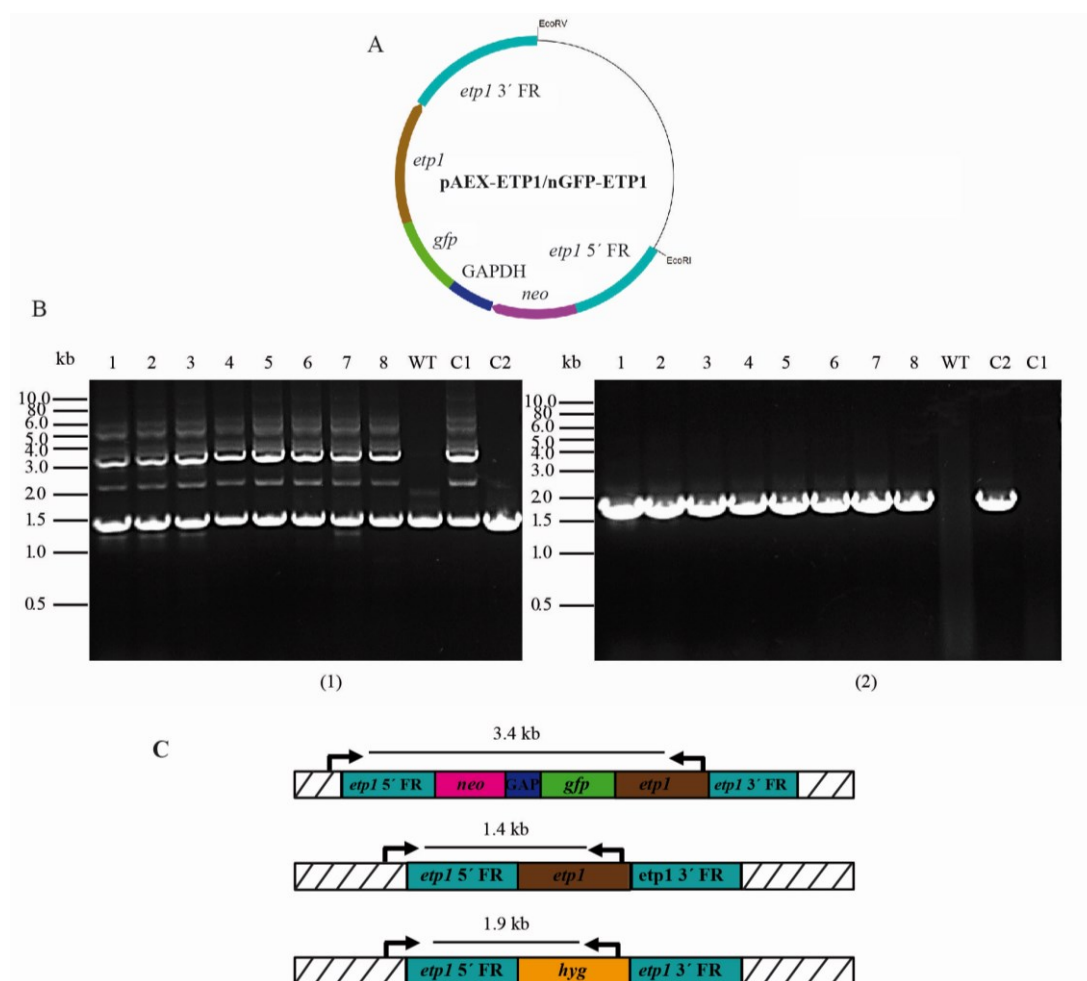
Figure 2.3-18: Testing the functionality of nGFP-ETP1 fusion protein by replacing in the Δ-ETP1^HYG SKO cell line the remaining *etp1* allele with the nGFP-ETP1 fusion protein. (A) A vector containing the 5′-FR of the *etp1* gene fused to the 5′-end of the *neo* followed by the GAPDH IR and then the *etp1* ORF fused to the C-terminus of EGFP followed by the 3′-FR of *etp1* gene (pAEX-ETP1/nGFP-ETP1 (pAdea048)) was constructed. Δ-ETP1^HYG SKO (Adea0071) was transfected with the restricted pAEX-ETP1/nGFP-ETP1. (B) Clonal cultures were selected and the insertion analyzed by PCR using primers that bind in the gDNA outside the cassette and inside *hyg* insertion. A band at 1.9 kb corresponding to *hyg* insertion was identified **(right panel)**. Moreover, PCR using primers that bind in the gDNA outside the cassette and inside *etp1* revealed a band at 3.4 kb corresponding to the *neo*-GAPDH IR-*gfp*-*etp1* insertion and a band at 1.4 kb corresponding to the retainment of the WT *etp1* allele **(left panel)**. C1, Δ-ETP1^nGFP-ETP1, NEO; **C2**, Δ-ETP1^HYG SKO; **striped rectangles**, genomic regions up and downstream of the insertion sites for the cassettes; **malachite green rectangle**, 5′- and 3′-flanking regions of *etp1* gene (etp1 FRs), that are present in the cassettes; **blue rectangle**, GAPDH IR; **green rectangle**, *gfp* ORF; **brown rectangle**, *etp1* ORF; **pink** and **orange rectangles**, *neo* and *hyg*, respectively.
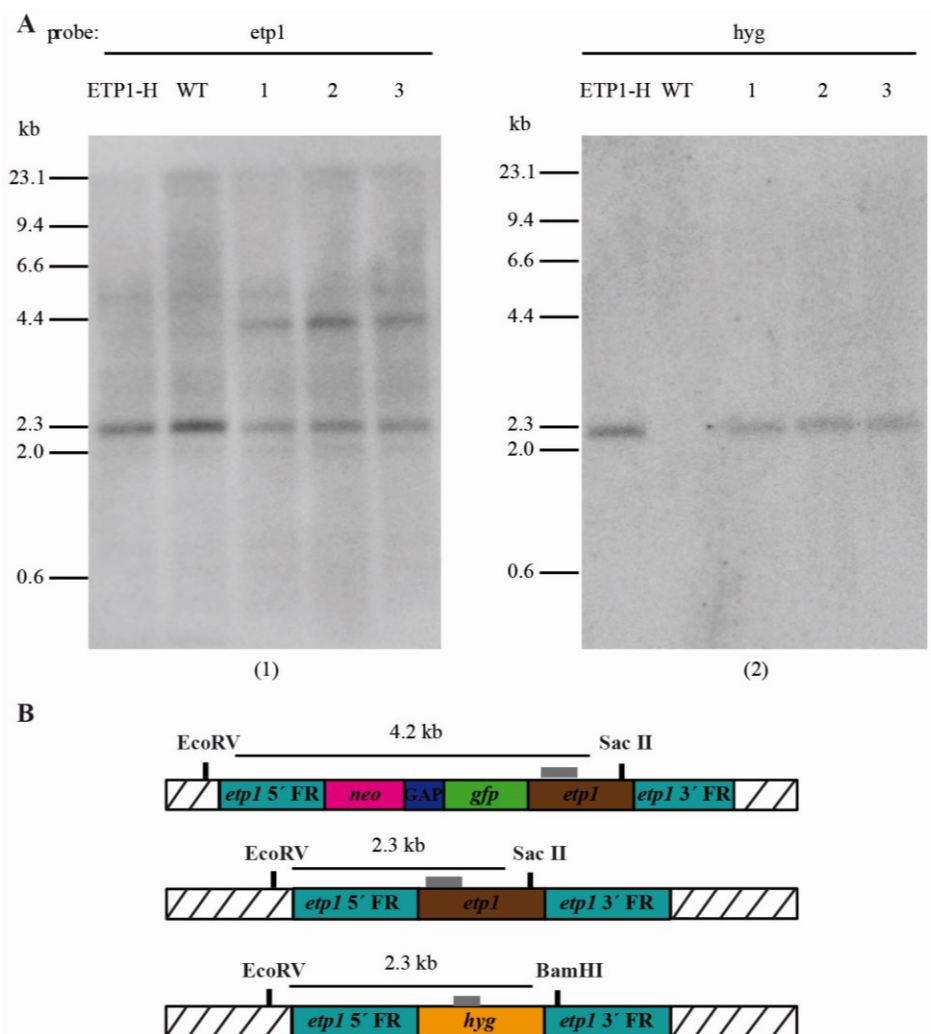
Figure 2.3-19: Southern blot analysis of Δ-ETP1^HYG/ Δ-ETP1^nGFP-ETP1, NEO transformants confirmed results of PCR analysis. Genomic DNA was restricted with the specified enzymes and analyzed by Southern blots using probes against *etp1* or *hyg* sequence as indicated above each blot. The Δ-ETP1^HYG SKO (**ETP1-H**), wild-type (**WT**) and three Δ-ETP1^HYG/ Δ-ETP1^nGFP-ETP1, NEO clones (**1, 2,** and **3**) were analyzed. (A) One band at 4.6 kb and one band at 2.3 kb could be readily observed in all three clones analyzed, indicating the insertion of *neo*-GAPDH IR-*gfp-etp1* in *etp1* allele and the retainment of *etp1* WT allele, respectively. (B) A single band at 2.3 kb corresponding to *hyg* insertion was identified in all three Δ-ETP1^HYG/Δ-ETP1^nGFP-ETP1, NEO clones analyzed. (C) Restriction maps are provided for the various blots. The **striped rectangles** represent genomic regions up and downstream of the insertion sites for the cassettes; **malachite green rectangle**, 5´- and 3´-flanking regions of *etp1* gene (etp1 FRs), that are present in the cassettes; **blue rectangle**, GAPDH IR; **green rectangle**, *gfp* ORF; **brown rectangle**, *etp1* ORF; **pink** and **orange rectangles**, *neo* and *hyg*, respectively.; **gray bar** highlights the binding region for the DNA probe used for hybridization.

3 Discussion

Parasitic and endosymbiotic bacteria share some common genetic aspects including genome size reduction, genetic isolation (both live in an intracellular environment remaining relatively isolated, a fact that favours the lack of horizontal gene transfer and DNA intake) and shift toward higher AT content, when compared to their free-living relatives (Moya *et al.*, 2009). They can be considered as two different products of a process that would begin the same way using the same mechanisms for the infection of the host cells, the avoidance of the host defense response and the multiplication within the host (Pérez-Brocal *et al.*, 2013). Nevertheless, whereas pathogenic bacteria manipulate their host cells through virulence factors according to their own needs, in endosymbiotic associations, the bacteria typically have lost their virulence factors and it seems to be the host that evolved mechanisms to control and manipulate its bacterial endosymbionts according to its own needs. Understanding the molecular basis of the mechanisms used by host cells to manipulate their bacterial endosymbionts will help to better comprehend the fascinating biology of endosymbiosis. Furthermore, this type of research might help eventually also to better understand immune reactions against intracellular parasitic bacteria and develop tools to control them.

3.1 Establishment of molecular biological tools for *A. deanei*

Despite the prevalence of bacterial endosymbionts across eukaryotic phyla the molecular mechanisms involved in host-endosymbiont associations are poorly understood. This gap of knowledge reflects the lack of effective molecular tools that enable functional studies of endosymbiotic systems. In the few insect/endosymbiont associations for which basic genetic tools are available, two approaches are used to disrupt gene function, insertional mutagenesis through untargeted transposon insertion (e.g. in *Sodalis glossinidius*, Dale *et al.*, 2001 and in *Burkholderia* symbionts, Ohbayashi *et al.*, 2015) and RNA interference (RNAi) (e.g. in *G.* m. *morsitans*, Wang *et al.*, 2009 and in *Sitophilus* spp., Vallier *et al.*, 2009). Untargeted insertional mutagenesis requires a substantial investment of time and resources, while RNAi works efficiently in only some groups of insects (Kikuchi, 2009; Cusson, 2008). For this reason, aim of this thesis was the development of molecular biological tools for *A. deanei*, a promising model system for studying endosymbiosis.

In this thesis, were established an electroporation protocol, hygromycin, neomycin and phleomycin resistances as markers for selection of transfectants and *γ*- and *δ-amastin* loci as

suitable genome sites for insertion and high-level expression of transgenes. The ability to obtain SKO and DKO cell lines in *A. deanei* allows for the knock-out of specific genes, the expression of endogenous proteins with altered amino acid sequences, the establishment of over-expresser cell lines and the testing of the function of proteins that are heterologously expressed. These fundamental capabilities will enable, for the first time, the time and cost effective in depth dissection of the molecular mechanisms that shape the interactions between a eukaryotic host and its vertically inherited endosymbiont.

However, in this thesis also the limitations of the so far developed tools became visible. Two rounds of knock-outs are needed to generate null mutants of a gene, and even more problems would arise for the deletion of multicopy genes. To overcome these limitations, the newly developed technology could be used to apply the CRISPR/Cas9 system to *A. deanei*. The CRISPR/Cas9 system has been established for genetic manipulation in different organisms including closely related *Leishmania* species (Sollelis *et al.*, 2015) and *T. cruzi* (Burle-Caldas *et al.*, 2015). In this system, non-coding small RNA (single guide RNA, sgRNA) binds to the complementary DNA sequence of the gene of interest (GOI) and directs the nuclease Cas9 to cause a double-stranded brake. More specifically, in *T. cruzi* a single transfection vector (pTREX plasmid), carrying both the Cas9 nuclease and sgRNA sequences transcribed from the rRNA promoter, was used and null mutant cell lines of the GOI were generated with high efficiency (Lander *et al.*, 2015). The same strategy could be applied in *A. deanei*. The CRISPR/Cas9 system could be an effective way to disrupt genes present in the *A. deanei* genome as multigene families (such as *α-tubulin*) and would allow the knock-out of multiple genes simultaneously (Jakočiūnas *et al.*, 2015), simplifying and speeding up the work even more.

3.2   An aposymbiotic *A. deanei* culture could not be obtained

A main aim of this thesis was to identify *A. deanei* genes involved in the symbiotic interaction with *Ca.* Kinetoplastibacterium crithidii. Comparison of the transcriptomes and proteomes of the symbiont-containing and the aposymbiotic *A. deanei* could be helpful for the identification of host genes that are involved in control of endosymbiosis. A previous study showed that *A. deanei* wild-type strain could lose its endosymbiont and generate an aposymbiotic strain after treatment with 0.4 µg/ml chloramphenicol (CAM) (Mundim and Roitman, 1977). In this thesis, treatment of *A. deanei* strain ATCC PRA-265 cells with different concentrations of either CAM or streptomycin and penicillin or ciprofloxacin failed to remove its

endosymbionts. Interestingly, loss of endosymbionts was initially observed in most treatments as evidenced by FISH stainings using a probe against the bacterial 16S rRNA. However, upon limiting dilution, only cells that retained their endosymbiont survived the treatment with the antibiotics. This finding indicated that the presence of the endosymbiont is obligatory for host survival and only cells whose endosymbionts showed resistance against high concentrations of the antibiotics used survived. This resistance could be conferred for example by mutations that cause reduced drug accumulation in the endosymbiont, either by a decreased uptake or by an increased efflux of the antibiotic (Hooper *et al.*, 1987; Fàbrega *et al.*, 2009). Mundim and Roitman used *A. deanei* strain ATCC 30255 as starting material from which to generate their aposymbiotic culture. However, the strain ATCC 30255 that was available until 2009 through the ATCC collection did not seem to represent the endosymbiont-bearing ancestral strain from which derived the so far only aposymbiotic *A. deanei* strain, as it was revealed by Yurchenko *et al.*, 2009, and it was removed from the collection. Thus, the nature of the ancestral strain used for aposymbiont generation remains unclear. A recent study revealed a marked genetic diversity within *A. deanei* and the identification of at least four genotypes (Dea1-Dea4) (Borghesan *et al.*, 2018). Therefore, the amount of genomic divergence between the strain used in this thesis (ATCC PRA-265) and the ancestral strain from which the aposymbiotic *A. deanei* cell line was obtained is unknown. Nevertheless, this genetic diversity could explain the difference in the importance of the presence of the endosymbiont for the survival of the host cells between the two strains.

Interestingly, a recent study compared the transcriptomes of the symbiont-containing *A. deanei* (ATCC 30255) and the aposymbiotic strain (ATCC 044) (it is not specified in the study if this is the strain generated by Mundim and Roitman) using large-scale RNA sequencing technology (Penha *et al.*, 2016). This comparison revealed that the genes downregulated in the absence of the symbiont were those involved in oxidation-reduction process, ATP hydrolysis coupled proton transport and glycolysis. In contrast, among the upregulated genes were those involved in proteolysis, microtubule-based movement, and cellular metabolic process. This up or down regulation of specific gene categories in the aposymbiotic strain suggests that the presence of the endosymbiont intervenes in the gene expression of several host cell pathways (Penha *et al.*, 2016). Future studies focusing on these gene categories may increase our understanding of the symbiotic process in *A. deanei.*

3.3 Host-encoded ETP1 is targeted to the endosymbiont of *A. deanei*

Since the treatment of wild-type ATCC PRA-265 *A. deanei* cells with different antibiotics failed to generate an aposymbiotic strain, an alternative approach was chosen to identify host genes that might control endosymbiosis. Proteomic analysis of the whole cell lysate versus the purified endosymbiont fraction that was previously conducted in our lab by Dr. Jorge Morales, revealed the enrichment of 8 host-encoded proteins in the endosymbiont fraction. Among them, Endosymbiont-targeted Protein 1 (ETP1) was the protein that showed the highest enrichment (Morales *et al.*, 2016). Using the newly developed tools, it was shown that heterologous expression of a fusion protein comprised of ETP1 and EGFP resulted in a clear fluorescence signal that co-localized with the endosymbiont (Firgure 2.3-3). Further analysis using confocal microscopy and stimulated emission depletion (STED) microscopy showed localization of ETP1 only in the periphery of the endosymbiont (Master thesis, Georg Ehret). Moreover, preliminary analysis of *A. deanei* nGFP-ETP1 expressing cells by immunogold EM showed that ETP1 seems to localize in the periplasm of the endosymbiont but this needs to be confirmed further (in progress, Georg Ehret).

Regarding the signal sequences that target ETP1 to the endosymbiont, expression of truncated forms of ETP1 missing up to 150 amino acids (aa) from N-or C-terminus fused to EGFP still showed endosymbiont localization, indicating the absence of any apparent N-or C-terminal targeting signal in ETP1. Interestingly, expression of an ETP1 truncated form missing 200 aa from the N-terminus fused to EGFP resulted in complete loss of the signal, suggesting that the 50 aa sequence between the 150 and 200 truncated forms is important for the stability and the targeting of the protein. Expression of this 50 aa sequence fused to EGFP was not sufficient for endosymbiont targeting and cytosolic localization of EGFP signal was detected (Master thesis, Georg Ehret).

Targeting signals vary between different endosymbiotic systems. In insects with bacteriocyte-housed endosymbionts, proteins targeted to the endosymbionts contain an N-terminal signal peptide indicating that these proteins enter the secretory pathway (Nakabachi *et al.*, 2014). The proteins synthesized in the endoplasmic reticulum may be transported, via the Golgi, to the extracellular space. This is consistent with the fact that the endosymbionts are housed within vesicles called symbiosomes, which consist of host-derived membranes and electron microscopic observations have shown that these symbiosomal membranes are sometimes closely associated with the bacteriocyte endoplasmic reticulum (Griffiths and Beck, 1975). As the inner space of the symbiosome is topologically equivalent to the external space of the

plasma membrane, the signal peptide should be capable of leading the protein to the symbiosomes. However, the mechanism that controls that the protein is targeted specifically to the symbiosome and how the protein is transported to the endosymbiont after entering the symbiosome is unclear (Shigenobu and Stern, 2013; Nakabachi *et al.*, 2014). A similar mechanism might exist in legume plants in which endosymbiotic rhizobial bacteria are housed in symbiosomes and the proteins targeted to the endosymbionts contain an N-terminal signal peptide that targets proteins to the secretory pathway (Willem Van de Velde *et al.*, 2010; Coba de la Peña *et al.*, 2018). In the amoeba *P. chromatophora*, where the cyanobacterium-derived chromatophore lies freely in the cytoplasm, short nuclear-encoded proteins that are targeted to the chromatophore are missing a signal peptide, whereas nuclear-encoded proteins longer than 268 amino acids that are also targeted to the chromatophore carry a conserved N-terminal sequence of approximately 200 amino acids (Singer *et al.*, 2017). Additionally, import of most nucleus-encoded proteins into mitochondria and plastids is mediated by N-terminal presequences (Diekert *et al.*, 1999; Patron and Waller, 2007). Interestingly, all proteins of the mitochondrial outer membrane and some proteins of the intermembrane space and the inner membrane are devoid of such signals (Diekert *et al.*, 1999). More specifically, in β-barrel proteins, like porin, the targeting signal is not restricted to a specific region but is distributed throughout the polypeptide sequence and it has been proposed that the targeting information is included in the secondary and/or tertiary structures (Rapaport, 2003; Walther *et al.*, 2009). A similar targeting signal might be responsible for the targeting of ETP1, since ETP1 is targeted to the periphery of the endosymbiont (Master Thesis, Georg Ehret).

3.4 Characterization of ETP1

3.4.1 Biochemical characterization

Considering the biochemical characterization of ETP1 only preliminary data were acquired during this thesis and there is the need to shed light on various aspects by further studies. ETP1 localizes in the periplasm of the endosymbiont (in progress, Georg Ehret) and porin, an outer membrane protein, is one of its possible interaction partners. Moreover, the difficulty to extract ETP1 into the soluble fraction with non-ionic detergents, such as Triton X-100, DDM or NP-40, indicates interaction with the membrane through covalent linkage. Considering that ETP1 has no transmembrane domain predicted, it could interact with the membrane through covalent attachment to a lipid group (e.g. palmitoylation, prenylation) or covalent linkage to an oligosaccharide molecule (e.g. N-glycosylation). Indeed, analysis of ETP1 sequence

predicted 2 possible sites for N-glycosylation and 2 potential sites for palmitoylation. The 2 positions that were predicted as possible N-glycosylation sites (amino acid position 19 and 234 in ETP1) were among the peptides of ETP1 identified by the MS analysis and no modification was observed. The 2 positions that were predicted as possible palmitoylation sites (amino acid position 176 and 179 in ETP1) were not identified by the MS analysis. Palmitoylation of ETP1 could explain the 14 kDa discrepancy between the expected and the observed size of ETP1. Thus, to study the possibility that ETP1 interacts with membrane lipids and to determine the identity of these lipids an approach that combines high-energy native mass spectrometry (HE-nMS) and solution-phase lipid profiling could be used (Gupta *et al.*, 2018). In the beginning, ETP1 needs to be expressed as a fusion with a protease-cleavable fusion, such as GFP containing a terminal His-tag in *A. deanei* cells. The GFP fusion provides visual confirmation of protein expression and the His-tag is needed for the purification of the membrane complex of ETP1. Then, the endosymbionts of this cell line should be isolated and lysed. Following that, their membranes are purified and solubilized, using the appropriate detergent that allows the stability of the membrane complexes and is compatible with MS analysis (Laganowsky *et al.*, 2013). The only detergent combination found so far to solubilize effectively ETP1 is the combination of 1% Triton X-100 and 0.1% SDS in the RIPA buffer. Nevertheless, SDS is not suitable for this approach so different detergents should be screened (Laganowsky *et al.*, 2013). Next the membrane complex of ETP1 should be purified by immobilized affinity chromatography (IMAC) and the buffer should be exchanged into MS-compatible buffer supplemented with the detergent that was used before. Then, HE-nMS platform is used that allows dissociation of the detergent micelle, the isolation of the protein–lipid complex and its successive fragmentation, which leads to identification of the bound lipid masses. Simultaneous coupling of this analysis with in-solution Liquid Chromatography (LC)-MS/MS-based quantitative lipidomic allows the determination of both the number and molecular identity of the lipids that co-purify with the ETP1 (Gupta *et al.*, 2018).

3.4.2 Generation of ETP1 null mutants

The absence of ETP1 homologs in any other organism, including other symbiont-harboring trypanosomatids of the Strigomonadinae, makes it impossible to predict the function of this protein. To characterize its function, this thesis aimed at generating ETP1 null mutants

through two consecutive rounds of *etp1* gene replacements through homologous recombination.

ETP1 SKOs could be easily generated by homologous recombination and were compared to wild-type cells using EM. Nevertheless, there were observed no obvious phenotypic differences between the two cell lines. So far it was not checked whether the levels of the mRNA that express ETP1 are decreased in the ETP1 SKOs. An mRNA expression analysis should be conducted comparing the expression of a housekeeping gene (e.g. *α-tubulin*) and *etp1* in both wild-type and ETP1-SKO cells. Thus, quantitative reverse transcription-PCR could be applied for this analysis using primers specific for the housekeeping gene and *etp1*.

Interestingly, in the Δ-ETP1^{NEO} (Adea0038) and Δ-ETP1^{HYG} (Adea0071) SKOs clonal cell lines that were generated it was found that specifically *etp1* encoded by the Q-rich scaffold was removed (Fig. 2.3-9). Moreover, generation of null mutants was not possible. These observations were initially interpreted to indicate a dispensable function of "ETP1 Q-rich" isoform while the "ETP1 Q-less" isoform seemed to be crucial for the survival of the cells. However, recently it was noticed by other members in our group that also the replacement of other *A. deanei* genes with resistance cassettes seems to consistently result in knock-out of the specific allele matching the flanking regions in the resistance cassette. This observation led to the hypothesis that slight nucleotide sequence differences in the flanking regions between the two alleles could be enough to drive homologous recombination specifically into one allele. In other trypanosomatids homologous recombination is not affected by slight nucleotide differences but it might be a species specific characteristic observed in *A. deanei*. Re-evaluation of the ETP1 knock-out strategy used in this thesis revealed that the *etp1* 5′ FR sequences used in both, pAdea-ETP1/NEO and pAdea-ETP1/HYG plasmids derive from the *etp1* Q-less allele, whereas the *etp1* 3′ FR sequences derived from the *etp1* Q-rich allele. *Etp1* 5′ FRs sequences differ slightly between the two alleles. *Etp1* 3′ FR sequence deriving from the *etp1* Q-less allele is not available yet so it is not known to what extend the sequences of the two alleles differ. If slight nucleotide differences between the target locus and the regions of homology in the resistance cassette indeed determine efficiency of HR, the presence of *etp1* 3′ FR sequence from the *etp1* Q-rich allele in both plasmids could drive HR specifically into the Q-rich allele and could explain the specific removal of "*etp1* Q-rich" form in all clonal cultures that were tested and the inability to generate ETP1 DKOs. For this reason, a plasmid should be constructed containing *etp1* 5′ and 3′ FRs from the *etp1* Q-less allele and a

resistance marker to repeat the transfections of the wild-type and ETP1 SKO *A. deanei* cell lines in order to test which allele will be deleted and if the generation of a DKO is possible.

The problem of generating null mutants of essential genes, like *etp1*, brought to the fore the need for the development of reliable protocols for gene knock-downs. RNA interference (RNAi) is a technique that could knock-down essential genes (e.g. *etp1*) to levels that would allow the cells to survive and simultaneously the observation of morphological alterations when compared to wild-type cells. The first step in RNAi system involves degradation of double-stranded RNA (dsRNA) (of exogenous or nuclear origin) into small interfering RNAs (siRNAs), 20- to 30- nucleotides long, by an RNase III-like activity. In the second step, the resulting siRNA is loaded onto an Argonaute family protein by a complex, which comprises of Dicer, a dsRBP protein, and an Argonaute protein. One of the two strands of the siRNA is cleaved and eliminated. The guide strand remains bound to Argonaute, forming the RNA-induced silencing complex (RISC). The RISC binds to complementary mRNA target sequences and it drives silencing of this mRNA via degradation (Wilson and Doudna, 2013; Agrawal *et al.*, 2003). A recent study demonstrated that the RNAi machinery is functional in *A. deanei* and knock-down of specific genes can be achieved by transfection of cells with dsRNAs (Catta-Preta *et al.*, 2016). However, an initial attempt to repeat gene knock-downs through RNAi in our lab was not successful. Another option would be the use of morpholino antisense oligonucleotides (MAOs) complementary to part of *etp1* sequence to block translation of etp1. MAOs are synthetic oligomers that bind to complementary sequences of RNA and can efficiently block translation by sterically blocking the translation initiation complex. MAOs have been previously used to decrease protein levels of a certain protein in the closely related *T. cruzi* (Hashimoto *et al.*, 2014). Alternatively, a technique known as protein knock-down that was established in *T. cruzi* could be applied (Ma *et al.*, 2015). Using this system, the expression of proteins is regulated at post-translational level. In the beginning, the gene of interest (GOI) needs to be expressed fused to the destabilization domain from dihydrofolate reductase (DDD). Proteins containing a DDD are stabilized by the addition of trimethoprim-lactate (TMP-lactate). Removal of TMP-lactate leads to protein degradation. That way it would be possible to generate a conditional knock-down of the GOI and investigate the changes in the phenotype of the cells from the time point that TMP-lactate is not added causing degradation of the protein of interest and compare it to wild-type cells. Any observed changes in the phenotype would provide insights into the function of the protein.

3.4.3 Putative interaction partners of ETP1

To investigate further the function of ETP1, the nGFP-ETP1 cell line was subjected to Co-IP assay coupled to MS analysis, in order to identify proteins that interact with ETP1. MS analysis of the proteins that were bound on the beads revealed according to the criteria set porin, Kinetoplastid membrane protein-11 (KMP-11), uncharacterized protein S9U3N9, protein FAM184A, 30S ribosomal protein S19 and ETP2 as putative interaction partners. Unfortunately, the results of the MS analysis were not satisfying and the intensities and spectral counts obtained by MS did not fit the distribution and amount of protein present in the samples as judged by the highly reproducible bands observed in the CBB and silver-stained gels (Figure 2.3-14). Therefore, it will be essential to repeat the MS analysis in order to draw a more robust conclusion regarding identity and distribution of the detected proteins over the various samples. However, there is a considerable overlap of proteins identified in the Co-IP approach and the two-dimensional CN/SDS-PAGE using ETP1-cV5 endosymbionts where porin, KMP-11 and ETP2 were found to co-migrate with the fusion protein. These findings strongly support the notion that these peptides are indeed interacting with ETP1 at the endosymbiont rather than this results being an artifact of the experimental procedure considering that the Co-IP and 2D analysis are very different approaches and that different tags were used in each procedure (GFP and V5, respectively).

The interactions between host-encoded proteins (i.e. ETP1, ETP2 and KMP-11) could be directly verified using the bimolecular fluorescence complementation (BiFC) assay, also referred to as "split-fluorescent protein" technology. BiFC is based on the *in vivo* reconstitution of fluorescence after two non-fluorescent halves of a fluorescent protein (FP) are brought together by a pair of interacting proteins (Kodama and Hu, 2012). As such, BiFC not only provides information on whether two proteins interact, but can also be used to determine the cellular and sub-cellular site of a protein-protein interaction event. In more detail, to prove for example the interaction of ETP1 and KMP-11 the FP (e.g. YFP) should be splitted into two non-fluorescent "halves", an N-terminal part/half of the protein (FP^N) and a C-terminal part/half of the protein (FP^C) and these parts should be fused to ETP1 and KMP-11 proteins, respectively. Then, the fusion proteins should be expressed in *A. deanei* using the molecular tools established. Reconstitution of the FP molecule upon interaction between ETP1 and KMP-11, would result in detection of fluorescence. The key elements of a BiFC assay are the selection of proper negative and positive controls and the fusion orientation.

Regarding the latter element protein-protein interactions are mediated by specific protein domains and fusing other proteins to a protein of interest can interfere with the interaction capacity of these domains (Horstman *et al.*, 2014). For this reason, all eight combinations of constructs in which the N- and C-terminal fragments of the FP are fused to the N- and C-terminus of the proteins of interest should be tested. Nevertheless, BiFC assay is not suitable to prove the interaction of ETP1 with porin, since porin is encoded by the endosymbiont and the genetic manipulation of the endosymbiont has not been established so far. An idea would be to utilize split-ubiquitin based Membrane Yeast Two-Hybrid (MYTH) system, which is applied for the identification of protein-protein interactions for full-length transmembrane proteins, using *Saccharomyces cerevisiae* as a host organism (Snider *et al.*, 2010). MYTH takes advantage of the observation that ubiquitin can be separated into two stable parts: the C-terminal half of yeast ubiquitin (C_{ub}) and the N-terminal half of the ubiquitin (N_{ub}). Porin could be fused to C_{ub} which is linked to an artificial transcription factor and ETP1 could be fused to the N_{ub} moiety and expressed in *S. cerevisiae*. Protein interaction between the porin and ETP1 would lead to reconstitution of the ubiquitin moieties, forming a full-length "pseudo-ubiquitin" molecule. This molecule would be then recognized by cytosolic deubiquitinating enzymes, resulting in cleavage of the transcription factor, and subsequent induction of a reporter gene expression (Snider *et al.*, 2010). The difficulty using this system arises from the fact that porin is a bacterial protein of the outer membrane that should be expressed in *S. cerevisiae.* It is not known if porin would insert into a suitable membrane in a suitable orientation in yeast in order to interact with the co-expressed ETP1. Alternatively, far-western blot analysis could be used (Wu *et al*., 2007). In more detail, *A. deanei* cell lysate could be separated by SDS-PAGE and then the proteins would be transferred to a membrane. A purified, tagged version of ETP1 (GFP or V5 tag) could be used to probe the membrane. Once bound, horseradish peroxidase-conjugated α-GFP or α-V5 antibody that targets the tag of ETP1 could be used to label the interaction, and it would be then detected by enzymatic chemiluminescence. A band at the size of porin (45 kDa) should be detected to confirm the interaction of ETP1 with porin.

Porin is the candidate interaction partner that shows the highest enrichment in both analyses and a prominent band at the expected size of 45 kDa is visible in CBB and silver-stained gels in the Co-IP eluates. In a previous study a single-copy gene sequence encoding a 45 kDa porin-like protein was retrieved from the *A. deanei* endosymbiont genome database (GenBank accession no. HM480845) and it was proved that it localizes only in the endosymbiont of *A. deanei* and more specifically its outer membrane (Andrade *et al.*, 2011). The endosymbiont

porin represents an 18 β- strands porin, as it was revealed by its secondary structure analysis in the same study, but little is known about its specific function. 18 strands porins are trimeric and they are classified as substrate specific (Galdiero *et al.*, 2012). Porins are commonly found in the outer membrane of bacteria and in archaea, as well as in organelles of symbiotic origin (Burghardt *et al.*, 2007). Interestingly, Tom40 in mitochondria and Toc75 in chloroplasts are components of the organellar protein importers which derive from bacterial β-barrel pores (Dolezal *et al.*, 2006; Shi and Theg, 2013). Together with eukaryotic proteins they assemble into multimeric protein import machineries that span the inner and outer membranes of the organelles. The mitochondrial porins also include a voltage-dependent anion-selective channel (VDAC), the most abundant protein of the mitochondrial OM, which constitutes the main channel to transport metabolites and substrates (Colombini, 2004). Following the same strategy, endosymbiont metabolites might pass the outer bacterial membrane through porin to complement essential metabolic pathways of the host cell. The opposite way, molecules produced by *A. deanei* might pass through porin and become available to the endosymbiont (Andrade *et al.*, 2011). The molecular basis for metabolite transport across the inner endosymbiont membrane is not explored yet. Nevertheless, considering that porin is a possible interaction partner of ETP1 and that there is evidence that ETP1 localizes to the periphery of the endosymbiont (Master Thesis, Georg Ehret) ETP1 might play a role in regulating porin-based metabolite transport from different pathways across the endosymbiont outer membrane. Therewith, ETP1 might be involved in controlling metabolite flux between the two partners thus, establishing a homeostatic and synergistic relationship between the endosymbiont and the host protozoan.

Another interesting candidate interaction partner of ETP1 is the host-encoded KMP-11. It is widely distributed among kinetoplastids and it was found in association with membrane structures at cell surface, flagellar pocket and intracellular vesicles (de Mendonçaet *et al.*, 2015). Additionally, it is suggested that it has an important role in parasite membrane organization and function (Stebeck *et al.*, 1995). In fact, our group has generated a cell line that expresses a fusion of KMP-11 with GFP and the resultant peptide has a dual localization between the basal body of the parasite and the periphery of the endosymbiont in a way that resembles ETP1 at this location (Dr. Jorge Morales, unpublished data). Interestingly, in trypanosomatids cell division starts by the basal body and it has been shown in *T. brucei* that RNA interference of KPM-11 inhibits segregation of this organelle and cytokinesis (Li and Wang, 2008). It is plausible to speculate that the putative association of KMP-11 with ETP1 is somehow related to synchronization of the cell cycle of the host with its endosymbiont. To

gain information about the specific function and the importance of KMP-11 in *A. deanei*, single and double (if possible) gene knock-outs should be generated. Moreover, the conditional protein knock-down could be applied as described before. KMP-11 could be fused with a different destabilization domain than ETP1, like the destabilization domain of rapamycin binding protein (ddFKBP). This protein could be stabilized by the addition of the synthetic ligand Shield-1 (Burle-Caldas *et al.*, 2015). That way it would be possible to degrade ETP1 or KMP-11 or both depending on which stabilizing agent is used and check the effect on the cells. Degrading both partners could test if ETP1 and KMP-11 cause any synergistic effect. In this context in a cell line that expresses a tagged version of ETP1 or KMP-11 the other partner could be knocked out and any change in the localization of the other partner could be investigated. Additionally, the mRNA levels of one protein should be checked when the other partner is knocked-down/out by quantitative reverse transcription-PCR. It might be the case that when one partner is downregulated the expression of the other partner is affected. Certainly, for the assays mentioned above new selection markers are needed. Antibiotics that are used for genetic manipulation in other trypanosomatids like zeocin, tunicamycin or blasticidin should be tested (Burle-Caldas *et al.*, 2015).

Last but not least, ETP2 is also a candidate interaction partner of ETP1. ETP2 is one of the 8 host-encoded proteins that were found to be enriched in the endosymbiont fraction in the proteomic analysis that was conducted by Dr. Jorge Morales. ETP2, like ETP1, has no homologs to other organisms and is specifically targeted to the endosymbiont of *A. deanei,* as it was shown by previous work in our lab (Master thesis, Georg Ehret). In contrast to ETP1, ETP2 is distributed among the endosymbiont heterogeneously, showing a prominent localization over the endosymbiont division site. Considering the localization of ETP2 and that ETP1, KMP-11 and ETP2 are putative interaction partners it could be hypothesized that these partners have a significant role to the synchronization of the cell cycle of the host with its endosymbiont and the cell division. Similar approaches like the ones that were used for ETP1 and should be used for KMP-11 could be applied in order to investigate the function of ETP2 (Co-IP analysis, gene knock-out and/or protein knockdown).

Additionally, uncharacterized protein S9U3N9 was co-purified with nGFP-ETP1 in all replicates in the Co-IP assay and it was not identified in any of the controls. On the other hand, it was not identified by the 2-D CN/SDS-PAGE analysis. It might be worth trying to express a fusion of the host-encoded S9U3N9 with EGFP or V5 and check its subcellular

localization. In the case that it is targeted to the endosymbiont, the same approach as for ETP1, KMP-11 and ETP2 should be followed.

Regarding nGFP-ETP1 fusion protein that was used for the CoIP assay different approaches were applied in order to prove its functionality but all approaches were unsuccesful. In the last approach was tested whether in the Δ-ETP1^{HYG} SKO background it would be possible to substitute the second *etp1* allele by the nGFP-ETP1 fusion protein. The double-resistant clones that were generated were analyzed by PCR analysis, using primers that anneal to gDNA outside of the FR sequence and inside the resistance markers. Surprisingly, this PCR analysis demonstrated *hyg* and *neo*-GAPDH IR-*gfp-etp1* insertions in the desired *etp1* locus and the simultaneous conservation of wild-type *etp1* (Figure 2.3-18). This result was also confirmed by Southern Blot analysis of the double-resistant clones (Figure 2.3-19). At this point two scenarios are possible and could explain this result. On one hand, there is the case that nGFP-ETP1 was inserted in a third *etp1* locus, indicating that there is a second copy of *etp1* in the *A. deanei* genome. This could be tested by restriction of WT *A. deanei* gDNA with different combinations of restriction enzymes and subsequent Southern blot analysis of these samples using the etp1 probe. The identification of more than one band in any of the restriction enzyme combinations would be an indication of the presence of additional *etp1* copies. On the other hand, there might be the case that *A. deanei* cells have undergone gene rearrangements involving the *etp1* sequence. Similar events have been described frequently in *Leishmania* species (Boitz and Ullman, 2010) and rarely in *T. cruzi,* where gene amplification and overexpression of sequences have been observed after disruption of essential genes (Cardoso *et al.*, 2013). In the latter study, it was suggested that the amplification of the sequence of their gene of interest involved the production of episomal DNA molecules (Cardoso *et al.*, 2013). Until today the mechanisms that are employed by the parasites in order to achieve these genomic rearrangements have not been identified. If this is the case it would mean that nGFP-ETP1 is not functional and for this reason *etp1* was amplified in a different genomic location. This scenario could be checked by restriction of gDNA isolated from the WT and a double resistant clone with different combinations of restriction enzymes and subsequent Southern blot analysis of these samples using the etp1 probe. The identification of different banding patterns in the double resistant clone compared to the WT sample, would be an indication of genomic rearrangements.

3.4.4 Outlook

In this thesis a molecular toolbox was developed that allows for the generation of null mutants of specific genes by homologous recombination and the heterologous expression of transgenes in an endosymbiotic system for the first time. Using this newly established genetic tool, ETP1 was the first host-encoded protein proved to target the endosymbiont. Moreover, the fact that ETP1, porin, KMP-11 and ETP2 were identified as interaction partners by two independent methods and the indications that these four proteins target the periphery of the endosymbiont are really promising. It is still early to propose that ETP1 has any role in the transport of different metabolites, in the organization and function of the endosymbiont membrane or in cell division. Nevertheless, these first findings give some evidence on how host-encoded proteins might shape symbiosis with bacterial endosymbionts in *A. deanei*. Considering that *A. deanei* is a newly established model organism, this thesis lead to major improvements and certainly there is still a lot of room for optimization in order to get detailed insights into the host-endosymbiont interactions.

4 Material & Methods

4.1 Materials

4.1.1 Enzymes and Kits

Table 4.1-1: Enzymes used in this study

Name	Function	Company
DNase I	nonspecifically cleaves DNA	Thermo Scientific
Lysozyme	breaks down the bacterial cell wall	Sigma Aldrich
Phusion DNA Polymerase	Amplification of DNA molecules	New England Biolabs or prepared in the lab
Restriction enzymes	DNA restriction	New England Biolabs
RNase A	specifically degrades single stranded RNA at C and U residues.	Thermo Scientific
T4-DNA ligase	DNA ligation	New England Biolabs

Table 4.1-2: Protease inhibitors used in this study

Name	Application	Company	Preparation
cOmplete™ EDTA-free Protease Inhibitor Cocktail	Preparation of cell extracts and supernatant samples (inhibits a broad spectrum of serine and cysteine proteases)	Roche	10x stock solution in dH$_2$O
E64	Preparation of cell extracts and supernatant samples (irreversibly inhibits cysteine proteases).	Sigma-Aldrich	1 mM stock solution in dH$_2$O
Na-tosyl-L-lysine chloromethyl ketone (TLCK)	Preparation of cell extracts and supernatant samples (serine protease inhibitor)	Sigma-Aldrich	100 mM stock solution in dH$_2$O
Pepstatin A	Preparation of cell extracts and supernatant samples (reversibly inhibits aspartic acid proteases)	Sigma-Aldrich	1 mM stock solution in methanol

Phenylmethylsulfon ylfluorid (PMSF)	Preparation of cell extracts and supernatant samples (serine protease inhibitor)	Applichem GmbH	100 mM stock solution in methanol

Table 4.1-3: Kits used in this study

Name	Application	Company
CDP-Star®	Chemiluminescent substrate for alkaline phosphatase in Southern blot analysis	Roche
CloneJET PCR cloning Kit	Cloning of blunt- or sticky-ended DNA fragments in pJET	Thermo Fisher Scientific
Mix2Seq Kit	Sanger Sequencing of DNA molecules	Eurofins
PCR DIG Labeling Mix	Digoxigenin-labelling of PCR products (for DNA probes used in Southern blots)	Roche
Plasmid Mini Kit	Purification of plasmids	Qiagen
QIAprep Spin Midiprep Kit	Purification of plasmids	Qiagen
QIAquick Gel Extraction Kit	Elution and purification of DNA fragments from agarose gels	Qiagen

Table 4.1-4: Protein- and DNA size standards used in this study

Name	Application	Company
GeneRuler 1 kb DNA Ladder	DNA Gel electrophoresis	Thermo Fisher Scientific
GeneRuler 100 bp DNA Ladder	DNA Gel electrophoresis	Thermo Fisher Scientific
NativeMark™ Unstained Protein Standard (20-1,236 kDa)	Native gels	Thermo Fisher Scientific
PageRuler™ Prestained Protein Ladder (10-180 kDa)	SDS-PAGE	Thermo Fisher Scientific

4.1.2 Media

4.1.2.1 Cultivation of *Angomonas deanei*

A. deanei strain ATCC PRA-265 obtained from the American Type Culture Collection was grown at 28℃ without shaking in Brain Heart Infusion (BHI) Broth (37 g/L) supplemented with 10% horse serum and hemin (10 µg/ml) (Motta *et al.*, 2013).

Table 4.1-5: Media for cultivation of *A. deanei*

Name	Components and stock solutions
BHI powder	5 g/L beef heart (infusion from 250 g), 12.5 g/L calf brains (infusion from 200 g), 2.5 g/L disodium hydrogen phosphate, 2 g/L D(+)-glucose, 10 g/L peptone, 5 g/L sodium chloride
Hemin	2.5 mg/ml stock solution in dH$_2$O
Horse serum	inactivated by heating at 56℃ for 30 min
Preparation	Addition of BHI (37 g/L) in dH$_2$O. Followed by autoclaving, addition of 10% of heat inactivated horse serum (Sigma-Aldrich) and 10 µg/ml hemin (Sigma-Aldrich).

Table 4.1-6: Antibiotics used in *A. deanei* cultivation

Antibiotic	Stock solution concentration	Company
Chloramphenicol	50 mg/ml stock solution in 95% EtOH	Sigma Aldrich
Ciprofloxacin	25 mg/ml stock solution in 0.1N HCl	Sigma Aldrich
G418 (Geneticin)	50 mg/ml stock solution in dH$_2$O	Sigma Aldrich
Hygromycin	50 mg/ml stock solution (ready to use)	Roth
Penicillin	100 mg/ml stock solution in dH$_2$O	Sigma Aldrich
Phleomycin	20 mg/ml stock solution in dH$_2$O	InvivoGen
Puromycin	10 mg/ml stock solution (ready to use)	InvivoGen
Streptomycin	25 mg/ml stock solution in dH$_2$O	Sigma Aldrich

4.1.2.2 Cultivation of *Escherichia coli*

E. coli Top10 cells were grown in LB medium supplemented with 100 µg/ml Ampicillin (Roth) at 37° C either on a rotary wheel in glass tubes or in shaking flasks. All cultures were inoculated with single colonies from selective LB agar plates.

Genotype: F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*araleu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G

Table 4.1-7: Media for cultivation of *E. coli*.

Name	Ingredients	Preparation
Lysogeny broth (LB)	10 g tryptone 5 g yeast extract 5 g sodium chloride	Add 1 L H$_2$O Autoclave
LB agar plates	10 g tryptone 5 g yeast extract 5 g sodium chloride 20 g agar	Add 1 L H$_2$O Autoclave

4.1.3 Oligonucleotides

Table 4.1-8: Oligonucleotides used in this study.

Name	Nucleotide sequence 5´-3´	Project
Ama5'UTRsc2 FW	ACTCCTCCACCACTACCACC	Checking insertion in *δ-amastin* locus
Neo internal FW	CGAAACATCGCATCGAGCG	Checking insertion of neo resistance gene
Neo internal RV	ATCGACAAGACCGGCTTCC	Checking insertion of neo resistance gene
Ama3'UTRsc RV	AGAGAGAGGTGGGTACATGCAA	Checking insertion in *δ-amastin* locus
gama-Ama-region FW	CTTTCTGCCATCTGCCTCAT	Checking insertion in *γ-amastin* locus
Hyg internal FW	TGTGTATCACTGGCAAACTGTG	Checking insertion of hygro resistance gene
Hyg internal RV	AAAGCATCAGCTCATCGAGAG	Checking insertion of

		hygro resistance gene
gama-Ama-region RV	CATCCTTACGATCTTCTTATTTTTGG	Checking insertion in *γ-amastin* locus
5'neonostopBsaI FW	GGTCTCGCCTGCAGTGCGTCGCCCGGCTA	
pAEXCand1-nGFP REV (2)	GGTCTCCGGCTTGTACAGCTCGTCCAT	
pAEXCand1-nGFP FWD (3)	GGTCTCGAGCCCGCCGCGCTGCCCCCC	Construction of pAEX-nGFP-ETP1 (pAdea039)
pAEXCand1-n REV (3)	GGTCTCGCTTTTACGCATCCTTCTTCTCTACAACATCCGA	
pAEXCand1-n FWD (4)	GGTCTCCAAAGAGGGGGGAGAGAGAC	
5'neonostopBsaI FW	GGTCTCGCCTGCAGTGCGTCGCCCGGCTA	
pAEXCand1-c REV(2)	GGTCTCCTTTGGATAACTGTGTTTTTTGATGAAAGAAGAC	
pAEXCand1-c FWD (3)	GGTCTCCCAAATGCCCGCCGCGCTGCCC	Construction of pAEX-ETP1-cGFP (pAdea040)
pAEX Cand1-cREV(3)	GGTCTCCCACCGCATCCTTCTTCTCTACA	
pAEX Cand1-c FWD(4)	GGTCTCCGGTGAGCAAGGGCGAGGAG	
3-pUMAmaBsaI RV	GGTCTCGCTGCGTCGCGGGGGCTGTCGCA	
Neo probe fw	TTG TCA AGA CCG ACC TGT C	
Neo probe rev	CAA GAA GGC GAT AGA AGG C	
Amast probe fw	ATC TAC TGC ATC CTC CAA TGC	
Amast probe rev	GAC GAT GAT GCA GTT GAT GA	
d-Amastin probe fw	ACT ACT GCG TGA CCC TCT GG	
d-Amastin probe rev	AGT CGT CGT GCT TGT TGT TG	
Hyg probe fw	ATC TTA GCCAGA CGA GCG	Probes for Southern blot analysis
Hyg probe rev	CAC TAT CGG CGA GTA CTT CTA CA	
g Amastin probe fw	CCG CTT TTG AGA TTG TCT CC	
g Amastin probe rev	ACG ATG ACT TGC AGA CAC CA	
ETP1 probe 1 fw	TGC CGT CTA CTA CGG CGG TC	
ETP1 probe 1 rv	TGC AGC CTC CTT CCT TCG TA	
ETP1 probe 2 fw	CTC GCT GGC ATC AAG CGC	
ETP1 probe 2 rv	GGC ACC TTC CTC GCC CTC	
ETP1 seq pr fw	TCT CAG CCA GCC ACG ACG GC	Sequencing of *etp1* 3´

Etp1 seq pr rev	AGG GAC GCT TAT CGG TCA TT	FR
Etp1 SKO FWD (1)	GGTCTCGCAGCAGGTCTAGATATCGG ATC	
Etp1 SKO REV (1)	GGTCTCGCACAGGTGAGCTCGAATTC A	
Etp1 SKO FWD (2)	GGTCTCCTGTGACTGTCAGAAAGCGG C	
Etp1 SKO REV (2)	GGTCTCCTGATGGGGTGTACGAGGCG	Construction of pAEX-ETP1/NEO (pAdea042)
Etp1 SKO FWD (3)	GGTCTCGATCATGATTGAACAAGATG GATTGC	
Etp1 SKO REV (3)	GGTCTCGTCCTTAGAAGAACTCGTCA AGAAGG	
Etp1 SKO FWD (4)	GGTCTCCAGGAAAAGATAAATAGAC GTTAAAAAAG	
Etp1 SKO REV (4)	GGTCTCCGCTGGTGTTATGTGACTAG AGTAA	
ETP1 DKO FWD (1) final	GGTCTCGGTGCAGGTCTAGATATCGG ATC	
ETP1 DKO REV (1)	GGTCTCGTGATGGGGTGTACGAGGCG	
ETP1 DKO FWD (2)	GGTCTCCATCATGAAAAAGCCTGAAC TCAC	Construction of pAEX-ETP1/HYG (pAdea063)
ETP1 DKO REV (2) FIN	GGTCTCCTCCTTATTTCTTTGCCCTCG GA	
ETP1 DKO FWD (3) fin	GGTCTCCaGGAAAAGATAAATAGACG TTAAAAAAG	
ETP1 DKO REV (3) fin	GGTCTCCGCACAAATAATGGTTTTGA GGCT	
ETP1-ph FWD (2) ETP1 5´	GGTCTCGCTGCTGACTGTCAGAAAGC GGC	
ETP1-ph FW2 ETP1 5´ alt	GGTCTCGCCTGTGACTGTCAGAAAGC GGC	
ETP1-ph REV2 ETP1 5´	GGTCTCGATGCGTACTTGAATAGCTT TTAATAAT	Construction of pAEX-ETP1/BLE (pAdea083)
ETP1-ph FW 3 phleo	GGTCTCGGCATGGCCAAGTTGACCAG T	
ETP1-ph RV 3 phleo	GGTCTCGTCCTCAGTCCTGCTCCTCGG C	
ETP1-ph FW 4 ETP1 3´	GGTCTCGAGGAAAAGATAAATAGAC GTTAAAAAAG	
ETP1-ph RV 4 ETP1 3´	GGTCTCGctgCACAAATAATGGTTTTG	

	AGGCT		
ETP1 phleo etp1 FR REV2	GGTCTCCCTTCATCATCAGTCCTGCTC	Construction of pAEX-	
ETP1 phle etp1 FR RV2´	GGTCTCCCTTCAGTCCTGCTCCTC	ETP1/ BLE [2]	
ETP1 phle etp1 FR FD3	GGTCTCCGAAGACCCCTGCCGCCGAG C	(pAdea089)	
nGFP ETP1 enlocRV3	GGT CTC CTC CTT ACG CAT CCT TCT TCT CTA C	Construction of pAEX- ETP1/nGFP-ETP1	
nGFP ETP1 enloc FW 4	GGT CTC CAG GAA AAG ATA AAT AGA CGT TAA AAA AG	(pAdea048)	

4.1.4 Plasmids

Plasmid generation was performed by standard molecular cloning techniques as described by Sambrook *et al*., 1989 or by Golden Gate cloning as described by Terfrüchte *et al*., 2014. Oligonucleotides used for cloning purposes are listed in Tab. 4.1-8. Sequencing of all constructs was performed by Eurofins Genomics and GATC-Biotech.

Table 4.1-9: Plasmids generated and used in this study

Name	pAdea number	Description	Selection
pAEX-ETP1-cGFP	pAdea040	Contains *A. deanei* candidate ETP1 c-term. fusion with eGFP	Amp (*E. coli*) Neo (*A. deanei*)
pAEX-nGFP-ETP1	pAdea039	Contains *A. deanei* candidate ETP1 n-term. fusion with eGFP	Amp *(E. coli)* Neo (*A. deanei*)
pAdea-ETP1/NEO	pAdea042	Contains *A. deanei etp1* 5´ and 3´FRs and *neo* res gene (ETP1 knock-out)	Amp (*E. coli*) Neo (*A. deanei*)
pAEX-ETP1/nGFP-ETP1	pAdea048	Contains *A. deanei etp1* full (scaf 1854) n-term. fusion with eGFP (expr. in endogenous locus)	Amp (*E. coli*) Neo (*A. deanei*)
pAdea-ETP1/HYG	pAdea063	Contains *A. deanei etp1* 5´ and 3´FRs and *hyg* res gene (ETP1 knock-out)	Amp (*E. coli*) Hygro (*A. deanei*)

pAdea-ETP1/ BLE	pAdea083	Contains *A. deanei etp1* 5´ and 3´FRs and *ble* res gene (ETP1 knock-out)	Amp (*E. coli*) Ble (*A. deanei*)
pAdea-ETP1/ BLE [2]	pAdea089	Contains *A. deanei etp1* 5´ FR, last 700 bp of *etp1* ORF and *ble* res gene (ETP1 knock-out)	Amp (*E. coli*) Ble (*A. deanei*)

4.1.5 Strains

Table 4.1-10: Strains generated and used in this study

Strain *E. coli*	Description	Resistance	Integrated Plasmid	Progenitor
Top 10	F- mcrA Δ(mrrhsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(araleu) 7697 galE15 galK16 rpsL(StrR) endA1 λ-	*strepto*	No plasmid	K-12
Ec_pAdea001	*δ-amastin* replaced by *neo*	*amp*	pAdea001	Top 10
Ec_pAdea002	*δ -amastin* replaced by *hyg*	*amp*	pAdea002	Top 10
Ec_pAdea003	*γ-amastin* replaced by *neo*	*amp*	pAdea003	Top 10
Ec_pAdea004	*γ-amastin* replaced by *hyg*	*amp*	pAdea004	Top 10
Ec_pAdea009	*γ-amastin* replaced by *hyg*/ *γ-amastin* replaced by *neo*	*amp*	pAdea003	Top 10
Ec_pAdea0039	*δ -amastin* replaced by *neo*/GAPDH IR/*gfp*/*etp1*	*amp*	pAdea039	Top 10
Ec_pAdea0040	*δ -amastin* replaced by *neo*/GAPDH IR/*etp1*/*gfp*	*amp*	pAdea040	Top 10
Ec_pAdea0042	*etp1* replaced by *neo*	*amp*	pAdea042	Top 10
Ec_pAdea0048	*etp1* replaced by *neo*/GAPDH IR/*gfp*/*etp1*	*amp*	pAdea048	Top 10
Ec_pAdea0057	*etp1* replaced by *neo* and *etp1*/*gfp*	*amp*	pAdea057	Top 10
Ec_pAdea063	*etp1* replaced by *hyg*	*amp*	pAdea063	Top 10
Strain	**Description**	**Resistance**	**Integrated Plasmid**	**Progenitor**

A. deanei

Adea0001	δ-amastin replaced by neo	neo	pAdea001	ATCC PRA-265
Adea0002	δ -amastin replaced by hyg	hyg	pAdea002	ATCC PRA-265
Adea0003	γ-amastin replaced by neo	neo	pAdea003	ATCC PRA-265
Adea0004	γ-amastin replaced by hyg	hyg	pAdea004	ATCC PRA-265
Adea0009	γ-amastin replaced by hyg/ γ-amastin replaced by neo	neo hyg	pAdea003	Adea0004
Adea0035	δ -amastin replaced by neo/GAPDH IR/gfp/etp1	neo	pAdea039	ATCC PRA-265
Adea0036	δ -amastin replaced by neo/GAPDH IR/etp1/gfp	neo	pAdea040	ATCC PRA-265
Adea0038	etp1 replaced by neo	neo	pAdea042	ATCC PRA-265
Adea0039	etp1 replaced by neo and gfp-etp1	neo	pAdea048	ATCC PRA-265
Adea0051	etp1 replaced by neo and etp1-gfp	neo	pAdea057	ATCC PRA-265
Adea0052	etp1 replaced by hyg (δ -amastin replaced by neo/GAPDH IR/ gfp/etp1)	neo hyg	pAdea063	Adea0035
Adea0071	etp1 replaced by hyg	hyg	pAdea063	ATCC PRA-265
Adea0072	etp1 replaced by phleo (δ-amastin replaced by neo/GAPDH IR/ gfp/etp1 and etp1 replaced by hyg)	hyg neo phleo	pAdea083	Adea0052
Adea0075	etp1 replaced by ble (not full ORF) (δ-amastin replaced by neo/GAPDH IR/ gfp/etp1 and etp1 replaced by hyg	hyg neo phleo	pAdea089	Adea0052
Adea0076	etp1 replaced by ble (δ-ama replaced by neo/GAPDH IR/nV5ETP1)	neo phleo	pAdea083	Adea0041
Adea0077	etp1 replaced by ble (δ-ama replaced by neo/GAPDH IR/ETP1cV5)	neo phleo	pAdea083	Adea0040
Adea0101	etp1 replaced by neo/GAPDH IR/ gfp/etp1 (etp1 replaced by hyg)	neo hyg	pAdea048	Adea 0071

4.1.6 Bioinformatic Tools

The programs Benchling (Benchling, RRID:SCR_013955 https://benchling.com) and Serial Cloner 2.1 were used to generate and edit DNA sequences of plasmids. Furthermore, cloning strategies and oligonucleotides were designed with these programs.

The analysis of nucleic acid and protein sequences was performed with BioEdit (Hall, 1999). Sequence alignments were made using ClustalW version 2.0 (Larkin *et al.*, 2007).

For the generation of illustrations the programs Adobe Illustrator version 16.0 and Adobe Photoshop version 11.0 were used.

4.2 Methods

4.2.1 Molecular Biology Methods.

4.2.1.1 Isolation and purification of nucleic acids

DNA Extraction from *A. deanei*

For PCR analysis the gDNA of different strains of *A. deanei* was obtained by DNAzol (Thermo Scientific) treatment according to the following procedure. Homogenization of 200 µl sample from a late-log phase culture of *A. deanei* was performed by the addition of 1 ml of DNAzol reagent and by pipetting the suspension up and down several times. After pelleting the insoluble material (10000 g for 10 min at 4°C or RT) the supernatant was transferred to a fresh tube. Next followed the addition of 0.5 ml of 100% EtOH to the supernatant, the storage for 1-3 min at RT and the spin down of the DNA at 4000 g for 3 min. Supernatant was discarded and the pellet was washed two times with 75% ethanol (4000 g for 3 min). The pellet was resuspended in 20 µl ddH$_2$O. For Southern blot analysis the gDNA of different strains of *A. deanei* was obtained using DNAeasy Blood & Tissue kit (Qiagen), according to the manufactures' instructions.

Isolation of plasmid DNA from *E. coli*

Isolation of plasmid DNA from *E. coli* cells was performed using the Plasmid Mini Kit (Qiagen) for isolation of up to 20 µg high-purity plasmid or the QIAprep Spin Midiprep Kit (Qiagen) for isolation of up to 200 µg high-purity plasmid, according to the manufactures' instructions.

4.2.1.2 Determination of DNA concentration

Nucleic acid concentrations were measured photometrical with the NanoDrop 2000c spectral photometer (Thermo Fisher Scientific). An OD260 = 1 refers to a concentration of 50 μg/ml double stranded DNA at 1 cm layer thickness. Purity was determined by calculating the A260/A280 quotient with an optimum of 1.8.

4.2.1.3 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) was used to amplify DNA-fragments. Depending on size and application, PCR programs were modified. Reactions were performed in a PTC200, MJ Research or a SensoQuest Labcycler. PCR reaction was set as given below:

Table 4.2-1: PCR reaction protocol

Components	Volume to 25 μl reaction	Final concentration
Nuclease-free water	Up to 25 μl	
5X Phusion HF or GC Buffer	5 μl	1X
25 mM MgCl$_2$	0.75 μl	0.75 mM
10 mM dNTPs	0.5 μl	200 μM
10 μM Forward Primer	1.25 μl	0.5 μM
10 μM Reverse Primer	1.25 μl	0.5 μM
5 M Betaine	6.25 μl	1.25 M
Template DNA	Variable	< 250 ng
Phusion DNA Polymerase	0.2 μl	0.5 units

Table 4.2-2: Program used for reactions in thermo cycler

Step	Temperature	Time
Initial Denaturation	98°C	30 seconds
	98°C	10 seconds
30 cycles	45-72°C	10 seconds
	72°C	30 seconds per kb of product
Final Extension	72°C	5 minutes
Hold	4-10°C	

Colony PCR: For this PCR, some cell material of *E. coli* cells was suspended directly to the PCR reaction mixture. The PCR reaction set up was the same as before.

Site-directed mutagenesis PCR: In order to exchange single bases in DNA-fragments, PCR site-directed mutagenesis was performed as described by Zheng *et al.*, 2004.

4.2.1.4 Construction of plasmids

The pAdea-δ-AMA/NEO, δ-AMA/HYG, γ-AMA/NEO, and γ-AMA/HYG plasmids were generated by Gibson cloning using the pGEM-Teasy vector (Promega) as the backbone. The 5′- and 3′-FRs of the δ- and γ-*amastin* genes were amplified from *A. deanei* gDNA. The *neo* and *hyg* resistance genes were amplified from pEF1V5_HisA (Invitrogen) and pcDNA3.1/Hygro(+) (Invitrogen), respectively. The one-pot cloning reaction using the Gibson Assembly Master Mix (New England Biolabs) contained 0.15 pmol of each purified PCR product for the 5′- and 3′-FRs, the resistance marker gene (*neo* or *hyg*), and the *Nco*I/*Pst*I-restricted pGEM-Teasy vector (pGEM-Teasy manual, Promega) (done by Dr. Eva Nowack). To construct the pAEX series of expression plasmids, the glyceraldehydes-3-phosphate dehydrogenase intergenic region (GAPDH IR) was amplified from *A. deanei* gDNA. The resulting fragment was sub-cloned into pJET (Thermo Scientific) to generate the pJET-GAPDH IR and sequenced. Generation of the pAEX-EGFP expression plasmid was accomplished by amplifying the δ-amastin 5′-FR-*neo* and 3′-FR fragments from pAdea-δ-AMA/NEO, the *GAPDH* IR from pJET-GAPDH IR, and the *egfp* gene from the vector pUMA 1445 (Terfrüchte *et al.*, 2014). A total of 20 fmol of each of the fragments and the pUMA 1467 backbone (Terfrüchte *et al.*, 2014) were joined using the Golden Gate cloning system following restriction with *Bsa*I (Engler *et al.*, 2008) (done by Dr. Jorge Morales). For

the construction of pAEX-nGFP-ETP1 and pAEX-ETP1-cGFP, the *etp1* ORF was amplified from *A. deanei* gDNA, sub-cloned into pJET, and sequenced. An internal *Bsa*I site in the *etp1* gene was removed by introducing a synonymous point mutation (G to A) at position 501 of the ORF. The *Bsa*I-free *ETP1* fragment was used to generate the vectors encoding the N- or C-EGFP tagged fusion proteins. Construction of all other plasmids followed the same strategy.

4.2.1.5 Sequencing of plasmid DNA

Sequencing of all constructs was performed by Eurofins Genomics or GATC-Biotech. When sequencing was performed by Eurofins 1 μg purified plasmid DNA and 20 pmol of the respective sequencing oligonucleotide were mixed and sent to the sequencing facility. When sequencing was performed by GATC Biotech 500 ng purified plasmid DNA and 25 pmol of the respective sequencing oligonucleotide were sent to the sequencing facility. The subsequent analysis of the raw data was performed using BioEdit software.

4.2.1.6 Southern Blot analysis

This method is modified from Southern, 1975. *A. deanei* gDNA samples (1 μg) were digested into smaller fragments using appropriate restriction enzymes. Digested DNA samples were loaded on 0.8% agarose gels and DNA fragments were separated by electrophoresis at 90 V. Then the agarose gel was incubated for 20 min in 0.25 M HCl. This step resulted in partial depurination of the DNA fragments, which in turn led to strand cleavage. The length reduction improves the transfer of longer molecules. This step was followed by incubation for 20 min in denaturation solution to separate double-stranded DNA into single strands and finally for 20 min in Neutralization solution to neutralize pH (Table 4.2-4). The transfer of the separated DNA fragments from an agarose gel to a nylon membrane was then performed using capillary force. This transfer was mediated by the transfer buffer (20xSSPE) running through a Whatman filter paper salt-bridge. On top of the salt-bridge the agarose gel, nylon membrane (Nytran N Nylon Blotting Membrane, 0.45 μm; GE Healthcare Life Sciences) and Whatman filter paper were stacked. The blot was performed overnight. Then, the DNA was fixed on the membrane using 120 mJ UV-irradiation using CL-508.M (Clever Scientific). Gene-specific probes were labeled by digoxigenin-11-dUTP (DIG, Roche) during PCR as described below. The reaction was performed in a thermocycler.

Table 4.2-3: Generation of a labeled probe by DIG PCR

Components	Volume of 50 µl reaction	Final concentration
Nuclease-free water	Up to 50 µl	
5X Phusion HF Buffer	10 µl	1X
25 mM MgCl$_2$	1.5 µl	0.75 mM
10 mM dNTPs non labeled	0.5 µl	100 µM
10 mM dNTPs labeled	0.5 µl	100 µM
10 µM Forward Primer	2.5 µl	0.5 µM
10 µM Reverse Primer	2.5 µl	0.5 µM
5 M Betaine	12.5 µl	1.25 M
Template DNA		100 ng of gDNA
Phusion DNA Polymerase	0.4 µl	1 unit

The blotted membranes were preincubated for saturation of nonspecific binding sites with Southern hybridization buffer for 20 min at 65°C. The probes were denatured by boiling at 98°C for 5 min and rapidly cooling on ice. Then the probes were added in pre-warmed hybridization buffer. This probe solution was added to the membranes and hybridized at the appropriate hybridization temperature under continuous agitation overnight. The appropriate hybridization temperature was calculated by the following equation.

$$T_m = 49.82 + 0.41(\% \text{ G+C}) - (600/\,l)$$
$$T_{opt} = T_m - 20 \text{ to } 25^{o}C$$

Where: T_m = melting point of probe-target hybrid, (% G + C) = % of G and C residues in probe sequence, l = length of target hybrid in base pairs, T_{opt}= optimal temperature for hybridization of probe to target.

After hybridization the membrane was washed for each 15 min in Southern-wash I, Southern-wash II, and Southern wash III at 65°C. The following steps were performed at RT. First the membrane was washed for 5 min in DIG-wash and then blocked with DIG-2 solution for 30 min. Subsequently, the membrane was incubated with anti-DIG antibody solution (1:10,000 of Anti-Digoxigenin-AP Fab fragments (Roche) in DIG-2) and afterwards washed twice in

DIG-wash for 15 min. The membrane was then equilibrated in DIG-3 solution for 5 min followed by incubation with 3 mL CDP-Star solution (Roche) (CDP star 1:100 in DIG-3). Chemiluminescence was detected using a LAS4000 (GE Healthcare).

Table 4.2-4: Southern Blot Buffers

Name	Components
Denaturation solution	1.5 M NaCl
	0.4 M NaOH
Neutralization solution	0.5 M Tris, pH 7.0
	3.0 M NaCl
20x SSPE buffer (Transfer buffer)	3 mM NaCl
	227 mM NaH2PO4 x H20
	20 mM Na2EDTA x 2H2O, Set pH to 7.4 with NaOH
Denhardt's solution	2% (w/v) BSA fraction V
	2% (w/v) Ficoll
	2% (w/v) PVP (Polyvinyl Pyrrolidone, SIGMA PVP-360)
Southern Hybridization buffer	26% (v/v) SSPE (20x)
	5% (v/v) Denhardt's solution
	5% (v/v) SDS (10%)
Southern wash Buffer I	2x SSPE (20x)
	0.1% SDS (10%)
Southern wash Buffer II	1x SSPE (20x)
	0.1% SDS (10%)
Southern wash Buffer III	0.1x SSPE (20x)
	0.1% SDS (10%)
DIG-1	0.1 M maleic acid
	0.15 M NaCl, Set pH to 7.5 using NaOH
DIG-wash	0.3% (v/v) Tween20 in DIG1 solution
DIG-2	1% (w/v) skimmed milk powder in DIG1
DIG-3	0.1 M Tris-HCl
	0.1 M NaCl, Set pH to 9.5

4.2.2   Microbiological Methods

4.2.2.1   Generation of competent *E. coli* Top 10 cells

The competence was mediated by a treatment with $RbCl_2$ and $CaCl_2$ as described before (Cohen *et al.*, 1972) with some modifications. A single bacterial colony was grown in 5 ml LB medium overnight at 37°C. Then 1 ml of the of the overnight culture was transferred into 100 ml LB medium and incubated at 37°C with vigorous agitation, monitoring the growth of the culture to an $OD_{600}= 0.5$. Next cells were pelleted by centrifugation at 3000 g for 15 min at 4°C, resuspended in 33 ml of ice cold RF1 solution (100 mM RbCl, 50mM $MnCl_2$ x 2 H_2O, 30 mM potassium acetate, 10 mM $CaCl_2$, 15% Glycerin, pH 5.8) and incubated on ice for 30-60 min. Then cells were collected by centrifugation at 3000 g for 15 min at 4°C, resuspended in 5 ml of ice-cold RF2 solution (10 mM MOPS, 10 mM RbCl, 75 mM $CaCl_2$, 15% Glycerin, pH 6.8) and incubated on ice for 15 min. Cell suspension was dispensed in 50 µl aliquots, snap frozen in liquid nitrogen and stored at -80°C.

4.2.2.2   Transformation of *E. coli*

For transformation, 50 µl of competent cells were thawed on ice for 5 min and 50-200 ng of plasmid was added. The cells were incubated on ice for 30 min. Cells were then heat shocked at 42°C for 30 s and placed on ice for 5 min. Supplementation with 200 µL of LB medium and incubation at 37°C shaking for 30 min ensured regeneration of the cells and expression of the ampicillin resistance marker. Transformed cells were plated on ampicillin containing agar plates and incubated overnight at 37°C.

4.2.2.3   Transfection of *A. deanei*

Cells grown to mid-log phase were collected (2000 g for 10 min) and resuspended to 1 x 10^8 cells/ml in buffer for transfection (25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.6, 120 mM KCl, 0.15 mM $CaCl_2$, 10 mM K_2HPO_4, 2 mM EDTA, and 5 mM $MgCl_2$). A total of 100 µl of cells resuspended in this buffer were mixed with 10 µg of restricted cassette and electroporation was carried out using the program X-001 in the Nucleofector 2b (Lonza). Alternatively, a total of 2 x 10^6 cells were resuspended in 20 µl of solution for primary cells P3 (Lonza) and 5 µg of restricted cassette and electroporated using the program FP-158 in the Nucleofector 4D (Lonza). Following transfection, cells were

transferred into 10 ml of fresh BHI growth medium and after a 22 h recovery period at 28°C, antibiotics were added to the medium and were present throughout all the downstream steps unless specifically indicated otherwise. After approximately 4 days, the transfected cells recovered and were sub-cultured in fresh medium every 3 days. At the third passage, the flagellates were diluted to a density of 1 cell/ml and aliquots of 200 μl were separated in 96-well plates. Around 7 days later, clonal cultures recovered and were transferred to 10 ml of fresh medium.

4.2.2.4 Fluorescence microscopy

For visualization of EGFP-expressing cell lines, an aliquot of approximately 20 μl (2 x 10^5 cells) was spread on a polylysine-coated slide and incubated for 10 min at RT. The same volume of freshly prepared ice-cold 4% w/v paraformaldehyde (PFA) in phosphate-buffered saline (PBS) was added to the spread cells. After 15 min, slides were washed 3X with PBS and stained for 5 min in PBS containing 1 μg/ml 4′,6-diamidino-2-phenylindole (DAPI). Excess dye was removed by 3 washes with PBS.

4.2.2.5 Fluorescence *in situ* hybridization coupled to immunofluorescence assay

For fluorescence *in situ* hybridization coupled to immunofluorescence assay (FISH-IFA), cells were collected (2000 g for 10 min), resuspended in 4% w/v PFA solution in PBS (adjusted to 1 x 10^7 cells/ml) and incubated for 30 min at RT. After fixation cells were washed 2 times with PBS. Aliquots of 20 μl (2 x 10^5 cells) were spotted onto gelatin-coated slides, air dried and dehydrated for 3 min each in 50, 80 and 100% ethanol. FISH was performed as described before (Cottrell and Kirchman, 2000, Amann and Fuchs, 2008) using the probe Eub338 (5'-GCTGCCTCCCGTAGGAGT-3') against the 16S bacterial rRNA coupled at its 5′-end to Cy3 (Amann *et al*., 1990). The 5′ Cy3-labeled probe was diluted to 50 ng DNA/μl working solution. For the hybridization mixtures the probe working solution was diluted with hybridization buffer (900 mM NaCl, 20 mM Tris/HCl, pH 7.2, 35% Formamide, 0.01% SDS) (ratio 1:9). Then the hybridization mix was added to the dehydrated cells and slides were incubated at 46°C for 90 min. Subsequently the slides were incubated with the washing buffer (80 mM NaCl, 20 mM Tris/HCl, pH 7.2, 0.01% SDS) for 25 min at 48°C. After FISH, the slides were thoroughly washed with PBS and blocked with the same buffer containing 5% v/v horse serum (HS) for 30 min followed by an additional 30 min incubation

with anti-GFP rabbit IgG (Chromotek) at a dilution of 1:1000. After 3 washes of 5 min each with PBS + 5% v/v HS, slides were incubated with Alexa Fluor 488 goat anti-rabbit IgG (ThermoFisher) at a dilution of 1:250 for 30 min, washed 3 times as before and stained with DAPI as described above. Images for localization studies were acquired using a laser-based epifluorescence microscope (Zeiss Axio Observer.Z1) and image manipulation and measurements were performed using the Metamorph software package (version 7, Molecular Devices).

4.2.2.6 Preparation of *A. deanei* cells for transmission electron microscopy

5 ml of a fully-grown culture (1-3x10^8 cells/ml) were harvested at 2000 g for 10 min and resuspended in fixation solution containing 0.1% OsO_4, 2% PFA, 0.5% glutaraldehyde, dissolved in BHI medium. After 45 min incubation on ice, cells were washed three times with PBS and resuspended in 10 µl PBS. 6% low melting point agarose was boiled in dH_2O and cooled to 37°C. A small stopper of the 6% agarose solution was placed in a 1.5 ml reaction tube and cooled on ice until it completely solidified. 0.5 ml of 6% agarose solution was added on top of the stopper and quickly 5 µl of cells in PBS were injected into the middle of the liquid agarose. While retracting the pipette, the initial insertion channel was destroyed by gentle circular movement of the pipette tip. This step was necessary to seal the cells in the agarose. After the agarose became solid the encapsulated cells were cut out of the reaction tube and as much as possible of the surrounding agarose was removed, without piercing into cells.

This pellet was dehydrated in rising concentrations of ethanol (15%, 30%, 45%, on ice and 60%, 75%, 90%, 100% at -20°C) for 30 min per dehydration step. The last dehydration step was over night in 100% ethanol at -20°C. The following morning the pellet was incubated for 1h at RT in fresh 100% ethanol. Ethanol was then exchanged with rising concentrations of LR white resin in ethanol (20%, 33%, 50%, 67%, 80%, 100%, medium grade acrylic resin, London resin company). Each step was performed on a roller for 1 h. The final infiltration step with 100% LR white resin was done overnight. Gelatin capsules were warmed up to 60°C overnight to remove residual condensation. The following day the pellet was transferred in a gelatin capsule and positioned at the very bottom and after the capsule was filled up with fresh 100% LR white resin it was sealed with a gelatin cap. To polymerize the infiltrated pellet/LR white resin, gelatin capsules were incubated for three days at 50°C. Pellet was cut with a diamond knife into 50 nm thin sections and sections were fixed on nickel grids as they do not corrode during immunogold staining.

4.2.2.7 Purification of the endosymbiont

Endosymbionts were isolated from *A. deanei* cells essentially as described before (Novak *et al.*, 1988) with some minor modifications. All steps were performed at 4°C and centrifugation steps for the gradients were carried out in a Beckman JS-21 centrifuge using the swinging bucket rotor JS-13.1. A total of 600 ml of *A. deanei* culture grown to late-log phase was collected at 5000 *g* for 10 min, resuspended in 30 ml buffer A (10 mM Hepes and 0.5 mM EDTA, pH 7.4) containing 150 mM of sucrose and 1x cOmplete™ Protease Inhibitor Cocktail (Roche). After sonication, the lysate was centrifuged at 7600 *g* for 15 min and the pellet resuspended in 16 ml of buffer A containing 250 mM of sucrose (buffer B). A total of 4 ml of the suspension was loaded on top of a discontinuous gradient of 5 ml 0.8 M sucrose and 10 ml 0.4 M sucrose in buffer A in 30 ml glass tubes (Corex) and centrifuged at 760 *g* for 30 min. The lower whitish band obtained in the 0.4 M sucrose layer of the gradient was collected and centrifuged at 7600 *g* for 15 min. The resulting pellet was resupended in 4 ml of buffer B, and loaded on top of a percoll gradient of 4 ml each 80%, 70%, 60%, 50%, and 40% percoll (Percoll plus, GE Healthcare), adjusted with the buffer A to 250 mM final concentration of sucrose; the gradient was made in 30 ml glass tubes (Corex). After centrifugation at 10050 *g* for 1 h, the lower two bands located between 70-60% percoll were collected, diluted 3 times in buffer B and centrifuged at 7600 g for 15 min. The same steps of dilution in buffer B and centrifugation at 7600 g for 15 min were repeated 3 more times in order to remove all residual percoll. Then the pellet of the pure endosymbiont was treated as described below.

4.2.2.8 Lysozyme digestion of endosymbionts and formation of spheroplasts

In order to disrupt *A. deanei* endosymbionts, freshly purified endosymbionts were treated as described before (Hill and Sillence, 1997) with some minor modifications. Briefly, the bacterial cell wall of freshly purified endosymbionts (0.2 g of dry pellet) was digested by resuspending the endosymbiont pellet in 1.5 ml of buffer S1 (100 mM Tris–HCl, pH 7.8, 0.75 M sucrose, 1.5 mM EDTA and 180 µg lysozyme (Human, Sigma Aldrich). The suspension was maintained at 16°C and stirred for 20 min. Then, cells were subjected to an osmotic shock by supplementing the stirred suspension with 1.5 ml of buffer S2 (100 mM Tris–HCl, pH 7.8, 1.5 mM EDTA, 30 units DNase I, 0.5 mg RNase A, and 1x cOmplete™ Protease Inhibitor Cocktail (Roche)) and stirred at 6°C for 20 min. The suspension was then centrifuged at 17200 g for 10 min at 4°C. The supernatant was discarded and the pellet was

resuspended by vigorous aspiration in 1.5 ml of buffer S3 (10 mM Tris–HCl, pH 7.8, 1.5 mM EDTA, 30 units DNase I, 0.5 mg RNase A, 1x cOmplete™ Protease Inhibitor Cocktail and 1 mM PMSF). This step resulted in the formation of spheroplasts. The suspension was stirred vigorously for 15 min at 4°C. After centrifugation at 1000 g for 10 min at 4°C, the supernatant was collected. The same step of pellet resuspension in 1.5 ml buffer S3 and centrifugation at 1000 g for 10 min at 4°C was repeated one more time. The supernatants from these two steps were collected, divided into 6 aliquots of 500 µl, and stored at -20°C for further experiments.

4.2.3   Protein Biochemical Methods

4.2.3.1   Sample preparation for Clear-native electrophoresis

Proteins of *A. deanei* endosymbionts, purified from ETP1-cV5 cell line, were solubilized using 1% DDM (n-Dodecyl-beta-Maltoside, Glycon Biochemicals GmbH) and Clear native PAGE was used for separation of DDM solubilized proteins as described by Wittig *et al.*, 2007 with some modifications. All steps were performed at 4°C in the presence of 1x cOmplete™ Protease Inhibitor Cocktail (Roche), 1 mM PMSF, 10 µM E64, 100 µM TLCK and 1 µM Pepstatin A. One 500 µl aliquot of endosymbionts (∼200 µg of protein), that was treated as described in paragraph 4.2.2.7 and stored at -20°C after the addition of 1% DDM, was thawed, supplemented with 0.05% Sodium deoxycholate (DOC) and incubated on ice for 30 min (vortex every 10 min). After centrifugation at 20000 g for 10 min at 4°C the supernatant was diluted with buffer containing 50 mM NaCl, 50 mM Imidazole/ HCl, pH 7.2 mM aminohexanoic acid, 1 mM EDTA, 0.05% DOC and 0.1% DDM (ratio 1:10), and concentrated to 200 µl using Amicon® centrifugal filter (Ultra-30, PLTK Ultracel-PL membrane) (Merck Millipore). The same step of dilution (ratio 1:10) and concentration of the supernatant to 200 µl was repeated twice.

4.2.3.2   Clear-native (CN) electrophoresis

The procedure was done essentially as described by Wittig *et al.*, 2007 with some modifications. 200 µl of the concentrated supernatant, prepared as described in the previous paragraph, were supplemented with 25 µl of 50% glycerol and 0.1% Ponceau S (from 2% stock in 30% acetic acid, Sigma Diagnostics). Equal sample volumes (30 µl) were loaded into the wells of 8x8 cm polyacrylamide gels for clear native PAGE (NativePAGE 4-16% Bis-Tris

gels, Thermo Fisher Scientific). Anode and Cathode buffers were prepared as described in table 4.2-5. Gels were run at 4°C and the initial voltage was set to 100 V with the current limited to 15 mA. When the samples had entered the sample gel, the voltage was raised to 300 V.

Table 4.2-5: Anode and Cathode buffer components

Anode buffer components	Final concentration in Anode buffer
Imidazole/HCl pH 7.0	25 mM
Cathode buffer components	**Final concentration in Cathode buffer**
Tricine	50 mM
Imidazole	7.5 mM
DOC	0.05%
DDM	0.02%

4.2.3.3   SDS- PAGE

To separate proteins according to their molecular weight, standard SDS-PAGE (denaturing SDS-polyacrylamide-gel-electrophoresis) was used (Laemmli, 1970). Samples were supplemented with 5x concentrated SDS sample buffer (0.2 M Tris-HCl pH 6.8, 10% (v/v) SDS, 10 mM β-Mercaptoethanol, 20% (v/v) glycerol and 0.05% (w/v) Bromophenol blue) to a final concentration of 1x and boiled for 5 min at 98°C. After centrifugation (1 min, 1000 g), the samples were loaded to Bolt 4-12% Bis-Tris Plus gels (Thermo Fisher Scientific) and SDS-PAGE was performed using 1x MES running buffer (50 mM MES (2-[N-morpholino]ethanesulfonic acid), 50 mM Tris, 1 mM EDTA, 0.1% (w/v) SDS, pH 7.4) at 180 V until an appropriate separation was achieved.

4.2.3.4   Western blot analysis

After SDS-PAGE, proteins were transferred onto methanol-activated PVDF membrane (Amersham Hybond-P, GE Healthcare) according to the following procedure. Three Whatman filter papers were soaked in transfer buffer (12 mM Tris, 96 mM glycine, 20% methanol). On top of these, the activated PVDF membrane was placed, followed by the gel from SDS-PAGE and 3 Whatman filter paper soaked in transfer buffer. Transfer was

performed at 60 mA per membrane for 1 h, using a semi-dry Western blot chamber (846-015-200, Biometra). Then the SNAP i.d.® 2.0 Protein Detection System (Merck Millipore) was used for blocking, washings and antibody incubations following the protocol that was provided by the company. After blocking with 0.1% skimmed milk powder in PBST the membranes were incubated with the GFP antibody [3H9] (1:1000) (Chromotek) or with V5 antibody (1:1000) (Sigma-Aldrich), depending on the application, for 10 min at RT. The membrane was then washed with PBST, and incubated with HRP-conjugated secondary antibody (Rabbit anti-Rat IgG Secondary Antibody, HRP by Thermo Fisher Scientific) or (anti-Rabbit IgG Secondary Antibody, HRP by Thermo Fisher Scientific) in blocking solution (1:5000) for 10 min at RT, followed by washing with PBST. After incubating the membrane for 1 min in HRP-Substrate Luminata^{TM} Classico (Millipore), chemiluminescence was detected using the ImageQuant LAS-4000 mini (GE Healthcare).

Table 4.2-6: PBST buffer components

Components	Final concentration in PBST buffer
NaCl	137 mM
KCl	2.7 mM
Na$_2$HPO$_4$	4.3 mM
KH$_2$PO$_4$	1.47 mM
Tween-20	0.1% (v/v)
in H$_2$O$_{bid}$, pH 7.4	

4.2.3.5 Semidry electroblotting of native gels

Native gels were incubated for 20 min in buffer (300 mM Tris, 100 mM acetic acid, 1% SDS, pH 8.6) turning the gel several times to denature proteins. Then, the Western blot analysis was performed as described in paragraph 4.2.3.4.

4.2.3.6 Coomassie staining of SDS-polyacrylamide gels

Coomassie brilliant blue staining was used to stain polyacrylamide gels. Gels were incubated in Coomassie staining solution for 1 h and then washed with H$_2$O. After incubation in destaining solution, the gel was washed with H$_2$O.

Table 4.2-7: Buffers for coomassie staining

Name	Components
Coomassie staining solution	0.05% (w/v) Coomassie Brilliant Blue R250 15% (v/v) acetic acid 15% (v/v) methanol in H_2O_{bid}
Destaining solution	15% (v/v) acetic acid 15% (v/v) methanol in H_2O_{bid}

4.2.3.7 Silver staining

The procedure was done essentially as described by Blum *et al.*, 1987.

Table 4.2-8: Procedure of silver staining of proteins

Steps	Components of solutions	Incubation
Fix	10% acetic acid, 50% Methanol	1 h
Wash	30% Ethanol	30 min
Pretreat	0.2 g/L $Na_2S_2O_3$	1 min
Rinse	H_2O	
Impregnate	2 g/L $AgNO_3$, 0.075 mL/L 37% HCOH	20 min
Rinse	H_2O	
Develop	60 g/L Na_2CO_3 ; 0.5 mL/L 37% HCOH; 4 mg/L $Na_2S_2O_3$	until the desirable result was achieved
Rinse	H_2O	
Stop	75% Methanol 18% acetic acid	10 min
Preservation	20% Glycerol	

4.2.3.8   Co-Immunoprecipitation

Endosymbiont fraction (~500 µg protein) that was treated as described in paragraph 4.2.2.6 was supplemented with RIPA buffer (1% Triton X-100, 0.1% SDS and 1% deoxycholate) and stored at -20°C. After 16 h it was thawed and incubated on ice for 30 min. All steps were performed at 4°C in the presence of 1x cOmplete™ Protease Inhibitor Cocktail (Roche), 1 mM PMSF, 10 µM E64, 100 µM TLCK and 1 µM Pepstatin. After centrifugation at 150000 g for 1 h at 4°C the supernatant was transferred to a precooled tube and supplemented with dilution buffer (10 mM Tris/HCL pH 7.5, 150 mM NaCl, 0.5 mM EDTA) up to 1 ml. The GFP-Trap®_MA beads (Chromotek) were vortexed and 25 µl beads slurry were transferred into 500 µl ice-cold dilution buffer. The beads were magnetically separated until supernatant was clear. Supernatant was discarded and the wash was repeated twice. Then, the cell lysate was added to the equilibrated GFP-Trap®_MA beads and tumble end-over-end for 1 hour at 4°C. After incubation the beads were magnetically separated until the supernatant appeared clear. GFP-Trap®_MA beads with bound target proteins were washed six times in 500 µl dilution buffer and magnetically separated until supernatant was clear. Next, followed resuspension of the beads in 50 µl of 2x SDS-sample buffer, boiling of the beads for 10 min at 95°C to dissociate immunocomplexes from GFP-Trap®_MA beads and analysis of the sample by SDS-PAGE and mass spectrometry (MS).

4.2.3.9 Mass spectrometric analysis of protein samples

The mass spectrometric analysis was conducted by Prof. Dr. Kai Stühler and Dr. Gereon Poschmann in the Molecular Proteomics Laboratory (MPL), at Heinrich Heine University, Düsseldorf. Samples for MS analysis were prepared as described by Poschmann *et al.*, 2014. Peptides produced by trypsin digestion were separated on an Ultimate 3000 Rapid Separation Liquid Chromatography system (RSLC, Thermo Scientific, Dreiech, Germany) and analyzed on a Q Exactive quadrupole-orbitrap mass spectometer (Thermo Scientific, Bremen, Germany) coupled online via a nano-electrospray source. Data-dependent tandem mass spectra were acquired in positive mode and analyzed within the MaxQuant software environment (version 1.5.3.8, MPI for Biochemistry, Planegg, Germany) using standard parameters; label-free quantification was enabled and the UniProtKB proteome datasets for *A. deanei* (UP000015341) and its endosymbiont *Ca.* Kinetoplastibacterium crithidii

(UP000010479), which consist of 14,609 and 739 entries, respectively, were considered for protein identification.

4.2.3.10 Treatment of samples with different concentrations of reducing agents

Cells grown to mid-log phase (200 µl) were spun down (at 5000 g for 5 min) and after the removal of the medium were incubated at 95°C for 10 or 60 min in buffer containing 75 mM Tris/HCL pH 7.8, 15% glycerol, 6% SDS and bromophenol blue in the absence or in the presence of 100 or 600 mM of dithiothreitol/ 5 or 10% β-mercaptoethanol and subjected to SDS-PAGE and Western blot analysis as described in paragraph 4.2.3.4.

4.2.3.11 Optimization of conditions for solubilization of endosymbiont fractions

Endosymbiont fraction (∼120 µg protein) that was treated as described in paragraph 4.2.2.6 was supplemented and incubated with 0.5 or 1% Tergitol-type NP-40/ 2% n-Dodecyl β-D-maltoside/ 1% Triton X-100/ 1% Trition X-100, 0,1% SDS and 1% deoxycholate (RIPA) at -20°C for 16 h in the presence of 1x cOmplete™ Protease Inhibitor Cocktail (Roche), 1 mM PMSF, 10 µM E64, 100 µM TLCK and 1 µM Pepstatin. After centrifugation of the endosymbiont lysate the supernatant was withdrawn and the pellet was resuspended in the same volume as the supernatant. Same volumes of supernatants and pellets were loaded on 4-12% polyacrylamide gel, resolved by SDS-PAGE and subsequently subjected to Western blot analysis as described in paragraph 4.2.3.4.

5   References

Agrawal, N. *et al.* (2003) 'RNA interference: biology, mechanism, and applications.', *Microbiology and molecular biology reviews : MMBR*. American Society for Microbiology (ASM), 67(4), pp. 657–85. doi: 10.1128/MMBR.67.4.657-685.2003.

Altschul, S. F. *et al.* (1997) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.', *Nucleic acids research*, 25(17), pp. 3389–402. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9254694 (Accessed: 24 May 2018).

Alves, J. M. *et al.* (2013) 'Endosymbiosis in trypanosomatids: the genomic cooperation between bacterium and host in the synthesis of essential amino acids is heavily influenced by multiple horizontal gene transfers', *BMC Evolutionary Biology*. BioMed Central, 13(1), p. 190. doi: 10.1186/1471-2148-13-190.

Alves, J. M. P. *et al.* (2013) 'Genome Evolution and Phylogenomic Analysis of Candidatus Kinetoplastibacterium, the Betaproteobacterial Endosymbionts of Strigomonas and Angomonas', *Genome Biology and Evolution*, 5(2), pp. 338–350. doi: 10.1093/gbe/evt012.

Amann, R. *et al.* (1996) 'In situ visualization of high genetic diversity in a natural microbial community.', *Journal of bacteriology*. American Society for Microbiology, 178(12), pp. 3496–500. doi: 10.1128/JB.178.12.3496-3500.1996.

Amann, R. and Fuchs, B. M. (2008) 'Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques', *Nature Reviews Microbiology*, 6(5), pp. 339–348. doi: 10.1038/nrmicro1888.

Amann, R. I., Krumholz, L. and Stahl, D. A. (1990) 'Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology.', *Journal of bacteriology*, 172(2), pp. 762–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1688842 (Accessed: 28 September 2017).

Andrade, I. d. S. *et al.* (2011) 'Characterization of a porin channel in the endosymbiont of the trypanosomatid protozoan Crithidia deanei', *Microbiology*. Microbiology Society, 157(10), pp. 2818–2830. doi: 10.1099/mic.0.049247-0.

Anselme, C. *et al.* (2008) 'Identification of the Weevil immune genes and their expression in the bacteriome tissue', *BMC Biology*. BioMed Central, 6(1), p. 43. doi: 10.1186/1741-7007-6-43.

Araújo, P. R. *et al.* (2011) 'Development of a dual reporter system to identify regulatory cis-acting elements in untranslated regions of Trypanosoma cruzi mRNAs.', *Parasitology international*, 60(2), pp. 161–9. doi: 10.1016/j.parint.2011.01.006.

Araújo, P. R. and Teixeira, S. M. (2011) 'Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in Trypanosoma cruzi: a review', *Memórias do Instituto Oswaldo Cruz*. Fundação Oswaldo Cruz, 106(3), pp. 257–266. doi: 10.1590/S0074-02762011000300002.

de Azevedo-Martins, A. C. *et al.* (2007) 'Phosphatidylcholine synthesis in *Crithidia deanei* : the influence of the endosymbiont', *FEMS Microbiology Letters*, 275(2), pp. 229–236. doi: 10.1111/j.1574-6968.2007.00892.x.

Bagos, P. G. *et al.* (2004) 'PRED-TMBB: a web server for predicting the topology of -barrel outer membrane proteins', *Nucleic Acids Research*, 32(Web Server), pp. W400–W404. doi: 10.1093/nar/gkh417.

Barrett, M. P. *et al.* (2003) 'The trypanosomiases', *The Lancet*. Elsevier, 362(9394), pp. 1469–1480. doi: 10.1016/S0140-6736(03)14694-6.

Bary, A. (1879) *Die Erscheinung der Symbiose : Vortrag*. Strassburg: Verlag von Karl J. Trübner. Available at: http://www.worldcat.org/title/erscheinung-der-symbiose-vortrag/oclc/65143305 (Accessed: 13 February 2018).

Bhattacharya, D. *et al.* (2012) 'Single cell genome analysis supports a link between

phagotrophy and primary plastid endosymbiosis.', *Scientific reports*. Nature Publishing Group, 2, p. 356. doi: 10.1038/srep00356.

Blum, H., Beier, H. and Gross, H. J. (1987) 'Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels', *Electrophoresis*, 8(2), pp. 93–99. doi: 10.1002/elps.1150080203.

Boitz, J. M. and Ullman, B. (2010) 'Amplification of Adenine Phosphoribosyltransferase Suppresses the Conditionally Lethal Growth and Virulence Phenotype of *Leishmania donovani* Mutants Lacking Both Hypoxanthine-guanine and Xanthine Phosphoribosyltransferases', *Journal of Biological Chemistry*, 285(24), pp. 18555–18564. doi: 10.1074/jbc.M110.125393.

Boratyn, G. M. *et al.* (2012) 'Domain enhanced lookup time accelerated BLAST', *Biology Direct*, 7(1), p. 12. doi: 10.1186/1745-6150-7-12.

Borghesan, T. C. *et al.* (2018) 'Genetic Diversity and Phylogenetic Relationships of Coevolving Symbiont-Harboring Insect Trypanosomatids, and Their Neotropical Dispersal by Invader African Blowflies (Calliphoridae).', *Frontiers in microbiology*. Frontiers Media SA, 9, p. 131. doi: 10.3389/fmicb.2018.00131.

Boucher, N. *et al.* (2002) 'A common mechanism of stage-regulated gene expression in Leishmania mediated by a conserved 3'-untranslated region element.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 277(22), pp. 19511–20. doi: 10.1074/jbc.M200500200.

Burghardt, T. *et al.* (2007) 'The dominating outer membrane protein of the hyperthermophilic Archaeum Ignicoccus hospitalis: a novel pore-forming complex', *Molecular Microbiology*, 63(1), pp. 166–176. doi: 10.1111/j.1365-2958.2006.05509.x.

Burle-Caldas, G. de A. *et al.* (2015) 'Expanding the tool box for genetic manipulation of Trypanosoma cruzi', *Molecular and Biochemical Parasitology*. Elsevier, 203(1–2), pp. 25–33. doi: 10.1016/J.MOLBIOPARA.2015.10.004.

Cardoso, M. S. *et al.* (2013) 'Identification and Functional Analysis of Trypanosoma cruzi Genes That Encode Proteins of the Glycosylphosphatidylinositol Biosynthetic Pathway', *PLoS Neglected Tropical Diseases*. Edited by P. W. Denny, 7(8), p. e2369. doi: 10.1371/journal.pntd.0002369.

Carvalho, A. L. and Deane, M. P. (1974) 'Trypanosomatidae isolated from Zelus leucogrammus (Perty, 1834) (Hemiptera, Reduviidae), with a discussion on flagellates of insectivorous bugs.', *The Journal of protozoology*, 21(1), pp. 5–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4594242 (Accessed: 11 March 2018).

Castanys-Muñoz, E. *et al.* (2012) 'Leishmania mexicana metacaspase is a negative regulator of amastigote proliferation in mammalian cells', *Cell Death & Disease*. Nature Publishing Group, 3(9), pp. e385–e385. doi: 10.1038/cddis.2012.113.

Catta-Preta, C. M. C. *et al.* (2015) 'Endosymbiosis in trypanosomatid protozoa: the bacterium division is controlled during the host cell cycle.', *Frontiers in microbiology*. Frontiers Media SA, 6, p. 520. doi: 10.3389/fmicb.2015.00520.

Catta-Preta, C. M. C. *et al.* (2016) 'Reduction of Tubulin Expression in *Angomonas deanei* by RNAi Modifies the Ultrastructure of the Trypanosomatid Protozoan and Impairs Division of Its Endosymbiotic Bacterium', *Journal of Eukaryotic Microbiology*. Wiley/Blackwell (10.1111), 63(6), pp. 794–803. doi: 10.1111/jeu.12326.

CAVALIER-SMITH, T. and LEE, J. J. (1985) 'Protozoa as Hosts for Endosymbioses and the Conversion of Symbionts into Organelles [1],[2]', *The Journal of Protozoology*. Wiley/Blackwell (10.1111), 32(3), pp. 376–379. doi: 10.1111/j.1550-7408.1985.tb04031.x.

Charles, H. *et al.* (1997) 'Genome Size Characterization of the Principal Endocellular Symbiotic Bacteria of the Weevil Sitophilus oryzae, using Pulsed Field Gel Electrophoresis', *Insect Biochem. Molec. Biol*, 27(4), pp. 345–350. Available at: https://ac.els-cdn.com/S0965174897000052/1-s2.0-S0965174897000052-main.pdf?_tid=320b79bd-9cad-40ec-840f-8860ece0483b&acdnat=1520443135_97cbf75be29b317b80117d989dfffd1a (Accessed: 7 March 2018).

Chauhan, J. S., Rao, A. and Raghava, G. P. S. (2013) 'In silico Platform for Prediction of N-, O- and C-Glycosites in Eukaryotic Protein Sequences', *PLoS ONE*. Edited by S. C. E Tosatto. Public Library of Science, 8(6), p. e67008. doi: 10.1371/journal.pone.0067008.

Coba de la Peña, T. *et al.* (2018) 'The Symbiosome: Legume and Rhizobia Co-evolution toward a Nitrogen-Fixing Organelle?', *Frontiers in Plant Science*. Frontiers, 8, p. 2229. doi: 10.3389/fpls.2017.02229.

Cohen, S. N., Chang, A. C. Y. and Hsu, L. (1972) 'Nonchromosomal Antibiotic Resistance in Bacteria: Genetic Transformation of Escherichia coli by R-Factor DNA* (CaCI2/extrachromosomal DNA/plasmid)', 69(8), pp. 2110–2114. Available at: http://thewatchers.us/EPA/10/1972-Cohen-Antibiotic-resistance.pdf (Accessed: 28 September 2017).

Colombini, M. (2004) 'VDAC: The channel at the interface between mitochondria and the cytosol', *Molecular and Cellular Biochemistry*. Kluwer Academic Publishers, 256(1/2), pp. 107–115. doi: 10.1023/B:MCBI.0000009862.17396.8d.

Cottrell, M. T. and Kirchman, D. L. (2000) 'Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence in situ hybridization.', *Applied and environmental microbiology*, 66(12), pp. 5116–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11097877 (Accessed: 28 September 2017).

Cristina Motta, M. M. (2010) 'Endosymbiosis in Trypanosomatids as a Model to Study Cell Evolution', *The Open Parasitology Journal*, 4, pp. 139–147. Available at: https://pdfs.semanticscholar.org/4b8f/b5dabbf12d5f720c8cad5be0770826dca893.pdf (Accessed: 11 March 2018).

Cusson, M. (2008) 'The Molecular Biology Toolbox and Its Use in Basic and Applied Insect Science', *BioScience*. Oxford University Press, 58(8), pp. 691–700. doi: 10.1641/B580806.

d'Avila-Levy, C. M. *et al.* (2008) 'Crithidia deanei: Influence of parasite gp63 homologue on the interaction of endosymbiont-harboring and aposymbiotic strains with Aedes aegypti

midgut', *Experimental Parasitology*, 118(3), pp. 345–353. doi: 10.1016/j.exppara.2007.09.007.

Dale, C. *et al.* (2002) 'Type III secretion systems and the evolution of mutualistic endosymbiosis.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 99(19), pp. 12397–402. doi: 10.1073/pnas.182213299.

Delaye, L., Valadez-Cano, C. and Pérez-Zamorano, B. (2016) 'How Really Ancient Is Paulinella Chromatophora?', *PLoS Currents*. Public Library of Science. doi: 10.1371/currents.tol.e68a099364bb1a1e129a17b4e06b0c6b.

Diekert, K. *et al.* (1999) 'An internal targeting signal directing proteins into the mitochondrial intermembrane space.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 96(21), pp. 11752–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10518522 (Accessed: 20 May 2018).

Dolezal, P. *et al.* (2006) 'Evolution of the Molecular Machines for Protein Import into Mitochondria', *Science*, 313(5785), pp. 314–318. doi: 10.1126/science.1127895.

Engler, C., Kandzia, R. and Marillonnet, S. (2008) 'A One Pot, One Step, Precision Cloning Method with High Throughput Capability', *PLoS ONE*. Edited by H. A. El-Shemy. Public Library of Science, 3(11), p. e3647. doi: 10.1371/journal.pone.0003647.

Fàbrega, A. *et al.* (2009) 'Mechanism of action of and resistance to quinolones.', *Microbial biotechnology*. Wiley-Blackwell, 2(1), pp. 40–61. doi: 10.1111/j.1751-7915.2008.00063.x.

Ferrière, R., Gauduchon, M. and Bronstein, J. L. (2007) 'Evolution and persistence of obligate mutualists and exploiters: competition for partners and evolutionary immunization', *Ecology Letters*, 10(2), pp. 115–126. doi: 10.1111/j.1461-0248.2006.01008.x.

Gadelha, C. *et al.* (2005) 'Cryptic paraflagellar rod in endosymbiont-containing kinetoplastid protozoa.', *Eukaryotic cell*. American Society for Microbiology (ASM), 4(3), pp. 516–25.

doi: 10.1128/EC.4.3.516-525.2005.

Galdiero, S. *et al.* (2012) 'Microbe-host interactions: structure and role of Gram-negative bacterial porins.', *Current protein & peptide science*. Bentham Science Publishers, 13(8), pp. 843–54. doi: 10.2174/138920312804871120.

Gil, R., Latorre, A. and Moya, A. (2004) 'Bacterial endosymbionts of insects: insights from comparative genomics', *Environmental Microbiology*, 6(11), pp. 1109–1122. doi: 10.1111/j.1462-2920.2004.00691.x.

Griffiths, G. W. and Beck, S. D. (1975) 'Uhrastructure of Pea Aphid Mycetocytes: Evidence for Symbiote Secretion *', *Cell Tiss. Res*, 159, pp. 351–367. Available at: https://link.springer.com/content/pdf/10.1007/BF00221782.pdf (Accessed: 20 June 2018).

Guerin, P. J. *et al.* (2002) 'Visceral leishmaniasis: current status of control, diagnosis, and treatment, and a proposed research and development agenda.', *The Lancet. Infectious diseases*, 2(8), pp. 494–501. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12150849 (Accessed: 17 May 2018).

Gupta, K. *et al.* (2018) 'Identifying key membrane protein lipid interactions using mass spectrometry', *Nature Protocols*. Nature Publishing Group, 13(5), pp. 1106–1120. doi: 10.1038/nprot.2018.014.

Haanstra, J. R. *et al.* (2016) 'Biogenesis, maintenance and dynamics of glycosomes in trypanosomatid parasites', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. Elsevier, 1863(5), pp. 1038–1048. doi: 10.1016/J.BBAMCR.2015.09.015.

Haile, S. and Papadopoulou, B. (2007) 'Developmental regulation of gene expression in trypanosomatid parasitic protozoa', *Current Opinion in Microbiology*. Elsevier Current Trends, 10(6), pp. 569–577. doi: 10.1016/J.MIB.2007.10.001.

Hashimoto, M. *et al.* (2015) 'Antisense Oligonucleotides Targeting Parasite Inositol 1,4,5-Trisphosphate Receptor Inhibits Mammalian Host Cell Invasion by Trypanosoma cruzi',

Scientific Reports. Nature Publishing Group, 4(1), p. 4231. doi: 10.1038/srep04231.

Heddi, A. *et al.* (1998) 'Molecular characterization of the principal symbiotic bacteria of the weevil Sitophilus oryzae: a peculiar G + C content of an endocytobiotic DNA.', *Journal of molecular evolution*, 47(1), pp. 52–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9664696 (Accessed: 1 June 2018).

Heddi, A. (2003) 'Endosymbiosis in the weevil of the genus Sitophilus: Genetic, physiological, and molecular interactions among associated genomes.', *Insect Symbiosis. Bourtzis, K. and Miller T., eds. CRC Press, Boca Raton-London-New York-Washington DC*, pp. 67–82. Available at: http://www.programamoscamed.mx/EIS/biblioteca/libros/libros/Bourtzis and Miller 2003_2 .pdf.

Hill, R. A. and Sillence, M. N. (1997) 'Improved Membrane Isolation in the Purification of β2-Adrenoceptors from TransgenicEscherichia coli', *Protein Expression and Purification*. Academic Press, 10(1), pp. 162–167. doi: 10.1006/PREP.1997.0732.

Hooper, D. C. *et al.* (1987) 'Mechanisms of action of and resistance to ciprofloxacin.', *The American journal of medicine*, 82(4A), pp. 12–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3034057 (Accessed: 18 June 2018).

Horstman, A. *et al.* (2014) 'A cautionary note on the use of split-YFP/BiFC in plant protein-protein interaction studies.', *International journal of molecular sciences*. Multidisciplinary Digital Publishing Institute (MDPI), 15(6), pp. 9628–43. doi: 10.3390/ijms15069628.

Jakočiūnas, T. *et al.* (2015) 'Multiplex metabolic pathway engineering using CRISPR/Cas9 in Saccharomyces cerevisiae', *Metabolic Engineering*, 28, pp. 213–222. doi: 10.1016/j.ymben.2015.01.008.

Kelly, J. M. *et al.* (no date) 'A shuttle vector which facilitates the expression of transfected genes in Trypanosoma cruzi and Leishmania', *Nucleic Acids Research*, 20(15), pp. 3963–3969. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC334073/pdf/nar00226-

0155.pdf (Accessed: 24 May 2018).

Kikuchi, Y. (2009) 'Endosymbiotic bacteria in insects: their diversity and culturability.', *Microbes and environments*, 24(3), pp. 195–204. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21566374 (Accessed: 18 June 2018).

Kodama, Y. and Hu, C.-D. (2012) 'Bimolecular fluorescence complementation (BiFC): A 5-year update and future perspectives', *BioTechniques*, 53(5), pp. 285–98. doi: 10.2144/000113943.

Kostopoulou, O. N. *et al.* (2011) 'Insights into the Chloramphenicol Inhibition Effect on Peptidyl Transferase Activity, Using Two New Analogs of the Drug', *The Open Enzyme Inhibition Journal*, 4, pp. 1–10. Available at: https://benthamopen.com/contents/pdf/TOEIJ/TOEIJ-4-1.pdf (Accessed: 18 May 2018).

Kostygov, A. Y. *et al.* (2016) 'Novel Trypanosomatid-Bacterium Association: Evolution of Endosymbiosis in Action', *mBio*, 7(2), pp. e01985-15. doi: 10.1128/mBio.01985-15.

Kozlowski, L. P. (2016) 'IPC - Isoelectric Point Calculator.', *Biology direct*. BioMed Central, 11(1), p. 55. doi: 10.1186/s13062-016-0159-9.

LAEMMLI, U. K. (1970) 'Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4', *Nature*. doi: 10.1038/227680a0.

Laganowsky, A. *et al.* (2013) 'Mass spectrometry of intact membrane protein complexes', *Nature Protocols*. Nature Publishing Group, 8(4), pp. 639–651. doi: 10.1038/nprot.2013.024.

Lan-XinShi, S. M. T. (2013) 'The chloroplast protein import system: From algae to trees', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. Elsevier, 1833(2), pp. 314–331. doi: 10.1016/J.BBAMCR.2012.10.002.

Lander, N. *et al.* (2015) 'CRISPR/Cas9-Induced Disruption of Paraflagellar Rod Protein 1 and 2 Genes in Trypanosoma cruzi Reveals Their Role in Flagellar Attachment.', *mBio*. American

Society for Microbiology (ASM), 6(4), p. e01012. doi: 10.1128/mBio.01012-15.

Larkin, M. A. *et al.* (2007) 'Clustal W and Clustal X version 2.0', *Bioinformatics*. Oxford University Press, 23(21), pp. 2947–2948. doi: 10.1093/bioinformatics/btm404.

Lefèvre, C. *et al.* (2004) 'Endosymbiont Phylogenesis in the Dryophthoridae Weevils: Evidence for Bacterial Replacement', *Molecular Biology and Evolution*. Oxford University Press, 21(6), pp. 965–973. doi: 10.1093/molbev/msh063.

Li, Z. and Wang, C. C. (2008) 'KMP-11, a basal body and flagellar protein, is required for cell division in Trypanosoma brucei.', *Eukaryotic cell*. American Society for Microbiology (ASM), 7(11), pp. 1941–50. doi: 10.1128/EC.00249-08.

Login, F. H. *et al.* (2011) 'Antimicrobial Peptides Keep Insect Endosymbionts Under Control', *Science*, 334(6054), pp. 362–365. doi: 10.1126/science.1209728.

Ma, Y., Weiss, L. M. and Huang, H. (2015) 'Inducible suicide vector systems for Trypanosoma cruzi', *Microbes and Infection*, 17(6), pp. 440–450. doi: 10.1016/j.micinf.2015.04.003.

Magnelli, P. E., Bielik, A. M. and Guthrie, E. P. (2011) 'Identification and characterization of protein glycosylation using specific endo- and exoglycosidases.', *Journal of visualized experiments : JoVE*. MyJoVE Corporation, (58), p. e3749. doi: 10.3791/3749.

Marin, B., M. Nowack, E. C. and Melkonian, M. (2005) 'A Plastid in the Making: Evidence for a Second Primary Endosymbiosis', *Protist*, 156(4), pp. 425–432. doi: 10.1016/j.protis.2005.09.001.

Martínez-Calvillo, S. *et al.* (2010) 'Gene expression in trypanosomatid parasites.', *Journal of biomedicine & biotechnology*. Hindawi, 2010, p. 525241. doi: 10.1155/2010/525241.

McCutcheon, J. P. and Moran, N. A. (2012) 'Extreme genome reduction in symbiotic bacteria', *Nature Reviews Microbiology*. Nature Publishing Group, 10(1), pp. 13–26. doi:

10.1038/nrmicro2670.

Michaeli, S. (2011) '*Trans* -splicing in trypanosomes: machinery and its impact on the parasite transcriptome', *Future Microbiology*, 6(4), pp. 459–474. doi: 10.2217/fmb.11.20.

Michels, P. A. *et al.* (1986) 'Two tandemly linked identical genes code for the glycosomal glyceraldehyde-phosphate dehydrogenase in Trypanosoma brucei.', *The EMBO journal*, 5(5), pp. 1049–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3013612 (Accessed: 13 March 2018).

Morales, J. *et al.* (2016) 'Development of a toolbox to dissect host-endosymbiont interactions and protein trafficking in the trypanosomatid Angomonas deanei', *BMC Evolutionary Biology*. BioMed Central, 16(1), p. 247. doi: 10.1186/s12862-016-0820-z.

Motta, M. C. *et al.* (1997) 'Ultrastructural and biochemical analysis of the relationship of Crithidia deanei with its endosymbiont.', *European journal of cell biology*, 72(4), pp. 370–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9127737 (Accessed: 10 March 2018).

Motta, M. C. M. *et al.* (2010) 'The Bacterium Endosymbiont of Crithidia deanei Undergoes Coordinated Division with the Host Cell Nucleus', *PLoS ONE*. Edited by D. Lew, 5(8), p. e12415. doi: 10.1371/journal.pone.0012415.

Motta, M. C. M. *et al.* (2013) 'Predicting the Proteins of Angomonas deanei, Strigomonas culicis and Their Respective Endosymbionts Reveals New Aspects of the Trypanosomatidae Family', *PLoS ONE*. Edited by J. Parkinson. Public Library of Science, 8(4), p. e60209. doi: 10.1371/journal.pone.0060209.

MOTTA, M. C. M. *et al.* (1997) 'Detection of Penicillin-binding Proteins in the Endosymbiont of the Trypanosomatid Crithidia deanei', *The Journal of Eukaryotic Microbiology*. Blackwell Publishing Ltd, 44(5), pp. 492–496. doi: 10.1111/j.1550-7408.1997.tb05729.x.

Moya, A., Gil, R. and Latorre, A. (2009) 'The evolutionary history of symbiotic associations

among bacteria and their animal hosts: a model', *Clinical Microbiology and Infection*. Elsevier, 15, pp. 11–13. doi: 10.1111/J.1469-0691.2008.02689.X.

MUNDIM, M. H. and ROITMAN, ISAACMUNDIM, M. H. and ROITMAN, I. (1977) 'Extra Nutritional Requirements of Artificially Aposymbiotic Crithidia deanei *', The Journal of Protozoology. Blackwell Publishing Ltd, 24(2), pp. 329–331. doi: 10. 1111/j. 155.-7408. 1977. tb00988. x. (1977) 'Extra Nutritional Requirements of Artificially Aposymbiotic Crithidia deanei *', *The Journal of Protozoology*. Blackwell Publishing Ltd, 24(2), pp. 329–331. doi: 10.1111/j.1550-7408.1977.tb00988.x.

Nakabachi, A. *et al.* (2014) 'Aphid gene of bacterial origin encodes a protein transported to an obligate endosymbiont', *Current Biology*. Cell Press, 24(14), pp. R640–R641. doi: 10.1016/J.CUB.2014.06.038.

Nikoh, N. *et al.* (2010) 'Bacterial Genes in the Aphid Genome: Absence of Functional Gene Transfer from Buchnera to Its Host', *PLoS Genetics*. Edited by G. P. Copenhaver, 6(2), p. e1000827. doi: 10.1371/journal.pgen.1000827.

Noronha, M. C., De Menezes, D. and Roitmanz, I. (1991) 'Nutritional Requirements of Blastocrithidia culicis, a Trypanosomatid with an Endosymbiont', 38(2), pp. 122–123. Available at: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1550-7408.1991.tb06030.x (Accessed: 1 June 2018).

NOVAK, E. *et al.* (1988) 'Protein Synthesis in Isolated Symbionts from the Flagellate Protozoon *Crithidia deanei* [1]', *The Journal of Protozoology*. Blackwell Publishing Ltd, 35(3), pp. 375–378. doi: 10.1111/j.1550-7408.1988.tb04110.x.

Nowack, E. C. M. *et al.* (2011) 'Endosymbiotic Gene Transfer and Transcriptional Regulation of Transferred Genes in Paulinella chromatophora', *Molecular Biology and Evolution*, 28(1), pp. 407–422. doi: 10.1093/molbev/msq209.

Nowack, E. C. M. *et al.* (2016) 'Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of *Paulinella chromatophora*', *Proceedings*

of the National Academy of Sciences, 113(43), pp. 12214–12219. doi: 10.1073/pnas.1608016113.

Nowack, E. C. M. and Grossman, A. R. (2012) 'Trafficking of protein into the recently established photosynthetic organelles of Paulinella chromatophora', *Proceedings of the National Academy of Sciences*, 109(14), pp. 5340–5345. doi: 10.1073/pnas.1118800109.

Nowack, E. C. M., Melkonian, M. and Glöckner, G. (2008) 'Chromatophore Genome Sequence of Paulinella Sheds Light on Acquisition of Photosynthesis by Eukaryotes', *Current Biology*, 18(6), pp. 410–418. doi: 10.1016/j.cub.2008.02.051.

Nozaki, T. and Cross, G. A. (1995) 'Effects of 3' untranslated and intergenic regions on gene expression in Trypanosoma cruzi.', *Molecular and biochemical parasitology*, 75(1), pp. 55–67. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8720175 (Accessed: 8 March 2018).

Oakeson, K. F. *et al.* (2014) 'Genome degeneration and adaptation in a nascent stage of symbiosis.', *Genome biology and evolution*. Oxford University Press, 6(1), pp. 76–93. doi: 10.1093/gbe/evt210.

ODA, L. M. *et al.* (1984) 'Surface Anionic Groups in Symbiote-Bearing and Symbiote-Free Strains of *Crithidia deanei* [1]', *The Journal of Protozoology*. Blackwell Publishing Ltd, 31(1), pp. 131–134. doi: 10.1111/j.1550-7408.1984.tb04301.x.

Ohbayashi, T. *et al.* (2015) 'Insect's intestinal organ for symbiont sorting.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 112(37), pp. E5179-88. doi: 10.1073/pnas.1511454112.

de Paiva, R. M. C. *et al.* (2015) 'Amastin Knockdown in Leishmania braziliensis Affects Parasite-Macrophage Interaction and Results in Impaired Viability of Intracellular Amastigotes', *PLOS Pathogens*. Edited by S. M. Beverley. Public Library of Science, 11(12), p. e1005296. doi: 10.1371/journal.ppat.1005296.

Patron, N. J. and Waller, R. F. (2007) 'Transit peptide diversity and divergence: A global

analysis of plastid targeting signals', *BioEssays*, 29(10), pp. 1048–1058. doi: 10.1002/bies.20638.

Penha, L. L. *et al.* (2016) 'Symbiont modulates expression of specific gene categories in Angomonas deanei.', *Memorias do Instituto Oswaldo Cruz*. Instituto Oswaldo Cruz, 111(11), pp. 686–691. doi: 10.1590/0074-02760160228.

Pérez-Brocal, V. *et al.* (2013) 'Symbionts and Pathogens: What is the Difference?', *Current Topics in Microbiology and Immunology*, 358, pp. 215–243. doi: 10.1007/82_2011_190.

Rapaport, D. (2003) 'Finding the right organelle. Targeting signals in mitochondrial outer-membrane proteins.', *EMBO reports*. European Molecular Biology Organization, 4(10), pp. 948–52. doi: 10.1038/sj.embor.embor937.

Sachs, J. L., Skophammer, R. G. and Regus, J. U. (no date) 'Evolutionary transitions in bacterial symbiosis'. doi: 10.1073/pnas.1100304108.

Sela, D. *et al.* (no date) 'Unique Characteristics of the Kinetoplast DNA Replication Machinery Provide Potential Drug Targets in Trypanosomatids Reevaluating the Kinetoplast as a Potential Target for Anti-Trypanosomal Drugs'. Available at: https://pdfs.semanticscholar.org/773f/a52c49e786264c850fa4a3195199e4190ec0.pdf (Accessed: 17 May 2018).

Shigenobu, S. and Stern, D. L. (2013) 'Aphids evolved novel secreted proteins for symbiosis with bacterial endosymbiont.', *Proceedings. Biological sciences*. The Royal Society, 280(1750), p. 20121952. doi: 10.1098/rspb.2012.1952.

Singer, A. *et al.* (2017) 'Massive Protein Import into the Early-Evolutionary- Stage Photosynthetic Organelle of the Amoeba Paulinella chromatophora'. doi: 10.1016/j.cub.2017.08.010.

Snider, J. *et al.* (2010) 'Split-ubiquitin based membrane yeast two-hybrid (MYTH) system: a powerful tool for identifying protein-protein interactions.', *Journal of visualized experiments :*

JoVE. MyJoVE Corporation, (36). doi: 10.3791/1698.

Sollelis, L. *et al.* (2015) 'First efficient CRISPR-Cas9-mediated genome editing in *L eishmania* parasites', *Cellular Microbiology*, 17(10), pp. 1405–1412. doi: 10.1111/cmi.12456.

Sonnhammer, E. L. L., Von Heijne, G. and Krogh, A. (1998) 'A hidden Markov model for predicting transmembrane helices in protein sequences'. AAAI Press, (1), pp. 175–182. Available at: http://people.binf.ku.dk/~krogh/publications/ps/SonnhammerEtal98.pdf (Accessed: 24 May 2018).

Stebeck, C. E. *et al.* (1995) 'Kinetoplastid membrane protein-11 (KMP-11) is differentially expressed during the life cycle of African trypanosomes and is found in a wide variety of kinetoplastid parasites.', *Molecular and biochemical parasitology*, 71(1), pp. 1–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7630374 (Accessed: 2 May 2018).

Steinert, M., Hentschel, U. and Hacker, J. (2000) 'Symbiosis and pathogenesis: evolution of the microbe-host interaction.', *Die Naturwissenschaften*, 87(1), pp. 1–11. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10663126 (Accessed: 13 February 2018).

Stouthamer, R., Breeuwer, J. A. J. and Hurst, G. D. D. (1999) '*Wolbachia Pipientis* : Microbial Manipulator of Arthropod Reproduction', *Annual Review of Microbiology*, 53(1), pp. 71–102. doi: 10.1146/annurev.micro.53.1.71.

Teixeira, M. M. G. *et al.* (2011) 'Phylogenetic Validation of the Genera Angomonas and Strigomonas of Trypanosomatids Harboring Bacterial Endosymbionts with the Description of New Species of Trypanosomatids and of Proteobacterial Symbionts', *Protist*, 162(3), pp. 503–524. doi: 10.1016/j.protis.2011.01.001.

Terfrüchte, M. *et al.* (2014) 'Establishing a versatile Golden Gate cloning system for genetic engineering in fungi.', *Fungal genetics and biology : FG & B*, 62, pp. 1–10. doi: 10.1016/j.fgb.2013.10.012.

Votýpka, J. *et al.* (2014) 'Kentomonas gen. n., a New Genus of Endosymbiont-containing

Trypanosomatids of Strigomonadinae subfam. n.', *Protist*, 165(6), pp. 825–838. doi: 10.1016/j.protis.2014.09.002.

Walther, D. M. *et al.* (2009) 'Signals in bacterial beta-barrel proteins are functional in eukaryotic cells for targeting to and assembly in mitochondria.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 106(8), pp. 2531–6. doi: 10.1073/pnas.0807830106.

Wang, J. *et al.* (2009) 'Interactions between mutualist Wigglesworthia and tsetse peptidoglycan recognition protein (PGRP-LB) influence trypanosome transmission.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 106(29), pp. 12133–8. doi: 10.1073/pnas.0901226106.

Webster, N. S. (2014) 'Cooperation, communication, and co-evolution: grand challenges in microbial symbiosis research', *Frontiers in Microbiology*. Frontiers, 5, p. 164. doi: 10.3389/fmicb.2014.00164.

Wernegreen, J. J. (no date) 'Endosymbiont evolution: Predictions from theory and surprises from genomes'. doi: 10.1111/nyas.12740.

Wheeler, R. J., Gluenz, E. and Gull, K. (2013) 'The limits on trypanosomatid morphological diversity', *PLoS ONE*, 8(11). doi: 10.1371/journal.pone.0079581.

Wilson, R. C. and Doudna, J. A. (2013) 'BB42CH10-Doudna Molecular Mechanisms of RNA Interference Argonaute: a protein capable of binding short ssRNAs and, in some cases, cleaving a bound complementary strand', *Annu. Rev. Biophys*, 42, pp. 217–39. doi: 10.1146/annurev-biophys-083012-130404.

Wittig, I., Karas, M. and Schägger, H. (2007) 'High resolution clear native electrophoresis for in-gel functional assays and fluorescence studies of membrane protein complexes.', *Molecular & cellular proteomics : MCP*. American Society for Biochemistry and Molecular Biology, 6(7), pp. 1215–25. doi: 10.1074/mcp.M700076-MCP200.

Wu, Y., Li, Q. and Chen, X.-Z. (2007) 'Detecting protein–protein interactions by far western blotting', *Nature Protocols*. Nature Publishing Group, 2(12), pp. 3278–3284. doi: 10.1038/nprot.2007.459.

Xie, Y. *et al.* (2016) 'GPS-Lipid: a robust tool for the prediction of multiple lipid modification sites', *Scientific Reports*. Nature Publishing Group, 6(1), p. 28249. doi: 10.1038/srep28249.

Yurchenko, V. Y. *et al.* (2009) 'Selective recovery of the cultivation-prone components from mixed trypanosomatid infections: a case of several novel species isolated from Neotropical Heteroptera', *International Journal of Systematic and Evolutionary Microbiology*, 59(4), pp. 893–909. doi: 10.1099/ijs.0.001149-0.

Zheng, L., Baumann, U. and Reymond, J.-L. (2004) 'An efficient one-step site-directed and site-saturation mutagenesis protocol', *Nucleic Acids Research*, 32(14), pp. e115–e115. doi: 10.1093/nar/gnh110.

6   Appendix

Figure A1: Sequence analysis of *etp1* 3′ FR of scaffold 1854. Genomic DNA (gDNA) extracted from *A. deanei* WT cells was used as a template for PCR with primers binding to the gDNA close to end of *etp1* ORF and in the beginning of the next fragment of scaffold 1854. The PCR product obtained was sequenced and the alignment shows the resulting sequence.
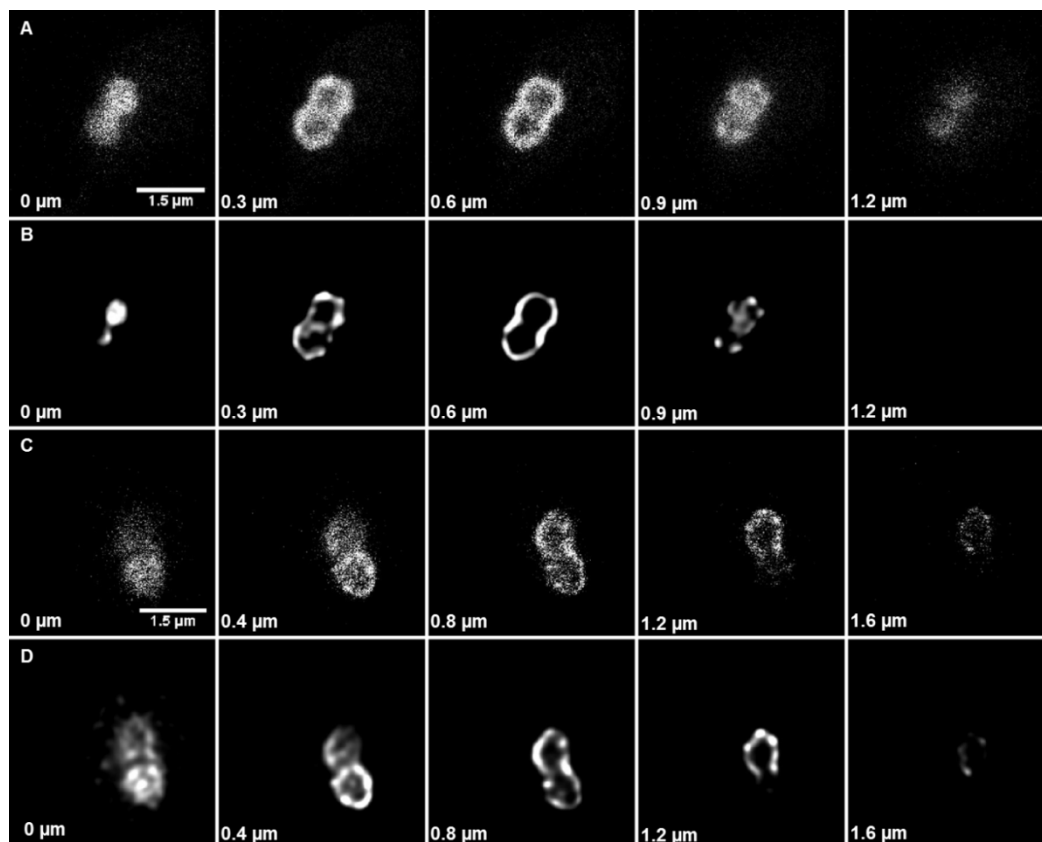
Figure A2: ETP1-EGFP and SNAP-ETP1 signals were detected only in the membrane of the endosymbiont. (A) EGFP-ETP1 fluorescence at 488 nm. Signal was detected by emission at 498-582 nm. Slices of 0.2 μm depth were resliced to 0.3 μm; (C) Δ-δ-AMA$^{SNAP-ETP1, NEO}$ (Adea0046) cells stained with SNAP-cell 647 SiR. Excitation at 647 nm and depletion at 776 nm. Slice depth at 0.4 μm. (B) and (D): deconvoluted data of A and C, respectively. All samples were embedded in prolong diamond and covered with 1.5H coverslips. Scale bar: 1.5 μm. (Master thesis, Georg Ehret)

Table A1: The immunoprecipitated samples (from three biological replicates) acquired by Co-IP assay were subjected to solution trypsin digestion and analyzed by tandem mass spectrometry (MS) (as described in 4.2.3.9). Proteins marked with red are considered as potential interaction partners of ETP1.

Protein ID[a]	Protein Name	Organism	Averaged intensities[b]		
			WT control	**GFP contol**	**nGFP-ETP1**
M1LUL4	Porin	*C. K. crithidii* TCC036E	1012400	41252567	1390886667
S9TVC7	Uncharacterized protein	*A. deanei*	49919000	77493333	79756666
P42212	GFP	*A. victoria*	0	54115100	10346100

					(Jellyfish)
S9VIB5	Kinetoplastid membrane protein-11	*A. deanei*	0	0	10111380
S9U3N9	Uncharacterized protein	*A. deanei*	0	0	8065967
ETP1	ETP1	*A. deanei*	0	0	7639273
M1LQ13	Outer membrane protein A	*C. K. crithidii* TCC036E	0	175406	4374193
M1LQ20	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex	*C. K. crithidii* TCC036E	0	383773	3192443
M1LW78	Succinate--CoA ligase subunit beta	*C. K. crithidii* TCC036E	0	805656	1962233
S9WUQ5;S9V8N2;S9V6E3	Uncharacterized protein	*A. deanei*	927560	1102313	763653
M1M5R8	Succinate-CoA ligase subunit alpha	*C. K. crithidii* TCC036E	0	246116	1385807
M1LW90	3-isopropylmalate dehydratase large subunit	*C. K. crithidii* TCC036E	0	209381	1207766
M1LNS1	Histidinol dehydrogenase	*C. K. crithidii* TCC036E	0	114760	1258770
M1M6B8	Phospho-2-dehydro-3-deoxyheptonate aldolase	*C. K. crithidii* TCC036E	0	313963	1023133
M1LTH4	60 kDa chaperonin	*C. K. crithidii* TCC036E	0	272700	984570
S9VJN5;S9VEA1;S9WN06;S9U8P5;S9WM11	Tubulin alpha	*A. deanei*	774730	733893	689190
S9V852	Protein FAM184A	*A. deanei*	0	0	1067380
S9V2U7	Uncharacterized protein	*A. deanei*	0	0	1012666
M1LV02	L-threonine dehydratase	*C. K. crithidii* TCC036E	0	106000	887700

M1M5C8	Putative ATPase of the AAA+ superfamily	*C. K. crithidii* TCC036E	0	116133	872940
M1L578	Ketol-acid reductoisomerase	*C. K. crithidii* TCC036E	0	445170	513133
M1LQY0	30S ribosomal protein S19	*C. K. crithidii* TCC036E	0	0	918330
M1LUN1	Chaperone protein DnaK (HSP70)	*C. K. crithidii* TCC036E	0	135480	732313
S9UV10	Uncharacterized protein	*A. deanei*	0	0	508866
M1L4J1	Phosphoenolpyruvate synthase	*C. K. crithidii* TCC036E	0	118540	718423
M1LWT2	Adenylosuccinate synthetase	*C. K. crithidii* TCC036E	0	0	282420
M1M6U3	Acetolactate synthase	*C. K. crithidii* TCC036E	0	0	259923
S9VL77	Uncharacterized protein	*A. deanei*	0	20296	32892

[a] Protein identifier corresponding to the Uniprot Database (http://www.uniprot.org).

[b] Average deviations were calculated from three biological replicates for the GFP control and nGFP-ETP1 samples. On the other hand, in the case of the WT control one out of three replicates was analyzed by Mass Spectrometry.

Table A2: Tandem MS analysis of the band that contained ETP1-cV5 and was extracted from the Clear native gel. Proteins marked with red are considered as potential interaction partners of ETP1, since they were also identified by the Co-IP analysis.

Protein ID[a]	Protein Name	Intensity
M1LUL4	Porin	65137000000
ETP1	ETP1	6925000000
M1LWZ5	Malate dehydrogenase	3130100000
M1LTD0	ATP synthase subunit alpha	2405800000

M1LUE3	Dihydrolipoyl dehydrogenase	2000500000
M1LP27;S9VRM9	Carbamoyl-phosphate synthase large chain	1967400000
M1LQ13	Outer membrane protein A	1940200000
M1M6U0	Polyribonucleotide nucleotidyltransferase	1842000000
M1M5V9	Aminotransferase	1617500000
M1LW72	Phosphoribosylamine--glycine ligase	1604600000
M1L5G6	Serine hydroxymethyltransferase	1554300000
M1LUN1	Chaperone protein DnaK (HSP70)	1261300000
M1M5H3	Transketolase	1146600000
M1L5V6;M1M7I9	Elongation factor Tu	1139000000
M1L4Y2	2-oxoglutarate dehydrogenase E1 component	1080600000
M1L3K6	2-isopropylmalate synthase	1029200000
M1LP54	Aspartokinase	920900000
M1LVU3	Dihydroxy-acid dehydratase	919740000
M1LV02	L-threonine dehydratase	918610000
M1M6S7	Peroxiredoxin	858010000
M1LQ20	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex	819640000
M1LQV4	Diaminopimelate decarboxylase	675890000
M1L4D0	Superoxide dismutase	671970000
M1L3P4	ATP synthase subunit beta	633480000
M1M7H6	30S ribosomal protein S7	549730000
M1LQ27	Enolase	536100000
M1L5W1	DNA-directed RNA polymerase su	462760000
M1M5J1	Transaldolase	437550000
M1LNW3	Phosphoenolpyruvate carboxylase 30S ribosomal protein S1	407330000
S9VBF7	Phosphoglycerate kinase	405620000
M1LWX8	30S ribosomal protein S1	400460000

M1L4J1	Phosphoenolpyruvate synthase	393100000
M1M7I2	50S ribosomal protein L10	384440000
S9VKS1	Aldehyde dehydrogenase	370210000
M1LWJ9	Phosphoribosylformylglycinamidine synthase	362520000
M1L578	Ketol-acid reductoisomerase	330570000
M1LUP5	Bifunctional purine biosynthesis protein PurH	323520000
M1LTD9	Thiazole synthase	317130000
M1L5T8	30S ribosomal protein S5	294760000
M1M5R8	Succinate--CoA ligase [ADP-forming] subunit alpha	282620000
M1LWT2	Adenylosuccinate synthetase	276050000
M1M634	Dihydrolipoyl dehydrogenase	274370000
M1LW78	Succinate--CoA ligase [ADP-forming] subunit beta	274160000
M1LX60	NADH-quinone oxidoreductase	266720000
M1M765	Cytochrome bd-I oxidase subunit I	265800000
M1M6E7	ATP phosphoribosyltransferase regulatory subunit	264040000
M1M6U3	Acetolactate synthase	257350000
M1LWK3	Amidophosphoribosyltransferase	248970000
M1LTJ2	Glyceraldehyde-3-phosphate dehydrogenase	242940000
M1L4P5	GTP cyclohydrolase FolE2	241690000
M1L5H1	30S ribosomal protein S9	240110000
M1L3L8	ATP phosphoribosyltransferase	227480000
M1L5M1	HPr kinase/phosphorylase	219780000
M1LQN1	Glutamate-1-semialdehyde 2,1-aminomutase	209960000

M1M5E9	Ribose-phosphate pyrophosphokinase	207450000
M1LP11	Phosphoglycerate kinase	202250000
M1L466	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase	196000000
M1LXN6	Aspartate--tRNA(Asp/Asn) ligase	193360000
M1L4E1	Pyruvate dehydrogenase E1 component	187780000
M1LW90	3-isopropylmalate dehydratase large subunit	177870000
M1L4X0	Histidinol-phosphate aminotransferase	167760000
S9VAI5	Uncharacterized protein	163680000
S9V2H7;S9VNV5	Aconitate hydratase	145170000
S9V0H9;S9WTI9;S9VKL2;S9VYC8;S9VHB2;S9WJL9;S9UNN6;S9VFG9;S9VKF1	Molecular chaperone DnaK	141410000
M1LTS7	3-isopropylmalate dehydrogenase	139640000
M1M5S4	Inorganic pyrophosphatase	139570000
M1LUL8	30S ribosomal protein S15	136790000
M1L442	Chorismate synthase	119810000
M1LV64	DNA-directed RNA polymerase subunit alpha	112010000
M1LQY7	30S ribosomal protein S10	111270000
S9VIB5	Kinetoplastid membrane protein KMP-11	110270000
M1L3N9	ATP synthase subunit c	13202000
M1LPI7	Bifunctional protein FolD	13083000
M1LVR1	Anthranilate synthase component II	12761000
M1L563	NADH-quinone oxidoreductase subunit H	12472000
M1M7G7	50S ribosomal protein L4	12341000

S9V2U7	ETP2	12329000
M1LXP0	Single-stranded DNA-binding protein	12269000

[a] Protein identifier corresponding to the Uniprot Database (http://www.uniprot.org).

7   Manuscript

The following manuscript was published on November 11, 2016 in the journal BMC Evolutionary Biology:

"Development of a toolbox to dissect host-endosymbiont interactions and protein trafficking in the trypanosomatid *Angomonas deanei"*

Authors: Jorge Morales*, Sofia Kokkori*, Diana Weidauer, Jarrod Chapman, Eugene Goltsman, Daniel Rokhsar, Arthur R. Grossman and Eva C. M. Nowack.

*Contributed equally

Below are listed the illustrations from the manuscript that were used modified as illustrations in this dissertation (with the corresponding image number):

Number of figure in manuscript	Number of figure in Dissertation
Figure 1	Figure 2.1-1
Figure 2	Figure 2.1-4
Figure 3	Figure 2.1-5
Figure 4a	Figure 2.1-7
Figure 4b	Figure 2.1-6
Figure 6	Figure 2.3-3
Figure S2	Figure 2.1-2
Figure S3 a,b	Figure 2.1-3
Figure S4	Figure 2.3-4

Genome and transcriptome assemblies presented in this manuscript were generated by Dr. Eva C. M. Nowack in collaboration with Daniel Rokhsar's lab at the DOE, Joint Genome Institute in Walnut Creek, California. The growth curve showing the sensitivity of WT *A. deanei* cells in G418 was constructed by Diana Weidauer (Figure 1a). Moreover, Dr. Jorge Morales generated Δ-δ-AMA^{NEO} (Adea0001) and Δ-γ-AMA^{HYG}/Δ-δ-AMA^{NEO} (Adea0008) cell lines and I conducted the PCR and Southern blot analyses using these cell lines. Aditionally, results shown in Figure 5 were generated by Dr. Jorge Morales. Gereon

The following manuscript was published on November 11, 2016 in the journal BMC Evolutionary Biology:

Poschmann and Kai Stühler carried out the mass spectrometric analysis and provided the raw data for Table 1, whereas the raw data were evaluated by Dr. Jorge Morales. All the other results presented in this publication were produced by me. Dr. Jorge Morales and Dr. Eva C. M. Nowack wrote the manuscript.

I hereby confirm the accuracy of this information.

16.07.2018, Düsseldorf

Acknowledgements

It gives me great pleasure to express my sincere and heartfelt thanks to all those who helped me during the period of my research and writing my thesis.

First and foremost, I would like to thank my supervisor *Eva Nowack* for giving me the great opportunity to be part of her research group and introducing me to the fascinating field of endosymbiosis. During my research, I received inspiration and encouragement, when the experiments did not work as it was planed, through her guidance. Thank you for being always there.

I also thank my second supervisor *Prof. Dr. Johannes Hegemann* for the fruitful discussions that we had in our meetings and his valuable suggestions for my project.

Moreover, I would like to thank Molecules of Infection (MOI) *Manchot Graduate School* for funding my thesis.

It gives me pleasure to express my gratitude to *Jorge Morales*. Jorge taught me a lot of techniques and many of my data would not have been generated without his help. Thanks for the personal support and interesting discussions (about science, politics and many other topics). Special thanks to my friend *Anna Singer*, who was always there supporting me. Thanks for cheering me up when I was depressed with my research progress and for all the fun and the nice discussions (about science, movies and not only) that we had. I am thankful to *Linda, Georg, Jan, Diana, Luis, Valentin* and *Eugenia* for the great atmosphere that we had in the lab. You all made my stay in the group a memorable one.

I would like to extend my heartfelt gratitude towards *all Feldi members*. Thanks for the nice discussions during seminars and for being there whenever I needed something, answering every question. You are the best.

At this point I would like to thank my friends, *Eva* and *Maria*. Thank you for your honest and deep interest in me and my work. Heartfelt thanks to my husdand, *Thodoris Latsis*, for understanding me and supporting my decision to pursue my PhD abroad. Your support means the world to me. Above all I would like to thank my *father, mother* and my siblings, *Maria* and *Ilias*. You always believe in me and encourage me in the difficult times. *Dear parents* you are my inspiration and this work is dedicated to you.

Curriculum Vitae

Sofia Kokkori

Date of birth: 17/08/1987

Nationality: Greek

EDUCATION AND TRAINING

07/2014-09/2018

PhD Candidate: Institute of Microbiology, Heinrich-Heine-University of Düsseldorf, Germany

Scholarship holder of The Manchot Graduate School "Molecules of Infection".

Title of dissertation: Molecular tools for the trypanosomatid *Angomonas deanei* help to dissect host/endosymbiont interactions. Laboratory and technical skills acquired during PhD: Analysis of DNA sequences, Gibson and Golden Gate cloning, DNA/RNA isolation, PCR analysis, agarose gel electrophoresis, SDS-PAGE, clear-native (CN) electrophoresis, 2-D CN/SDS-PAGE, Western/Southern blot analysis, Fluorescence In Situ Hybridization (FISH), Immunofluorescence microscopy, Co-Immunoprecipitation assay, transformation of *E. coli*, genetic manipulation of *A. deanei* (transfection, knock-out, heterologous expression of genes, RNA interference).

09/2011-01/2014

Master in Biomedical Sciences: Department of Medicine, University of Patras, Greece

Title of thesis: Linezolid-dependent structure and function adaptation of ribosomes from nosocomial strains of *Staphylococcus epidermidis*. Laboratory skills acquired during Master thesis: Cultivation of *S. epidermidis,* isolation of active ribosomal particles, in vitro assays of translation (binding of tRNAs at the A-, P-, E-site, reconstitution of 70S ribosomes, peptide bond formation, puromycin reaction).

09/2005-07/2011

Bachelor in Biology: Department of Biology, University of Patras, Greece

Title of Diploma thesis: Functionality of 70S ribosome in the presence of ketolide K- 1602.

WORKING EXPERIENCE

10/2017-11/2017 and 10/2016-11/2016

Teaching assistant in Laboratory of Microbiology, Heinrich-Heine-University of Düsseldorf, Germany

PERSONAL SKILLS

Mother tongue: Greek

Other Languages	Understanding	Speaking	Writing
English	C2	C2	C2
German	B1 (B2 soon)	B1 (B2 soon)	B1(B2 soon)

PUBLICATIONS

- Morales, J.*, **Kokkori, S.***, Weidauer, D., Chapman, J., Goltsman, E., Rokhsar, D., Grossman, A.R., and Nowack, E.C.M. (2016) Development of a toolbox to dissect host-endosymbiont interactions and protein trafficking in the trypanosomatid *Angomonas deanei.* BMC Evol Biol *16*: 247.(* authors contributed equally)

- **Kokkori, S**., Apostolidi, M., Tsakris, A., Pournaras, S., Stathopoulos, C., and Dinos, G. (2014) Linezolid-dependent function and structure adaptation of ribosomes in a *Staphylococcus epidermidis* strain exhibiting linezolid dependence. Antimicrob Agents Chemother *58*: 4651.

WORKSHOPS

- Good scientific practice (15/09/2014).
- Presenting Science-How to own the stage on (international) conferences (26-27/02/2015).
- Get into teaching (10-11/03/2016).
- Career Planning in Business-How to shape your future (04-05/04/2016).
- Leadership skills (16-17/02/2017).
- Preparing for Potential Conflicts (30-31/03/2017).

CONFERENCES

- **Kokkori S.,** Morales J., Nowack E. (2017). Genetic tools for the trypanosomatid *Angomonas deanei* help to dissect host/endosymbiont interactions. ISE-G Conference 2017, September 3-6, 2017, Konstanz/ Germany.
- **Kokkori S.,** Apostolidi M., Pournaras S., Stathopoulos C., Dinos G. (2013). Enhanced peptidyltransferase activity in mutant ribosomes from nosocomial strains of *S. epidermidis* in the presence of linezolid. 35th Scientific conference of Hellenic society for Biological Sciences, May 23-25, 2013, Nafplio/ Greece.
- **Kokkori S**., Apostolidi M., Tsakris A., Pournaras S., Stathopoulos C., Dinos G. (2013).Linezolid-dependent structure and function adaptation of ribosomes from nosocomial strains of *S. epidermidis*. 64th Scientific conference of Hellenic society for Biochemistry and Molecular Biology, December 6-8, 2013, Athens/ Greece.