CellPress

Current Biology

Article

A *neo*-functionalized homolog of host transmembrane protein controls localization of bacterial endosymbionts in the trypanosomatid *Novymonas esmeraldas*

Alexandra Zakharova,[1,8] Daria Tashyreva,[2,8] Anzhelika Butenko,[1,2,3] Jorge Morales,[4,6] Andreu Saura,[1] Michaela Svobodová,[2] Gereon Poschmann,[5] Satish Nandipati,[2,3] Alena Zakharova,[1] David Noyvert,[1,7] Ondřej Gahura,[2] Jiří Týč,[2] Kai Stühler,[4,5] Alexei Y. Kostygov,[1,*] Eva C.M. Nowack,[4] Julius Lukeš,[2,3] and Vyacheslav Yurchenko[1,9,*]

[1]Life Science Research Centre, Faculty of Science, University of Ostrava, 710 00 Ostrava, Czech Republic
[2]Institute of Parasitology, Biology Centre, Czech Academy of Sciences, 370 05 České Budějovice (Budweis), Czech Republic
[3]Faculty of Sciences, University of South Bohemia, 370 05 České Budějovice (Budweis), Czech Republic
[4]Institute of Microbial Cell Biology, Heinrich Heine University, 40225 Düsseldorf, Germany
[5]Institute of Molecular Medicine, Heinrich Heine University, 40225 Düsseldorf, Germany
[6]Present address: Envigo RMS GmbH, Düsseldorf, Germany
[7]Present address: University of Cambridge, Cambridge, UK
[8]These authors contributed equally
[9]Lead contact
*Correspondence: kostygov@gmail.com (A.Y.K.), vyacheslav.yurchenko@osu.cz (V.Y.)
https://doi.org/10.1016/j.cub.2023.04.060

SUMMARY

The stability of endosymbiotic associations between eukaryotes and bacteria depends on a reliable mechanism ensuring vertical inheritance of the latter. Here, we demonstrate that a host-encoded protein, located at the interface between the endoplasmic reticulum of the trypanosomatid *Novymonas esmeraldas* and its endosymbiotic bacterium *Ca.* Pandoraea novymonadis, regulates such a process. This protein, named TMP18e, is a product of duplication and *neo*-functionalization of the ubiquitous transmembrane protein 18 (TMEM18). Its expression level is increased at the proliferative stage of the host life cycle correlating with the confinement of bacteria to the nuclear vicinity. This is important for the proper segregation of bacteria into the daughter host cells as evidenced from the TMP18e ablation, which disrupts the nucleus-endosymbiont association and leads to greater variability of bacterial cell numbers, including an elevated proportion of aposymbiotic cells. Thus, we conclude that TMP18e is necessary for the reliable vertical inheritance of endosymbionts.

INTRODUCTION

The transition of a free-living bacterium into a fully integrated organelle is accompanied by massive structural, physiological, and genetic adaptations.[1,2] Although mitochondria and plastids originated from bacteria, over a billion years of co-evolution with the host resulted in a situation where organelle and the surrounding cell cannot be regarded as independent organisms anymore. Instead, the organelles' proteome composition, metabolic state, the timing of division, positioning within the cell, etc. are largely under the genetic control of the nucleus.[2]

More recently established endosymbiotic associations provide an opportunity to observe snapshots of intermediate stages of the process during which a prokaryote becomes integrated into a eukaryotic cell. Hence, the dissection of molecular mechanisms underlying these host-endosymbiont interactions holds the promise to shed light on possible scenarios and molecular underpinnings of organellogenesis. In most cases, endosymbionts are distributed throughout the host cytoplasm, usually enclosed in host-derived vacuoles, although in a few systems endosymbionts reside in the nucleus or are tightly associated with specific host compartments (e.g., the hydrogenosomes of anaerobic ciliates).[3] In some endosymbiotic associations, such as those of legume plants with nitrogen-fixing Rhizobia or deep-sea tube worms with chemosynthetic bacteria, the microbes can live independently and every host generation has to acquire them from an environmental pool.[4,5] However, in the majority of systems, endosymbionts are vertically transmitted between host generations. Genomes of these obligate intracellular symbionts underwent and are undergoing changes similar to those of present-day organelles. They are characterized by low GC content, elevated evolutionary rate, shrinkage of metabolic capacities, gene transfer into the host nucleus, and the loss of genes providing no benefit for the host.[6–9] The growing dependence of the host on the biological functions of the endosymbiont dictates that a reliable inheritance of the latter must be ensured. In multicellular organisms, vertical endosymbiont transmission can involve complex developmental processes. For instance, *Buchnera* endosymbionts of pea aphids are exocytosed from the maternal bacteriocytes, endocytosed by the syncytial cytoplasm of the blastula, and concentrated in the specific region of the developing embryo giving rise to new bacteriocytes.[10] Many protists have also established vertical endosymbiont transmission. In host cells that are densely packed with

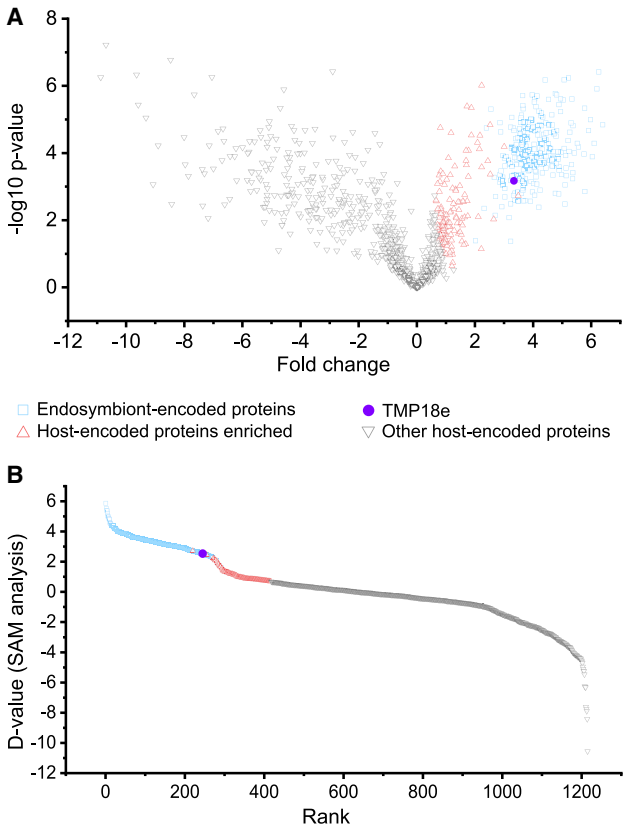


Figure 1. Identification of putative endosymbiont-targeted proteins by mass spectrometry

(A) Volcano plot visualizing abundance differences between purified endosymbionts and whole-cell lysates (fold change calculated as the difference of mean values of $\log_2$-normalized intensities).

(B) Rank plot showing ordered proteins according to the D value combining fold change and $-\log_{10}$ p value.

See also Table S1.

hundreds or thousands of symbionts, the segregation of bacteria to the daughter cell upon host cell division might represent a passive, stochastic process. This can be exemplified by the cellulolytic parabasalid *Pseudotrichonympha grassii* with 10^3–10^4 Bacteroidetes endosymbionts per host cell[11] and the ciliate *Pseudoblepharisma tenue*, which houses hundreds of purple bacterial endosymbionts along with green algae.[12] Other associations of protists with prokaryotes are characterized by a strict control of the number of endosymbionts per host cell. Among those are the cercozoan *Paulinella* with two cyanobacterium-derived photosynthetic "chromatophores,"[13] and trypanosomatids of the subfamily Strigomonadinae harboring a single β-proteobacterium.[14] The molecular mechanisms underlying the host's control over bacterial localization, abundance, and segregation are largely unknown. This, at least partially, can be explained by the lack of suitable model systems for which efficient molecular tools are available.

Trypanosomatid flagellates, a parasitic lineage of Euglenozoa, are a group of protists for which a wide spectrum of forward and reverse genetic methods is available.[15,16] Its endosymbiont-carrying members grow axenically to high densities, can be cryopreserved, and are readily amenable to genetic manipulations.[17,18] These recently established tools in *Angomonas deanei* (Strigomonadinae) enabled the discovery of nucleus-encoded machinery forming a ring structure around the endosymbiont-division site that controls endosymbiont division in a similar fashion to that of the mitochondria and plastids via a dynamin-based mechanism.[19]

The acquisition of an intracellular bacterium by the ancestor of Strigomonadinae had been considered a unique case in the evolution of trypanosomatids[20] until the discovery of another endosymbiotic association of the bacterium "*Ca.* Pandoraea novymonadis" (*Ca.* P. novymonadis) with *Novymonas esmeraldas*.[21] In both trypanosomatid flagellates, the endosymbiont compensates for the host's inability to synthesize certain vitamins, amino acids, and nucleosides.[22] Although this association is presumed to have been established relatively recently, *Ca.* P. novymonadis already underwent pronounced reductive genome evolution. The number of endosymbionts per host cell does not seem to be strictly controlled and varies in culture from zero to over a dozen.[21] This observation makes *N. esmeraldas* an appropriate model to explore the question of whether endosymbiont number and segregation can be actively controlled by the host cell at this relatively early stage of integration.

In this work, we identified several *N. esmeraldas* proteins, which co-purify with *Ca.* P. novymonadis and, thus, might be involved in host-symbiont interaction. One of these proteins that emerged by duplication of the widespread transmembrane protein 18 (TMEM18) with an unknown function (henceforth, TMP18) was selected for a detailed analysis. We describe the first known case in which the ablation of a protistan host-derived protein has an impact on the prokaryotic endosymbionts, resulting in their altered localization and asynchronization of their division with that of the host cell in a stage-specific manner.

RESULTS

Identification of putative endosymbiont-targeted proteins

To identify host proteins involved in the interaction of *N. esmeraldas* and *Ca.* P. novymonadis, we compared the lysates from isolated endosymbionts and the holobiont by liquid chromatography-tandem mass spectrometry (LC-MS/MS). In these samples, 3,548 proteins were identified with high confidence (Table S1). Of these, 2,879 and 669 proteins were identified as host- and endosymbiont-encoded, which corresponded to 31% and 69% of the protein-coding genes in these organisms, respectively. Besides endosymbiont-encoded proteins, there were 144 host-encoded proteins significantly enriched in the endosymbiont fraction (Figure 1; Table S1). However, many of these proteins based on their annotation and/or predicted targeting signals were peroxisomal, mitochondrial, or endoplasmic reticulum (ER) proteins suggesting a fair amount of contamination in the analyzed endosymbiont fractions.

Phylogenetic analysis of TMP18 proteins

One nucleus-encoded protein enriched in the endosymbiont fraction, namely NESM_000205400, attracted our attention. It was the second most enriched nucleus-encoded protein in the endosymbiont fraction when compared with the holobiont

lysate, and more importantly, BLAST analysis revealed its in-paralog, NESM_000205500, which was not enriched in the endosymbiont fraction. Notably, in the genomes of all previously analyzed trypanosomatids, the corresponding gene is not duplicated. Both proteins were annotated as TMEM18, so we designated the variant present in all analyzed genomes as TMP18, and the endosymbiont-associated one as TMP18e.

Despite the short alignment length (190 aa [amino acids] after trimming), the phylogenetic inference resulted in a reasonable topology (Figure 2A), in most details coinciding with that of the species tree based on phylogenomic data.[23] The sequences of the two *N. esmeraldas* in-paralogs, TMP18 and TMP18e, formed a monophyletic group with absolute supports, indicating that they diverged within this lineage. However, the level of TMP18 sequence divergence was comparable to that between the sister genera *Porcisia* and *Endotrypanum* indicating that duplication may not be a recent event or that it was followed by accelerated sequence evolution. Of note, the evolution of different parts of this protein across the tree is uneven, and, therefore, a model accounting for heterotachy had to be used for the topology inference. We conclude that TMP18e of *N. esmeraldas* is a unique, duplicated, and *neo*-functionalized copy of TMP18, which is present across trypanosomatids.

Bioinformatic analyses of TMP18 proteins in trypanosomatids

The search in the Conserved Domain Database identified the TMEM18 family domain in both TMP18 variants of *N. esmeraldas*. The sequence of this domain appears to be conserved in Trypanosomatidae (Figure S1), although three species (*Herpetomonas muscarum*, *Strigomonas galati*, and *Angomonas deanei*) possess non-overlapping internal insertions of different lengths (20, 6, and 2 aa, respectively). The sequence of TMP18e was distinct from all other homologs by the presence of a 25 aa-long N-terminal extension. At the same time, it was shorter at the C terminus than most TMP18 proteins of other trypanosomatids. We employed TMHMM and Phobius to predict the transmembrane helices (TMHs), inherent to the abovementioned protein family. These programs detected two and three such helices, respectively. The first (N-terminal) TMH1 has also been detected by TMHMM but was rejected due to the low estimated probability (~0.4). The estimated borders of the TMH predicted by both programs showed significant overlap (Figure S1).

A comparison of the predicted tertiary structures of the two *N. esmeraldas* TMP18 variants to each other and to orthologs of *Leishmania infantum* and *Trypanosoma cruzi* revealed a high level of structure conservation. All four proteins feature three TMH (Figures 2B and S1), with TMH3 being longer than the other two. For *N. esmeraldas*, we estimated that in the TMP18e and TMP18 proteins, this segment should extend beyond the membrane bilayer by approximately 3 and 6 nm, respectively. The shortened C terminus of TMP18e is remarkable since in the majority of other trypanosomatids there is a hypervariable region followed by a charged tail. Interestingly, the same shortening as in TMP18e is observed in TMP18 of the most divergent member of the family Trypanosomatidae—*Paratrypanosoma confusum*. *Strigomonas galati* also features a deletion in the C terminus, but the charged tail with the characteristic motif SKKXQ is preserved, although misaligned by MAFFT (Figure S1). Although conserved in other proteins considered here, the loop between TMH2 and TMH3 in TMP18e is divergent and predicted to contain a single α-helical turn (Figures 2B and S1). All four proteins have a long N-terminal segment, which in *Novymonas* and *Leishmania* consists of three helices, whereas in *T. cruzi*, the second and third helices are joined. TMP18e bears a unique N-terminal extension, which does not have any predicted secondary structure (Figure 2B).

TMP18e co-localizes with bacterial endosymbionts in *N. esmeraldas*

N. esmeraldas has two distinct developmental stages, which can be observed *in vitro* (Figure 3). The young culture is dominated by attached cells forming rosettes, which in the process of their growth produce the swimming stage. Although the density of both types increases, the attached stages are restricted by the available surface, while the swimmers have at their disposal the whole column of the medium and start to prevail at about day 4 (Figure S2). The swimming stage is represented by a continuum of shapes ranging from short ones, reminiscent of typical choanomastigotes that appear first in the culture (hereafter called "early swimmers") to typical elongated promastigotes ("late swimmers") observable after day 4. The latter stage develops from the early swimmers and becomes prevalent later on (Figures 3 and S2). The main proliferating cell type is the attached stage.

To investigate the subcellular localization of TMP18e, we endogenously tagged the protein with HA_3 in endosymbiont-containing and aposymbiotic cell lines, creating strains TMP18e-HA ES+ and TMP18e-HA ES−, respectively (Figure S3). The mRNA level of the tagged *tmp18e* gene in the TMP18e-HA ES+ strain was about 2-fold higher than that in the wild type (WT), due to distinct untranslated regions (from *dhfr* and *tmp18e* genes, respectively), while a further ~2-fold increase was observed in the aposymbiotic cells (Figure 4A). The difference in RNA levels was reflected in the expression of the tagged protein in TMP18e-HA ES+ and TMP18e-HA ES− strains (Figure 4B). Regardless of the strain, the swimming and attached stages showed similar *tmp18e* mRNA levels (Figure 4A), but the corresponding protein was considerably more abundant in the latter cells (Figure 4B) implying that its expression in *N. esmeraldas* is regulated mainly post-transcriptionally.

Immunofluorescence microscopy demonstrated co-localization of TMP18e with bacteria in the attached and swimming stage. This co-localization was tighter in the former and looser in the latter cell stage, especially in the late swimmers, whereas the distribution of fluorescent signal became more uniform in both stages upon ablation of bacteria (Figure 4C). Importantly, according to the TrypTag database (www.tryptag.org), in *Trypanosoma brucei* the TMP18 protein is associated with the ER.[24] This entailed an inevitable question about the subcellular localization of bacteria within the cell of *N. esmeraldas*.

Endosymbionts localize in the perinuclear space and ER lumen

In the attached stage of *N. esmeraldas*, the endosymbionts typically surround the nucleus, whereas in the swimming stage, they drift away, usually not farther than one nucleus diameter, and become oriented either longitudinally or obliquely (Figure 4C).

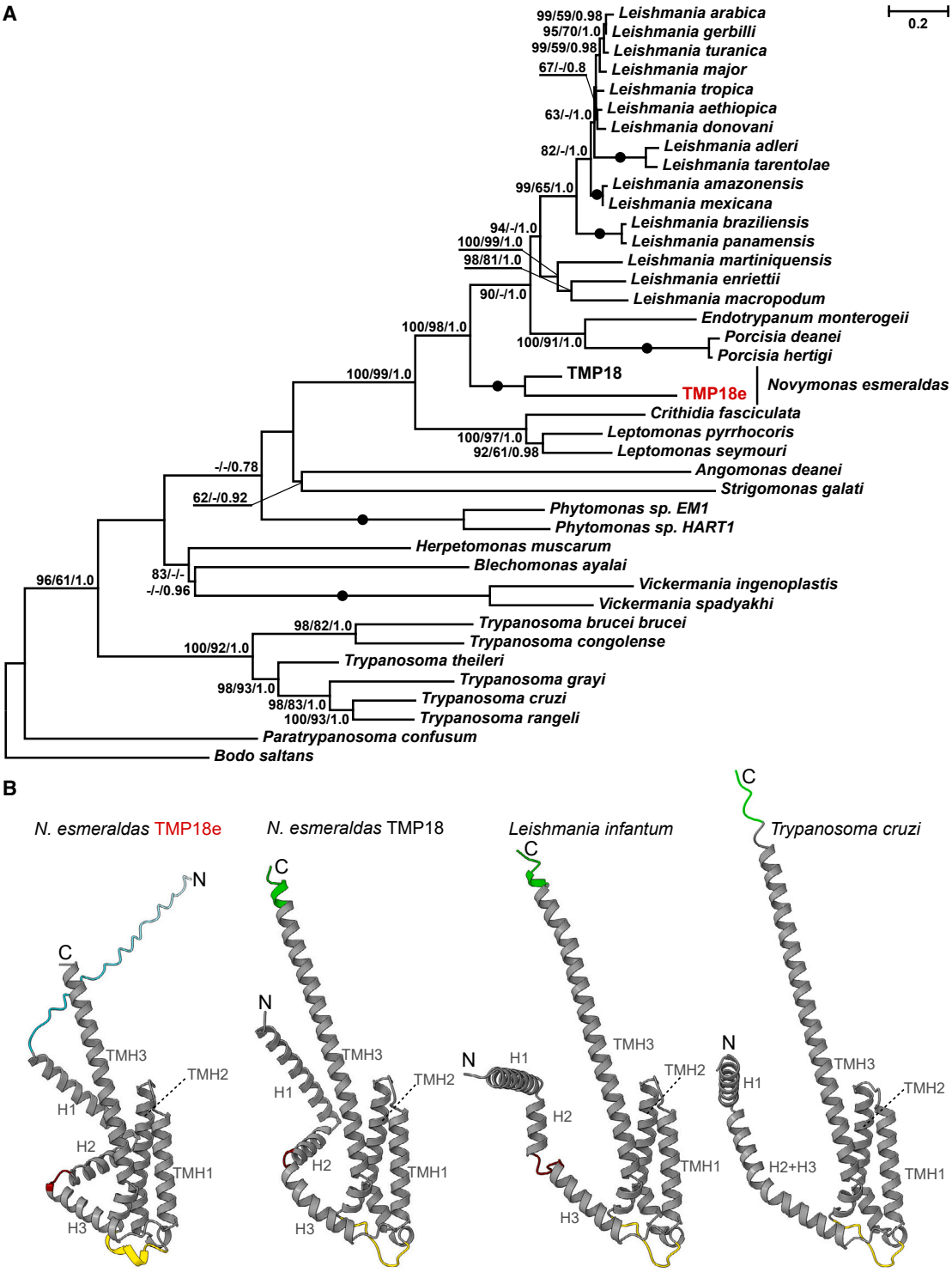


Figure 2. TMP18 proteins in Trypanosomatidae

(A) Maximum likelihood phylogenetic tree. The numbers at branches indicate ultrafast bootstraps/standard bootstraps/approximate Bayes test with values below 50 or 0.5 replaced with dashes or omitted. Black circles stand for absolute support by all methods. Scale bar represents the number of substitutions per site. TMP18e is in red. Please note that the evolutionary rates between codon positions are considerably different, explaining the inferred lengths of some branches over 1.

(B) AlphaFold2-predicted structures of TMP18e and TMP18 of *N. esmeraldas*, as well as TMP18 of *L. infantum* and *T. cruzi*. Labels H and TMHs denote helices and transmembrane helices, respectively. The loops between TMH2 and TMH3, and H2 and H3 are in yellow and red, correspondingly. The unique N-terminal extension of *N. esmeraldas* TMP18e is in turquoise.

See also Figure S1.

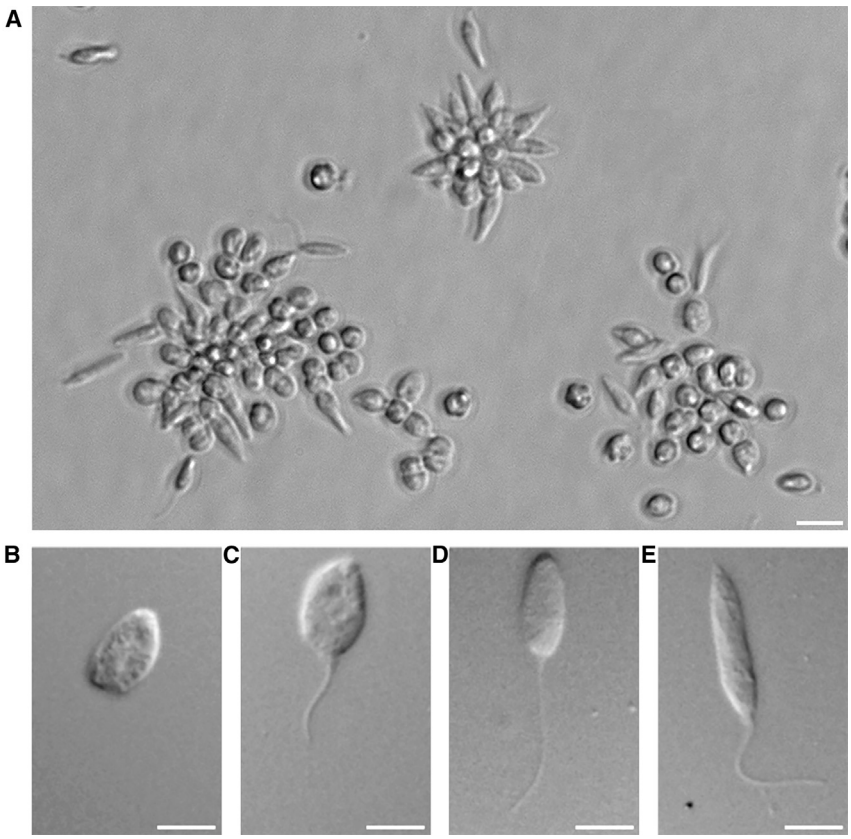


Figure 3. Cell morphotypes in the culture of *N. esmeraldas*

(A) Overview of an early stage of culture growth in a cultivation flask (integrated modulation contrast) showing single and grouped in rosettes attached cells, as well as free swimmers.

(B–E) Close-up view of different cell types isolated from the culture (DIC). (B) single attached cell; (C–E) different variants of swimming cells from the early to the late ones. Scale bars: 10 μm in (A) and 5 μm in (B)–(E).

See also Figure S2.

Examination with transmission electron microscopy (TEM) revealed that in the attached stage, bacteria are localized within the ER lumen, connected to the perinuclear space by short tubule(s) or directly in the perinuclear space (Figures 5A and 5B). In the swimming stage, where bacteria are positioned farther from the nucleus (Figure 5C), these relationships are more difficult to establish. However, the continuity of ER membrane with that of the symbiontophorous vacuole (Figure 5D) and the association of ribosomes with the latter (Figure 5E) indicate that in these cells bacteria still reside within the ER lumen. Thus, the subcellular localization of the endosymbionts coincides with that of the prototypic TMP18 protein, a trait likely inherited by the divergent TMP18e.

Ablation of TMP18e changes the number and distribution of endosymbionts in *N. esmeraldas*

To investigate the role of the TMP18e protein, two additional strains have been established: one, where the corresponding gene was completely ablated (TMP18e-KO, hereafter abbreviated as KO), and another, where it was added back on the KO background (TMP18e-AB, hereafter abbreviated as AB). A correct integration and expression have been confirmed by PCR, southern and western blotting (Figure S4). The comparison of these strains with the WT revealed that while their overall morphology did not change, there were significant differences in the growth rates (Figure S4E), the number of endosymbionts per cell, as well as in their intracellular distribution (Figure 6). Importantly, the pattern and extent of differences were stage-dependent. In the attached stage, the average number of bacteria per host cell substantially increased after the gene ablation (from 3.67 to 6.60 on average) (Figures 6A and 6C). Moreover, while in the WT, the attached stage predominantly contained even numbers of endosymbionts (2, 4, 6, and 8, typically observed in pairs), in the KO strain, a much smoother distribution was accompanied by a higher variance and increased proportion of bacteria-free flagellates (from 2% to 18%) (Figure 6C). In the swimming stage, the average number of endosymbionts did not significantly change after the gene ablation (4.58 versus 4.67), whereas their distribution showed higher skewness, and the proportion of aposymbiotic cells increased from 6% to 20% (Figures 6B and 6C). In both life stages, the gene add-back resulted in less skewed distribution and decreased proportion of bacteria-free cells (Figures 6C and 6D). Nevertheless, the overall shape of the distribution was still significantly different from that in WT.

Another consequence of the *tmp18e* gene ablation was the increased distance between the endosymbionts and the host nucleus. This was documented for both stages: in the attached one, the averaged distances of bacteria from the nucleus were 0.79, 1.05, and 0.73 μm for the WT, KO, and AB cells, respectively, whereas for the swimmers, the corresponding numbers were 1.43, 1.81, and 1.40 μm (Figure 6D). We also confirmed this observation using 3D serial block-face scanning electron microscopy (SBF-SEM) for the swimming stage (Figure S5). The observed effect in the swimmers was partially blurred because of the mixed nature of this cell category. It appeared to be more pronounced in the late swimmers, but in the absence of strict criteria for discriminating them from the early ones, it was not possible to show this numerically.

In the WT, both daughter cells obtain equal numbers of endosymbionts, whereas in the KO strain, the mechanism ensuring this equal distribution is disrupted, leading to heterogeneous numbers of bacteria in the progeny (Figure 6E). In the AB strain, both equal and unequal distribution of bacteria between products of cell division can be observed (Figure 6E, left and right subpanels, respectively). This ambiguity explains why for this strain the plots in Figure 6C have shapes intermediate between those for WT and KO.

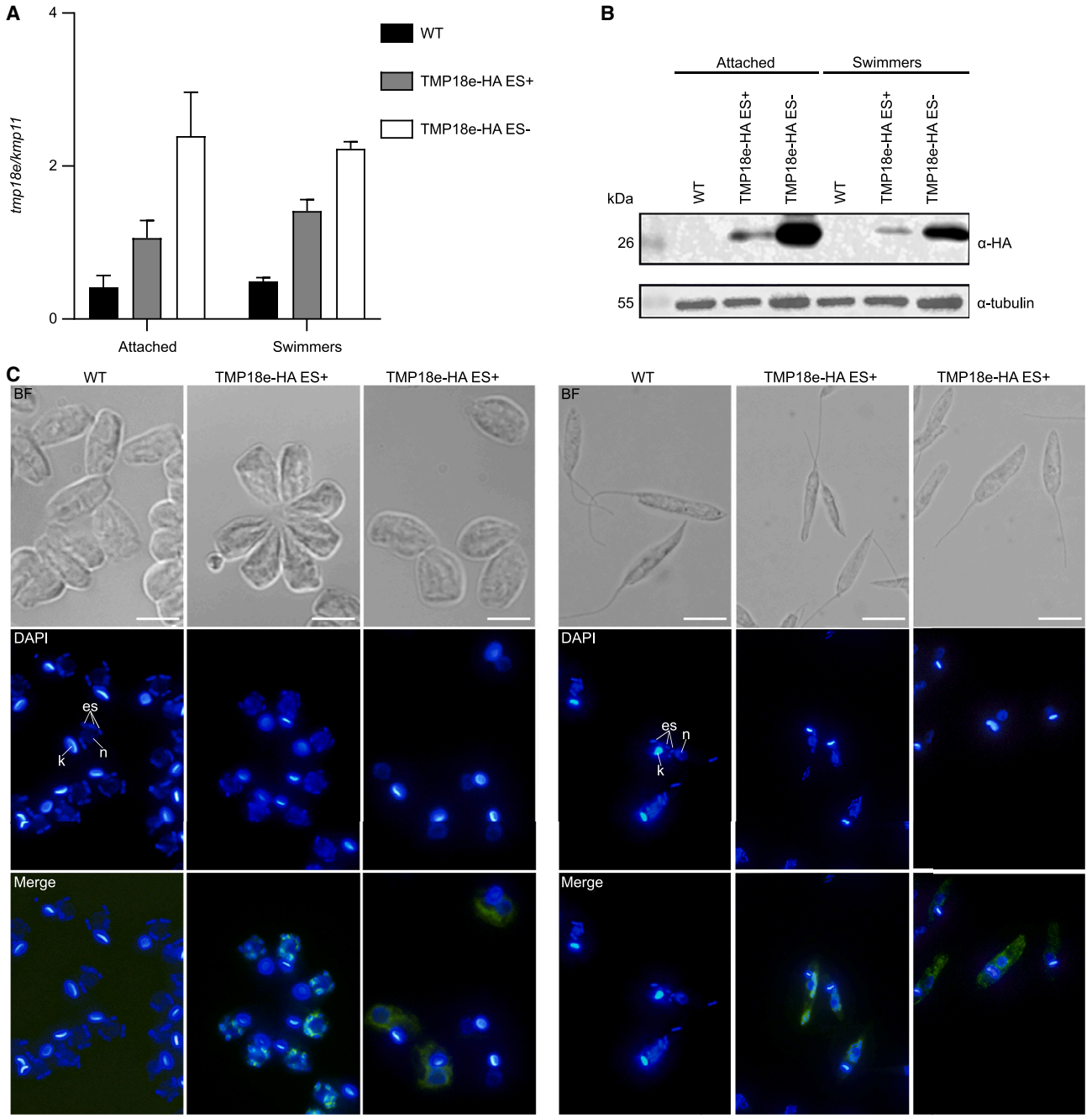


Figure 4. TMP18e localization in *N. esmeraldas*

(A) Real-time quantitative PCR analysis of recombinant *tmp18e* expression in endosymbiont-positive (ES+) and -negative (ES−) *N. esmeraldas* cells and comparison to *tmp18e* expression in WT cells. Data represent averages and standard deviations of the three independent biological replicates.

(B) A representative western blot with anti-HA antibodies (α-HA) confirming TMP18E-HA expression in endosymbiont-positive and -negative *N. esmeraldas* cells. Tubulin served as a loading control.

(C) TMP18e localization in the attached (left) and swimming (right) WT, TMP18e-HA ES+, and TMP18e-HA ES− *N. esmeraldas* cells analyzed by immunofluorescence assay (IFA). Images of bright field (BF), 4′,6-diamidino-2-phenylindole (DAPI)-staining of DNA (DAPI), and the merged (Merge) blue and green fluorescent channels representing DAPI and IFA using α-HA antibody, respectively, are presented in the corresponding rows. Scale bars, 5 μm. Labels "es," "k," and "n" define endosymbiont, kinetoplast, and nucleus, respectively.

See also Figure S3.

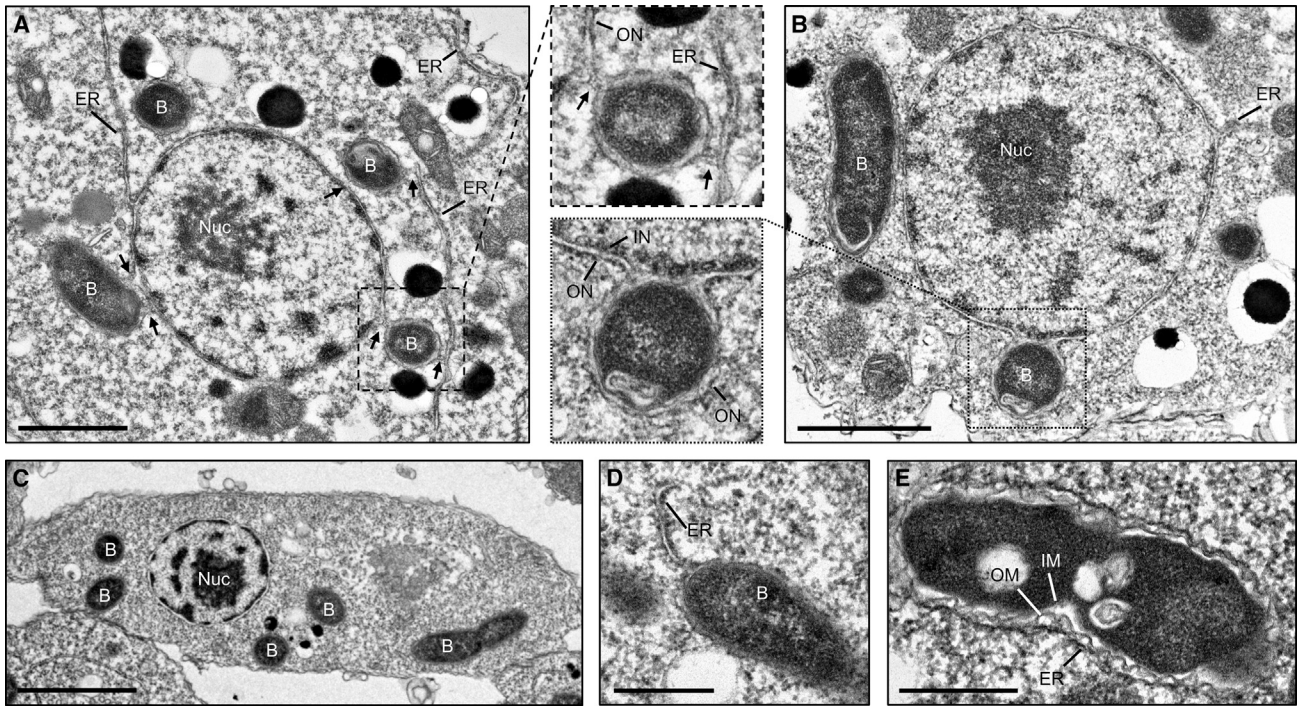


Figure 5. Transmission electron micrographs of *N. esmeraldas*

(A) Attached cell with bacteria (labeled B) localized in the endoplasmic reticulum (labeled ER), whose lumen is connected to the perinuclear space by short tubules (arrows).

(B) Attached cell with a bacterium inside the perinuclear space between the outer (labeled ON) and inner (labeled IN) nuclear membranes.

(C) Swimming cell with bacteria distant from the nucleus.

(D) ER tubule connected to the membrane of the symbiontophorous vacuole in a swimming cell.

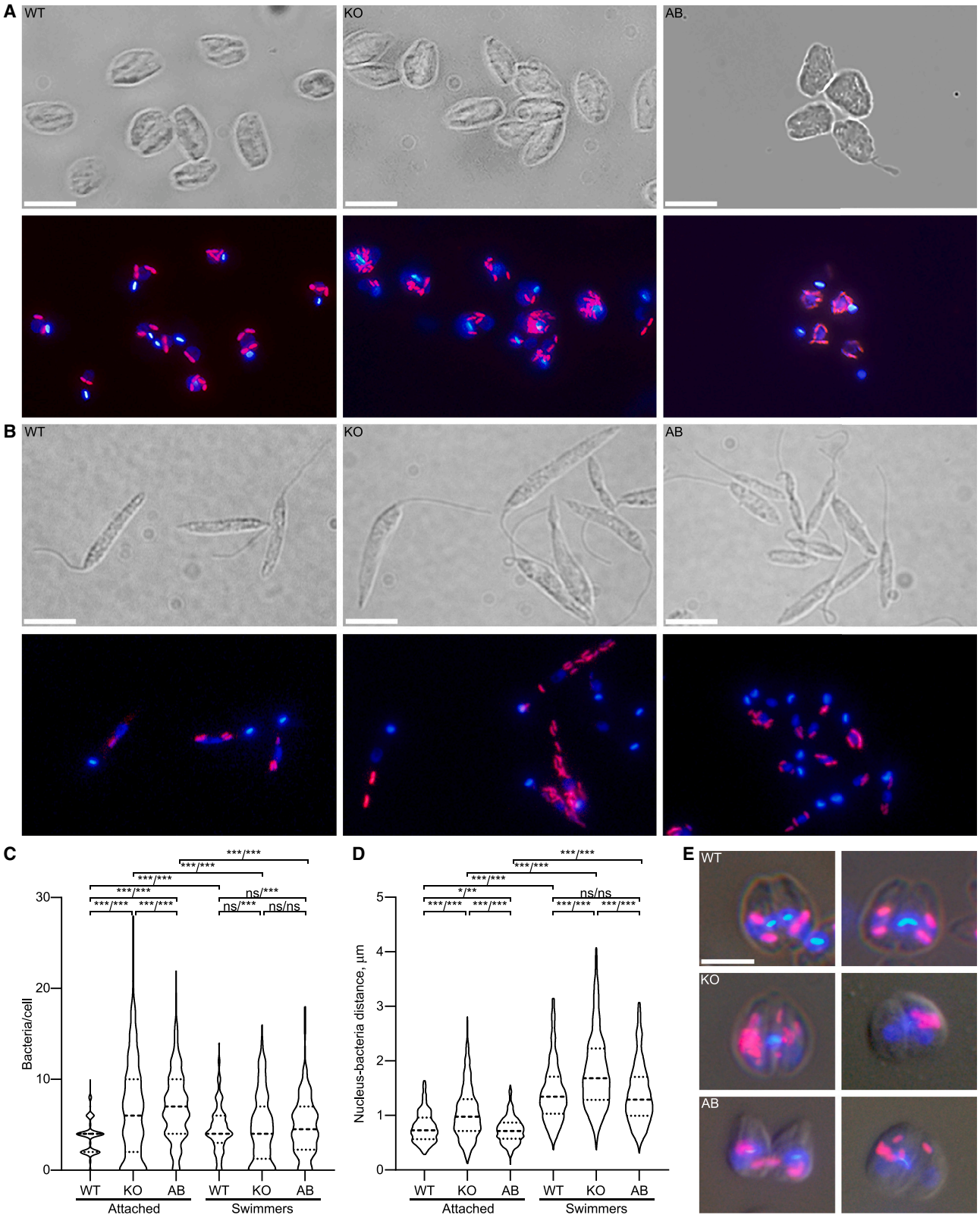
(E) Three membranes surrounding the endosymbiont: inner (labeled IM) and outer (labeled OM) bacterial membranes and that of the rough ER. Scale bars: 1 μm in (A) and (B), 2 μm in (C), and 500 nm in (D) and (E).

DISCUSSION

In this work, we aim to provide novel insights into the largely unexplored molecular interactions of a genomically reduced bacterial endosymbiont with its protistan host. Previous studies of the association between the trypanosomatid *N. esmeraldas* and its endosymbiont *Ca.* P. novymonadis revealed their tight metabolic cooperation and suggested that there was a certain control over the number of bacteria per host cell.[21,22] The investigation of *N. esmeraldas in vitro* development with detailed characterization of the attached and swimming stages presented here greatly refined this picture. Our results indicate that the attached stage plays a key role in the control of the bacterial cell number, as evidenced by the fact that over 80% of the host cells carry two or four symbionts (likely representing optimal numbers). Such a pattern is consistent with the clonal proliferation of the attached cells in rosettes, with the endosymbionts being evenly distributed between the daughter cells during each binary division. This leads to the formation of clonal colonies, all cells of which have equal numbers of bacteria. By contrast, the swimming stage serves for dispersal and, therefore, does not divide but grows in size. The increased variability of the bacterial load in this stage is explained by the loss of control upon the division of the endosymbionts, as well as their digestion by lysosomes, which has been demonstrated before.[21] The latter process also leads to the appearance of disadvantageous bacteria-free phenotype.

The correct segregation of bacteria during the attached stage cell division is likely facilitated by their localization in the nuclear vicinity (in the perinuclear space or in the ER lumen continuous with it). In this respect, it is important to note that trypanosomatids undergo closed mitosis, with their nuclear envelope remaining intact throughout the whole-cell cycle.[25] This enables the nucleus to serve as a guide for bacterial cell division. A similar situation has been documented in another trypanosomatid, *Strigomonas culicis*, but its endosymbiont lies freely in the host cytoplasm, and there is no information about the underlying molecular mechanism.[26]

To explore the intricate host-endosymbiont interactions in *N. esmeraldas* at the molecular level, we delineated 144 nucleus-encoded proteins that are enriched in the endosymbiont fractions. This large number probably reflects, at least partially, the tight endosymbiont-ER association, as the ER represents the main hub for membrane contact sites with other cellular compartments.[27] It is, therefore, not surprising that a considerable proportion of proteins that co-purified with the bacteria were predicted to come from mitochondria or peroxisomes, as fragments of these organelles might remain attached to the endosymbiont-surrounding ER membranes and become enriched in the corresponding fraction. In addition, the overall dominance



(legend on next page)

of peroxisomal and mitochondrial contaminants over typical ER proteins in the endosymbiont fraction may reflect one or both of the following: (1) similar sedimentation behavior of peroxisomes and mitochondrion-derived fragments with the endosymbionts and (2) the fact that these proteins are synthesized in the ER and only then transported to those specific organelles. Other proteins in this dataset are likely involved in mediating host-endosymbiont interactions and, thus, represent priority targets for further experimental studies.

One of the identified proteins, TMP18e, harbors an interesting hallmark: it originated as a result of the unique *N. esmeraldas*-specific duplication of a gene coding for the conserved TMP18 protein present in all trypanosomatids. TMP18 is homologous to TMEM18, which is known for its role in the regulation of carbohydrates and lipids storage in animals and, in particular, body weight control in mammals.[28–30] In all organisms investigated in this respect, TMEM18-encoding gene is either invariably single-copy (most opisthokonts, plants, and trypanosomatids) or completely lost (some fungi).[28] In metazoan cells, the TMP18 protein was shown to be localized to the nuclear membrane, where it appears to interact with several components of the nuclear pore complex.[29] In this work, we decided to use a separate name TMP18 instead of TMEM18 since we do not have any evidence that this protein in trypanosomatids has the same function as in other phylogenetically remote groups of eukaryotes.

Significant changes in the TMP18e sequence and structure suggested its *neo*-functionalization associated with the presence of endosymbionts. The observed fluorescence signal of the recombinant TMP18e around endosymbionts supports its involvement in the host-symbiont interaction. The most parsimonious assumption is that TMP18e is localized in the ER membrane and mediates, via its unique N-terminal extension, interaction with the endosymbiont. Indeed, as judged by structural predictions, the N-terminal extension faces the lumen and may therefore be responsible for the direct contact with the intraluminal bacteria allowing the host to sense and position the endosymbionts.

The disruption of the bacteria-nucleus association following TMP18e ablation leads to an unguided and, therefore, frequently uneven distribution of the endosymbionts during host cell division. As a likely consequence, the attached stage of the *tmp18e* KO strain demonstrates an elevated dispersion of variants and a significantly increased proportion of the aposymbiotic flagellates. We assume that the lack of complete functional rescue, observed in the AB cell line, could be caused by the differences in expression levels of TMP18e and/or interference with the introduced antibiotic-resistance genes. Importantly, we observed that the median number of endosymbionts increased in the attached stage of both KO and AB strains. It would be tempting to conclude that the proximity to the nucleus ensured by TMP18e in the attached stage is also responsible for the coordinated host and endosymbiont divisions, but there are two facts favoring an alternative explanation. It has been reported previously that the number of endosymbionts increased upon application of antibiotics (ampicillin, kanamycin, or chloramphenicol) at high concentrations, which inhibited the growth of flagellates.[21] This implies that bacteria can divide independently of the host. Consequently, there is no strict mechanism controlling bacterial division. Similarly, the increase in bacterial load in KO and AB strains can be explained by their slower growth rates as compared with the WT (Figure S4E). Therefore, we believe that TMP18e governs bacterial segregation, but not coordination of their division with that of the host.

The differentiation from the attached to the swimming stage is accompanied by reduced TMP18e expression correlating with a looser association between the endosymbionts and the nucleus (Figure 7). This is reminiscent of the situation in the attached cells of the KO strain, where the segregation of bacteria into daughter host cells also becomes uncontrolled. The *tmp18e* ablation does not influence the median endosymbiont number per cell in the swimming stage, but, as in the attached cells, it leads to a significant increase of variance, with a substantially elevated proportion of aposymbiotic cells. Nevertheless, as compared with the attached stage, the effect is less pronounced.

The differential expression of one particular gene and the resulting strength of the endosymbiont division control depending on the cell type is remarkable. Why do the non-proliferating swimming cells loosen this control posing a risk to the stability of symbiotic association? Further studies of the life cycle of *N. esmeraldas* and its relationship with the insect host(s) will certainly shed light on this question. We can speculate that bacteria may be used as a food reserve if the swimmers find themselves in a nutrient-poor environment. The dynamics of bacterial proliferation and their digestion creates heterogeneity ensuring that substantial proportion of the cells possesses an optimal bacterial number, although in some flagellates, the bacterial load is elevated. Cells of the first kind can immediately attach and start new colonies, whereas the second kind may represent survivors able to colonize new insect hosts.

In conclusion, our study illuminates an early step in the control of the vertical inheritance of endosymbionts, which stabilizes the symbiotic partnership between a protist and its bacteria. Here, for the first time, we show that the subcellular localization of the bacteria is governed in a stage-specific manner by a novel protein that originated via duplication of a

Figure 6. Role of TMP18e in controlling the number and position of endosymbionts

(A and B) Localization of the bacterial endosymbionts in the attached (A) and swimming (B) stages of *N. esmeraldas* WT, KO, and AB strains. Blue channel: DAPI-stained DNA; red channel: Cy3-labeled Eub338 probe against the bacterial 16S rRNA visualized following fluorescent *in situ* hybridization (FISH). The bright field image and a merge of the red and blue fluorescence channels are presented separately. Scale bars, 5 μm.

(C) Comparison of distributions of the number of bacteria per host cell.

(D) Comparison of distributions of the distance between the host nucleus and bacteria. Asterisks show significant pairwise differences (*$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$), ns, not significant. The value before and after the dash correspond to Mann-Whitney and F tests, respectively. Dashed line—median, dotted lines delineate the interquartile range.

(E) Distribution of endosymbiont between the daughter cells during the division of WT, KO, and AB strains of *N. esmeraldas* (two examples for each strain).

See also Figures S4 and S5.

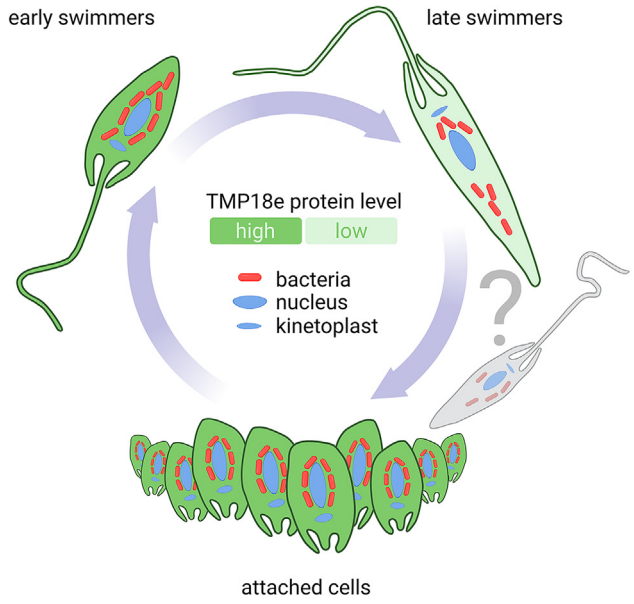


Figure 7. Stages of *N. esmeraldas* life cycle observed *in vitro*
The rosettes-forming attached cells have a higher expression of TMP18e; the level of this protein in swimming cells gradually decreases during the transition from the "early" to the "late" stage (defined by the color tonality). The question mark refers to the proportion of swimmers that can reattach completing the life cycle.

conserved eukaryotic gene followed by a *neo*-functionalization of one of its copies. The spatial confinement of the endosymbionts in the proximity to the nucleus at the proliferative stage of the trypanosomatid is conducive to controlling their division and equal distribution during cytokinesis, resulting in their reliable vertical inheritance.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cultivation and growth curves
- METHOD DETAILS
  - Preparation of cell lysates and endosymbiont isolation
  - Proteome analysis
  - Analyses of TMP18 orthologs in Trypanosomatidae
  - Establishment of experimental strains
  - cDNA synthesis and RT-qPCR
  - Fluorescent in situ hybridization (FISH), immunofluorescence assay (IFA), and microscopy
  - 3D serial block-face scanning electron microscopy (SBF-SEM)
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2023.04.060.

ACKNOWLEDGMENTS

This work was supported by the Grant Agency of the Czech Republic (project 20-07186S to V.Y. and J.L.), Czech Ministry of Education ERD Funds (projects OP VVV 16_013/0001775 to D.T., J.T., and J.L. and 16_019/0000759 to V.Y., J.L., A.Y.K., A.Z., and A.S.), Czech-BioImaging project LM2018129, Deutsche Forschungsgemeinschaft (grants 1090/1-1 and 1090/1-2 to E.C.M.N.), and the Gordon and Betty Moore Foundation GBMF#9354 to J.L. We thank Marie Vancová and Martina Tesařová (Institute of Parasitology) for their assistance with advanced microscopy.

AUTHOR CONTRIBUTIONS

Conceptualization, V.Y., J.L., and E.C.M.N.; methodology, Alexandra Zakharova, D.T., and A.Y.K.; validation, A.B., J.M., G.P., A.Y.K., and K.S.; formal analysis, Alena Zakharova and A.Y.K.; investigation, Alexandra Zakharova, D.T., A.B., J.M., A.S., M.S., S.N., D.N., O.G., J.T., and A.Y.K.; data curation, V.Y., J.L., G.P., K.S., A.Y.K., and E.C.M.N.; writing – original draft, V.Y.; writing – review & editing, all authors; visualization, Alexandra Zakharova, D.T., A.S., and A.Y.K.; supervision, V.Y., J.L., and E.C.M.N.; project administration, V.Y.; funding acquisition, V.Y., J.L., and E.C.M.N.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 17, 2023
Revised: March 27, 2023
Accepted: April 25, 2023
Published: May 17, 2023

REFERENCES

1. Roger, A.J., Muñoz-Gómez, S.A., and Kamikawa, R. (2017). The origin and diversification of mitochondria. Curr. Biol. *27*, R1177–R1192.
2. Bhattacharya, D., Archibald, J.M., Weber, A.P., and Reyes-Prieto, A. (2007). How do endosymbionts become organelles? Understanding early events in plastid evolution. BioEssays *29*, 1239–1246.
3. Husnik, F., Tashyreva, D., Boscaro, V., George, E.E., Lukeš, J., and Keeling, P.J. (2021). Bacterial and archaeal symbioses with protists. Curr. Biol. *31*, R862–R877.
4. Nussbaumer, A.D., Fisher, C.R., and Bright, M. (2006). Horizontal endosymbiont transmission in hydrothermal vent tubeworms. Nature *441*, 345–348.
5. Gibson, K.E., Kobayashi, H., and Walker, G.C. (2008). Molecular determinants of a symbiotic chronic infection. Annu. Rev. Genet. *42*, 413–441.
6. Moran, N.A., McCutcheon, J.P., and Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. Annu. Rev. Genet. *42*, 165–190.
7. McCutcheon, J.P., and Moran, N.A. (2011). Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. *10*, 13–26.
8. Husnik, F., and McCutcheon, J.P. (2018). Functional horizontal gene transfer from bacteria to eukaryotes. Nat. Rev. Microbiol. *16*, 67–79.
9. Toft, C., and Andersson, S.G. (2010). Evolutionary microbial genomics: insights into bacterial host adaptation. Nat. Rev. Genet. *11*, 465–475.
10. Koga, R., Meng, X.Y., Tsuchida, T., and Fukatsu, T. (2012). Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. Proc. Natl. Acad. Sci. USA *109*, E1230–E1237.
11. Hongoh, Y., Sharma, V.K., Prakash, T., Noda, S., Toh, H., Taylor, T.D., Kudo, T., Sakaki, Y., Toyoda, A., Hattori, M., et al. (2008). Genome of an

endosymbiont coupling N_2 fixation to cellulolysis within protist cells in termite gut. Science *322*, 1108–1109.

12. Muñoz-Gómez, S.A., Kreutz, M., and Hess, S. (2021). A microbial eukaryote with a unique combination of purple bacteria and green algae as endosymbionts. Sci. Adv. *7*, eabg4102.
13. Nowack, E.C.M. (2014). *Paulinella chromatophora* - rethinking the transition from endosymbiont to organelle. Acta Soc. Bot. Pol. *83*, 387–397.
14. Motta, M.C. (2010). Endosymbiosis in trypanosomatids as a model to study cell evolution. Open Parasitol. J. *4*, 139–147.
15. Kostygov, A.Y., Karnkowska, A., Votýpka, J., Tashyreva, D., Maciszewski, K., Yurchenko, V., and Lukeš, J. (2021). Euglenozoa: taxonomy, diversity and ecology, symbioses and viruses. Open Biol. *11*, 200407.
16. Lukeš, J., Butenko, A., Hashimi, H., Maslov, D.A., Votýpka, J., and Yurchenko, V. (2018). Trypanosomatids are much more than just trypanosomes: clues from the expanded family tree. Trends Parasitol. *34*, 466–480.
17. Zakharova, A., Saura, A., Butenko, A., Podešvová, L., Warmusová, S., Kostygov, A.Y., Nenarokova, A., Lukeš, J., Opperdoes, F.R., and Yurchenko, V. (2021). A new model trypanosomatid *Novymonas esmeraldas*: genomic perception of its "*Candidatus* Pandoraea novymonadis" endosymbiont. mBio *12*, e0160621.
18. Morales, J., Kokkori, S., Weidauer, D., Chapman, J., Goltsman, E., Rokhsar, D., Grossman, A.R., and Nowack, E.C. (2016). Development of a toolbox to dissect host-endosymbiont interactions and protein trafficking in the trypanosomatid *Angomonas deanei*. BMC Evol. Biol. *16*, 247.
19. Morales, J., Ehret, G., Poschmann, G., Reinicke, T., Maurya, A.K., Kröninger, L., Zanini, D., Wolters, R., Kalyanaraman, D., Krakovka, v, et al. (2023). Host-symbiont interactions in Angomonas deanei include the evolution of a host-derived dynamin ring around the endosymbiont division site. Curr. Biol. *33*, 28–40.e7.
20. Catta-Preta, C.M., Brum, F.L., da Silva, C.C., Zuma, A.A., Elias, M.C., de Souza, W., Schenkman, S., and Motta, M.C. (2015). Endosymbiosis in trypanosomatid protozoa: the bacterium division is controlled during the host cell cycle. Front. Microbiol. *6*, 520.
21. Kostygov, A.Y., Dobáková, E., Grybchuk-Ieremenko, A., Váhala, D., Maslov, D.A., Votýpka, J., Lukeš, J., and Yurchenko, V. (2016). Novel trypanosomatid – bacterium association: evolution of endosymbiosis in action. mBio *7*, e01985.
22. Kostygov, A.Y., Butenko, A., Nenarokova, A., Tashyreva, D., Flegontov, P., Lukeš, J., and Yurchenko, V. (2017). Genome of *Ca.* Pandoraea novymonadis, an endosymbiotic bacterium of the trypanosomatid *Novymonas esmeraldas*. Front. Microbiol. *8*, 1940.
23. Kostygov, A.Y., Frolov, A.O., Malysheva, M.N., Ganyukova, A.I., Chistyakova, L.V., Tashyreva, D., Tesařová, M., Spodareva, V.V., Režnarová, J., Macedo, D.H., et al. (2020). *Vickermania* gen. nov., trypanosomatids that use two joined flagella to resist midgut peristaltic flow within the fly host. BMC Biol. *18*, 187.
24. Dean, S., Sunter, J.D., and Wheeler, R.J. (2017). TrypTag.org: a trypanosome genome-wide protein localisation resource. Trends Parasitol. *33*, 80–82.
25. Martínez-Calvillo, S., Florencio-Martínez, L.E., and Nepomuceno-Mejía, T. (2019). Nucleolar structure and function in trypanosomatid protozoa. Cells *8*, 421.
26. Brum, F.L., Catta-Preta, C.M., de Souza, W., Schenkman, S., Elias, M.C., and Motta, M.C. (2014). Structural characterization of the cell division cycle in *Strigomonas culicis*, an endosymbiont-bearing trypanosomatid. Microsc. Microanal. *20*, 228–237.
27. Santos, H.J., and Nozaki, T. (2021). Interorganellar communication and membrane contact sites in protozoan parasites. Parasitol. Int. *83*, 102372.
28. Almén, M.S., Jacobsson, J.A., Shaik, J.H., Olszewski, P.K., Cedernaes, J., Alsiö, J., Sreedharan, S., Levine, A.S., Fredriksson, R., Marcus, C., et al. (2010). The obesity gene, TMEM18, is of ancient origin, found in majority of neuronal cells in all major brain regions and associated with obesity in severely obese children. BMC Med. Genet. *11*, 58.
29. Larder, R., Sim, M.F.M., Gulati, P., Antrobus, R., Tung, Y.C.L., Rimmington, D., Ayuso, E., Polex-Wolf, J., Lam, B.Y.H., Dias, C., et al. (2017). Obesity-associated gene TMEM18 has a role in the central control of appetite and body weight regulation. Proc. Natl. Acad. Sci. USA *114*, 9421–9426.
30. Wiemerslage, L., Gohel, P.A., Maestri, G., Hilmarsson, T.G., Mickael, M., Fredriksson, R., Williams, M.J., and Schiöth, H.B. (2016). The *Drosophila* ortholog of TMEM18 regulates insulin and glucagon-like signaling. J. Endocrinol. *229*, 233–243.
31. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. *47*, D442–D450.
32. Grube, L., Dellen, R., Kruse, F., Schwender, H., Stühler, K., and Poschmann, G. (2018). Mining the secretome of C2C12 muscle cells: data dependent experimental approach to analyze protein secretion using label-free quantification and peptide based analysis. J. Proteome Res. *17*, 879–890.
33. Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.
34. Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA *98*, 5116–5121.
35. Fukasawa, Y., Tsuji, J., Fu, S.C., Tomii, K., Horton, P., and Imai, K. (2015). MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. Mol. Cell. Proteomics *14*, 1113–1126.
36. Teufel, F., Almagro Armenteros, J.J., Johansen, A.R., Gíslason, M.H., Pihl, S.I., Tsirigos, K.D., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat. Biotechnol. *40*, 1023–1025.
37. Almagro Armenteros, J.J., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson, A., and Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. Life Sci. Alliance *2*, e201900429.
38. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.
39. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.
40. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview, version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.
41. Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. Ser. *41*, 95–98.
42. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. *37*, 1530–1534.
43. Crotty, S.M., Minh, B.Q., Bean, N.G., Holland, B.R., Tuke, J., Jermiin, L.S., and Haeseler, A.V. (2020). GHOST: recovering historical signal from heterotachously evolved sequence alignments. Syst. Biol. *69*, 249–264.
44. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. *35*, 518–522.
45. Anisimova, M., Gil, M., Dufayard, J.F., Dessimoz, C., and Gascuel, O. (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst. Biol. *60*, 685–699.
46. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021).

Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589.

47. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682.
48. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci. *30*, 70–82.
49. Dean, S., Sunter, J., Wheeler, R.J., Hodkinson, I., Gluenz, E., and Gull, K. (2015). A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids. Open Biol. *5*, 140197.
50. Ishemgulova, A., Hlaváčová, J., Majerová, K., Butenko, A., Lukeš, J., Votýpka, J., Volf, P., and Yurchenko, V. (2018). CRISPR/Cas9 in *Leishmania mexicana*: a case study of LmxBTN1. PLoS One *13*, e0192723.
51. Ishemgulova, A., Kraeva, N., Hlaváčová, J., Zimmer, S.L., Butenko, A., Podešvová, L., Leštinová, T., Lukeš, J., Kostygov, A., Votýpka, J., et al. (2017). A putative ATP/GTP binding protein affects *Leishmania mexicana* growth in insect vectors and vertebrate hosts. PLoS Negl. Trop. Dis. *11*, e0005782.
52. Záhonová, K., Hadariová, L., Vacula, R., Yurchenko, V., Eliáš, M., Krajčovič, J., and Vesteg, M. (2014). A small portion of plastid transcripts is polyadenylated in the flagellate *Euglena gracilis*. FEBS Lett. *588*, 783–788.
53. Tashyreva, D., Prokopchuk, G., Votýpka, J., Yabuki, A., Horák, A., and Lukeš, J. (2018). Life cycle, ultrastructure, and phylogeny of new diplonemids and their endosymbiotic bacteria. mBio *9*, e02447-02417.
54. Amann, R.I., Krumholz, L., and Stahl, D.A. (1990). Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. J. Bacteriol. *172*, 762–770.
55. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.
56. Durante, I.M., Butenko, A., Rašková, V., Charyyeva, A., Svobodová, M., Yurchenko, V., Hashimi, H., and Lukeš, J. (2020). Large-scale phylogenetic analysis of trypanosomatid adenylate cyclases reveals associations with extracellular lifestyle and host-pathogen interplay. Genome Biol. Evol. *12*, 2403–2416.
57. Yurchenko, V., Votýpka, J., Tesařová, M., Klepetková, H., Kraeva, N., Jirků, M., and Lukeš, J. (2014). Ultrastructure and molecular phylogeny of four new species of monoxenous trypanosomatids from flies (Diptera: Brachycera) with redefinition of the genus *Wallaceina*. Folia Parasitol. *61*, 97–112.
58. Belevich, I., Joensuu, M., Kumar, D., Vihinen, H., and Jokitalo, E. (2016). Microscopy Image Browser: a platform for segmentation and analysis of multidimensional datasets. PLOS Biol. *14*, e1002340.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Monoclonal Anti-HA antibody produced in mouse, clone HA-7, ascites fluid	MilliporeSigma	Catalog number H9658; RRID: AB_260092
Alexa FluorTM 488 goat anti-mouse IgG (H+L)	Thermo Fisher Scientific/ Invitrogen	Catalog number A11001; RRID: AB_2534069
Anti-Mouse IgG (whole molecule)-Peroxidase antibody produced in rabbit	MilliporeSigma	Catalog number A9044; RRID: AB_258431
Pierce anti-HA Magnetic beads	Thermo Fisher Scientific	Catalog number 88836; RRID: AB_2749815
Bacterial and virus strains		
XL1-Blue Supercompetent Cells	Agilent Technologies	Catalog number 200236
Chemicals, peptides, and recombinant proteins		
Brain heart infusion	MilliporeSigma	Catalog number 53286
Horse serum	MilliporeSigma	Catalog number H1270
OptiPrep – Iodixanol	Progen	Catalog number 1114542
Phleomycin	InvivoGen	Catalog number ant-ph-5
Azithromycin	MilliporeSigma	Catalog number 75199
Hemin	Jena Bioscience	Catalog number ML-108
Penicillin-Streptomycin Solution 100x	BioWest	Catalog number MS00YH100G
L-Biopterin	Cayman Chemical Company	Catalog number 10007662
Fetal Bovine Serum (South America)	BioSera	Catalog number FB-1001H
Deposited data		
Leishmania arabica LEM1108	TriTryp DB	release 52
Leishmania gerbilli LEM452	TriTryp DB	release 52
Leishmania turanica LEM423	TriTryp DB	release 52
Leishmania major Friedlin	TriTryp DB	release 52
Leishmania tropica L590	TriTryp DB	release 52
Leishmania aethiopica L147	TriTryp DB	release 52
Leishmania donovani BPK282A1	TriTryp DB	release 52
Leishmania adleri HO174	NCBI Assembly DB	GCA_902369305
Leishmania tarentolae Parrot-TarII	TriTryp DB	release 52
Leishmania amazonensis M2269	TriTryp DB	release 52
Leishmania mexicana U1103	TriTryp DB	release 52
Leishmania braziliensis M2904	TriTryp DB	release 52
Leishmania panamensis L13	TriTryp DB	release 52
Leishmania martiniquensis LEM2494	NCBI SRA DB	SRX5006816
Leishmania enriettii LV90	NCBI SRA DB	SRX5006814
Leishmania macropodum LV756	NCBI SRA DB	SRX5006815
Endotrypanum monterogeii ATCC 30507	NCBI Assembly DB	GCA_018683865
Porcisia deanei TCC258	NCBI Assembly DB	GCA_018683835
Porcisia hertigi TCC260	NCBI Assembly DB	GCA_019345635
Crithidia fasciculata Cf-Cl	TriTryp DB	release 52
Leptomonas pyrrhocoris H10	TriTryp DB	release 52
Leptomonas seymouri ATCC 30220	TriTryp DB	release 52
Angomonas deanei TCC036E	NCBI Assembly DB	GCA_000482225
Strigomonas galati TCC219	NCBI Assembly DB	GCA_000482125
Phytomonas sp. EM1	NCBI Assembly DB	GCA_000582765
Phytomonas sp. HART1	NCBI Assembly DB	GCA_000982615

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Herpetomonas muscarum TCC001E	NCBI Assembly DB	GCA_000482205
Blechomonas ayalai B08-376	TriTryp DB	release 52
Vickermania ingenoplastis CP021	NCBI Assembly DB	GCA_010157825
Trypanosoma brucei brucei TREU927	TriTryp DB	release 52
Trypanosoma congolense IL3000	TriTryp DB	release 52
Trypanosoma theileri Edinburgh	TriTryp DB	release 52
Trypanosoma grayi ANR4	TriTryp DB	release 52
Trypanosoma cruzi CL Brener Esmeraldo-like	TriTryp DB	release 52
Trypanosoma rangeli AM80	NCBI Assembly DB	GCA_003719475
Paratrypanosom confusum CUL13	TriTryp DB	release 52
Bodo saltans Konstanz	TriTryp DB	release 52
Novymonas esmeraldas E262	NCBI Assembly DB	GCA_019188245
Raw proteomic data	This paper	PXD019306
Experimental models: Organisms/strains		
Novymonas esmeraldas E262	Yurchenko Lab	N/A
Novymonas esmeraldas E262 aposymbiotic	Yurchenko Lab	N/A
Novymonas esmeraldas E262 TMP18e-HA ES+	Yurchenko Lab	N/A
Novymonas esmeraldas E262 TMP18e-HA ES-	Yurchenko Lab	N/A
Novymonas esmeraldas E262 TMP18e KO	Yurchenko Lab	N/A
Novymonas esmeraldas E262 TMP18e AB	Yurchenko Lab	N/A
Recombinant DNA		
Vector: pLEXSY-neo2.1	Jena Bioscience	Catalog number EGE-273
Plasmid: pLEXSY-neo2.1 - U6 promoter-sgRNA with tracrRNA-U6 terminator	This paper	N/A
Plasmid: pLEXSY-blast	This paper	N/A
Plasmid: pLEXSY-blast - *tmp18e*-HA	This paper	N/A
Plasmid: pLEXSY-blast - *tmp18e*-HA (5 mutations)	This paper	N/A
Software and algorithms		
MAFFT v. 7.475	Katoh Lab	https://mafft.cbrc.jp/alignment/software/source.html
trimAl v 1.2	Gabaldón Lab	http://trimal.cgenomics.org/trimal
BioEdit v. 7.2.5	BioEdit Company	https://bioedit.software.informer.com/
IQ-Tree v. 2.1.3	University of Vienna	http://www.iqtree.org
MaxQuant v. 1.6.0.1	Max Planck Institute of Biochemistry	https://www.maxquant.org/
AlphaFold 2	DeepMind	https://alphafold.ebi.ac.uk/
UCSF ChimeraX	UCSF	https://www.cgl.ucsf.edu/chimerax/
Olympus cellSens v.1.6	Olympus	https://www.olympus-lifescience.com/en/software/cellsens/
ImageJ v.1.51n	NIH	https://imagej.nih.gov/ij/notes.html
Microscopy Image Browser v. 2.702	MIB	http://mib.helsinki.fi/
Amira v. 2020.2	Thermo Fisher Scientific	https://www.thermofisher.com/
GraphPad Prism v. 9	GraphPad Software	https://www.graphpad.com/features
Other		
Orbitrap Elite	Thermo Fisher Scientific	N/A
QExactive Plus Mass Spectrometer	Thermo Fisher Scientific	Catalog number IQLAAEGAAPFALGMBDK
UltiMate™ 3000 Rapid Separation System	Thermo Fisher Scientific	Catalog number IQLAAAGABHFAPBMBFE

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Pre-column: Acclaim PepMapRSLC, C18, 3 μm particle size, 100 Å pore size, 75 μm inner diameter	Thermo Fisher Scientific	Catalog number 164535
Main column: Acclaim PepMapRSLC, C18, 2 μm particle size, 100 Å pore size, 75 μm inner diameter	Thermo Fisher Scientific	Catalog number 164941

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Vyacheslav Yurchenko (vyacheslav.yurchenko@osu.cz)

Materials availability

Plasmids and cell lines generated in this study are available upon request from authors.

Data and code availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE repository[31] with the dataset identifier PRIDE: PXD019306. All datasets generated in this study are publicly available as of the date of publication.

This paper does not report original code.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cultivation and growth curves

Novymonas esmeraldas isolate E262 was cultivated in Brain-Heart Infusion medium (VWR) supplemented with 2 μg/ml hemin (Jena Bioscience), 100 units/ml of penicillin and 100 μg/ml of streptomycin (both from Life Technologies/ Thermo Fisher Scientific) at 23 °C. The previously established aposymbiotic strain[22] was cultivated in M199 (MilliporeSigma) supplemented with 2 mg/ml hemin (Jena Bioscience), 10% heat-inactivated fetal bovine serum (FBS; BioSera Europe), 2 mg/ml biopterin, and antibiotics as above at 23°C. For growth curves, *N. esmeraldas* cells were seeded at 10^5 cells/ml and counted every day in three independent biological replicates using Neubauer chamber. In order to break cell rosettes before counting, the aliquots were passed three times through a 18G syringe needle.

METHOD DETAILS

Preparation of cell lysates and endosymbiont isolation

N. esmeraldas was grown in 300 ml BHI medium supplemented with 10% horse serum (Life Technologies) and 10 μg/ml hemin at 28 °C until late-log phase. After centrifugation, cells were washed in a buffer containing 25 mM Tris-HCl, pH 7.5, 150 mM sucrose, 20 mM KCl, 2 mM EDTA, and a cocktail of protease inhibitors (Roche Life Science, Penzberg, Germany) and lysed by sonication. A sample of the lysate was stored at -80°C until further use. Endosymbionts were purified from the lysate essentially as described in Morales et al.[18] with a slight modification; the iodixanol gradient consisted of 2 ml-steps of 25, 22.5, 20, and 17.5% and endosymbionts were collected from the 22.5/20% and the 20/17.5% interphases. Purity of the endosymbiont fractions was assessed by light microscopy. Protein in four biological replicates of each isolated endosymbionts (ES) and holosymbiont (HS) lysate, was subjected to liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS).

Proteome analysis

Sample preparation for LC-MS/MS was carried out essentially as described previously.[32] Briefly, 5 μg of protein lysates from four replicates of each, HS, ES and AS (aposymbiont) samples were shortly separated (~5 mm running distance) in a polyacrylamide gel. After silver staining, protein-containing bands were cut out of the gel, de-stained and proteins reduced with dithiothreitol, alkylated with iodoacetamide, and digested with trypsin. Tryptic peptides were extracted from the gel, dried, and finally resuspended in 0.1% trifluoroacetic acid. Five hundred ng of peptides were subjected to LC-based separation using a 2 h gradient ran in an Ultimate 3000 Rapid Separation LC system (RSCL, Thermo Fisher Scientific) equipped with a 25 cm long analytical column (Acclaim PepMapRSLC, 2 μm C18 particle size, 100 Å pore size, 75 μm inner diameter, Thermo Fisher Scientific). Separated peptides were injected directly into a Q Exactive plus hybrid quadrupole-orbitrap mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated in positive, data-dependent mode. Firstly, survey scans were carried out in the orbitrap analyzer at a resolution of 70,000. Secondly, up to ten 2-3-fold charged precursors were selected by the quadrupole using a 2 m/z isolation

window, fragmented *via* higher-energy collisional dissociation and fragment spectra recorded at a resolution of 17,500. Recorded mass spectra were further processed by MaxQuant v. 1.6.0.1[33] for peptide and protein identification and label-free quantification (LFQ) using standard parameters if not stated otherwise. 9,299 protein sequences from *N. esmeraldas* (NCBI BioProject accession: PRJNA681813) and 971 from "*Ca.* P. novymonadis" (GenBank accession MUHY00000000) were used as input for searches.[17,22] Protein N-terminal acetylation and methionine oxidation were considered as variable and carbamidomethylation of cysteines as fixed modifications. Label-free quantification was enabled, as well as the 'match between runs' functionality. Peptides and proteins were accepted at a false discovery rate of 1% and only proteins, which were identified based on at least 2 different peptides and three valid values in at least one of the investigated sample groups.

To detect host proteins enriched in the endosymbiont fraction, we carried out a Student's t-test based significance analysis[34] on $\log_2$ transformed label-free quantification (LFQ) intensity values. Only 1,215 out of 3,548 identified total proteins were selected for this based on the following criteria: detected with at least three valid values both in the endosymbiont and homogenate fraction using an S_0=0.8 and a false discovery rate of 5%. Out of 417 proteins were enriched in the endosymbiont fraction, 144 were host-encoded. For the resulting set we predicted putative proteins' localization based on their annotations and/or presence of signal peptides detected by MitoFates v. 1.2,[35] SignalP v. 6.0[36] and TargetP v. 2.0.[37] The TMP18e was selected for further analysis based on its high enrichment, no association with either glycosomes or mitochondria, and presence of a non-enriched homolog (Table S1).

Analyses of TMP18 orthologs in Trypanosomatidae

The search for homologs of TMP18 (hereafter, the protein and gene names are capitalized and italicized, respectively) was performed by BlastP against all proteins predicted in the genome of *N. esmeraldas* with the following thresholds: 10^{-20} for *e*-value and 0.5 for the ratio of alignment and query sequence lengths. The same was done for the available genomes of 36 additional Trypanosomatidae species and that of *Bodo saltans*, the closest known free-living relative of this family (see the key resources table).

The collected protein sequences were aligned using MAFFT v. 7.475 with L-INS-I algorithm and BLOSUM45 substitution matrix.[38] The alignment was then trimmed with trimAl v 1.2 in the "gappyout" mode[39] and visualized in Jalview v. 2.11.1.4.[40] BioEdit v. 7.2.5[41] was used to calculate sequence similarity. Phylogenetic tree was constructed using IQ-Tree v. 2.1.3.[42] with LG substitution matrix and GHOST model with four classes.[43] Branch support was estimated by approximate Bayes test and standard and ultrafast bootstrap methods with 1,000 replicates.[44,45]

The tertiary structures of TMP18 proteins in *N. esmeraldas* were predicted *de novo* by AlphaFold 2[46] using the ColabFold platform.[47] The structures of *Trypanosoma cruzi* and *Leishmania infantum* proteins from were downloaded from the AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk). Models were visualized in UCSF ChimeraX.[48]

Establishment of experimental strains

To investigate the localization, expression level, and potential function of the TMP18e (endosymbiont-targeted variant of TMP18) protein, four experimental strains were established.

I) TMP18e-HA ES+ (endosymbiont-bearing) strain was obtained by C-terminal tagging of the protein with triple hemagglutinin ($HA_{\times 3}$) using a conventional homologous recombination approach[17] and 600 bp- and 1,070 bp-long arms of homology for 3'- and 5'-, respectively (Figure S3). The neomycin resistance gene and dihydrofolate reductase intergenic region were amplified separately and then assembled by fusion PCR.[49] All primers used in this work are listed in Table S2. A total of 5×10^7 mid-log phase cells were electroporated with 5 μg of the purified construct using Nucleofector 2b (Lonza). Cells were recovered in the antibiotic-free BHI medium for 16 hours at 23 °C and subsequently selected for 10 days using 100 μg/ml neomycin (Thermo Fisher Scientific). Tagging and expression were confirmed by PCR, Western blotting with anti-HA antibodies, and RT-qPCR.

II) TMP18e-HA ES- (endosymbiont-negative) strain was derived from TMP18e-HA ES+ by azithromycin treatment as described before.[22]

III) For the KO (knockout) strain, *tmp18e* was ablated in the wild type (WT) cell line by replacing it with the phleomycin-resistance gene using the CRISPR/Cas9 method, as described previously.[17,50] The U6 promoter, sgRNA with tracrRNA, and the U6 terminator were amplified, fused, and cloned into pLEXSY-neo2.1 (Jena Bioscience). Gene deletion was confirmed by PCR (expected fragment sizes for wild type and knock-out are 735 and 1,760 bp, respectively), by RT-qPCR, and by Southern blotting with probes against *tmp18e* and phleomycin resistance genes (Figure S4). In total, 1×10^8 mid-log phase cells were electroporated with 6 to 8 μg of purified linearized DNA construct using ECM 630 Exponential Decay Wave Electroporator (BTX). Cells were recovered as above and selected for 3 weeks with 200 μg/ml neomycin and 650 μg/ml phleomycin (both from Life Technologies) for sgRNA integration and donor DNA integration, respectively.

IV) the add-back (AB) strain was established on the KO background *via* restoration of the ablated gene using homologous recombination.[17] To avoid interruption of the target gene by the CRISPR/Cas9 system, five synonymous mutations were introduced into the gRNA recognition sequence in *tmp18e* by site-directed mutagenesis (Figure S4). The blasticidin resistance gene was amplified and cloned into pLEXSY-neo2.1 (Jena Bioscience), replacing the neomycin resistance gene. The mutated *tmp18e*-HA was amplified and cloned into pLEXSY-blast. Trypanosomatids were transfected and selected as above, but with 500 μg/ml blasticidin instead of phleomycin. The expression of the restored gene was confirmed by Western blotting with anti-HA antibodies (Figure S4D) and RT-qPCR. For at least two passages before the experiments all produced strains were cultivated in the absence of the selective antibiotics to minimize potential by-effects of the latter.

Probes for *tmp18e* and phleomycin resistance gene were amplified using PCR DIG probe synthesis kit (Roche Life Science). For Southern blotting, 20 μg of DNA was digested with *Sac*I and *Spe*I, separated on a 0.75 % agarose gel, transferred onto a membrane and probed as described previously.[51]

cDNA synthesis and RT-qPCR

Total RNA was isolated with TRI Reagent (MilliporeSigma) and cDNA was synthesized with cDNA First Strand Synthesis Kit (Thermo Fisher Scientific) following the manufacturers' instructions. The expression of *tmp18e* was measured by RT-qPCR in the LightCycler480 (Roche Life Science) as described previously.[52] All experiments were performed in biological and technical triplicates, using kinetoplast membrane protein 11 (*kmp11*) expression for normalization.

Fluorescent in situ hybridization (FISH), immunofluorescence assay (IFA), and microscopy

Trypanosomatid cells were fixed for 30 min with 4% paraformaldehyde in phosphate buffered saline (PBS) at room temperature and processed as described elsewhere.[53] Bacterial endosymbionts were visualized by FISH using the standard Eubacteria-specific probe Cy3-Eub338.[54] The cells were stained with 4′,6-diamidino-2-phenylindole (DAPI) in Mounting Medium (Thermo Fisher Scientific) and observed in the Olympus microscope BX-53 (Olympus) equipped with the Olympus DP73 digital camera. Images were taken with Olympus cellSens v.1.6 software and then merged and analyzed in ImageJ v.1.51n.[55] For immunofluorescence microscopy, cells were fixed as above and visualized with anti-HA antibodies (MilliporeSigma) as specified previously.[56] The processing of samples for transmission electron microscopy (TEM) was described before.[57]

3D serial block-face scanning electron microscopy (SBF-SEM)

For SBF-SEM, cells were re-suspended in a modified Karnovsky fixative containing 2.5% glutaraldehyde, 2% paraformaldehyde and 2 mM $CaCl_2$ in 150 mM cacodylate buffer, pH 7.4, embedded in 2% agarose, and cut into small pieces. Samples were stained using the modified OTO method, dehydrated in an ascending acetone series, and embedded in the Hard Plus 812 Resin (Electron Microscopy Sciences). The blocks were trimmed and imaged using the Apreo 2 SEM scanning electron microscope equipped with the VolumeScope (Thermo Fisher Scientific). The microscope acquisition settings were set to 3.5 kV, 0.1 nA, and low vacuum (10, 20, and 50 Pa for the wild type, knock-out, and add-back, respectively), dwell time per pixel at 2 μs, and serial images resolution of 7.5 nm (X,Y) with 100 nm slice thickness (Z). Cells were processed, segmented, measured, and visualized in 3D using Microscopy Image Browser v. 2.702[58] and Amira v. 2020.2 (Thermo Fisher Scientific) software packages.

QUANTIFICATION AND STATISTICAL ANALYSIS

The statistical analysis was performed using GraphPad Prism v. 9 (GraphPad Software). The Kolmogorov–Smirnov test was used to determine whether values show normal distribution for bacterial load and distance between bacteria and nucleus. The pairwise comparisons of median and variance values of these distributions were performed using Mann-Whitney U-test and the Fisher's F-test, respectively.