



Bodo saltans (Kinetoplastida) is dependent on a novel *Paracaedibacter*-like endosymbiont that possesses multiple putative toxin-antitoxin systems

Samriti Midha¹ · Daniel J. Rigden² · Stefanos Siozios¹ · Gregory D. D. Hurst¹ · Andrew P. Jackson¹

Received: 22 July 2020 / Revised: 3 December 2020 / Accepted: 9 December 2020 / Published online: 15 January 2021
© The Author(s) 2021. This article is published with open access

Abstract

Bacterial endosymbiosis has been instrumental in eukaryotic evolution, and includes both mutualistic, dependent and parasitic associations. Here we characterize an intracellular bacterium inhabiting the flagellated protist *Bodo saltans* (Kinetoplastida). We present a complete bacterial genome comprising a 1.39 Mb circular chromosome with 40.6% GC content. Fluorescent in situ hybridisation confirms that the endosymbiont is located adjacent to the nuclear membrane, and a detailed model of its intracellular niche is generated using serial block-face scanning electron microscopy. Phylogenomic analysis shows that the endosymbiont belongs to the Holosporales, most closely related to other α -proteobacterial endosymbionts of ciliates and amoebae. Comparative genomics indicates that it has a limited metabolism and is nutritionally host-dependent. However, the endosymbiont genome does encode diverse symbiont-specific secretory proteins, including a type VI secretion system and three separate toxin-antitoxin systems. We show that these systems are actively transcribed and hypothesize they represent a mechanism by which *B. saltans* becomes addicted to its endosymbiont. Consistent with this idea, attempts to cure *Bodo* of endosymbionts led to rapid and uniform cell death. This study adds kinetoplastid flagellates to ciliates and amoebae as hosts of *Paracaedibacter*-like bacteria, suggesting that these antagonistic endosymbioses became established very early in Eukaryotic evolution.

Introduction

Eukaryotes commonly live in intimate associations with microbes [1]. For microeukaryotes, intracellular microbes live as endosymbionts passing to progeny cells following fission [2], whereas endosymbionts of multicellular species live within host tissues and pass to progeny during reproduction, commonly inside eggs [3]. Heritable host-microbe

interactions have arisen multiple times and involve diverse eubacteria and life strategies [4].

The impact of endosymbioses on host and microbe vary widely. In some cases, the host captures a microbe for its own benefit. For instance, *Paramecium* captures *Chlorella* algae from the environment, utilises them to facilitate a mixotrophic rather than heterotrophic lifestyle, but controls symbiont numbers such that *Chlorella* replication is higher when free-living than in the symbiotic state [5]. In other cases, the microbe can be parasitic. This can occur when the symbiont is biparentally inherited, or uniparentally inherited, when it distorts reproduction towards the transmitting sex, for example with *Wolbachia* [6]. Finally, and perhaps most commonly, symbiont and host mutually benefit from the interaction. For instance, symbionts such as *Buchnera* and *Euplotes* provide protective and nutritional benefits to their host that promote host survival and reproduction, and in doing so increase their fitness [7].

In many cases, host survival and reproduction depend on functions provided by the symbiont. For instance, loss of the anabolic capacity of *Buchnera* leaves their aphid host with insufficient essential amino acids that causes host

Supplementary information The online version of this article (<https://doi.org/10.1038/s41396-020-00879-6>) contains supplementary material, which is available to authorized users.

✉ Samriti Midha
S.midha@liv.ac.uk

¹ Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Ic2 Liverpool Science Park, 146 Brownlow Hill, Liverpool L3 5RF, UK

² Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Crown St., Liverpool L69 7ZB, UK

sterility [7, 8]. Host dependency may be due also to coadaptation of cellular and developmental processes [9]. The defensive *Burkholderia* symbiont of the fungus *Rhizopus* is required for completion of the host's sexual phase [10], a system mirroring the requirement of *Asobara tabida* wasps for one of their *Wolbachia* symbionts to complete oogenesis [11]. In these cases, coadaptation over deep evolutionary time to the presence of the symbiont means that the host fails when the symbiont is removed, irrespective of any benefit the symbiont might confer.

There is another potential route to dependency that arises through selection on the symbiont to addict the host to its presence. Addiction mechanisms using toxin-antitoxin systems are a common retention mechanism for plasmids in bacteria [12] and are also observed in a variety of selfish genetic elements, such as the peel-zeel system of *C. elegans* [13]. These systems typically function through delivery of a long-lived toxin molecule alongside an antitoxin with a shorter half-life. Should the element carrying the system not be inherited then the toxin becomes active and kills the non-bearer individual. Bacterial endosymbionts are currently not known to addict hosts through toxin-antitoxin systems, although they are known to exploit similar 'rescue' based systems in creating the phenotype of cytoplasmic incompatibility [13, 14].

In this study, we investigated the symbiosis between the flagellate *Bodo saltans* (Kinetoplastida) and its intracellular bacterial symbiont. *B. saltans* is a heterotrophic, phagocytic bacterivore found in freshwater and marine habitats worldwide, and is among the closest free-living relatives of trypanosomatid parasites [15]. Within the trypanosomatids, three parasitic lineages, *Novymonas*, *Phytomonas* and the Strigomonadinae, are known to contain a β -proteobacterial symbionts [16–19]. These symbionts are mutualistic, cooperating in various metabolic pathways that provide essential amino acids and vitamins to the host [20–22]. Although these endosymbionts are all mutualists, it is clear that parasitic kinetoplastids have acquired endosymbionts independently and relatively often [23]. *B. saltans* is a free-living kinetoplastid, but microscopic examination has established that it too possesses endosymbiotic bacteria [24–26], whose precise identity is unknown.

We characterised the cytoplasmic bacterium inhabiting *B. saltans*, which we name *Candidatus* Bodocaedibacter vickermanii gen. nov., sp. nov., using a combination of genome sequencing, epifluorescence microscopy and serial block-face scanning electron microscopy (SBF-SEM). We then examined the endosymbiont genome sequence for putative function and identified a type VI secretion system and three chromosomal operons that are predicted to function as toxin-antitoxin systems. We therefore tested whether *Bodo* was dependent on its symbiont, potentially driven by these toxin-antitoxin systems, showing that antibiotic treatment results in rapid death of *B. saltans*, consistent with an addiction hypothesis.

Material and methods

DNA sequencing and genome assembly

Bodo saltans strain Lake Konstanz cells were grown in 0.05% yeast extract in the presence of *Klebsiella pneumoniae* ATCC 13883 as prey. To exclude extracellular bacteria from the DNA extraction, cells were sorted by size using a BD FACS Aria cell sorter (Becton- Dickinson, USA). DNA was extracted from sorted cells using a Qiagen MagAttract HMW DNA Kit (Qiagen Inc., Germany). DNA quality was assessed using a Qubit fluorometer (Thermo Fisher Scientific, USA) and Nanodrop instrument (Thermo Fisher Scientific, USA). A 20Kb insert library was prepared and sequenced on the PacBio Sequel platform (Pacific Biosciences, USA). To assess *K. pneumoniae* contamination, reads were mapped using Bwa v0.7 to the published *B. saltans* genome sequence (CYKH01000000) and analysed with Samtools v0.1.18 [27]. DNA sequence assembly was carried out using Canu v1.5 [28]. To identify all distinct bacterial sequences present, rRNA reads were filtered using SortMeRNA v2.1 [29], which identified a single candidate endosymbiont. *De novo* assembly showed the presence of only one contig belonging to the same bacterium with terminal overlaps, indicating a complete circular chromosome. The endosymbiont genome sequence was circularized using the Amos package v3.1.0 [30] and constituent reads were mapped back to check for mis-assembly.

Endosymbiont genome annotation

Genome annotation was carried out using Prokka v1.12 [31] complemented with BLASTp and InterProScan matches obtained using BLAST2GO v5.0 [32]. Presence of prophage sequences were checked with PHAST [33]. Genes containing a canonical signal peptide were predicted with SignalP 4.1 [34]. Proteins with eukaryotic-like domains were identified with EffectiveELD [35]. Secretome P was used to predict genes encoding proteins secreted by non-classical pathway [36]. Membrane transporters were predicted using TransportTP [37] and transporters were classified using the BLAST tool of the Transport Classification DataBase [38]. HHsearch [39] was used to search for distant relationships between putative toxin-antitoxin system components and known structures at the Protein Data Bank (PDB) [40]; or protein families in the Pfam database [41]. Three iterations of jackhmmer [42] searching against the UniClust database [43] were used to generate the sequence profiles for comparison with database entries. 16S rRNA gene sequences were predicted using RNAmmer v1.2 and EzBioCloud was used for taxonomical identification of the bacterial lineage [44]. Prevalence of this bacteria in diverse ecological environments was searched using IMGNS webserver [45].

Phylogenetic analysis and comparative genomics

OrthoMCL v2.0.9 was used to cluster orthologous genes shared by the *Bodo* endosymbiont with 12 related alpha-proteobacterial genome sequences [46]. Protein sequences encoded by 187 conserved genes from the 14 bacterial genomes were aligned using ClustalW [47]. Prottest v3.0 was used to check for the optimal amino acid substitution model for phylogeny estimation [48]. An LG + I + G model was applied in RaxML to estimate a maximum likelihood tree with 1000 non-parametric bootstrap replicates [49]. BLAST-based Average Amino Acid Identity (AAI) values amongst the 13 genomes were calculated using AAI calculator [50]. MAPLE 2.3.0 was used to identify and compare the completeness of various metabolic pathways in these endosymbiont genomes [51]. Genomagle 2.3.2 was used to compare the metabolic pathways of endosymbiont and *B. saltans* [51].

Fluorescent in situ hybridisation (FISH)

The endosymbiont 16S rRNA gene sequence was integrated into the SILVA database using the web-based SINA aligner [52, 53]. This alignment was further merged with SILVA Ref NR 99 [54] using the ARB package and the probe design tool was used to select FISH probes specific for the endosymbiont genome [55]. The probes were tested for mismatch analysis using mathFISH, probeCheck and BLAST [56, 57]. *B. saltans* cells were fixed using 4% paraformaldehyde and spotted on 0.1% gelatin-coated slides. Slides were dehydrated with 50%, 80% and 100% ethanol consecutively before hybridisation with the probe (50 ng/ μ l) at 46 °C. The probe was labelled with cyanine dye (([CY3]CGAAGTGAAATCTACGTCTCCGT)) and hybridized with 15% formamide in three independent replicates. Counterstaining of the cells was achieved using VECTASHIELD Antifade Mounting Medium with DAPI.

Electron microscopy

Cells were fixed in 2.5% glutaraldehyde (Wt/Vol) in 0.1 M phosphate buffer (pH7.4) in Pelco Biowave (Ted Pella Inc., USA) and washed twice in 0.1 M PB before embedding in 3% agarose. Agarose embedded cell pellets were post-fixed and stained as described previously [58], except for use of 0.1% thiocarbohydrazide as a mordant. For TEM, after UA staining samples were embedded with TAAB medium Premix resin in silicone moulds and Beem capsules. Ultrathin serial section (70–75 nm) were cut on an UC6 ultra-microtome (Leica, Austria) collected on formvar coated copper grids, before viewing at 120KV in a FEI Tecnai G2 Spirit. Images were taken using a MegaView III

camera and multiple Image Alignment (MIA) was used to create a high-resolution overview of areas of interest.

For SBF-SEM, cell pellets were embedded with TAAB hard premix resin in plastic dishes. Excess resin was removed before the block was mounted onto a cryo pin, cell side up, using silver conductive epoxy. Targeted trimming created a block face of cells 500 μ m \times 500 μ m. Samples were painted and dissected as previously described [58]. To mitigate charge build up and maximize image quality, imaging conditions were as follows: low vacuum mode with a chamber pressure of 70 Pa. Low accelerating voltage (1.7 kV), dwell time per pixel (14 μ s), magnification (5040 \times) pixel size 5.4 nm in *x* and *y*, frame width 6144 \times 6144, section thickness 100 nm over 104 sections. Amira 6.5.0 was used to analyse the SBF-SEM images and generate a 3D model for the cell.

Attempt to cure the symbiont with antibiotic treatment

Besides *B. saltans*, two parasitic and asymbiotic kinetoplastids (*Trypanosoma theileri* and *Zelonina costaricensis* [59]), and one parasitic kinetoplastid with nutritional endosymbiont (*Novymonas esmeraldas* [16]) were treated with rifampicin (20 μ g/ml), alongside control populations. *T. theileri* and *Z. costaricensis* were chosen as representative organisms of Kinetoplastida certain to be without endosymbionts. Their inclusion in this analysis was to show that the rifampicin effect we observed on *Bodo*, is not a general feature of kinetoplastids. To avoid impacts of rifampicin on *Bodo* mediated through changes in bacterial prey abundance, a rif-resistant prey strain of *Klebsiella* was developed for these experiments. The number of kinetoplastid cells was counted at 0 h and after 24 h of treatment in non-treated and treated flasks. The experiment was repeated three times and statistical significance of changes in cell number between species were analysed using ratio paired *T*-test. This antibiotic treatment experiment was repeated using streptomycin (100 μ g/ml) on *B. saltans* (in the presence of strep-resistant *Klebsiella*) and *N. esmeraldas*. For these experiments, rifampicin (100 mg/ml) and streptomycin (100 mg/ml) stock solutions were prepared in DMSO and H₂O respectively and diluted in 0.05% yeast extract media.

Bodo saltans RNA-sequencing

B. saltans cells were treated with gentamicin (50 μ g/ml) before RNA extraction to get rid of extracellular bacteria. Total RNA was extracted from cells and rRNA depleted using RiboZero rRNA depletion kit (Epidemiology). A strand-specific library was constructed with the NEBNext Ultra Directional RNA library preparation kit (NEB, UK). Paired

end 2×150 bp sequencing was carried out on the HiSeq 4000 platform (Illumina, USA). Raw reads were mapped on to the endosymbiont genome sequence using bwa v0.7.12 and low quality read alignments were removed using samclip (<https://github.com/tseemann/samclip>) and default parameters. Sort read alignments were visualised with the IGV 2.8.6 tool, while strand-specific read quantitation was performed with the SeqMonk program (<https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/>). To account for the presence of low DNA contamination in our library, a Difference Quantitation correction was applied. Read counts for each gene were calculated as the read counts originated from the opposite strand (template strand) minus the read counts from the same strand (coding strand) and normalized to the length of each gene (RPK).

Results

A *Bodo saltans* endosymbiont of the Holosporales

A single PacBio Sequel sequencing run generated 697,281 reads; of these, ~78% mapped to the *B. saltans* genome. Multiple proteobacterial genomes were discovered among the assembled contigs (Supplementary Fig. 1). Analysis of prokaryotic rRNA reads identified four bacterial taxa; *Klebsiella*, *Cupriavidus*, *Delftia* (each 100% identical with environmental sequences and assumed to derive from prey bacteria in the cell culture), and an unknown bacterium with greatest sequence identity (98.6%) to uncultured bacterial sequences belonging to Family Paracaedibacteraceae, followed by *Paracaedibacter* (86.9%). Given the symbiotic association of Family Paracaedibacteraceae with other protists and low sequence identity of this taxon with any defined genus, we propose that it represents the endosymbiont and hereafter refer to it as *Candidatus Bodocaedibacter vickermanii* (Cbv) sp. nov. (vic.ker.ma'ni.i. N.L. gen. n. vickermanii, named after Keith Vickerman (1933–2016), author of seminal microscopic studies of diverse kinetoplastids, including *B. saltans*).

We explored various metagenomics datasets for prevalence of this bacterium in diverse environments (Supplementary Table 1). Based on 16 S rRNA gene identity, *Ca. B. vickermanii* is more abundant in samples extracted from various water sources compared to soil and human origin.

General features of the *Candidatus Bodocaedibacter vickermanii* genome

The genome containing the *Candidatus Bodocaedibacter vickermanii* rRNA sequence was assembled as a complete chromosome, 1.39 Mb in size and with an average coverage

of 212 \times . The GC skew graph (Fig. 1A) shows the pattern typical of a complete bacterial chromosome, transitional points referring to the origin and terminus of replication. The genome is 40.7% GC in content and encodes 1214 putative CDS, 40 tRNA and 2 rRNA-encoding operons (both rRNA operons have 16 S and 23 S rRNA genes in close vicinity, separated by two tRNAs). The genome also possesses two incomplete prophage elements (CPBP_00240-CPBP_00250; CPBP_00316-CPBP_00324).

Among sequenced genomes (Table 1), the *Ca. B. vickermanii* genome sequence is most closely related to a metagenomics sequence assembly from Canadian waste water (UBA6184) [60], and thereafter to various endosymbiont genomes of the Holosporales (86–87% 16 S rRNA gene sequence identity). 27% of gene sequences in the *Bodo* symbiont genome encode uncharacterized proteins found mostly in *Ca. B. vickermanii* and the UBA6184 metagenome (Fig. 1B). A majority (~67%) of putative coding sequences were placed into one of 20 functional categories of Clusters of Orthologous Groups of proteins (COG), shown in Fig. 1C, while 18% of annotated proteins were categorized as having 'unknown function'.

Microscopy indicates this microbe represents an intracellular symbiont

To confirm that the *Ca. B. vickermanii* sequence is correctly attributed to an intracellular endosymbiont, we used Fluorescent in situ Hybridisation (FISH) and *Ca. B. vickermanii* 16 S rRNA gene sequence probes. Cy-3 labelled probes bound to bacterial DNA on fixed *B. saltans* cultures, which were visualized using DIC microscopy (Fig. 2A). Stained *B. saltans* nuclei and kinetoplast are seen in blue channel (Fig. 2B) and bacterial spots are seen in red channel (Fig. 2C). A merged image from (Fig. 2D) shows that the bacteria are found adjacent to the *B. saltans* nucleus and are not observed extracellularly.

Electron microscopy was used to further define the intracellular niche. In the TEM images (Fig. 2E), a large nucleus surrounded by double membrane occupies the cell centre. The kinetoplast is positioned consistently adjacent to basal bodies. Multiple mitochondrial sections are seen along the cellular periphery. Food vacuoles, enclosing engulfed bacteria, are usually evident towards the posterior end. In addition to these typically kinetoplastid features, we identified rod-shaped bacteria-like structures that are 0.9–1.2 μm in length and 0.3–0.4 μm in diameter and often surrounded by an electron lucid halo (Fig. 2F, G). These are consistent with endosymbiont cells previously observed in *B. saltans* [25, 26].

3-D models generated from SBF-SEM imaging further resolved the disposition of these rod-shaped structures.

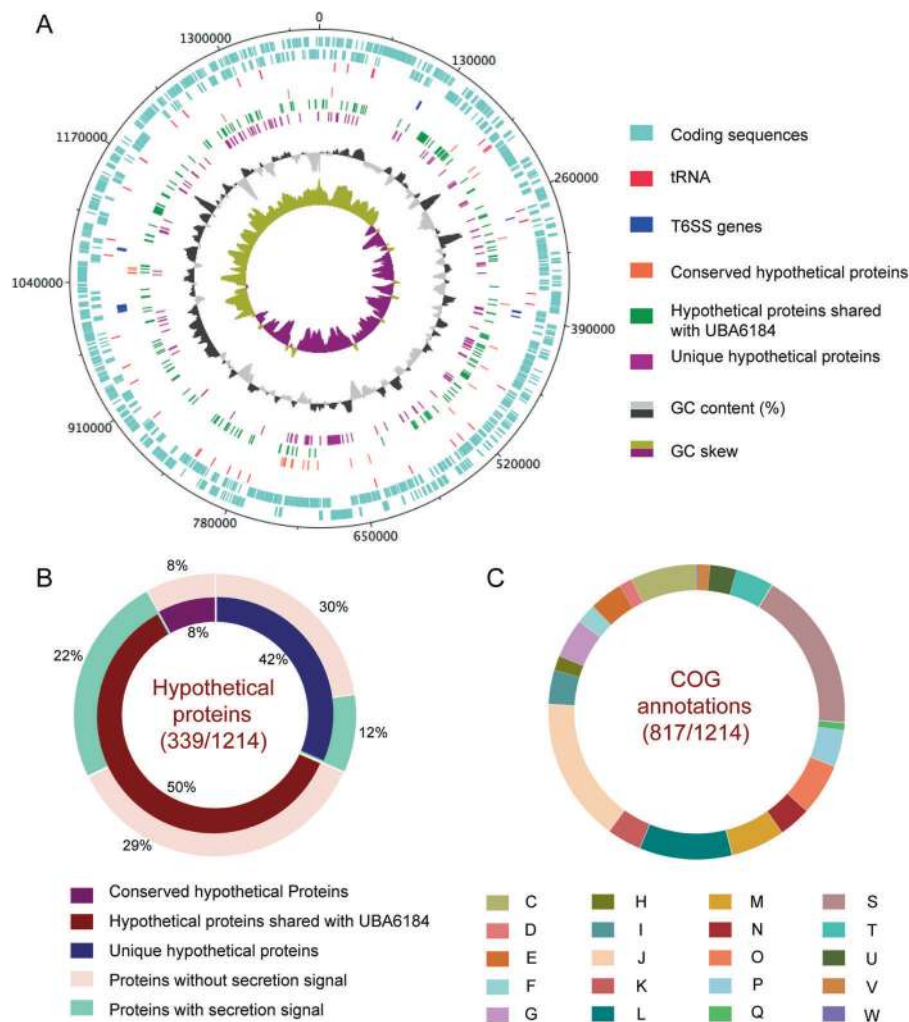


Fig. 1 Annotation features of *Bodo* endosymbiont. **A** Graphical representation of *Candidatus* Bodocaeidibacter vickermanii (Cbv) genome, generated using DNAPlotter v10.2 [93]. Moving inwards, the tracks represent forward and reverse CDS, tRNAs, type VI secretion system (T6SS) encoding genes, hypothetical proteins (conserved in the *Candidatus* Paracaeidibacteraceae bacterium UBA6184 genome and unique to Cbv), GC (%) plot and GC skew [(G-C)/(G+C)] plot. **B** Percentage of conserved and unique hypothetical proteins encoded in genome and presence of secretion signal is shown in pie chart. **C** Cluster of Orthologous Genes (COG) annotation categorization is shown in pie chart. Categories: Energy production and conversion (C), Cell cycle control, cell division, chromosome partitioning (D), Amino

acid transport and metabolism (E), Nucleotide transport and metabolism (F), Carbohydrate transport and metabolism (G), Coenzyme transport and metabolism (H), Lipid transport and metabolism (I), Translation, ribosomal structure and biogenesis (J), Transcription (K), Replication, recombination and repair (L), Cell wall/membrane/envelope biogenesis (M), Cell Motility (N), Post-translational modification, protein turnover, chaperone functions (O), Inorganic ion transport and metabolism (P), Secondary metabolites biosynthesis, transport and catabolism (Q), Function Unknown (S), Signal transduction mechanisms (T), Intracellular trafficking, secretion, and vesicular transport (U), Defence mechanisms (V), Extracellular structures (W).

Figure 2H-I and supplementary fig. 2 show the bean-shaped *B. saltans* cell with characteristic kinetoplastid features. In this 3-D rendering, rod-shaped bacterial cells are free in the cytoplasm, close to the nucleus. The number of bacterial cells varied from 3 to 10 ($N = 40$). Where a higher bacterial count was observed, a membrane enveloping those cells adjacent to the *Bodo* nucleus was often seen, similar to that previously observed [26]. This could be the nuclear membrane enveloping the endosymbionts prior to cell division and distribution into daughter cells.

Ca. Bodocaeidibacter is a novel genus in a clade of alpha-proteobacterial endosymbionts

A maximum likelihood phylogenetic tree was estimated from an alignment of 187 single-copy core genes from 13 genomes belonging to alpha-proteobacterial endosymbionts of protists of families Paracaeidibacteraceae, Caedimonadaceae, Midichloriaceae and Holosporaceae (Table 1). The genome of *Magnetococcus marinus*, belonging to basal lineage of alpha-proteobacteria, was used as an outgroup. The tree places the

Table 1 List of genomes used in the study: Annotation features, isolation source and accession numbers.

Organism	GC content (%)	Size (Mb)	CDS	tRNA	rRNA	Family	Order	Phylum	Host	Isolation source	Accession number
<i>Candidatus</i> Caedibacter acanthamoebae 30173	38.2	1.72	1359	42	3	Caedimonadaceae	Holosporales	Proteobacteria	<i>Acanthamoeba polyphaga</i> HN-3	human nasal swab	CP008936
<i>Caedibacter varicadens</i>	42.1	1.69	1726	41	3	Caedimonadaceae	Holosporales	Proteobacteria	<i>Paramecium biaurelia</i>	unknown source	BBVC000000000
<i>Candidatus</i> Paracaedibacter symbiosus PRA9	41.5	2.66	2199	43	6	Paracaedibacteraceae	Holosporales	Proteobacteria	<i>Acanthamoeba</i> sp. UWET39	soil	JQAK000000000
<i>Candidatus</i> Odyssella thessalonicensis L13	42	2.85	2398	40	3	Paracaedibacteraceae	Holosporales	Proteobacteria	<i>Acanthamoeba</i> sp.	water from air-conditioning unit	AWEF000000000
<i>Candidatus</i> Paracaedibacter acanthamoebae PRA3	41.0	2.47	2154	43	6	Paracaedibacteraceae	Holosporales	Proteobacteria	<i>Acanthamoeba</i> sp. UWC9	contact lens case	CP008941
<i>Candidatus</i> Bodocaedibacter vickermanii	40.7	1.39	1214	40	6	Bodocaedibacteraceae fam. nov.	Holosporales	Proteobacteria	<i>Bodo saltans</i> strain Lake Konstanz	lake water	CP054719 (this study)
<i>Candidatus</i> Paracaedibacteraceae bacterium UBA6184	40.1	1.47	1444	30	-	Bodocaedibacteraceae fam. nov.	Holosporales	Proteobacteria	MAG, Unknown source	waste water	DITM000000000
<i>Holospira curviuscula</i> 02AZ16	38.1	1.72	1534	42	3	Holosporaceae	Holosporales	Proteobacteria	<i>Paramecium bursaria</i>	unknown source from Astrakhan biosphere reserve	PHHC000000000
<i>Holospira obtusa</i> F1	35.2	1.33	1249	36	3	Holosporaceae	Holosporales	Proteobacteria	<i>Paramecium caudatum</i> strain C101	water	AWTR000000000
<i>Holospira undulata</i> HU1	36.1	1.40	1460	44	3	Holosporaceae	Holosporales	Proteobacteria	<i>Paramecium caudatum</i> strain 255	lake water	ARPM000000000
<i>Holospira elegans</i> E1	36.0	1.27	1454	37	3	Holosporaceae	Holosporales	Proteobacteria	<i>Paramecium caudatum</i>	water	BAUP000000000
Endosymbiont of <i>Acanthamoeba</i> UWC8	34.8	1.62	1494	36	3	Candidatus Midichloriaceae	Rickettsiales	Proteobacteria	<i>Acanthamoeba</i> sp. ATCC PRA2	human sample	CP004403
<i>Candidatus</i> Jidaibacter acanthamoeba UWC36	33.7	2.37	2267	36	3	Candidatus Midichloriaceae	Rickettsiales	Proteobacteria	<i>Acanthamoeba</i> sp. ATCC PRA6	cornea of a keratitis patient	JSWE000000000
<i>Magnetococcus marinus</i> MC-1	54.2	4.72	3766	45	9	Magnetococcaceae	Magnetococcales	Proteobacteria	Isolate, free living	marine water	CP000471

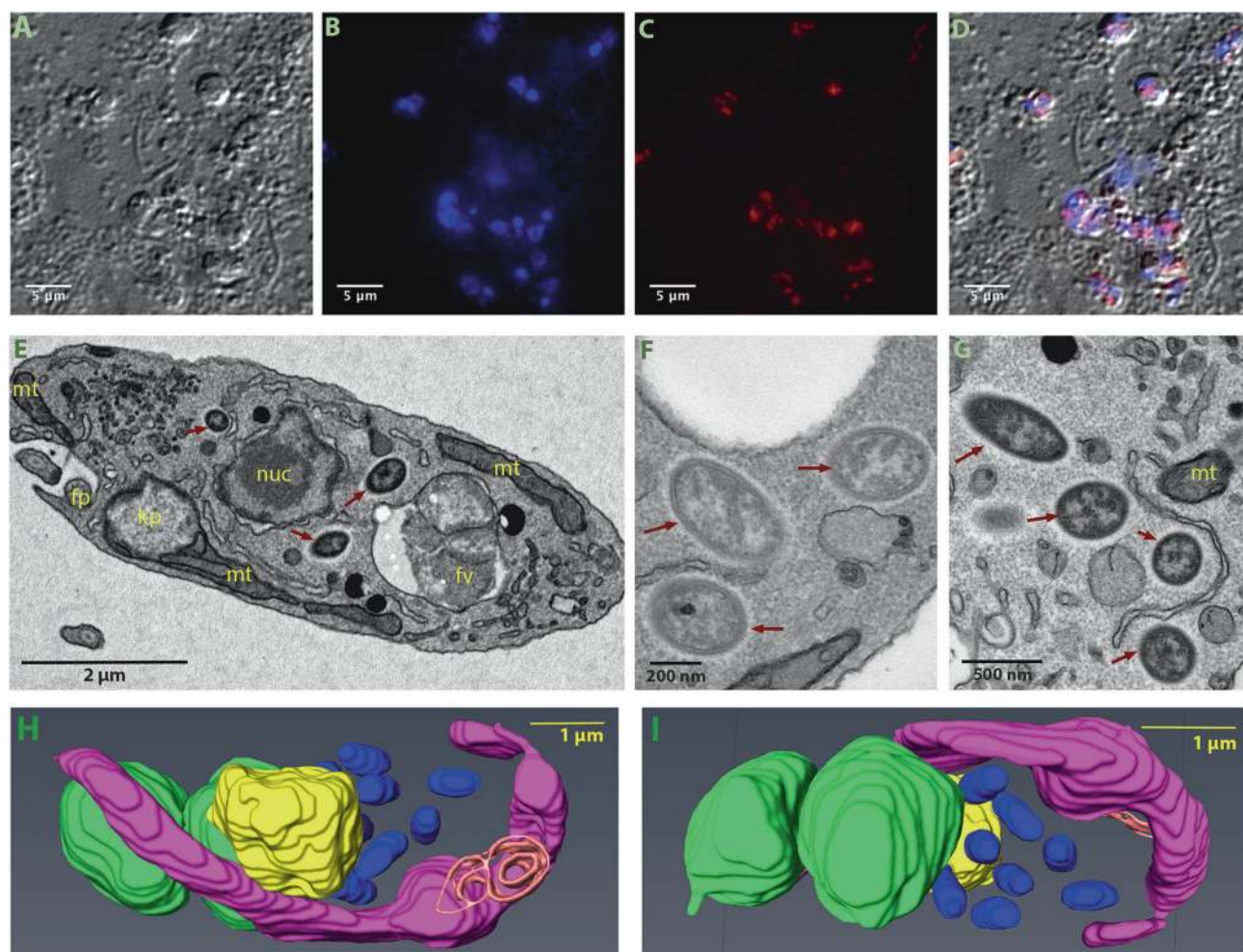


Fig. 2 Visualization of intracellular bacteria in *Bodo saltans*. Images from Fluorescent in situ hybridisation (FISH) experiment (A–D), transmission electron microscopy (TEM) experiment (E–G) and Serial Block Face-Scanning Electron Microscopy (H, I). Imaging of *B. saltans* cell cultures with differential interference contrast (A), staining with DAPI to visualize nucleus and kinetoplast (B), FISH staining with endosymbiont specific probe conjugated to Cy3 (C) and images from three channels overlaid (D). Images were acquired on a Zeiss Axio Observer Z1 (Carl Zeiss AG, Jena, Germany) equipped with 100×1.4NA objective, 2.5x optovar. Images captured were analyzed using ImageJ v2.0 [94]. Ultrastructure of the *B. saltans* cell (E) displaying nucleus (nuc), kinetoplast (kp), mitochondria (mt), food

vacuole (fv), flagellar pocket (fp) and three intracellular bacteria marked with dark red arrows, endosymbionts showing the presence of an electron lucid halo around cell membrane (F, G). 3D model of *B. saltans* cellular ultrastructure from two different angles (H, I). Bean shaped *B. saltans* cell is shown with flagellar pocket (in orange). Nucleus (in yellow) is situated in the centre of the cell and cytochromes (in blue) are distributed in close vicinity to the nucleus. A large, swirled mitochondrion (in magenta) is placed around the periphery of the cell, with a kinetoplast capsule just under the flagellar pocket. Multiple food vacuoles (in green) are present at the posterior end of the cell (colour figure online).

Bodo endosymbiont among other Holosporales and Rickettsiales endosymbionts of amoebae and ciliates, forming a clade with *Holospora*, *Caedibacter* and *Paracaedibacter*-like organisms (Fig. 3). As with the 16S rRNA gene homology analysis, the *Ca. B. vickermanii* genome grouped most closely with a wastewater metagenome (UBA6184). Branch lengths are long, consistent with these higher-level taxonomic comparisons. We examined the average amino acid identity (AAI) values between the *Bodo* endosymbiont and the 13 related genomes (Fig. 3) and, with the exception of UBA6184, AAI values for all related genomes are lower than

the conventional genus boundary of 55% [61], indicating that *Ca. B. vickermanii* and UBA6184 represent a novel genus. *Ca. B. vickermanii* forms a separate clade with UBA6184 genome. Based on their phylogeny and long branch lengths, we assign these two genomes to a novel family Bodocaedibacteraceae fam. nov.

Analysis of gene content using OrthoMCL [38] shows that these 13 endosymbiotic genomes share a core repertoire of 242 genes, and typically <50% of their total gene repertoire (Fig. 3). Mapping these gene sets to KEGG pathways indicates that many are involved in

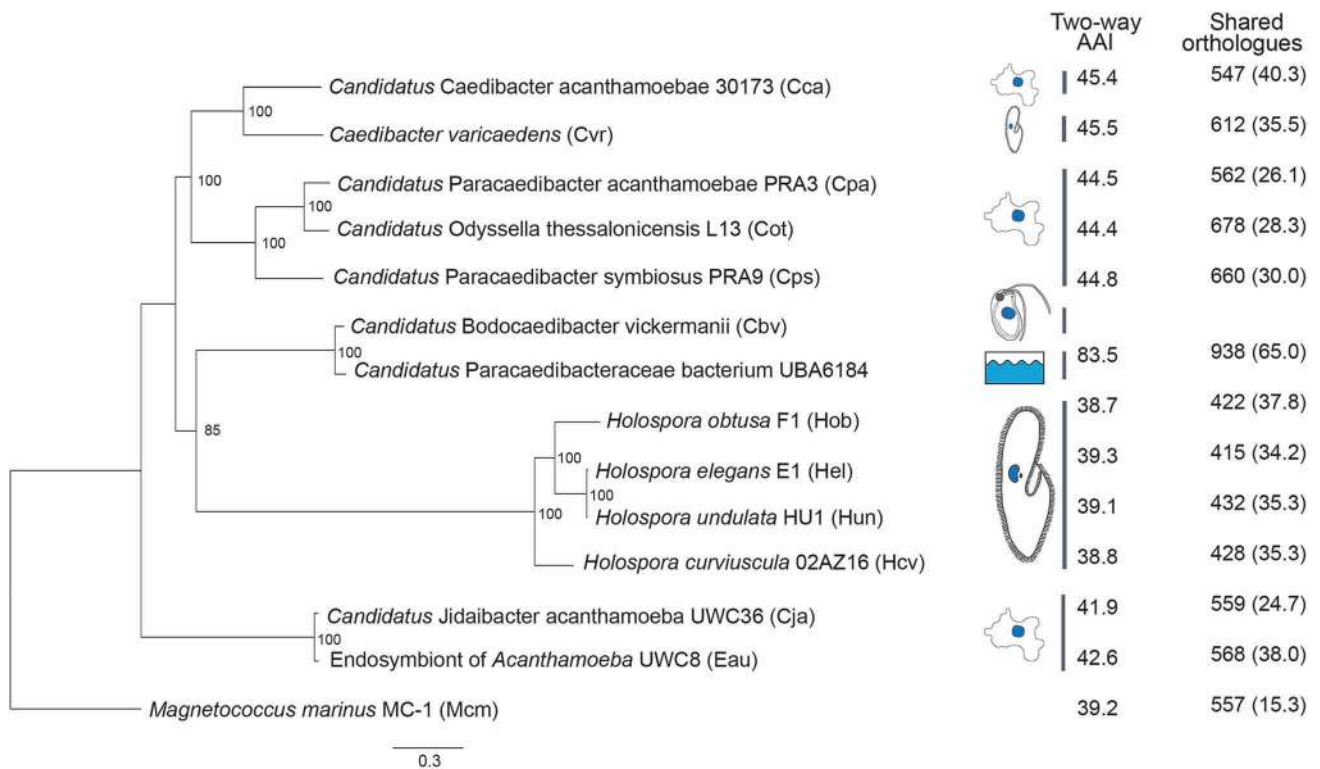


Fig. 3 Phylogenetic relationships of *Candidatus Bodocaedibacter vickermanii* and other alpha-proteobacterial endosymbionts of protists. A maximum likelihood tree estimated from a concatenated alignment of 187 protein sequences. The tree is rooted with an out-group (Mcm). Bootstrap values shown on the nodes are calculated from 1000 non-parametric replicates. The scale bar (0.3) is the number

of amino acid substitutions per site. Cartoons shown next to tree represent the isolation source (see Table 1 for details). Two columns on the right side depict the average amino acid identity (AAI) and number of orthologous gene clusters shared between *Ca. B. vickermanii* and each other genome sequence, with the corresponding value as a percentage of all genes shown in brackets.

central metabolism and information processing (Supplementary Fig. 3).

The *Ca. B. vickermanii* genome has limited metabolic capacity

The KEGG classifications of coding sequences in the *Ca. B. vickermanii* genome and 12 related endosymbiont sequences show that the *Bodo* endosymbiont has a small genome and relatively limited metabolic capacity (Table 1). We calculated Module Completion Ratios (MCR) for purine metabolism, pyrimidine metabolism, amino acid metabolism, polyamine biosynthesis, cofactor and vitamin biosynthesis pathways using MAPLE [51] (Supplementary Fig. 4). While often present in related genomes, several of these pathways are either completely or partially absent in *Ca. B. vickermanii*. For example, it lacks biotin, pantothenate and coenzyme A biosynthesis pathways entirely, and encodes only five out of nine genes in the ubiquinone biosynthesis pathway. Conversely, *Ca. B. vickermanii* possesses diverse membrane transporter genes for metabolite acquisition. The TransportTP [37] server identified 81 genes encoding membrane transporters in *Ca. B. vickermanii*, more than *Holospira*

genomes (56–67) but rather fewer than other *Caedibacter/Paracaedibacter* genomes (102–148).

A full account of all transporter genes, classified using the Transporter Classification Database [38], is shown in Supplementary Table 2. These include genes associated with amino acid transport, metabolite transport (e.g. drug/metabolite transporter (DMT) family ($n = 2$)), the major facilitator superfamily (MFS) ($n = 22$), and the ATP-binding cassette (ABC) superfamily ($n = 27$). There are also genes encoding for exporter proteins like the resistance-nodulation-cell division (RND) superfamily ($n = 6$) and the multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) flippase superfamily ($n = 4$).

B. saltans lacks the ability to synthesize some of the essential amino acids, vitamin and cofactors [62]. We therefore also examined the possibility that metabolic pathways are conducted cooperatively, with component genes being drawn from both host and endosymbiont genomes, which might indicate a mutualism. Only two pathways, lysine and threonine biosynthesis, appeared as potential candidates for cooperation, as shown in Supplementary Fig. 5. These two instances aside, most of the pathways associated with essential amino acids, vitamins and cofactor biosynthesis were

lacking in both organisms, quite unlike the mutualistic *Novymonas* and Strigomonadinae endosymbiosis [20–22], indicating that *B. saltans* remains firmly heterotrophic (Supplementary Fig. 6). Taken together, this does not suggest that *Ca. B. vickermanii* provides obvious metabolic benefit to *B. saltans*; indeed, the endosymbiont appears to be nutritionally dependent on its host.

The *Ca. B. vickermanii* genome encodes a Sec pathway and Type VI Secretion System (T6SS)

We examined the endosymbiont genome for various secretion systems that could facilitate communication with the host. This gram-negative bacterium contains the cellular components for a specific Type VI Secretion System (T6SS) and the general secretion (Sec) pathway. The *Ca. B. vickermanii* genome encodes most of the essential T6SS components, namely genes associated with membrane complex (*tssL* and *tssM*), baseplate complex (*tssA*, *tssE*, *tssF*, *tssG*, *tssK* and *tssI/vgrG*) and tail complex (*tssB* and *tssC*). Analysis with InterproScan also identified a *tssJ*-like gene (CPBP_00933) encoding a T6SS-associated lipoprotein and a gene (CPBP_00987) encoding a putative Hcp-like superfamily protein. RNA-seq analysis confirmed that these genes are actively transcribed, albeit at low levels (Supplementary table 3).

In Bodocaedibacteraceae, Paracaedibacteraceae and Caedimonadaceae genomes T6SS core genes are organised at four or more different genomic locations in a conserved arrangement: [*tssG-tssF*], [*tssB-tssC-hcp*], [*tssK-tssL-tssM-tssA*], and [*ywqK-vgrG*]. The *Ca. B. vickermanii* genome complies with this convention (Supplementary Fig. 7), indicating that the system could be functional in *Ca. B. vickermanii*, as in related endosymbionts [63–66].

The web-based server Bastion6 was used to predict for proteins potentially secreted by the T6SS [67]. This machine learning-based algorithm identified 117 proteins as putative type VI secretion effectors (T6SE) (Supplementary Table 4), which included various enzymes, flagellar proteins, membrane proteins and numerous uncharacterized proteins.

The Sec pathway is a major pathway for protein translocation across the cell membrane [68]. The *Ca. B. vickermanii* genome includes genes for a putative motor protein SecA, translocase complex SecYEG, and the auxillary components SecDF-YajC and YidC. It also encodes the proteins required for co-translational targeting (signal recognition particle proteins and receptor FtsY), as well as post-translational targeting (protein targeting component SecB). To discover proteins that may be targeted by the Sec pathway for integration or secretion outside the cell, we identified 158 coding sequences containing a signal sequence using SignalP 4.1 [34] (Supplementary Table 5). Of these, 105 proteins (66.4%) are uncharacterized and only

21 contain predicted transmembrane helices, indicating that most are probably secreted.

The genes predicted to encode for proteins with a secretion signal include various enzymes (xylanase, serine endoprotease, proteases, phospholipase, methyltransferase and acetyltransferase), flagellar proteins and membrane transporters. They also include two genes encoding proteins containing tetratricopeptide-repeats, and one gene for ankyrin-repeat containing protein, which are both typical eukaryotic protein domains involved in protein-protein interactions. A majority of these genes (130/158) have homologs only in the UBA6184 genome. The 105 genes encoding uncharacterized proteins with secretory signals are also limited to these two bacterial genomes, indicating this lineage has evolved a considerable repertoire of specific secreted proteins.

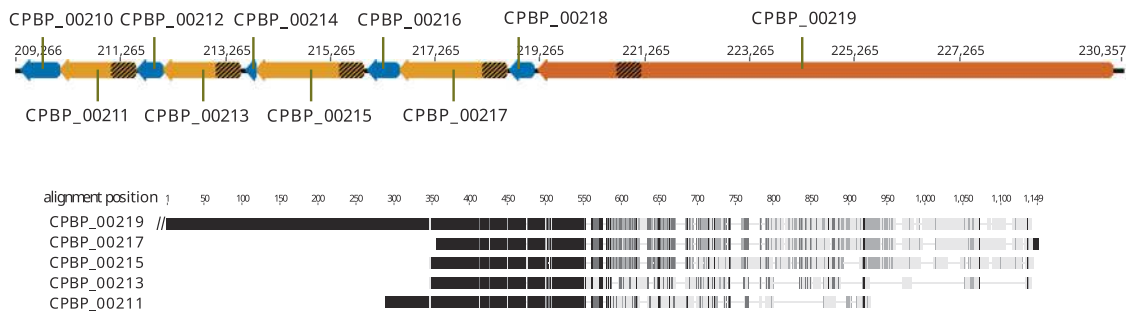
In case non-classical pathways are used to secrete *Ca. B. vickermanii* proteins, we analysed predicted protein sequences with SecretomeP [36]. There are 204 proteins that returned a score above the threshold, 59 of which also returned a significant result with SignalP (Supplementary Table 5). Amongst these cases were genes encoding proteins with homology to known cellular toxins, which led to the discovery of three putative novel toxin-antitoxin islands in the *Bodo* endosymbiont.

Multiple polymorphic toxin systems are a possible mechanism for addiction

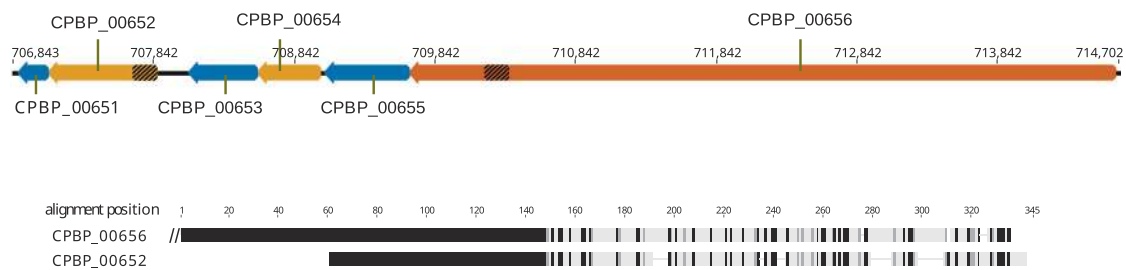
Bacterial Polymorphic Toxin Systems (PTS) such as the Rhs [69] and CDI [70] systems comprise genes encoding toxins and their cognate anti-toxins in characteristic operons [71]. Typically, a toxin gene encodes a large multi-domain protein contains an N-terminal secretion signal followed by a toxin domain. A toxin gene is immediately followed by a gene encoding an anti-toxin, (or ‘immunity protein’), capable of neutralising the toxin. This pair is often followed by additional ‘orphan modules’, each encoding an alternative toxin with its cognate antitoxin. Genes in these orphan modules may often contain a repeat region homologous to part of the toxin gene at the beginning of the operon. This repeat enables recombination between the full-length toxin gene and an orphan module to generate an alternative toxin protein.

We identified three genomic regions coding for putative PTS (Fig. 4; Supplementary Table 6). The putative toxin proteins identified here have no structural homology with known proteins, nor is there any homology between the three putative PTS. Nevertheless, the three regions show typical PTS characteristics. These include the presence of regions of homology between putative toxin proteins in the N-terminal part preceding the toxin domains (Fig. 4) and the organisation of putative toxin and antitoxin proteins across the loci (Supplementary Table 6), which leads us to conclude that they may represent novel PTS. In many cases, the

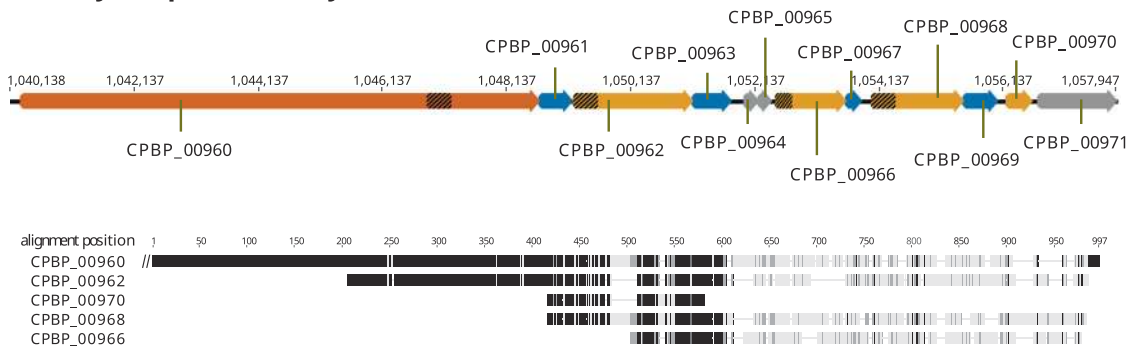
A) Polymorphic Toxin System I



B) Polymorphic Toxin System II



C) Polymorphic Toxin System III



- Complete multidomain toxins
- Alternative orphan toxin modules
- Anti-toxins
- Other
- Regions of homology

Fig. 4 Polymorphic toxin systems in *Ca. B. vickermanii*. Schematic diagrams of the organization of the three predicted polymorphic toxin systems in *Ca. B. vickermanii* (A–C). In each case, the putative, alternative orphan toxin modules are shown downstream of the main multidomain toxin. Beneath each is a protein multiple sequence alignment of the main toxin C-terminus and the alternative orphan sequences, generated using MAFFT program under default parameters as implemented in Geneious software v6.7. Grey scale gradient represents identical (black), similar (dark grey) and dissimilar (light grey) residues in the alignment based on the Blossum62 score matrix. Note that the specification of orphan

modules is least certain for PTS III. Although CPBP_00970 resembles the N-terminus of CPBP_00960, its coding sequence is abbreviated. This, and the metalloprotease (PF03410) match of CPBP_00971 that would, by expected PTS gene positioning encode an antitoxin, argue against this pair of genes representing an orphan module. Another pair of genes that should, by the PTS gene positioning scheme, be an orphan module are CPBP_00964 and CPBP_00965. However, these two sequences are short (75 and 67 residues respectively), homologous (around 34% identical) and CPBP_00964 lacks any homology to the N-terminus of CPBP_00960, arguing against these genes constituting an orphan module.

relationships used to infer toxin or antitoxin function are characterised by low degrees of sequence identity, but the HHsearch [39] probability scores are all >80% (and often much higher), which supports genuine homology.

In PTS I, the full-length toxin (CPBP_00219) terminates in an AHH nuclease domain. Accordingly, the following gene (in the coding direction) CPBP_00218 encodes a protein matching Immunity Protein 43 (Pfam: PF15570), predicted to act as antidote to the nuclease activity [72]. Interestingly, it also strongly matches two Pfam families, Gmx_para_CXXCG (PF09535) and DUF1629 (PF00791), suggesting that these too may encode antitoxins. These genes are followed by two orphan modules, each containing a toxin and an antitoxin (Supplementary Table 6). The next three genes encode proteins without recognisable homology to known structures or protein families, but CPBP_00210 clearly matches Immunity protein 49 (Pfam: PF15575), which is found in other PTS [72]. This suggests, given their context, that CPBP_00213- CPBP_00212 and CPBP_00211- CPBP_00210 may encode third and fourth orphan modules respectively with CPBP_00211 (matching Domain of Unknown Function DUF1837; PF08878) coding for a novel toxin domain.

The full-length toxin of PTS II (encoded by CPBP_00656) contains a Tox-HNH-HHH nuclease domain (PF15637) and is followed by its cognate antitoxin encoded by CPBP_00655. Two orphan modules follow, each containing toxin-antitoxin pairs that are unambiguously homologous to well-characterised families (Supplementary Table 6). However, the CPBP_00654 sequence contains no homology to the predicted N-terminal domain of the CPBP_00656 toxin; therefore, it is unclear whether the CPBP654- CPBP653 orphan module is functional.

Sequence homology in the predicted N-terminal regions of CPBP_00962, CPBP_00966, CPBP_00968, CPBP_00970 and the presumed (given its position) full-length toxin encoded by CPBP_00960 indicate the presence of a third PTS (Fig. 4). Unlike PTS I and II, most predicted proteins at the PTS III locus lack recognisable domains; only CPBP_00966 and CPBP_00967 display homologies to known PTS proteins, a Metallopeptidase toxin 5 (PF15641; [72]), and Colicin-like immunity protein (PF09204) respectively. Thus, the specification of functional orphan modules is least certain for this PTS, and it may be rich in novel toxin and antitoxin genes.

RNA-seq analysis of bulk *B. saltans* cell culture confirmed that genes comprising the putative PTS systems are all transcribed (Supplementary Table 3; Supplementary fig. 8). A few proteins of each of these polymorphic toxin systems were also predicted as type VI secretion effector in silico (Supplementary Table 4), suggesting the putative role of T6SS in transport and secretion of the involved toxin-antitoxin pairs.

Besides these three PTS, the endosymbiont genome also encodes for a toxin-antitoxin pair associated with T6SS.

It has an anti-toxin gene (YwqK family protein), upstream of one of the VgrG genes, as previously observed [63, 73].

Attempts to cure the symbiont result in host death

B. saltans cells were treated with antibiotics to cure them of endosymbiotic bacteria. Treatment with rifampicin (20 µg/ml) led to a decrease in cell count after 24 h of rifampicin treatment compared to the untreated cells. Cell counts from three independent experiments were compared and the difference in cell growth rate was found to be statistically significant (Paired *t*-test: $t = 5.9$, d.f. = 2, $p = 0.03$). By contrast, when the same treatment was applied to cell cultures of two other kinetoplastids that lack endosymbionts (*T. theileri* and *Z. costaricensis*), and a kinetoplastid with nutritional endosymbiont (*N. esmeraldas* [16]) rifampicin had no significant effect on cell number (Fig. 5A). Treatment of *B. saltans* and *N. esmeraldas* with streptomycin

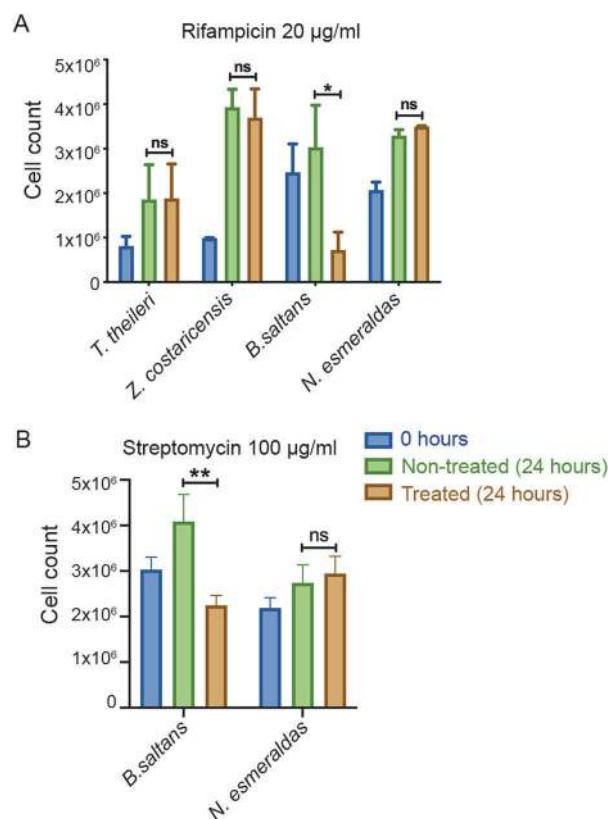


Fig. 5 Antibiotic treatment of kinetoplastids with and without endosymbionts. Four kinetoplastids were treated with rifampicin and observed for their behaviour on treatment with 20 µg/ml rifampicin (A). *B. saltans* and *N. esmeraldas* possess an endosymbiont, while *T. theileri* and *Z. costaricensis* do not. Two kinetoplastids, *B. saltans* and *N. esmeraldas* were treated with 100 µg/ml streptomycin (B). The graph shows number of cells before treatment (0 h), cells without any treatment (24 h) and with antibiotic treatment (24 h). The bars are average of three independent experiments. Statistical significance was calculated using ratio paired *T*-test.

(100 µg/ml) showed a similar effect as seen with rifampicin treatment (Fig. 5B).

Discussion

The *B. saltans* endosymbiont is identified here as an alpha-proteobacterium and a novel genus of the Holosporales, *Ca. Bodocaedibacter*. The complete genome indicates that the endosymbiont is incapable of synthesizing essential amino acids, vitamins and cofactors, but instead possesses an arsenal of membrane transporters for importing essential nutrients, a specialized secretory pathway and three putative PTS operons. Antibiotic treatment of *B. saltans* results in host cell death, consistent with *Ca. B. vickermanii* being essential for host viability, despite the meagre benefits it seems to offer. Certainly, the *Bodo* endosymbiont is unrelated, phylogenetically and physiologically, to the obligate bacterial endosymbionts of parasitic kinetoplastids such as *Ca. Kinetoplastibacterium* (Alcaligenaceae) [17, 74] and *Ca. Pandoraea* (Burkholderiaceae) [16]. These beta-proteobacterial symbionts can synthesize various essential amino acids independently (*Ca. Pandoraea*) [20], or in a cooperative manner with the host (*Ca. Kinetoplastibacterium*) [21], and also provide cofactors, vitamins and heme to their hosts [75]. Since *Ca. B. vickermanii* lacks the genes to perform such functions, we hypothesize that host dependency in its case arises from the expression of “addictive” bacterial toxin-antitoxin proteins.

This idea is immediately plausible when we consider the ecological strategies of related endosymbionts. While *Holospira* may increase host survival under adverse conditions [76, 77], it restrains host growth in normal conditions [78, 79]. An *Acanthamoeba* endosymbiont, *Ca. Amoebophilus asiaticus*, is parasitic; its genome lacks essential metabolic pathways but instead contains diverse genes shown to modulate host gene expression [64]. Another *Acanthamoeba* endosymbiont, *Ca. Jidaibacter acanthamoeba*, also encodes many proteins with eukaryotic-like domains thought to interact with the host [80]. Although the *Ca. B. vickermanii* genome encodes few eukaryotic-like domains, there are 339 uncharacterized proteins, most of which are lineage-specific and predicted to be secreted; such proteins could include factors for manipulating host physiology.

Perhaps the most pertinent comparison, however, is *Caedibacter*, an endosymbiont of *Paramecium* that is known for its ‘Killer trait’, which ensures its transmission at cell division [81]. *Caedibacter* provides a growth advantage to *Paramecium* cells [77] but also has an adaptation to ensure its spread through the population. A portion of the endosymbiont population forms ‘R-bodies’ that are secreted outside the host

cell, where they are lethal to *Paramecium* lacking the endosymbiont. In effect, the R-bodies become a toxin, against which *Caedibacter* provides protection, meaning that this is not a simple mutualism. The phylogenetic position of *Ca. B. vickermanii* among these various endosymbionts, with their ambiguous relationships with their hosts, makes it plausible that *Ca. B. vickermanii* too has an antagonistic mechanism for maintaining endosymbiosis.

Mechanisms like these are often described as evolved dependencies, exemplified by selfish genetic elements such as plasmids, which express a toxin-antitoxin (TA) system causing post-segregation killing or addiction [82]. TA systems in bacteria can also lead to programmed cell death [83] or persistence in response to stress conditions [84]. Another evolved dependency involves the wasp *Asobara tabida* and its symbiont *Wolbachia*, which the wasp requires for oogenesis and formation of a viable offspring. As *Wolbachia* is primarily transmitted through females, loss of the same strain of *Wolbachia* in males can result in cytoplasmic incompatibility, leading to offspring mortality [85, 86]. Similarly, in *C. elegans*, the peel-zeel system causes offspring to die if they do not carry the same allele as the sperm parent [13]. These instances of evolved dependency on a particular allele (peel-zeel), a genetic element (plasmids), or an endosymbiont (bacteria) carrying the relevant genes are fascinating examples of addiction, where losing the addictive element can lead to host death. The presence in *Ca. B. vickermanii* of three, actively transcribed PTS and a T6SS that could convey these effector proteins, coupled with an inability to survive in an aposymbiotic state, suggest that *B. saltans* is also subject to an evolved dependency.

Symbiosis is a well-known and intricate phenomenon found at various levels of biological organization and prokaryotic endosymbionts have been instrumental in the adaptive radiation of eukaryotes [2, 87]. The order Holosporales is associated with diverse protists, including Rhizaria, Excavates, Alveolates and Amoebozoans [88–91], and now kinetoplastid flagellates. Given the diversity of their hosts, and the evolutionary distances between them, this alpha-proteobacteria lineage would seem to have been associated with eukaryotes throughout their evolutionary diversification. Like other members of the lineage, the metabolic insufficiency, specialized secretory pathway and toxin-antitoxin systems of the *Ca. B. vickermanii* genome indicate that this long relationship has not been completely harmonious, the genomes bears witness to a struggle to retain the cooperation of their eukaryotic hosts. Besides, the *B. saltans* cells are also known to be infected by a giant virus [92]. In such a scenario of infection, the multi-level interactions that can occur between host, symbiont and virus and their impact on host would be another interesting attribute to investigate.

Data availability

The *Candidatus* *Bodocaeidibacter vickermanii* genome assembly has been submitted to NCBI under accession number CP054719. Raw PacBio genomic reads and Illumina RNA-Seq reads were submitted to Sequence Read Archive (SRA) database under the accession number SRR11932788 and SRR11935165 respectively.

Acknowledgements This research was supported by a Leverhulme Trust Research Grant to APJ (RPG-2014-005). We would like to thank Prof. Vyacheslav Yurchenko for kindly providing us *Novymonas* cultures. We thank Alison Beckett, Department of Cellular and Molecular Physiology, for helping in electron microscopy studies. We acknowledge Liverpool Centre for Cell Imaging (CCI) for providing access to microscopy equipment and technical assistance. We would also like to thank Prof. Bernhard Schink for Latin name assistance for bacteria.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bennett GM, Moran NA. Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. *Proc Natl Acad Sci USA*. 2015;112:10169–76.
- Nowack EC, Melkonian M. Endosymbiotic associations within protists. *Philos Trans R Soc Lond B Biol Sci*. 2010;365:699–712.
- Bright M, Bulgheresi S. A complex journey: transmission of microbial symbionts. *Nat Rev Microbiol*. 2010;8:218–30.
- Yurchenko V, Lukeš J. Parasites and their (endo) symbiotic microbes. *Parasitology*. 2018;145:1261–4.
- Lowe CD, Minter EJ, Cameron DD, Brockhurst MA. Shining a light on exploitative host control in a photosynthetic endosymbiosis. *Curr Biol*. 2016;26:207–11.
- Hurst GD, Frost CL. Reproductive parasitism: maternally inherited symbionts in a biparental world. *Cold Spring Harb Perspect Biol*. 2015;7:a017699.
- Douglas A. Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. *Annu Rev Entomol*. 1998;43:17–37.
- Houk E, Griffiths GW. Intracellular symbiotes of the Homoptera. *Annu Rev Entomol*. 1980;25:161–87.
- Miller IM. Bacterial leaf nodule symbiosis. In: Callow JA (eds). *Advances in botanical research*. Academic Press; 1990. pp 163–234.
- Lackner G, Partida-Martinez LP, Hertweck C. Endofungal bacteria as producers of mycotoxins. *Trends Microbiol*. 2009;17:570–6.
- Dedeine F, Vavre F, Fleury F, Loppin B, Hochberg ME, Boulétreau M. Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. *Proc Natl Acad Sci USA*. 2001;98:6247–52.
- Yarmolinsky MB. Programmed cell death in bacterial populations. *Science*. 1995;267:836–7.
- Seidel HS, Ailion M, Li J, van Oudenaarden A, Rockman MV, Kruglyak L. A novel sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in *C. elegans*. *PLoS Biol*. 2011;9:e1001115.
- Shropshire JD, On J, Layton EM, Zhou H, Bordenstein SR. One prophage WO gene rescues cytoplasmic incompatibility in *Drosophila melanogaster*. *Proc Natl Acad Sci USA*. 2018;115:4987–91.
- Hughes AL, Piontkivska H. Phylogeny of Trypanosomatidae and Bodonidae (Kinetoplastida) based on 18S rRNA: evidence for paraphyly of *Trypanosoma* and six other genera. *Mol Biol Evol*. 2003;20:644–52.
- Kostygov AY, Dobáková E, Grybchuk-Ieremenko A, Váhala D, Maslov DA, Votýpka J, et al. Novel trypanosomatid-bacterium association: evolution of endosymbiosis in action. *mBio*. 2016;7:e01985–15.
- Teixeira MM, Borghesan TC, Ferreira RC, Santos MA, Takata CS, Campaner M, et al. 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*. 2011;162:503–24.
- Alves JM, Serrano MG, Maia da Silva F, Voegtly LJ, Matveyev AV, Teixeira MM, et al. Genome evolution and phylogenomic analysis of *Candidatus* Kinetoplastibacterium, the betaproteobacterial endosymbionts of *Strigomonas* and *Angomonas*. *Genome Biol Evol*. 2013;5:338–50.
- Ganyukova AI, Frolov AO, Malysheva MN, Spodareva VV, Yurchenko V, Kostygov AY. A novel endosymbiont-containing trypanosomatid *Phytomonas borealis* sp. n. from the predatory bug *Picromerus bidens* (Heteroptera: Pentatomidae). *Folia Parasitol (Praha)*. 2020;67:004.
- Kostygov AY, Butenko A, Nenarokova A, Tashyreva D, Flegontov P, Lukeš J, et al. Genome of *Ca. Pandoraea novymonadis*, an endosymbiotic bacterium of the trypanosomatid *Novymonas esmeraldas*. *Front Microbiol*. 2017;8:1940.
- Alves JM, Klein CC, da Silva FM, Costa-Martins AG, Serrano MG, Buck GA, et al. 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 Evol Biol*. 2013;13:190.
- Klein CC, Alves JM, Serrano MG, Buck GA, Vasconcelos ATR, Sagot M-F, et al. Biosynthesis of vitamins and cofactors in bacterium-harboring trypanosomatids depends on the symbiotic association as revealed by genomic analyses. *PLoS ONE*. 2013;8:e79786.
- Harmer J, Yurchenko V, Nenarokova A, Lukeš J, Ginger ML. Farming, slaving and enslavement: histories of endosymbioses during kinetoplastid evolution. *Parasitology*. 2018;145:1311–23.
- Vickerman K, Preston TM. Comparative cell biology of the kinetoplastid flagellates. In: Lumsden WHR, Evans DA (eds). *Biology of the kinetoplastida*. New York: Academic Press; 1976. pp 35–130.
- Burzell LA. Fine structure of *Bodo curvifilus* Griessmann (Kinetoplastida: Bodonidae). *J Protozool*. 1975;22:35–39.
- Malysheva M, Karpova M, Frolov A. Mitosis of the free-living flagellate *Bodo saltans* strain Ps+ (Kinetoplastidea, Bodonida). *Cell Tiss Biol*. 2007;1:364–73.

27. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
28. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. 2017;27:722–36.
29. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*. 2012;28:3211–7.
30. Treangen TJ, Sommer DD, Angly FE, Koren S, Pop M. Next generation sequence assembly with AMOS. *Curr Protoc Bioinforma*. 2011;33:11.8.1–11.8.18.
31. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:2068–9.
32. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005;21:3674–6.
33. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PFAST: a fast phage search tool. *Nucleic Acids Res*. 2011;39 Suppl_2: W347–W352.
34. Petersen TN, Brunak S, Von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011;8:785–6.
35. Eichinger V, Nussbaumer T, Platzer A, Jehl M-A, Arnold R, Rattei T. EffectiveDB—updates and novel features for a better annotation of bacterial secreted proteins and type III, IV, VI secretion systems. *Nucleic Acids Res*. 2016;44:D669–D674.
36. Bendtsen JD, Kiemer L, Fausbøll A, Brunak S. Non-classical protein secretion in bacteria. *BMC Microbiol*. 2005;5:58.
37. Li H, Benedito VA, Udvardi MK, Zhao PX. TransportTP: a two-phase classification approach for membrane transporter prediction and characterization. *BMC Bioinforma*. 2009;10:418.
38. Saier MH Jr, Tran CV, Barabote RD. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res* 2006;34(suppl_1):D181–D186.
39. Söding J. Protein homology detection by HMM–HMM comparison. *Bioinformatics*. 2005;21:951–60.
40. wwPDBconsortium. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res*. 2019;47: D520–D528.
41. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47:D427–D432.
42. Eddy SR. A new generation of homology search tools based on probabilistic inference. *Genome Inform*. 2009;23:205–11.
43. Mirdita M, von den Driesch L, Galiez C, Martin MJ, Söding J, Steinegger M. Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res*. 2017;45:D170–D176.
44. Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H, et al. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol*. 2017;67:1613–7.
45. Lagkouvardos I, Joseph D, Kapfhammer M, Giritli S, Horn M, Haller D, et al. IMNGS: a comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Sci Rep*. 2016;6:33721.
46. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 2003;13:2178–89.
47. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinforma*. 2003;00:2.3.1–2.3.22.
48. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*. 2011;27:1164–5.
49. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30: 1312–3.
50. Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Prepr*. 2016;4:e1900v1.
51. Arai W, Taniguchi T, Goto S, Moriya Y, Uehara H, Takemoto K, et al. MAPLE 2.3. 0: an improved system for evaluating the functionomes of genomes and metagenomes. *Biosci Biotechnol Biochem*. 2018;82:1515–7.
52. Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics*. 2012;28:1823–9.
53. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res*. 2007;35:7188–96.
54. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41:D590–D596.
55. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhu-kumar, et al. ARB: a software environment for sequence data. *Nucleic Acids Res*. 2004;32:1363–71.
56. Loy A, Arnold R, Tischler P, Rattei T, Wagner M, Horn M. probeCheck—a central resource for evaluating oligonucleotide probe coverage and specificity. *Environ Microbiol*. 2008;10:2894–8.
57. Yilmaz LS, Parmerkar S, Noguera DR. mathFISH, a web tool that uses thermodynamics-based mathematical models for in silico evaluation of oligonucleotide probes for fluorescence in situ hybridization. *Appl Environ Microbiol*. 2011;77:1118–22.
58. Booth DG, Beckett AJ, Molina O, Samejima I, Masumoto H, Kouprina N, et al. 3D-CLEM reveals that a major portion of mitotic chromosomes is not chromatin. *Mol Cell*. 2016;64:790–802.
59. Kostygov AY, Yurchenko V. Revised classification of the subfamily Leishmaniinae (Trypanosomatidae). *Folia Parasitol (Praha)*. 2017;64:020.
60. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol*. 2017;2:1533–42.
61. Rodriguez-R LM, Konstantinidis KT. Bypassing cultivation to identify bacterial species. *Microbe*. 2014;9:111–8.
62. Opperdoes FR, Butenko A, Flegontov P, Yurchenko V, Lukeš J. Comparative metabolism of free-living *Bodo saltans* and parasitic trypanosomatids. *J Eukaryot Microbiol*. 2016;63:657–78.
63. George EE, Husnik F, Tashyreva D, Prokopchuk G, Horák A, Kwong WK, et al. Highly reduced genomes of protist endosymbionts show evolutionary convergence. *Curr Biol*. 2020;30:925–33.e3.
64. Schmitz-Esser S, Tischler P, Arnold R, Montanaro J, Wagner M, Rattei T, et al. The genome of the amoeba symbiont “*Candidatus Amoebophilus asiaticus*” reveals common mechanisms for host cell interaction among amoeba-associated bacteria. *J Bacteriol*. 2010;192:1045–57.
65. Brown A, Wasala SK, Howe DK, Peetz AB, Zasada IA, Denver DR. Comparative genomics of *Wolbachia-Cardinium* dual endosymbiosis in a plant-parasitic nematode. *Front Microbiol*. 2018;9:2482.
66. Yang JC, Madupu R, Durkin AS, Ekborg NA, Peadarallu CS, Hostetler JB, et al. The complete genome of *Teredinibacter turnerae* T7901: an intracellular endosymbiont of marine wood-boring bivalves (shipworms). *PLoS ONE*. 2009;4:e6085.
67. Wang J, Yang B, Leier A, Marquez-Lago TT, Hayashida M, Rocker A, et al. Bastion6: a bioinformatics approach for accurate prediction of type VI secreted effectors. *Bioinformatics*. 2018;34:2546–55.

68. Natale P, Brüser T, Driessen AJ. Sec-and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—distinct translocases and mechanisms. *Biochim Biophys Acta*. 2008;1778:1735–56.
69. Jackson AP, Thomas GH, Parkhill J, Thomson NR. Evolutionary diversification of an ancient gene family (rhs) through C-terminal displacement. *BMC Genomics*. 2009;10:584.
70. Ruhe ZC, Low DA, Hayes CS. Bacterial contact-dependent growth inhibition. *Trends Microbiol*. 2013;21:230–7.
71. Jamet A, Nassif X. New players in the toxin field: polymorphic toxin systems in bacteria. *mBio*. 2015;6:e00285–15.
72. Zhang D, de Souza RF, Anantharaman V, Iyer LM, Aravind L. Polymorphic toxin systems: comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol Direct*. 2012;7:18.
73. Ibrahim M, Subramanian A, Anishetty S. Comparative pan genome analysis of oral *Prevotella* species implicated in periodontitis. *Funct Integr Genomics*. 2017;17:513–36.
74. Votýpka J, Kostygov AY, Kraeva N, Grybchuk-Ieremenko A, Tesařová M, Grybchuk D, et al. *Kentomonas* gen. n., a new genus of endosymbiont-containing trypanosomatids of Strigomonadinae subfam. n. *Protist*. 2014;165:825–38.
75. Alves JM, Voegtly L, Matveyev AV, Lara AM, da Silva FM, Serrano MG, et al. Identification and phylogenetic analysis of heme synthesis genes in trypanosomatids and their bacterial endosymbionts. *PLoS ONE*. 2011;6:e23518.
76. Fujishima M, Kawai M, Yamamoto R. *Paramecium caudatum* acquires heat-shock resistance in ciliary movement by infection with the endonuclear symbiotic bacterium *Holospora obtusa*. *FEMS Microbiol Lett*. 2005;243:101–5.
77. Fujishima M. Infection and maintenance of *Holospora* species in *Paramecium caudatum*. In: Fujishima M (eds). *Endosymbionts in Paramecium*. Microbiology Monographs, vol 12. Berlin, Heidelberg: Springer; 2009. pp. 201–25.
78. Görtz H, Fujishima M. Conjugation and meiosis of *Paramecium caudatum* infected with the micronucleus-specific bacterium *Holospora elegans*. *Eur J Cell Biol*. 1983;32:86–91.
79. Restif O, Kaltz O. Condition-dependent virulence in a horizontally and vertically transmitted bacterial parasite. *Oikos*. 2006;114:148–58.
80. Schulz F, Martijn J, Wascher F, Lagkouvardos I, Kostanjšek R, Ettema TJ, et al. A Rickettsiales symbiont of amoebae with ancient features. *Environ Microbiol*. 2016;18:2326–42.
81. Kusch J, Görtz HD. Towards an understanding of the killer trait: *Caedibacter* endocytobionts in *Paramecium*. In: Overmann J (eds). *Molecular Basis of Symbiosis*. Prog Mol Subcell Biol., vol 41. Springer, Berlin, Heidelberg; 2005, pp 61–76.
82. Werren JH. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc Natl Acad Sci USA*. 2011;108 Supplement 2:10863–70.
83. Engelberg-Kulka H, Hazan R, Amitai S. mazEF: a chromosomal toxin-antitoxin module that triggers programmed cell death in bacteria. *J Cell Sci*. 2005;118:4327–32.
84. Page R, Peti W. Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nat Chem Biol*. 2016;12:208–14.
85. Dedeine F, Bouletreau M, Vavre F. *Wolbachia* requirement for oogenesis: occurrence within the genus *Asobara* (Hymenoptera, Braconidae) and evidence for intraspecific variation in *A. tabida*. *Heredity*. 2005;95:394–400.
86. Dedeine F, Vavre F, Shoemaker DD, Boulétreau M. Intra-individual coexistence of a *Wolbachia* strain required for host oogenesis with two strains inducing cytoplasmic incompatibility in the wasp *Asobara tabida*. *Evolution*. 2004;58:2167–74.
87. Aanen DK, Hoekstra RF. The evolution of obligate mutualism: if you can't beat 'em, join 'em. *Trends Ecol Evol*. 2007;22:506–9.
88. Hess S, Suthaus A, Melkonian M. “*Candidatus* Finniella” (Rickettsiales, Alphaproteobacteria), novel endosymbionts of viridiraptorid amoeboflagellates (Cercozoa, rhizaria). *Appl Environ Microbiol*. 2016;82:659–70.
89. Georgiades K, Madoui M-A, Phuong Le CR, Raoult D. Phylogenomic analysis of *Odysella thessalonicensis* fortifies the common origin of Rickettsiales, *Pelagibacter ubique* and *Reclinomonas americana* mitochondrion. *PLoS ONE*. 2011;6:e24857.
90. Kim E, Park JS, Simpson AG, Matsunaga S, Watanabe M, Murakami A, et al. Complex array of endobionts in *Petalomonas sphagnophila*, a large heterotrophic euglenid protist from *Sphagnum*-dominated peatlands. *ISME J*. 2010;4:1108–20.
91. Schrällhammer M, Castelli M, Petroni G. Phylogenetic relationships among endosymbiotic R-body producer: bacteria providing their host the killer trait. *Syst Appl Microbiol*. 2018;41:213–20.
92. Deeg CM, Chow C-ET, Suttle CA. The kinetoplastid-infecting *Bodo saltans* virus (BsV), a window into the most abundant giant viruses in the sea. *Elife*. 2018;7:e33014.
93. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics*. 2009;25:119–20.
94. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological image analysis. *Nat Methods*. 2012;9:676–82.