# The *Paracaedibacter*-like endosymbiont of *Bodo saltans* (Kinetoplastida) uses multiple putative toxin-antitoxin systems to maintain its host association

Samriti Midha[1]*, Daniel J. Rigden[2], Stefanos Siozios[1], Gregory D. D. Hurst[1], Andrew P. Jackson[1]

[1] Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Ic2 Liverpool Science Park, 146 Brownlow Hill, Liverpool, L3 5RF. United Kingdom.

[2] Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Crown St., Liverpool, L69 7ZB. United Kingdom.

Address for correspondence: S.midha@liv.ac.uk

**Running Title:** Host association of *Bodo* endosymbiont

**Keywords:** kinetoplastida, Holosporales, endosymbiosis, toxin, antitoxin, genome

# 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.

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, utilizes 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 (4). 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* (5). 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 (6).

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 sterility (6, 7). Host dependency may be due also to coadaptation of cellular and developmental processes (8). The defensive *Burkholderia* symbiont of the fungus *Rhizopus* is required for completion of the host's sexual

66 phase (9), a system mirroring the requirement of *Asobara tabida* wasps for one of their
67 *Wolbachia* symbionts to complete oogenesis (10). In these cases, coadaptation over deep
68 evolutionary time to the presence of the symbiont means that the host fails when the
69 symbiont is removed, irrespective of any benefit the symbiont might confer.
70
71 There is another potential route to dependency that arises through selection on the symbiont
72 to addict the host to its presence. Addiction mechanisms using toxin-antitoxin systems are a
73 common retention mechanism for plasmids in bacteria (11) and are also observed in a variety
74 of selfish genetic elements, such as the peel-zeel system of *C. elegans* (12). These systems
75 typically function through delivery of a long-lived toxin molecule alongside an antitoxin with
76 a shorter half-life. Should the element carrying the system not be inherited then the toxin
77 becomes active and kills the non-bearer individual. Bacterial endosymbionts are currently not
78 known to addict hosts through toxin-antitoxin systems, although they are known to exploit
79 similar 'rescue' based systems in creating the phenotype of cytoplasmic incompatibility (12,
80 13).
81
82 In this study, we investigated the symbiosis between the flagellate *Bodo saltans*
83 (Kinetoplastida) and its intracellular bacterial symbiont. *B. saltans* is a heterotrophic,
84 phagocytic bacteriovore found in freshwater and marine habitats worldwide, and is among
85 the closest free-living relatives of trypanosomatid parasites (14). Within the trypanosomatids,
86 three parasitic lineages, *Novymonas*, *Phytomonas* and the Strigomonadinae, are known to
87 contain a β-proteobacterial symbionts (15-17). These symbionts are mutualistic, cooperating
88 in various metabolic pathways that provide essential amino acids and vitamins to the host
89 (18-20). Although these endosymbionts are all mutualists, it is clear that parasitic
90 kinetoplastids have acquired endosymbionts independently and relatively often (21). *B.*

*saltans* is a free-living kinetoplastid, but microscopic examination has established that it too possesses endosymbiotic bacteria (22-24), whose precise identity is unknown.

We characterized the cytoplasmic bacterium inhabiting *B. saltans*, which we name *Candidatus* Bodocaedibacter vickermanii, 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* 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. 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). 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 (25). DNA

sequence assembly was carried out using Canu v1.5 (26). To identify all distinct bacterial sequences present, rRNA reads were filtered using SortMeRNA v2.1 (27), 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 (28) and constituent reads were mapped back to check for mis-assembly.

*Endosymbiont genome annotation*

Genome annotation was carried out using Prokka v1.12 (29) complemented with BLASTp and InterProScan matches obtained using BLAST2GO v5.0 (30). Presence of prophage sequences were checked with PHAST (31). Genes containing a canonical signal peptide were predicted with SignalP 4.1 (32). Proteins with eukaryotic like domains were identified with EffectiveELD (33). Secretome P was used to predict genes encoding proteins secreted by non-classical pathway (34). Membrane transporters were predicted using TransportTP (35) and transporters were classified using the BLAST tool of the Transport Classification DataBase (36). HHsearch (37) was used to search for distant relationships between putative toxin-antitoxin system components and known structures at the Protein Data Bank (PDB;(38)) or protein families in the Pfam database (39). Three iterations of jackhmmer (40) searching against the UniClust database (41) were used to generate the sequence profiles for comparison with database entries. 16S rRNA sequences were predicted using RNAmmer v1.2 and EzBioCloud was used for taxonomical identification of the bacterial lineage (42).

*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 (43). Protein sequences encoded by 187 conserved genes from the 14 bacterial genomes were aligned using ClustalW (44). Prottest v3.0 was used to check for the optimal amino acid substitution model for phylogeny estimation (45). An LG+I+G model was applied in RaxML to estimate a maximum likelihood tree with 1000 non-parametric bootstrap replicates (46). BLAST-based Average Amino Acid Identity (AAI) values amongst the 13 genomes were calculated using AAI calculator (47). MAPLE 2.3.0 was used to identify and compare the completeness of various metabolic pathways in these endosymbiont genomes (48). Genomaple 2.3.2 was used to compare the metabolic pathways of endosymbiont and *B. saltans* (48).

*Fluorescent in situ hybridization (FISH)*

The endosymbiont 16S rRNA sequence was integrated into the SILVA database using the web-based SINA aligner (49, 50). This alignment was further merged with SILVA Ref NR 99 (51) using the ARB package and the probe design tool was used to select FISH probes specific for the endosymbiont genome. (52). The probes were tested for mismatch analysis using mathFISH, probeCheck and BLAST (53, 54). *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 hybridization with the probe (50 ng/µl) 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 0.1M phosphate buffer (pH7.4) in Pelco Biowave (Ted Pella Inc.) and washed twice in 0.1M PB before embedding in 3% agarose. Agarose embedded cell pellets were post-fixed and stained as described previously (55), 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 (70-75nm) were cut on an UC6 ultra-microtome (Leica, Vienna) 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 500µm x 500µm. Samples were painted and dissected as previously described (55). 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 (1.7kV), dwell time per pixel (14µs), magnification (5040×) pixel size 5.4 nm in x and y, frame width 6144 × 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 Rifampicin treatment*

Besides *B. saltans*, two parasitic and asymbiotic kinetoplastids, *Trypanosoma theileri* and *Leptomonas costaricensis*, were treated with rifampicin (20 µg/ml), alongside control populations. 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 hours and after 24 hours of treatment in control 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.

*Bodo saltans RNA-sequencing*

*B. saltans* cells were treated with gentamicin (50 µg/ml) before RNA extraction. Total RNA was extracted from cells and rRNA was depleted using RiboZero rRNA depletion kit (Epidemiology). A strand-specific library was constructed with the NEBNext Ultra Directional RNA library preparation kit (NEB). Paired end 2x150 bp sequencing was carried out on the Illumina platform. Raw reads were mapped on to the endosymbiont genome sequence using bwa v0.7.12 and low quality read alignment 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).

*Data availability*

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

# Results

### A *Bodo saltans* endosymbiont of the Paracaedibacteraceae

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 figure 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 (98.59%) to uncultured bacterial sequences belonging to Family Paracaedibacteraceae, followed by *Paracaedibacter* (86.89%). Given that this fourth taxon has such low sequence identity with any defined genus, we propose that it represents the endosymbiont and hereafter refer to it as *Candidatus* Bodocaedibacter vickermanii (Cbv) gen. nov. sp. nov.. The organism is named after Keith Vickerman (1933-2016), author of seminal microscopic studies of diverse kinetoplastids, including *B*. *saltans*.

### 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 212x. 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 16S and 23S 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) (56), and thereafter to various endosymbiont genomes of the Holosporales (86-87% 16S rRNA 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* Hybridization (FISH) and *Ca*. B. vickermanii 16S rRNA 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-2G). These are consistent with endosymbiont cells previously observed in *B*. *saltans* (23, 24).

3-D models generated from SBF-SEM imaging further resolved the disposition of these rod-shaped structures. Fig. 3 shows 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 (24). This could be the nuclear membrane enveloping the endosymbionts prior to cell division and distribution into daughter cells.

**_Ca_. Bodocaedibacter 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 Paracaedibacteraceae, 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 *Bodo* endosymbiont among other Holosporales and Rickettsiales endosymbionts of amoebae and ciliates, forming a clade with *Holospora*, *Caedibacter* and *Paracaedibacter*-like organisms (Fig. 4). As with the 16S rRNA 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. 4) and, with the exception of UBA6184, AAI values for all related genomes are lower than the conventional genus boundary of 55% (57), indicating that *Ca*. B. vickermanii and UBA6184 represent a novel genus.

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. 4). Mapping these gene sets to KEGG pathways indicates that many are involved in central metabolism and information processing (Supplementary figure 2).

**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 (48) (Supplementary

Fig. 3). 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 (35) server identified 81 genes encoding membrane transporters in *Ca*. B. vickermanii, more than *Holospora* 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 (36), is shown in Supplementary Table 1. 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).

We 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 co-operation, as shown in Supplementary figure 4. 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 (18-20), indicating that *B*. *saltans* remains firmly heterotrophic (Supplementary figure 5). 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 2).

In 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. 6), indicating that the system could be functional in *Ca*. B. vickermanii, as in related endosymbionts (58-61).

The web-based server Bastion6 was used to predict for proteins potentially secreted by the T6SS (62). This machine learning-based algorithm identified 117 proteins as putative type VI

365 secretion effectors (T6SE) (Supplementary table 3), which included various enzymes,
366 flagellar proteins, membrane proteins and numerous uncharacterized proteins.
367
368 The Sec pathway is a major pathway for protein translocation across the cell membrane (63).
369 The *Ca*. B. vickermanii genome includes genes for a putative motor protein SecA, translocase
370 complex SecYEG, and the auxillary components SecDF-YajC and YidC. It also encodes the
371 proteins required for co-translational targeting (signal recognition particle proteins and
372 receptor FtsY), as well as post-translational targeting (protein targeting component SecB). To
373 discover proteins that may be targeted by the Sec pathway for integration or secretion outside
374 the cell, we identified 158 coding sequences containing a signal sequence using SignalP 4.1
375 (32) (Supplementary Table 4). Of these, 105 proteins (66.4%) are uncharacterized and only
376 21 contain predicted transmembrane helices, indicating that most are probably secreted.
377
378 The genes predicted to encode for proteins with a secretion signal include various enzymes
379 (xylanase, serine endoprotease, proteases, phospholipase, methyltransferase and
380 acetyltransferase), flagellar proteins and membrane transporters. They also include two genes
381 encoding proteins containing tetratricopeptide-repeats, and one gene for ankyrin-repeat
382 containing protein, which are both typical eukaryotic protein domains involved in protein-
383 protein interactions. A majority of these genes (130/158) have homologs only in the
384 UBA6184 genome. The 105 genes encoding uncharacterized proteins with secretory signals
385 are also limited to these two bacterial genomes, indicating this lineage has evolved a
386 considerable repertoire of specific secreted proteins.
387
388 In case non-classical pathways are used to secrete *Ca*. B. vickermanii proteins, we analysed
389 predicted protein sequences with SecretomeP (34). There are 204 proteins that returned a

score above the threshold, 59 of which also returned a significant result with SignalP (Supplementary Table 4). Also among 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 (64) and CDI (65) systems comprise genes encoding toxins and their cognate anti-toxins in characteristic operons (66). 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 5; Table 2). 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 5) and the organisation of putative toxin and antitoxin proteins across the loci (Table 2), which leads us to conclude that they may represent novel PTS. In many cases, the relationships used to infer toxin or

antitoxin function are characterised by low degrees of sequence identity, but the HHsearch (37) 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 (67). 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 (Table 2). 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 (67). 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 (Table 2). 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 5). 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; (67)), 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 2; Supplementary figure 7). A few proteins of each of these polymorphic toxin systems were also predicted as type VI secretion effector *in silico* (Supplementary table 3), 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 (58, 68).

**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 hours 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 = 5.906, d.f. = 2, p = 0.0275). By contrast, when the same treatment was applied to cell cultures of two other kinetoplastids that lack endosymbionts (*Trypanosoma theileri* and *Leptomonas costaricensis*), rifampicin had no effect on cell number (Fig. 6).

# 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, indicating that *Ca*. B. vickermanii is 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) (16, 69) and *Ca*. Pandoraea (Burkholderiaceae) (15). These beta-proteobacterial symbionts can synthesize various essential amino acids independently (*Ca*. Pandoraea) (18), or in a cooperative manner with the host (*Ca*. Kinetoplastibacterium) (19), and also provide cofactors, vitamins and heme to their hosts (70). 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 *Holospora* may increase host survival under adverse conditions (71, 72), it restrains host growth in normal conditions (73, 74). 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 (59). Another *Acanthamoeba* endosymbiont, *Ca*. Jidaibacter acanthamoeba, also encodes many proteins with eukaryotic-like domains thought to interact with the host (75). 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 (76). *Caedibacter* provides a growth advantage to *Paramecium* cells (72) 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 attitudes towards 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 (77). TA systems in bacteria can also lead to programmed

cell death (78) or persistence in response to stress conditions (79). 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 (80, 81). Similarly, in *C. elegans*, the peel-zeel system causes offspring to die if they do not carry the same allele as the sperm parent (12). 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, 82). The Family Paracaedibacteraceae is associated with diverse protists, including Rhizaria, Excavates and Amoebozoans (83-85), 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 exactly harmonious, the genomes bears witness to a struggle to retain the cooperation of their eukaryotic hosts.

## Acknowledgements

This research was supported by a Leverhulme Trust Research Grant to APJ (RPG-2014-005). 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.

## Competing Interests

The authors declare that there are no competing interests.

## References

1. Bennett GM, Moran NA. Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. Proceedings of the National Academy of Sciences. 2015;112(33):10169-76.

2. Nowack EC, Melkonian M. Endosymbiotic associations within protists. Philosophical Transactions of the Royal Society B: Biological Sciences. 2010;365(1541):699-712.

3. Bright M, Bulgheresi S. A complex journey: transmission of microbial symbionts. Nature Reviews Microbiology. 2010;8(3):218-30.

4. Lowe CD, Minter EJ, Cameron DD, Brockhurst MA. Shining a light on exploitative host control in a photosynthetic endosymbiosis. Current Biology. 2016;26(2):207-11.

560 5. Hurst GD, Frost CL. Reproductive parasitism: maternally inherited symbionts in a
561 biparental world. Cold Spring Harbor perspectives in biology. 2015;7(5):a017699.
562 6. Douglas A. Nutritional interactions in insect-microbial symbioses: aphids and their
563 symbiotic bacteria *Buchnera*. Annual review of entomology. 1998;43(1):17-37.
564 7. Houk E, and, Griffiths GW. Intracellular symbiotes of the Homoptera. Annual review
565 of entomology. 1980;25(1):161-87.
566 8. Miller IM. Bacterial leaf nodule symbiosis. Advances in botanical research. 17:
567 Elsevier; 1990. p. 163-234.
568 9. Lackner G, Partida-Martinez LP, Hertweck C. Endofungal bacteria as producers of
569 mycotoxins. Trends in microbiology. 2009;17(12):570-6.
570 10. Dedeine F, Vavre F, Fleury F, Loppin B, Hochberg ME, Boulétreau M. Removing
571 symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. Proceedings
572 of the National Academy of Sciences. 2001;98(11):6247-52.
573 11. Yarmolinsky MB. Programmed cell death in bacterial populations. Science.
574 1995;267(5199):836-8.
575 12. Seidel HS, Ailion M, Li J, van Oudenaarden A, Rockman MV, Kruglyak L. A novel
576 sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in
577 *C. elegans*. PLoS Biol. 2011;9(7):e1001115.
578 13. Shropshire JD, On J, Layton EM, Zhou H, Bordenstein SR. One prophage WO gene
579 rescues cytoplasmic incompatibility in *Drosophila melanogaster*. Proceedings of the National
580 Academy of Sciences. 2018;115(19):4987-91.
581 14. Hughes AL, Piontkivska H. Phylogeny of Trypanosomatidae and Bodonidae
582 (Kinetoplastida) based on 18S rRNA: evidence for paraphyly of *Trypanosoma* and six other
583 genera. Molecular biology and evolution. 2003;20(4):644-52.

584 15. Kostygov AY, Dobáková E, Grybchuk-Ieremenko A, Váhala D, Maslov DA, Votýpka
585 J, et al. Novel trypanosomatid-bacterium association: evolution of endosymbiosis in action.
586 MBio. 2016;7(2):e01985-15.
587 16. Teixeira MM, Borghesan TC, Ferreira RC, Santos MA, Takata CS, Campaner M, et
588 al. Phylogenetic validation of the genera *Angomonas* and *Strigomonas* of trypanosomatids
589 harboring bacterial endosymbionts with the description of new species of trypanosomatids
590 and of proteobacterial symbionts. Protist. 2011;162(3):503-24.
591 17. Alves JM, Serrano MG, Maia da Silva F, Voegtly LJ, Matveyev AV, Teixeira MM, et
592 al. Genome evolution and phylogenomic analysis of *Candidatus* Kinetoplastibacterium, the
593 betaproteobacterial endosymbionts of *Strigomonas* and *Angomonas*. Genome biology and
594 evolution. 2013;5(2):338-50.
595 18. Kostygov AY, Butenko A, Nenarokova A, Tashyreva D, Flegontov P, Lukeš J, et al.
596 Genome of *Ca*. Pandoraea novymonadis, an endosymbiotic bacterium of the trypanosomatid
597 *Novymonas esmeraldas*. Frontiers in microbiology. 2017;8:1940.
598 19. Alves JM, Klein CC, da Silva FM, Costa-Martins AG, Serrano MG, Buck GA, et al.
599 Endosymbiosis in trypanosomatids: the genomic cooperation between bacterium and host in
600 the synthesis of essential amino acids is heavily influenced by multiple horizontal gene
601 transfers. BMC evolutionary biology. 2013;13(1):190.
602 20. Klein CC, Alves JM, Serrano MG, Buck GA, Vasconcelos ATR, Sagot M-F, et al.
603 Biosynthesis of vitamins and cofactors in bacterium-harbouring trypanosomatids depends on
604 the symbiotic association as revealed by genomic analyses. PLoS One. 2013;8(11).
605 21. Harmer J, Yurchenko V, Nenarokova A, Lukeš J, Ginger ML. Farming, slaving and
606 enslavement: histories of endosymbioses during kinetoplastid evolution. Parasitology.
607 2018;145(10):1311-23.

608 22. Vickerman K, Preston TM. Comparative cell biology of the kinetoplastid flagellates.
609 In: Lumsden WHR, Evans DA, editors. Biology of the Kinetoplastida. 1. New York:
610 Academic Press; 1976. p. 35-130.
611 23. Burzell LA. Fine structure of *Bodo curvifilus* Griessmann (Kinetoplastida:
612 Bodonidae). The Journal of protozoology. 1975;22(1):35-9.
613 24. Malysheva M, Karpova M, Frolov A. Mitosis of the free-living flagellate *Bodo*
614 *saltans* strain Ps+ (Kinetoplastidea, Bodonida). Cell and Tissue Biology. 2007;1(4):364-73.
615 25. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence
616 alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
617 26. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu:
618 scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation.
619 Genome research. 2017;27(5):722-36.
620 27. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal
621 RNAs in metatranscriptomic data. Bioinformatics. 2012;28(24):3211-7.
622 28. Treangen TJ, Sommer DD, Angly FE, Koren S, Pop M. Next generation sequence
623 assembly with AMOS. Current Protocols in Bioinformatics. 2011;33(1):11.8. 1-.8. 8.
624 29. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics.
625 2014;30(14):2068-9.
626 30. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a
627 universal tool for annotation, visualization and analysis in functional genomics research.
628 Bioinformatics. 2005;21(18):3674-6.
629 31. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search
630 tool. Nucleic acids research. 2011;39(suppl_2):W347-W52.
631 32. Petersen TN, Brunak S, Von Heijne G, Nielsen H. SignalP 4.0: discriminating signal
632 peptides from transmembrane regions. Nature methods. 2011;8(10):785.

633 33. Eichinger V, Nussbaumer T, Platzer A, Jehl M-A, Arnold R, Rattei T. EffectiveDB—
634 updates and novel features for a better annotation of bacterial secreted proteins and Type III,
635 IV, VI secretion systems. Nucleic acids research. 2016;44(D1):D669-D74.
636 34. Bendtsen JD, Kiemer L, Fausbøll A, Brunak S. Non-classical protein secretion in
637 bacteria. BMC microbiology. 2005;5(1):58.
638 35. Li H, Benedito VA, Udvardi MK, Zhao PX. TransportTP: a two-phase classification
639 approach for membrane transporter prediction and characterization. BMC bioinformatics.
640 2009;10(1):418.
641 36. Saier Jr MH, Tran CV, Barabote RD. TCDB: the Transporter Classification Database
642 for membrane transport protein analyses and information. Nucleic acids research.
643 2006;34(suppl_1):D181-D6.
644 37. Söding J. Protein homology detection by HMM–HMM comparison. Bioinformatics.
645 2005;21(7):951-60.
646 38. wwPDBconsortium. Protein Data Bank: the single global archive for 3D
647 macromolecular structure data. Nucleic acids research. 2019;47(D1):D520-D8.
648 39. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam
649 protein families database in 2019. Nucleic acids research. 2019;47(D1):D427-D32.
650 40. Eddy SR. A new generation of homology search tools based on probabilistic
651 inference. Genome Informatics 2009: Genome Informatics Series Vol 23: World Scientific;
652 2009. p. 205-11.
653 41. Mirdita M, von den Driesch L, Galiez C, Martin MJ, Söding J, Steinegger M. Uniclust
654 databases of clustered and deeply annotated protein sequences and alignments. Nucleic acids
655 research. 2017;45(D1):D170-D6.

656 42. Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H, et al. Introducing EzBioCloud: a
657 taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies.
658 International journal of systematic and evolutionary microbiology. 2017;67(5):1613.
659 43. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for
660 eukaryotic genomes. Genome research. 2003;13(9):2178-89.
661 44. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW
662 and ClustalX. Current protocols in bioinformatics. 2003(1):2.3. 1-2.3. 22.
663 45. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit
664 models of protein evolution. Bioinformatics. 2011;27(8):1164-5.
665 46. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
666 large phylogenies. Bioinformatics. 2014;30(9):1312-3.
667 47. Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for
668 specialized analyses of microbial genomes and metagenomes. PeerJ Preprints; 2016. Report
669 No.: 2167-9843.
670 48. Arai W, Taniguchi T, Goto S, Moriya Y, Uehara H, Takemoto K, et al. MAPLE 2.3.
671 0: an improved system for evaluating the functionomes of genomes and metagenomes.
672 Bioscience, biotechnology, and biochemistry. 2018;82(9):1515-7.
673 49. Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple
674 sequence alignment of ribosomal RNA genes. Bioinformatics. 2012;28(14):1823-9.
675 50. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a
676 comprehensive online resource for quality checked and aligned ribosomal RNA sequence
677 data compatible with ARB. Nucleic acids research. 2007;35(21):7188-96.
678 51. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA
679 ribosomal RNA gene database project: improved data processing and web-based tools.
680 Nucleic acids research. 2012;41(D1):D590-D6.

681 52. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, et al. ARB: a
682 software environment for sequence data. Nucleic acids research. 2004;32(4):1363-71.
683 53. Loy A, Arnold R, Tischler P, Rattei T, Wagner M, Horn M. probeCheck–a central
684 resource for evaluating oligonucleotide probe coverage and specificity. Environmental
685 microbiology. 2008;10(10):2894-8.
686 54. Yilmaz LS, Parnerkar S, Noguera DR. mathFISH, a web tool that uses
687 thermodynamics-based mathematical models for in silico evaluation of oligonucleotide
688 probes for fluorescence in situ hybridization. Appl Environ Microbiol. 2011;77(3):1118-22.
689 55. Booth DG, Beckett AJ, Molina O, Samejima I, Masumoto H, Kouprina N, et al. 3D-
690 CLEM reveals that a major portion of mitotic chromosomes is not chromatin. Molecular cell.
691 2016;64(4):790-802.
692 56. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al.
693 Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of
694 life. Nature microbiology. 2017;2(11):1533-42.
695 57. Rodriguez-R LM, Konstantinidis KT. Bypassing cultivation to identify bacterial
696 species. Microbe. 2014;9(3):111-8.
697 58. George EE, Husnik F, Tashyreva D, Prokopchuk G, Horák A, Kwong WK, et al.
698 Highly reduced genomes of protist endosymbionts show evolutionary convergence. Current
699 Biology. 2020.
700 59. Schmitz-Esser S, Tischler P, Arnold R, Montanaro J, Wagner M, Rattei T, et al. The
701 genome of the amoeba symbiont "*Candidatus* Amoebophilus asiaticus" reveals common
702 mechanisms for host cell interaction among amoeba-associated bacteria. Journal of
703 bacteriology. 2010;192(4):1045-57.

704 60. Brown A, Wasala SK, Howe DK, Peetz AB, Zasada IA, Denver DR. Comparative
705 genomics of *Wolbachia–Cardinium* dual endosymbiosis in a plant-parasitic nematode.
706 Frontiers in microbiology. 2018;9:2482.
707 61. Yang JC, Madupu R, Durkin AS, Ekborg NA, Pedamallu CS, Hostetler JB, et al. The
708 complete genome of *Teredinibacter turnerae* T7901: an intracellular endosymbiont of marine
709 wood-boring bivalves (shipworms). PloS one. 2009;4(7).
710 62. Wang J, Yang B, Leier A, Marquez-Lago TT, Hayashida M, Rocker A, et al.
711 Bastion6: a bioinformatics approach for accurate prediction of type VI secreted effectors.
712 Bioinformatics. 2018;34(15):2546-55.
713 63. Natale P, Brüser T, Driessen AJ. Sec-and Tat-mediated protein secretion across the
714 bacterial cytoplasmic membrane—distinct translocases and mechanisms. Biochimica et
715 Biophysica Acta (BBA)-Biomembranes. 2008;1778(9):1735-56.
716 64. Jackson AP, Thomas GH, Parkhill J, Thomson NR. Evolutionary diversification of an
717 ancient gene family (rhs) through C-terminal displacement. BMC genomics. 2009;10(1):584.
718 65. Ruhe ZC, Low DA, Hayes CS. Bacterial contact-dependent growth inhibition. Trends
719 in microbiology. 2013;21(5):230-7.
720 66. Jamet A, Nassif X. New players in the toxin field: polymorphic toxin systems in
721 bacteria. MBio. 2015;6(3).
722 67. Zhang D, de Souza RF, Anantharaman V, Iyer LM, Aravind L. Polymorphic toxin
723 systems: comprehensive characterization of trafficking modes, processing, mechanisms of
724 action, immunity and ecology using comparative genomics. Biology direct. 2012;7(1):18.
725 68. Ibrahim M, Subramanian A, Anishetty S. Comparative pan genome analysis of oral
726 *Prevotella* species implicated in periodontitis. Functional & integrative genomics.
727 2017;17(5):513-36.

728 69. Votýpka J, Kostygov AY, Kraeva N, Grybchuk-Ieremenko A, Tesařová M, Grybchuk
729 D, et al. *Kentomonas* gen. n., a New Genus of Endosymbiont-containing Trypanosomatids of
730 Strigomonadinae subfam. n. Protist. 2014;165(6):825-38.
731 70. Alves JM, Voegtly L, Matveyev AV, Lara AM, da Silva FM, Serrano MG, et al.
732 Identification and phylogenetic analysis of heme synthesis genes in trypanosomatids and their
733 bacterial endosymbionts. PLos one. 2011;6(8).
734 71. Fujishima M, Kawai M, Yamamoto R. *Paramecium caudatum* acquires heat-shock
735 resistance in ciliary movement by infection with the endonuclear symbiotic bacterium
736 Holospora obtusa. FEMS microbiology letters. 2005;243(1):101-5.
737 72. Fujishima M. Infection and maintenance of Holospora species in *Paramecium*
738 *caudatum*. Endosymbionts in Paramecium: Springer; 2009. p. 201-25.
739 73. Görtz H, Fujishima M. Conjugation and meiosis of *Paramecium caudatum* infected
740 with the micronucleus-specific bacterium *Holospora elegans*. European journal of cell
741 biology. 1983;32(1):86-91.
742 74. Restif O, Kaltz O. Condition-dependent virulence in a horizontally and vertically
743 transmitted bacterial parasite. Oikos. 2006;114(1):148-58.
744 75. Schulz F, Martijn J, Wascher F, Lagkouvardos I, Kostanjšek R, Ettema TJ, et al. A
745 Rickettsiales symbiont of amoebae with ancient features. Environ Microbiol.
746 2016;18(8):2326-42.
747 76. Kusch J, Görtz H-D. Towards an understanding of the killer trait: *Caedibacter*
748 endocytobionts in Paramecium. Molecular Basis of Symbiosis: Springer; 2005. p. 61-76.
749 77. Werren JH. Selfish genetic elements, genetic conflict, and evolutionary innovation.
750 Proceedings of the National Academy of Sciences. 2011;108(Supplement 2):10863-70.

751 78. Engelberg-Kulka H, Hazan R, Amitai S. mazEF: a chromosomal toxin-antitoxin
752 module that triggers programmed cell death in bacteria. Journal of cell science.
753 2005;118(19):4327-32.
754 79. Page R, Peti W. Toxin-antitoxin systems in bacterial growth arrest and persistence.
755 Nature chemical biology. 2016;12(4):208.
756 80. Dedeine F, Bouletreau M, Vavre F. *Wolbachia* requirement for oogenesis: occurrence
757 within the genus *Asobara* (Hymenoptera, Braconidae) and evidence for intraspecific variation
758 in *A. tabida*. Heredity. 2005;95(5):394-400.
759 81. Dedeine F, Vavre F, Shoemaker DD, Boulétreau M. Intra-individual coexistence of a
760 *Wolbachia* strain required for host oogenesis with two strains inducing cytoplasmic
761 incompatibility in the wasp *Asobara tabida*. Evolution. 2004;58(10):2167-74.
762 82. Aanen DK, Hoekstra RF. The evolution of obligate mutualism: if you can't beat'em,
763 join'em. Trends in Ecology & Evolution. 2007;22(10):506-9.
764 83. Hess S, Suthaus A, Melkonian M. "*Candidatus* Finniella"(Rickettsiales,
765 Alphaproteobacteria), novel endosymbionts of viridiraptorid amoeboflagellates (Cercozoa,
766 rhizaria). Appl Environ Microbiol. 2016;82(2):659-70.
767 84. Georgiades K, Madoui M-A, Phuong Le CR, Raoult D. Phylogenomic analysis of
768 *Odyssella thessalonicensis* fortifies the common origin of Rickettsiales, *Pelagibacter ubique*
769 and *Reclimonas americana* mitochondrion. PloS one. 2011;6(9).
770 85. Kim E, Park JS, Simpson AG, Matsunaga S, Watanabe M, Murakami A, et al.
771 Complex array of endobionts in *Petalomonas sphagnophila*, a large heterotrophic euglenid
772 protist from Sphagnum-dominated peatlands. The ISME journal. 2010;4(9):1108-20.
773 86. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and
774 linear interactive genome visualization. Bioinformatics. 2009;25(1):119-20.

775 87. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al.
776 Fiji: an open-source platform for biological-image analysis. Nature methods. 2012;9(7):676-
777 82.

778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799

# Figure Legends

**Figure 1: Annotation features of Bodo endosymbiont**

**A.** Graphical representation of *Candidatus* Bodocaedibacter vickermanii (Cbv) genome, generated using DNAPlotter v10.2 (86). Moving inwards, the tracks represent forward and reverse CDS, tRNAs, type VI secretion system (T6SS) encoding genes, hypothetical proteins (conserved in the *Candidatus* Paracaedibacteraceae 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 2: Visualization of intracellular bacteria in *Bodo saltans***

Images from Fluorescent *in situ* hybridization (FISH) experiment (A-D) and transmission electron microscopy (TEM) experiment (E-G). Imaging of *B. saltans* cell cultures with differential interference contrast (A), staining with DAPI to visualize nucleus and kinteoplast

(B), FISH staining with endosymbiont specific probe conjugated to Cy3 (C) and images from three channels overlaid (D). 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). Images were acquired on a Zeiss Axio Observer Z1 (Carl Zeiss AG, Jena, Germany) equipped with 100x 1.4NA objective, 2.5x optovar. Images captured were analyzed using ImageJ v2.0 (87).

**Figure 3: 3D model of *Bodo saltans* cell generated using Serial Block Face-Scanning Electron Microscopy.**

Animations of *B*. *saltans* cellular ultrastructure in longitudinal section. Two cells are shown with differences in endosymbiont number and distribution: cell 1 (A-E) and cell 2 (F-J). For each cell, one image with cell envelope (A, F) and four longitudinal sections with consecutive 90-degree rotation around radial axis are shown (B-E and G-J). Cell envelope (in red) shows the bean shaped structure of *B*. *saltans* and two flagella emerging from flagellar pocket (in orange). Nucleus (in yellow) is situated in the centre of the cell and cytobionts (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.

**Figure 4: 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 outgroup (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.

**Figure 5: Polymorphic toxin systems in *Ca.* B. vickermanii**

Schematic diagrams of the organization of the three predicted polymorphic toxin systems in *Ca*. B. vickermanii. In each case, the putative, alternative orphan toxin modules are shown downstream of the main multi-domain 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. Gray scale gradient represents identical (black), similar (dark gray) and dissimilar (light gray) residues in the alignment based on the Blosum62 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.

**Figure 6: Rifampicin treatment of kinetoplastids with and without endosymbionts**

874 Three kinetoplastids were treated with rifampicin and observed for their behavior on
875 treatment. *B*. *saltans* possesses an endosymbont, *T*. *theileri* and *L*. *costaricensis* are parasitic
876 species that do not. The graph shows number of cells before treatment and after 24 hours of
877 treatment. The bars are average of three independent experiments. Statistical significance was
878 calculated using ratio paired T-test.
879

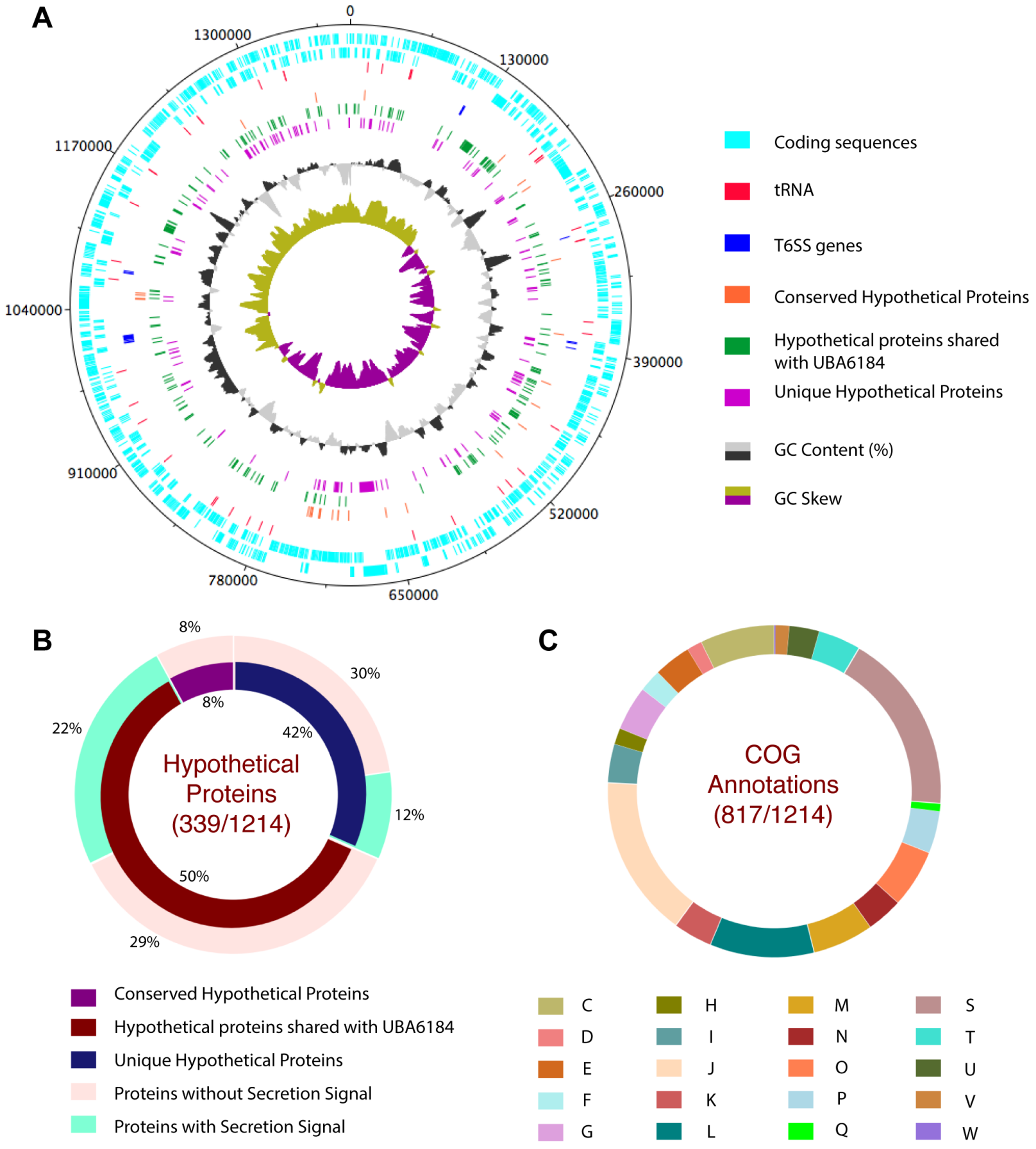


Figure 1

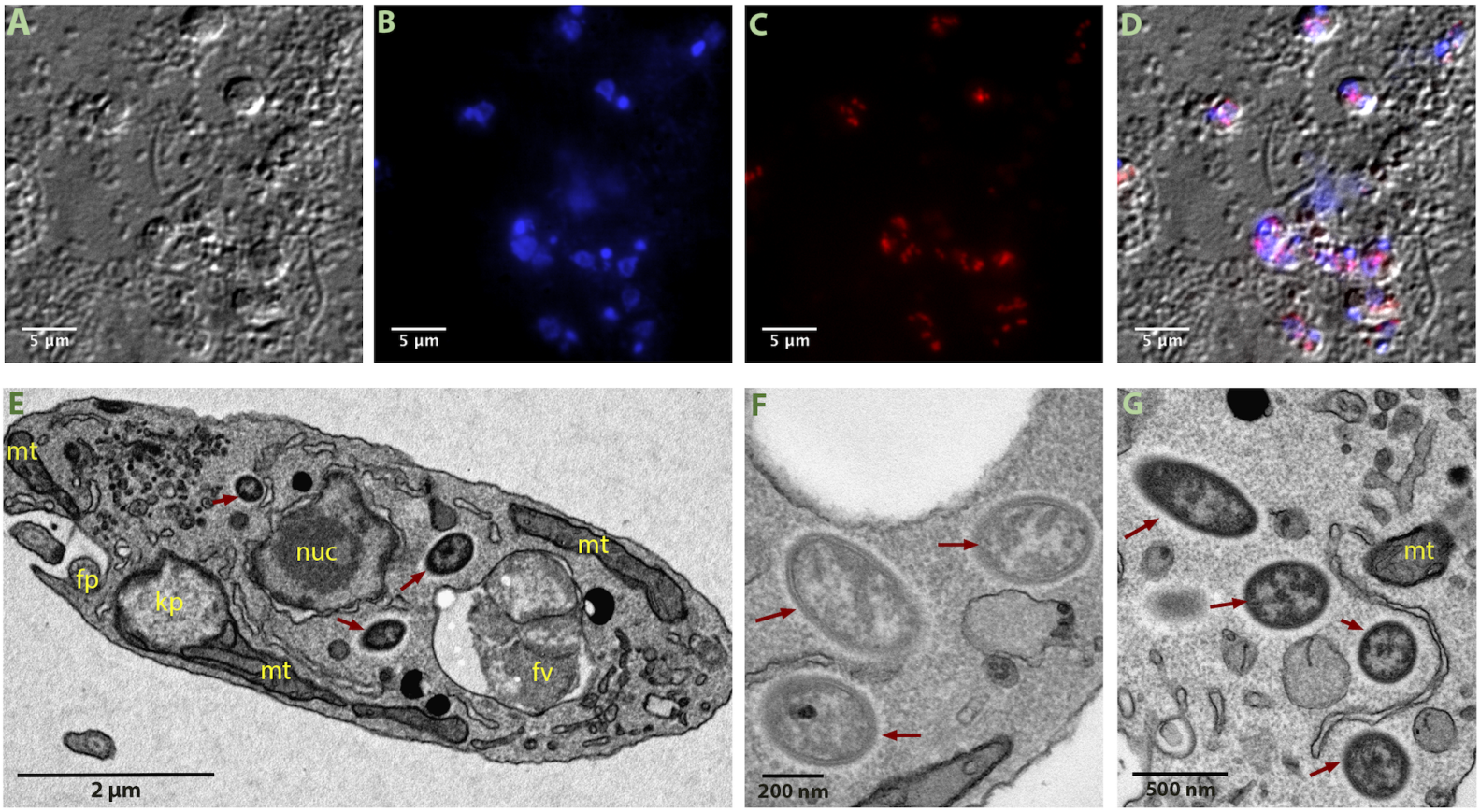


Figure 2

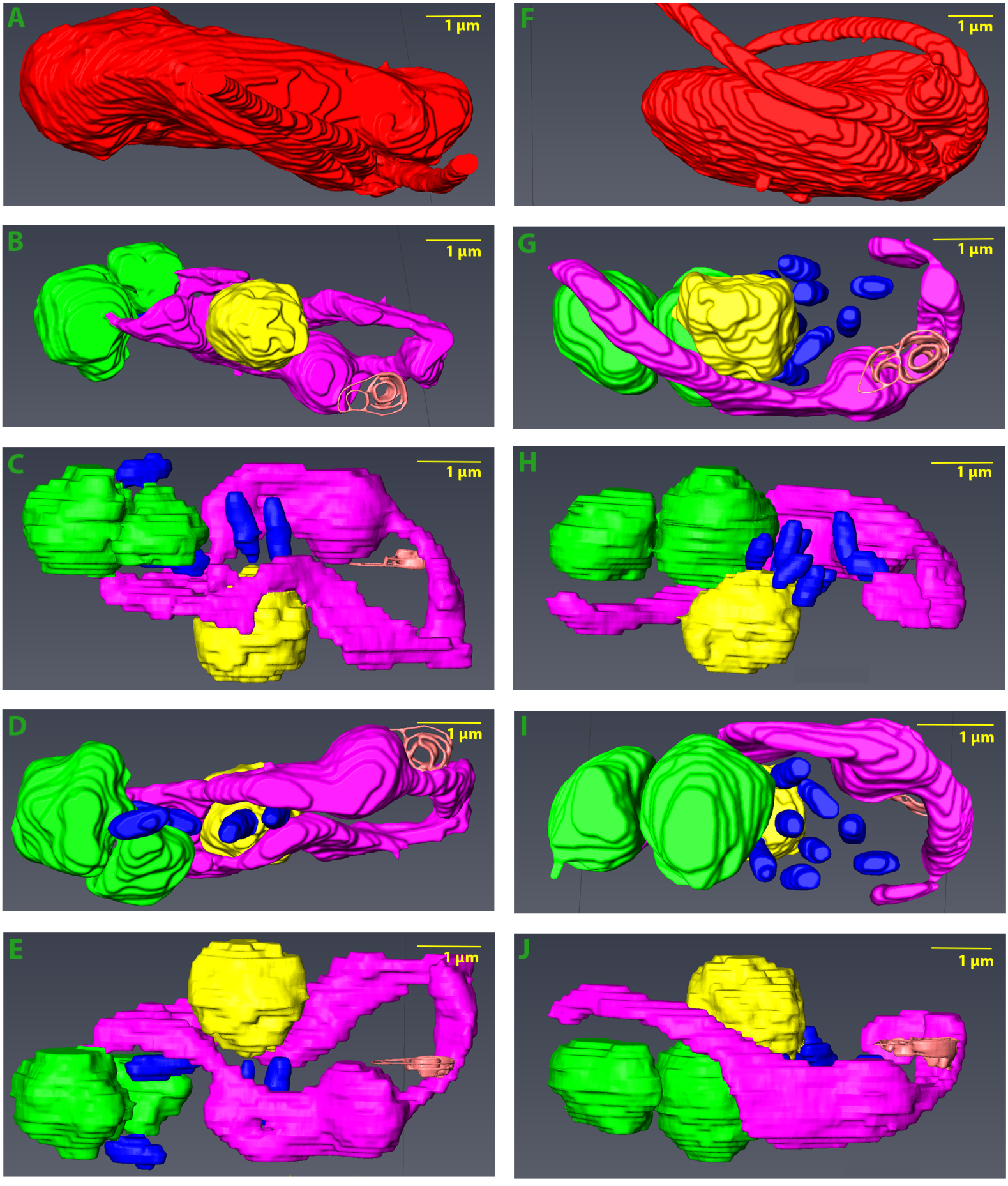


Figure 3

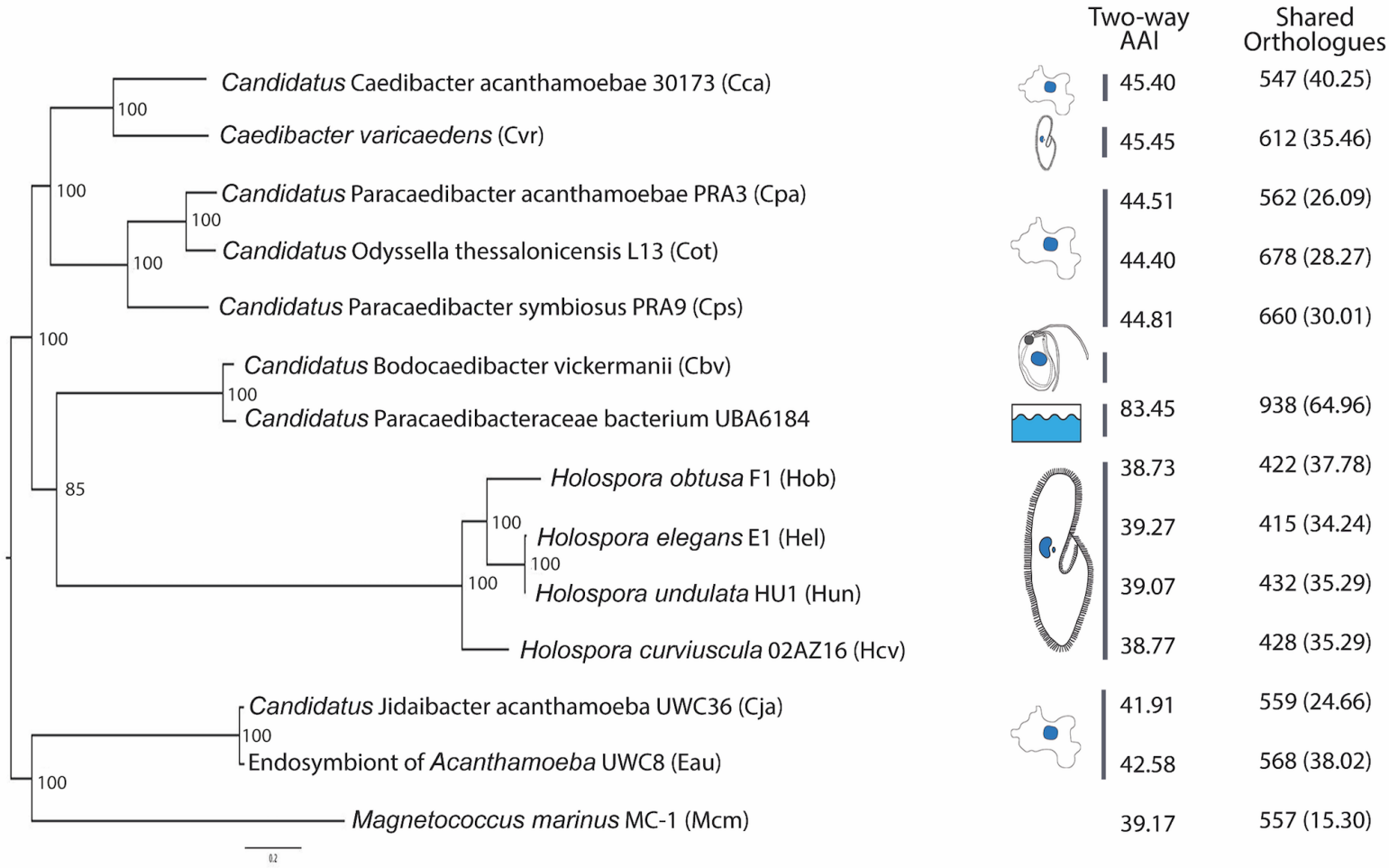


Figure 4

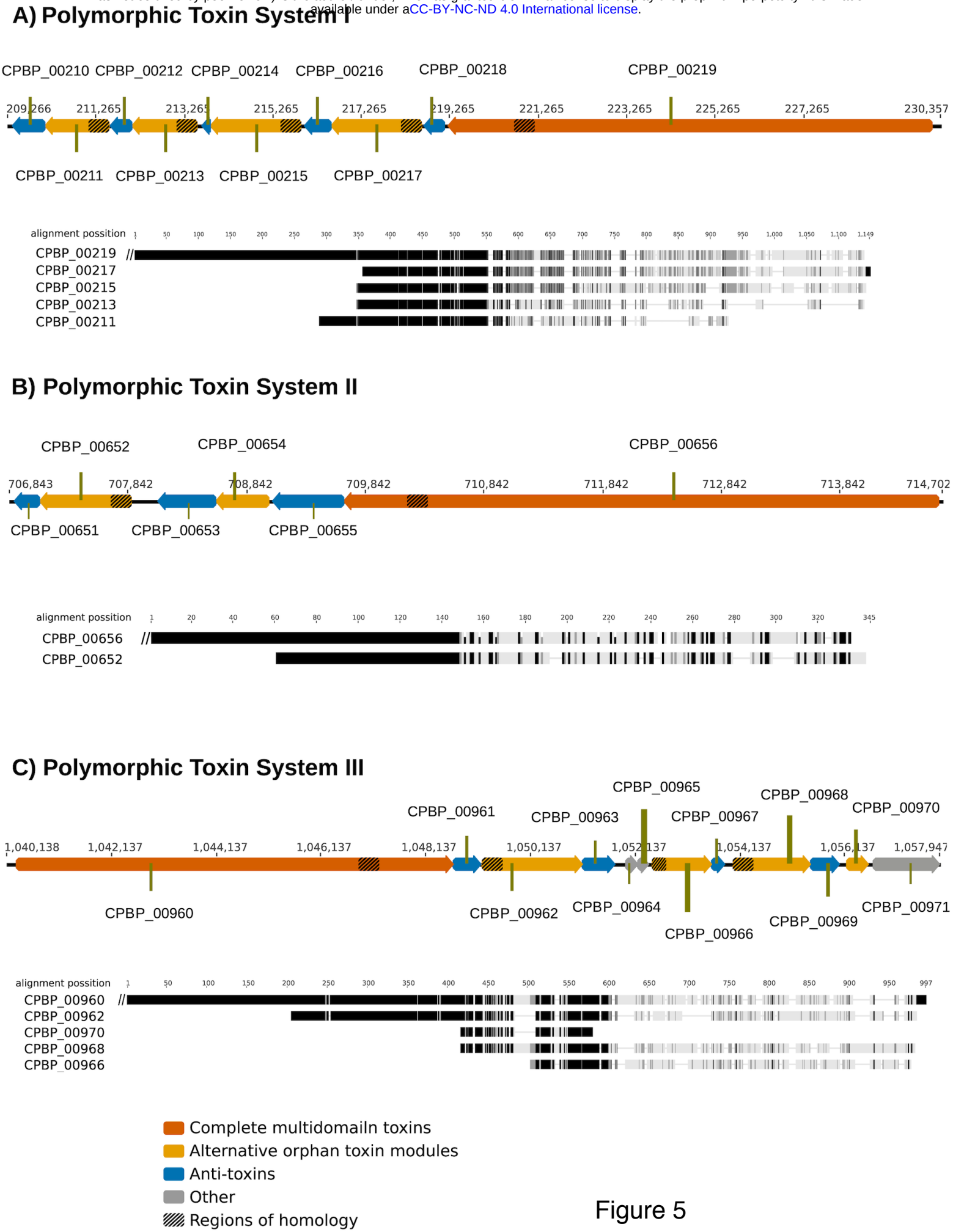


Figure 5

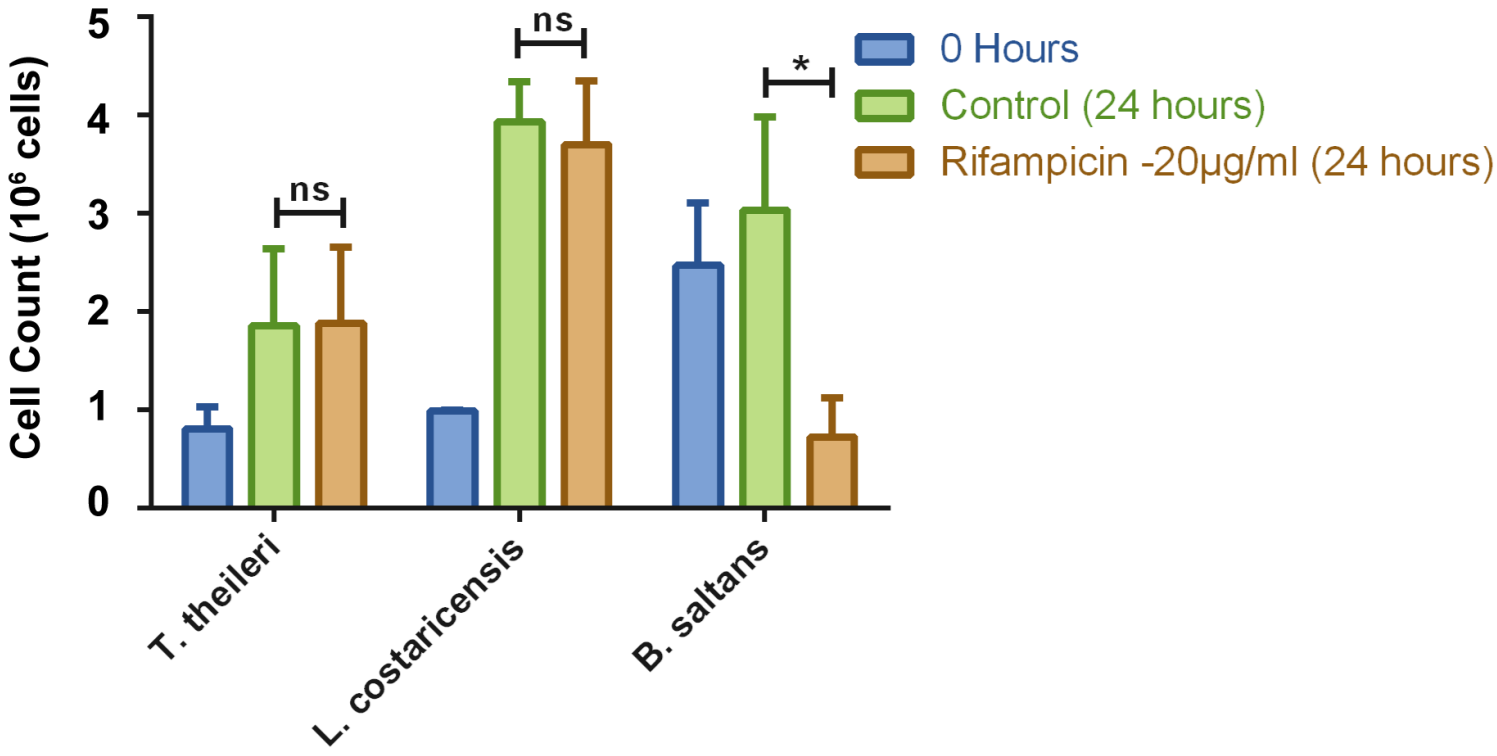


Figure 6

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

| S.No. | Organism | GC content (%) | Size (Mb) | CDS | tRNA | rRNA | Family | Order | Phylum | Host/Isolation source | Accession number |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | *Candidatus* Caedibacter acanthamoebae 30173 | 38.2 | 1.72 | 1359 | 42 | 3 | Caedimonadaceae | Holosporales | Proteobacteria | *Acanthamoeba polyphaga HN-3* | CP008936 |
| 2 | *Caedibacter varicaedens* | 42.1 | 1.69 | 1726 | 41 | 3 | Caedimonadaceae | Holosporales | Proteobacteria | *Paramecium biaurelia* | BBVC00000000 |
| 3 | *Candidatus* Paracaedibacter symbiosus PRA9 | 41.5 | 2.66 | 2199 | 43 | 6 | Paracaedibacteraceae | Holosporales | Proteobacteria | *Acanthamoeba sp. UWET39* | JQAK00000000 |
| 4 | *Candidatus* Odyssella thessalonicensis L13 | 42 | 2.85 | 2398 | 40 | 3 | Paracaedibacteraceae | Holosporales | Proteobacteria | *Acanthamoeba* | AEWF00000000 |
| 5 | *Candidatus* Paracaedibacter acanthamoebae PRA3 | 41.0 | 2.47 | 2154 | 43 | 6 | Paracaedibacteraceae | Holosporales | Proteobacteria | *Acanthamoeba* | CP008941 |
| 6 | *Candidatus* Bodocaedibacter vickermanii | 40.7 | 1.39 | 1214 | 40 | 6 | Paracaedibacteraceae | Holosporales | Proteobacteria | *Bodo saltans* | CP054719 (this study) |
| 7 | *Candidatus* Paracaedibacteraceae bacterium UBA6184 | 40.1 | 1.47 | 1444 | 30 | - | Paracaedibacteraceae | Holosporales | Proteobacteria | waste water | DITM00000000 |
| 8 | *Holospora curviuscula* 02AZ16 | 38.1 | 1.72 | 1534 | 42 | 3 | Holosporaceae | Holosporales | Proteobacteria | *Paramecium bursaria* | PHHC00000000 |
| 9 | *Holospora obtusa* F1 | 35.2 | 1.33 | 1249 | 36 | 3 | Holosporaceae | Holosporales | Proteobacteria | *Paramecium caudatum* | AWTR00000000 |
| 10 | *Holospora undulata* HU1 | 36.1 | 1.40 | 1460 | 44 | 3 | Holosporaceae | Holosporales | Proteobacteria | *Paramecium caudatum* | ARPM00000000 |
| 11 | *Holospora elegans* E1 | 36.0 | 1.27 | 1454 | 37 | 3 | Holosporaceae | Holosporales | Proteobacteria | *Paramecium caudatum* | BAUP00000000 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | Endosymbiont of *Acanthamoeba* UWC8 | 34.8 | 1.62 | 1494 | 36 | 3 | Candidatus Midichloriaceae | Rickettsiales | Proteobacteria | *Acanthamoeba* | CP004403 |
| 13 | *Candidatus* Jidaibacter acanthamoeba UWC36 | 33.7 | 2.37 | 2267 | 36 | 3 | Candidatus Midichloriaceae | Rickettsiales | Proteobacteria | *Acanthamoeba sp. UWC36* | JSWE00000000 |
| 14 | *Magnetococcus marinus* MC-*1* | 54.2 | 4.72 | 3766 | 45 | 9 | Magnetococcaceae | Magnetococcales | Proteobacteria | marine water | CP000471 |

## Table 2: Annotation of polymorphic toxin systems by distant homology detection with HHsearch

| Gene | Module position | Description based on position and/or matching families | Matching Pfam or PDB entry suggestive of toxin or antitoxin function | Matching probability (%), sequence identity of HHsearch alignment |
|---|---|---|---|---|
| ***Polymorphic Toxin System I*** | | | | |
| CPBP_00219 | | Complete multidomain toxin | AHH nuclease domain (PF14412) | 93, 23 |
| CPBP_00218 | | Antitoxin | Gmx_para_CXXCG (PF09535); DUF1629 (PF07791); Immunity protein 43 (PF15570) | 100, 14;<br>99, 15;<br>96, 16 |
| CPBP_00217 | Orphan module 1 | Alternate toxin domain | Tox-HNH-HHH nuclease domain (PF15637) | 100, 34 |
| CPBP_00216 | | Antitoxin | SUKH superfamily 5 (PF14567) | 88, 15 |
| CPBP_00215 | Orphan module 2 | Alternate toxin domain | Cloacin; Colicin-like bacteriocin tRNase domain (PF03515) | 99, 6 |
| CPBP_00214 | | Antitoxin | Colicin/pyocin immunity protein (PF01320) | 100,16 |
| CPBP_00213 | Probable orphan module 3 | Alternate toxin domain | *No strong match to any family or structure* | |
| CPBP_00212 | | Antitoxin | *No strong match to any family or structure* | |
| CPBP_00211 | Orphan module 4 | Alternate toxin domain | DUF1837 (PF08878) | 81, 16 |
| CPBP_00210 | | Antitoxin | Immunity protein 49 (PF15575) | 100,12 |

| | | | | |
|---|---|---|---|---|
| ***Polymorphic Toxin System II*** | | | | |
| CPBP_00656 | | Complete multidomain toxin | Tox-HNH-HHH nuclease domain (PF15637) | 99, 40 |
| CPBP_00655 | | Antitoxin | SUKH superfamily 5 (PF14567) | 91, 15 |
| CPBP_00654 | Orphan module 1 | Alternate toxin domain | DNase/tRNase domain of colicin-like bacteriocin (PF12639) | 100, 28 |
| CPBP_00653 | | Antitoxin | SUKH superfamily 5 (PF14567) | 100, 12 |
| CPBP_00652 | Orphan module 2 | Alternate toxin domain | S-type pyocin (PF06958) | 99, 9 |
| CPBP_00651 | | Antitoxin | Colicin/pyocin immunity protein (PF01320) | 100,13 |
| ***Polymorphic Toxin System III*** | | | | |
| CPBP_00960 | | Presumed complete multidomain toxin | *No strong match to any family or structure* | |
| CPBP_00961 | | Presumed antitoxin | *No strong match to any family or structure* | |
| CPBP_00962 | Possible Orphan module 1 | Alternate toxin domain | *No strong match to any family or structure* | |
| CPBP_00963 | | Antitoxin | DUF3885 (PF13021) | 100, 32 |
| CPBP_00964 | Not an orphan module | - | *Short, uninformative matches against a variety of Pfam and PDB entries* | |
| CPBP_00965 | | - | *Short, uninformative matches against a variety of Pfam and PDB entries* | |
| CPBP_00966 | Possible Orphan module 2 | Alternate toxin domain | Metallopeptidase toxin 5 (PF15641) | 98, 12 |
| CPBP_00967 | | Antitoxin | Colicin-like immunity protein (PF09204) | 92,13 |
| CPBP_00968 | Possible Orphan module 3 | Alternate toxin domain | *No strong match to any family or structure* | |
| CPBP_00969 | | Antitoxin | *No strong match to any family or structure* | |

| | | | | |
|---|---|---|---|---|
| CPBP_00970 | Unlikely Orphan module 4 | Alternate toxin domain | *No strong match to any family or structure* | |
| CPBP_00971 | | Antitoxin | Metalloprotease (PF03410) | |