Gene Transfer Agents in Bacterial Endosymbionts of Microbial Eukaryotes

Emma E. George [1,*], Daria Tashyreva[2], Waldan K. Kwong[1,3], Noriko Okamoto [1,4], Aleš Horák[2,5], Filip Husnik [1,6], Julius Lukeš [2,5], and Patrick J. Keeling[1]

[1]Department of Botany, University of British Columbia, Vancouver, Canada

[2]Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice (Budweis), Czech Republic

[3]Instituto Gulbenkian de Ciência, Oeiras, Portugal

[4]Hakai Institute, Quadra Island, British Columbia, Canada

[5]Faculty of Sciences, University of South Bohemia, České Budějovice (Budweis), Czech Republic

[6]Okinawa Institute of Science and Technology, Okinawa, Japan

*Corresponding author: E-mail: 3mma6eorg3@gmail.com.

Accepted: 18 June 2022

Abstract

Gene transfer agents (GTAs) are virus-like structures that package and transfer prokaryotic DNA from donor to recipient prokaryotic cells. Here, we describe widespread GTA gene clusters in the highly reduced genomes of bacterial endosymbionts from microbial eukaryotes (protists). Homologs of the GTA capsid and portal complexes were initially found to be present in several highly reduced alphaproteobacterial endosymbionts of diplonemid protists (*Rickettsiales* and *Rhodospirillales*). Evidence of GTA expression was found in polyA-enriched metatranscriptomes of the diplonemid hosts and their endosymbionts, but due to biases in the polyA-enrichment methods, levels of GTA expression could not be determined. Examining the genomes of closely related bacteria revealed that the pattern of retained GTA head/capsid complexes with missing tail components was common across *Rickettsiales* and *Holosporaceae* (*Rhodospirillales*), all obligate symbionts with a wide variety of eukaryotic hosts. A dN/dS analysis of *Rickettsiales* and *Holosporaceae* symbionts revealed that purifying selection is likely the main driver of GTA evolution in symbionts, suggesting they remain functional, but the ecological function of GTAs in bacterial symbionts is unknown. In particular, it is unclear how increasing horizontal gene transfer in small, largely clonal endosymbiont populations can explain GTA retention, and, therefore, the structures may have been repurposed in endosymbionts for host interactions. Either way, their widespread retention and conservation in endosymbionts of diverse eukaryotes suggests an important role in symbiosis.

Key words: gene transfer agent, endosymbiosis, *Rickettsiales*, *Holosporaceae*, protist, evolution.

Significance

Gene transfer agents (GTAs) provide a mechanism of gene transfer in free-living bacterial populations, but very little is known about GTAs in bacterial endosymbionts, especially endosymbionts of microbial eukaryotes. A small, subset of GTA genes is present in some endosymbionts with reduced genomes, but whether or not these GTAs remain functional is unknown. Here, we provide evidence for functional GTAs in bacterial endosymbionts with extremely reduced genomes and discuss the potential roles of GTAs in symbiosis.

Introduction

Interactions between bacteria and bacteriophages over billions of years have resulted in diverse outcomes, including the repurposing of phage structures by bacteria. Structures such as the Type VI Secretion Systems and tailocins derived from ancestral bacteriophages (phages) are used for multiple functions, including interspecific competition and infection of eukaryotic cells (Leiman et al. 2009; Ghequire and De Mot 2015). Another repurposed phage-like structure, called a gene transfer agent (GTA), packages and transfers bacterial DNA from donor to recipient cells within bacterial populations, and serves as an important mode of horizontal gene transfer (HGT) (McDaniel et al. 2010; Hynes et al. 2012; Lang et al. 2012; Westbye et al. 2017; Sherlock et al. 2019). Like other methods of genetic recombination in bacteria (e.g., transduction, conjugation, and transformation), GTAs can provide gains of function to recipient bacterial cells, but this comes with a cost: the donor bacterial cells are lysed during the release of GTA particles (Fogg et al. 2012; Westbye et al. 2013). Despite this detrimental aspect, GTAs have evolved multiple times independently in both Bacteria and Archaea, and they are common in *Alphaproteobacteria* where the best-studied GTA system, *Rhodobacter capsulatus* (RcGTA), has linked GTA production to cell regulatory processes (Sherlock et al. 2019).

Intriguingly, small clusters (SCs) of GTA genes have been detected in the genomes of certain *Rickettsiales* endosymbionts of animals (Lang and Beatty 2007; Shakya et al. 2017; Christensen and Serbus 2020). *Rickettsiales* are all obligate intracellular symbionts, except for one extracellular symbiont (Castelli et al. 2019), but what the function of GTAs or gene transfer in general might be in these highly reduced endosymbionts is unclear. The obligate endosymbionts are thought to reproduce in small, mostly clonal populations within a cell (Russell and Cavanaugh 2017), and deleterious mutations can accumulate and become fixed in these small populations through a process known as Muller's ratchet (Moran et al. 1996; Naito and Pawlowska 2016). GTAs may slow this process and provide gains of function through HGT, as suggested for the role of the non-related GTA in the facultative intracellular parasite, *Bartonella* (Québatte and Dehio 2019). The *Bartonella* GTA (BaGTA) is also linked to the regulation of parasite's pathogenicity and interactions with eukaryotic hosts (Québatte and Dehio 2019). Bacteriophage (phage) infection is another mechanism of HGT in obligate endosymbionts, but only a few phages are known to infect *Rickettsiales* endosymbionts, including the well-studied *Wolbachia*-infecting phage WO (Bordenstein and Bordenstein 2016).

Despite the unknown function of GTAs in obligate endosymbionts, GTAs are highly conserved in *Rickettsiales* endosymbionts of animals, and several species encode a limited set of GTA genes with terminase and head homologs along with many tail-related components. Some tail components are partly missing, whereas others appear to be pseudogenes in some taxa, altogether making the possibility that *Rickettsiales* GTAs are in the process of being reduced or eliminated impossible to rule out without additional functional data (Lang and Beatty 2007; Shakya et al. 2017; Christensen and Serbus 2020). Alternatively, the tail homologs may be too divergent for standard bioinformatics approaches to detect or the recruitment of non-related GTA proteins for tail component functions has occurred.

The best-studied *Rickettsiales* are pathogenic symbionts of animals (Sassera et al. 2006; Pilgrim et al. 2017; Zaila et al. 2017), but recent work has shown that most of the diversity of this group are endosymbionts of microbial eukaryotes (Husnik et al. 2021; Giannotti et al. 2022), and the presence of GTAs in these *Rickettsiales*, and in other predominantly endosymbiotic bacterial lineages, has not been explored. The *Rickettsiales* symbionts of protists infect diverse host groups including amoebae (Schulz et al. 2016), ciliates (Floriano et al. 2018), algae (Yurchenko et al. 2018), and various flagellates (George et al. 2020). Phage-related genes are reported from several *Rickettsiales* endosymbionts of protists (Floriano et al. 2018; Castelli et al. 2019, 2021), and are also found in another endosymbiotic group of *Alphaproteobacteria* called *Holosporaceae* (Garushyants et al. 2018; Midha et al. 2021; Castelli et al. 2022). *Holosporaceae* is a subgroup of the order *Rhodospirillales* (Muñoz-Gómez et al. 2019), and is also made up of obligate endosymbionts that infect diverse eukaryotes including animals (Nunan et al. 2013), amoebae (Schulz et al. 2014), ciliates (Dohra et al. 2014; Castelli et al. 2022), rhizarians (Muñoz-Gómez et al. 2019), and a variety of flagellates (George et al. 2020; Midha et al. 2021). Here we show that these phage genes are, in fact, homologues of GTAs, and report evidence for the expression of and selection on GTA genes across the diversity of *Rickettsiales* and *Holosporaceae* endosymbionts of eukaryotes.

Results and Discussion

Applying DELTA-BLAST (Altschul et al. 1990) and Hidden Markov Models (Finn et al. 2011) to the reduced genomes (605–632 kbp) of bacterial endosymbionts from marine microbial eukaryotes called diplonemids (Tashyreva et al. 2018; Prokopchuk et al. 2019; George et al. 2020), we identified GTA genes in the *Holosporaceae* (*Rhodospirillales*) endosymbionts *Cytomitobacter primus*, *C. indipagum*, and *Nesciobacter abundans*, as well as in the *Rickettsiaceae* (*Rickettsiales*) endosymbiont *Sneabacter namystus* (fig. 1; hereafter, bacterial taxa will be referred to

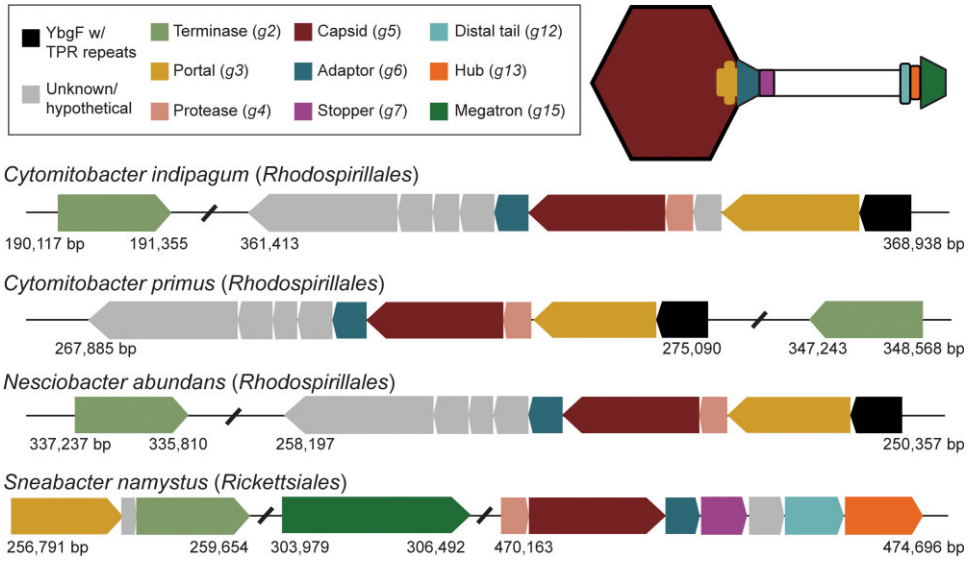

Fig. 1.—Gene transfer agent homologs in bacterial endosymbionts of diplonemids. GTA gene clusters encoded by four bacterial endosymbionts of diplonemids. All GTA clusters are downstream of the potential cell cycle gene, *ybgF*, in the *Rhodospirillales* endosymbionts. Multiple GTA clusters are encoded throughout the *Sneabacter namystus* (*Rickettsiales*) genome. Colors in the legend correspond to genes and GTA structural components. Missing or undetected components in the GTA structure are shown in white with black outlines.

without the *Candidatus* prefix). All endosymbionts of diplonemids encoded genes related to the GTA capsid and portal complex, but only the *Rickettsiaceae* endosymbiont, *S. namystus*, encoded identifiable tail-related homologs. A single GTA cluster was syntenic across the *Holosporaceae* endosymbiont genomes, whereas several GTA clusters were encoded throughout the *S. namystus* genome (fig. 1), a pattern common in *Rickettsiaceae* endosymbionts of animals (Lang and Beatty 2007; Shakya et al. 2017; Christensen and Serbus 2020). The large terminase in the *Holosporaceae* endosymbionts was not encoded near the main GTA cluster which is generally uncommon, but singlet homologs of GTA terminases have been observed in the genomes of *Rickettsiales* endosymbionts and free-living *Rhodospirillales* (Christensen and Serbus 2020).

The retention of the capsid–portal structure without other tail components like the tail tube suggests that the GTA structure is either reduced in the diplonemid endosymbionts or that the tail-related proteins are highly divergent and remain undetected (e.g., unknown genes downstream of the adapter may be tail genes, but have no detectable similarity, fig. 1). A third possibility is the recruitment of unrelated phage genes to fill the functional role of lost GTA components. However, no other phage-related genes were identified in the diplonemid endosymbiont genomes. Next, we estimated the proportion of diplonemid endosymbiont genomes with GTA genes in culture conditions by mapping genomic sequence reads to the endosymbiont genomes and calculating the coverage of GTA genes and housekeeping genes (e.g., *recA*, *dnaE*, and *polA*). The majority of GTA genes had similar coverage as the housekeeping genes, suggesting that a large proportion of the endosymbiont populations carry GTA genes (supplementary fig. S1 and table S1, Supplementary Material online).

Metatranscriptomics of the bacterial endosymbionts and their diplonemid hosts revealed evidence of GTA expression in the reduced endosymbionts. Publicly available (Kaur et al. 2020; Prokopchuk et al. 2022) and sequenced metatranscriptomic reads were mapped to the endosymbiont genomes, and the total RNA-seq reads mapped to each genome are reported in the Supplementary Material online along with the genes with highest transcript abundance (supplementary table S2, Supplementary Material online). Transcripts for the majority of GTA genes were identified in *C. primus*, *C. indipagum*, and *S. namystus*, and additional tail components including the distal tail, hub, and megatron transcripts were present in *S. namystus* (fig. 2*A*). These data only confirm the expression of GTAs in diplonemid endosymbionts and not the levels of GTA expression due to biases in the metatranscriptomic methods used to target eukaryotic RNA and enrich eukaryotic transcripts with polyA tails (Picelli et al. 2014). Despite this potential bias, only slight positive correlations between GC-content and transcript abundance of protein-coding genes were found in two out of five transcriptomes (supplementary fig. S2, Supplementary Material online), and the GC-content and transcript abundance of intergenic regions was only weakly correlated in the *N. abundans* transcriptome (supplementary fig. S3, Supplementary Material online). All other transcriptomes showed no

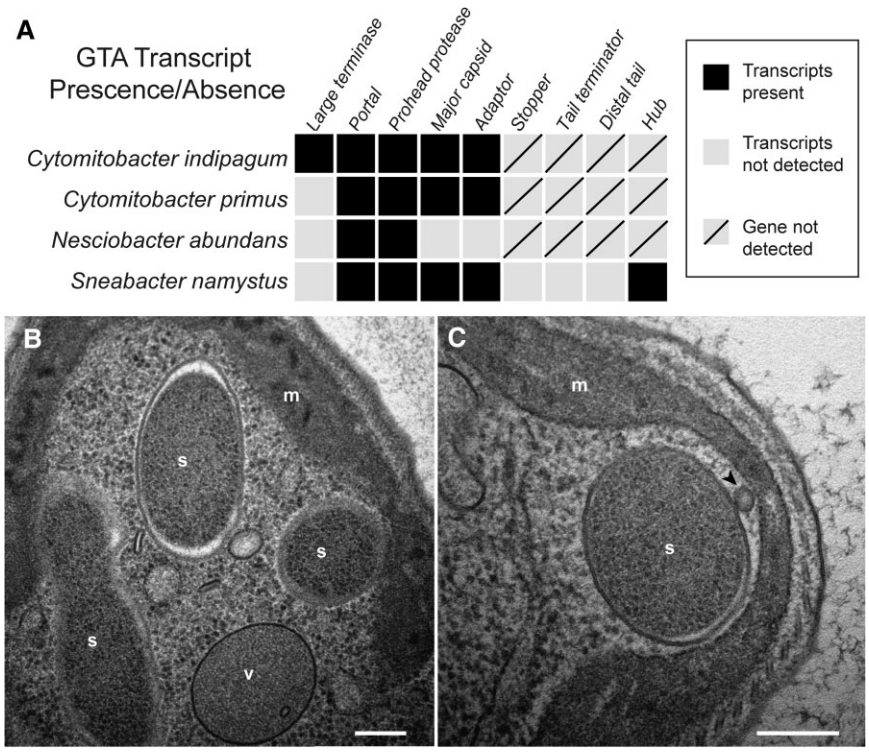

Fig. 2.—Metatranscriptomes and TEM of diplonemid endosymbionts. (*A*) Endosymbiont GTA transcripts identified in polyA-enriched metatranscriptomes of *Diplonema japonicum* and *Namystinia karyoxenos*. Due to biases in polyA-enriched metatranscriptomes, only presence/absence of bacterial transcripts are shown. (*B* and *C*) TEM micrographs of *Cytomitobacter primus* and *Nesciobacter abundans* in the cytoplasm of the host, *D. japonicum*. Scale bar is 0.2 µm; s, symbiont; m, mitochondria; v, host vacuole. Arrow points to host- or symbiont-derived vesicle.

significant relationships (supplementary figs. S2 and S3, Supplementary Material online).

To look for additional evidence of GTA expression in endosymbionts of diplonemids, transmission electron microscopy (TEM) imaging of over 850 endosymbiont cells from three different diplonemid hosts (fig. 2) was conducted. Both chemical and high-pressure freezing methods were applied, along with various staining techniques on several life stages and strains of endosymbiont-harboring diplonemids. However, no GTA-like particles were observed in the micrographs. Other structures such as vesicle-like particles located near the endosymbiont's outer-membrane surface were present in *Diplonema japonicum*, the diplonemid host of *C. primus* and *N. abundans*, but whether the particles were symbiont- or host-derived is unknown (fig. 2C). The lack of visible GTA structures may indicate that the GTA expression is relatively low and restricted to specific conditions of the endosymbionts or hosts. This observation coincides with free-living bacteria where only 1–3% of the population expresses GTAs (Fogg et al. 2012; Hynes et al. 2012), but no other GTA expression data from endosymbionts are available. Alternatively, the lack of observed GTA structures may be a result of the low-resolution and high-contrast background of the TEM imaging, and additional high resolution imaging (e.g., Cryo-EM) on enriched GTA particles isolated from the endosymbionts may be necessary to identify GTAs in diplonemid endosymbionts.

To form a broader picture of GTA evolution in protist endosymbionts, we searched for similar gene clusters across the trees of *Rickettsiales* and the *Holosporaceae*-containing *Rhodospirillales*. GTA gene clusters were found to be widespread throughout both orders (fig. 3, supplementary table S3, Supplementary Material online), as previously observed in *Rickettsiales* endosymbionts of animals (Lang and Beatty 2007; Shakya et al. 2017; Christensen and Serbus 2020). *Rickettsiales* contains mostly obligate endosymbionts and one extracellular symbiont, whereas *Rhodospirillales* is made up of free-living bacteria, ectosymbionts, and endosymbionts (*Holosporaceae*). Both *Rickettsiales* and *Holosporaceae* symbionts infect a wide range of hosts from unicellular algae and amoebae to ticks and shrimp. In *Rickettsiales*, many GTA components were not identified, but broad-scale conservation of the capsid/portal complexes along with the distal tail/hub and megatron components was apparent (fig. 3). *Neorickettsia*, an endosymbiont of flukes and mammals, encodes the most reduced GTA structure, with only the capsid and portal complex present. GTA genes were previously identified in several *Rickettsiales* endosymbionts of animals, including *Wolbachia*, where large species- and strain-level variation

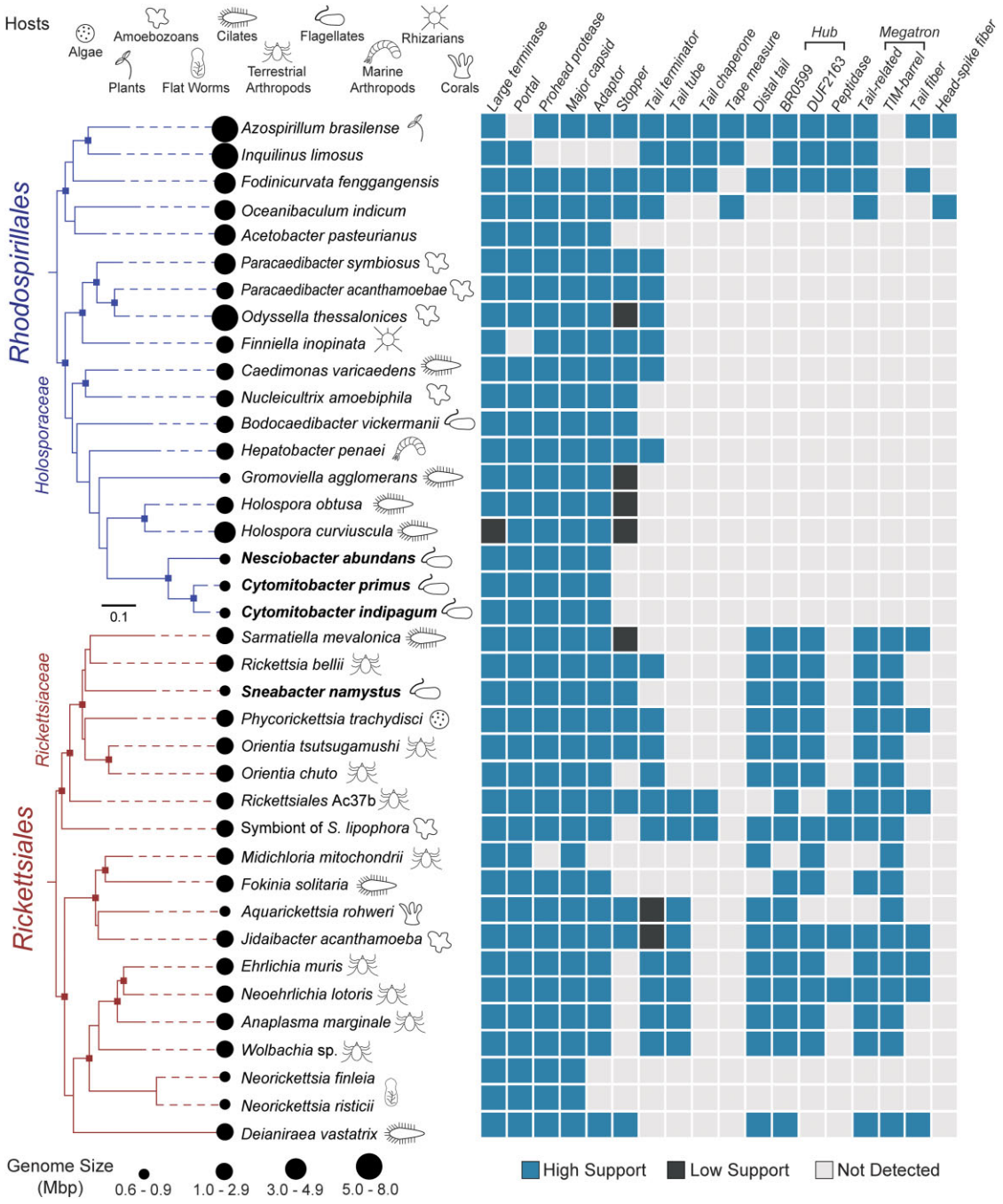

Fig. 3.—Distribution of GTA homologs in endosymbionts. All *Rickettsiales* taxa (red) are obligate endosymbionts, except the ectosymbiont *D. vastatrix*, whereas *Rhodospirillales* (blue) contains free-living bacteria, ectosymbionts (*Azospirillum brasilense*), and obligate endosymbionts (*Holosporaceae*). Diplonemid endosymbionts are in bold. Bacterial genome sizes are indicated by black circles, and icons depict the host of the endosymbionts. Phylogenetic trees are based on maximum likelihood trees (IQ-TREE) inferred under the GTR + I + G4 model from full-length 16S rRNA gene, and support values represent 1,000 bootstrap pseudoreplicates (squares indicate bootstrap scores >80; supplementary figs. S4 and S5, Supplementary Material online).

in GTA gene clusters was observed (Shakya et al. 2017; Christensen and Serbus 2020).

Within the *Holosporaceae* endosymbionts, the head and portal complexes were also highly conserved, and a clade containing endosymbionts of amoebae along with an endosymbiont of rhizarians encoded additional tail stopper and terminator components (fig. 3, supplementary table S3, Supplementary Material online). Several head and tail

components have previously been detected in free-living and ectosymbiotic members of *Rhodospirillales* (Lang and Beatty 2007; Shakya et al. 2017; Christensen and Serbus 2020), and in our analysis, the most complete GTA gene set was found in two taxa: *Azospirillum brasilense*, an ectosymbiont of plants, and the free-living *Fodinicurvata fenggangensis*. Other free-living members, such as *Acetobacter pasteurianus*, had undetected (divergent) or missing tail components. Therefore, the reduced GTA structure may have evolved in a free-living ancestor of *Holosporaceae* or alternatively, the loss of tail components has occurred multiple times independently in *Rhodospirillales*.

To ensure that the identified GTA homologs were not prophage-encoded viral genes, prophage regions were identified with PHASTER (PHAge Search Tool Enhanced Release) and checked for GTA proteins (supplementary table S4, Supplementary Material online). No GTA proteins were present in intact prophage regions, and intact prophages were identified in only three taxa: *Wolbachia* sp. that encoded two phage WO regions, *F. fenggangensis* that harbored an unknown prophage, and *A. brasilense* that encoded the phage Cd (ΦAb-Cd). GTA proteins were present in one questionable prophage region from *F. fenggangensis* (score = 70) and in incomplete prophage regions (score < 60) from 12 taxa, but few to no other phage proteins were present in these regions other than the GTA proteins (supplementary table S4, Supplementary Material online). Therefore, these regions were likely misclassified as incomplete prophages, and false positives caused by the presence of GTAs are common in prophage identification methods (Shakya et al. 2017).

Identifying this wide variety of GTAs allowed us to examine the primary mode of symbiotic GTA evolution, which provides further evidence for function. Specifically, we analyzed 20 *Rickettsiales* and 14 *Holosporaceae* symbionts for pairs of sufficiently closely related symbionts that would allow dN/dS analyses of all GTA protein domains and ten housekeeping genes involved in essential cellular processes (e.g., *dnaE*, *recA*, *thrS*). The majority of pairwise comparisons were found to have saturated dS values, but 4–7 comparisons for each GTA protein domain and housekeeping gene were informative (fig. 4, supplementary fig. S6, Supplementary Material online). The dN/dS values for the GTA domains were all found to be below 1, and the capsid- and portal-related domains exhibited markedly

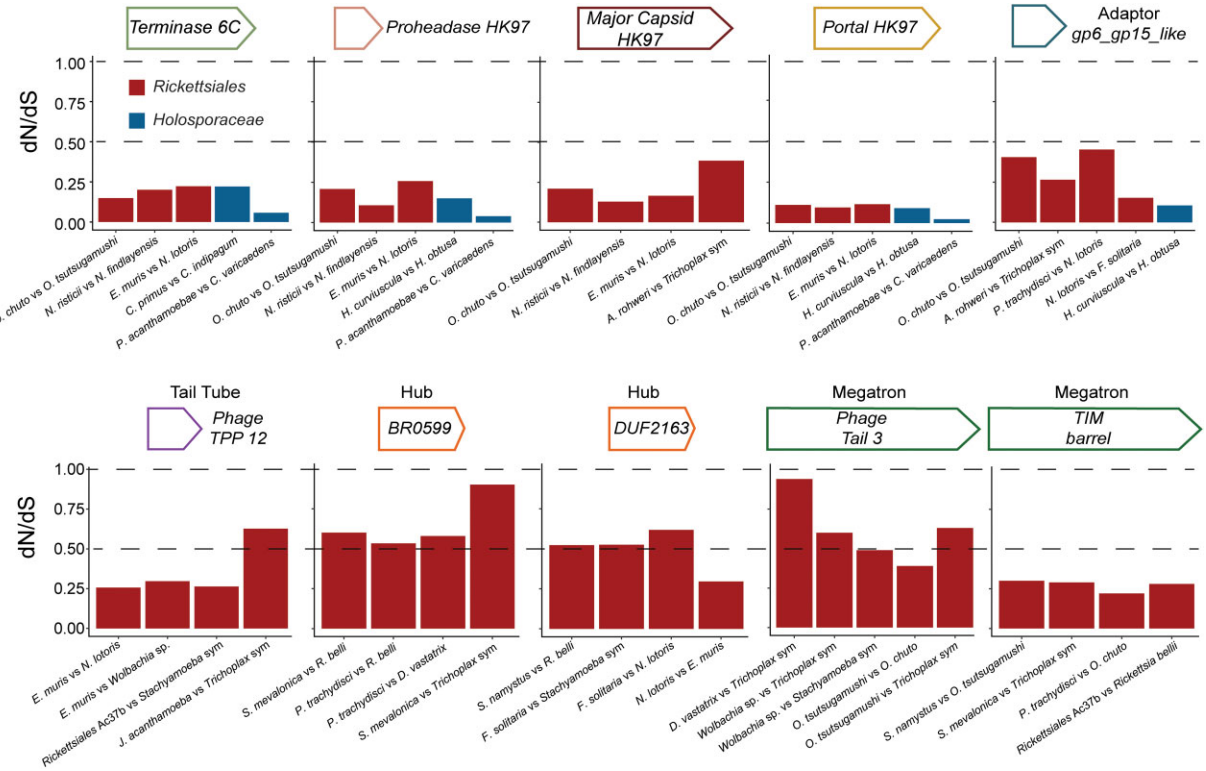

Fig. 4.—dN/dS analysis of GTA protein domains from *Rickettsiales* and *Holosporaceae* symbionts. Each GTA gene with the analyzed protein domain is shown above the bar chart, with head-related genes along the top row and tail-related genes in the bottom row. The taxa used in each pairwise comparison are listed below the bars. The top dashed line is at dN/dS = 1 and the bottom dashed line is at dN/dS = 0.5.

lower dN/dS values (0.04–0.60) than the tail-related domains (0.20–0.90, $P < 0.001$; fig. 4, supplementary fig. S6 and table S5, Supplementary Material online). The housekeeping genes had the lowest dN/dS values (0.02–0.42) compared with the head- or tail-related genes ($P = 0.002$ and $P < 0.001$, respectively), showing strong purifying selection on these essential genes (supplementary figs. S6, S7 and table S5, Supplementary Material online). Overall, GTA domains showed varying degrees of selection pressure, with stronger purifying selection on the head-related domains (e.g., terminase, prohead protease, portal, and capsid) than the tail domains (e.g., tail tube, hub domains, and phage tail 3 domain of the megatron; supplementary fig. S7, Supplementary Material online). Extremely low dN/dS values (0.02–0.1) were mostly found in *Holosporaceae* symbiont comparisons, but additional *Holosporaceae* comparisons—not possible now due to saturated dS values—will help determine whether *Holosporaceae* GTAs have stronger purifying selection compared with *Rickettsiales* GTAs. The higher dN/dS values of the *Rickettsiales* tail structures compared with the *Rickettsiales* capsid/portal components may represent a relaxation of selection pressure which would allow the pseudogenization and eventual loss of tail-related genes, a common pattern observed in *Rickettsiales* and *Holosporacaeae* symbionts (fig. 3). However, the lower dN/dS values of some tail-related domains (e.g., the megatron TIM barrel domain) demonstrate that purifying selection still plays a role in the evolution of symbiotic GTA tail structures and capsid complexes.

Previous dN/dS analyses on GTA genes in free-living bacteria, specifically *Rhodobacterales*, showed purifying selection (dN/dS = 0.06–0.08) acting upon the majority of GTA genes, including tail components (Lang et al. 2012). This may explain the conservation of tail and other GTA-related genes in *Rhodobacterales*, whereas the higher dN/dS values (0.20–0.90) of tail genes in *Rickettsiales* suggest a relaxation of selection in endosymbionts compared to free-living bacteria. *Rhodobacterales*, along with other closely related orders of *Alphaproteobacteria*, encode large clusters (LCs) of GTA genes, and evidence has been found for slower evolution of LCs compared with SCs of GTA genes present in *Rickettsiales* endosymbionts of animals and free-living *Rhodospirillales* (Hynes et al. 2016; Shakya et al. 2017). Altogether, the evidence points toward differences in the evolution of LCs and SCs, as well as endosymbiotic verses free-living GTAs.

The widespread conservation of GTAs in the reduced genomes of bacterial symbionts brings into question the potential functions of GTAs in symbiosis. The obvious possibility is that endosymbiont GTAs retain the same function found in their free-living relatives: packaging and transferring DNA from donor to recipient bacterial cells (Lang et al. 2017; Bárdy et al. 2020). Rates of genetic recombination are generally decreased in endosymbiont populations compared with free-living bacterial populations, due to both reduced genetic transfers (e.g., transduction, conjugation, and transformation), but also the loss of DNA repair and recombination mechanisms (Dale et al. 2003; McCutcheon and Moran 2012). GTAs could bring some relief to the first of these limits and, in doing so, provide a mode of genetic recombination to potentially slow Muller's ratchet, a process by which deleterious mutations become fixed in small, bacterial populations (Moran et al. 1996; Naito and Pawlowska 2016), as well as providing a mode of HGT whereby gains in function can occur (Québatte and Dehio 2019). In this case, GTAs in *Rickettsiales* and *Holosporaceae* endosymbionts may help alleviate some of the risks of genome reduction and contribute to the success of these symbiotic groups that infect a large diversity of eukaryotes. However, this is only true if endosymbiont populations mix with genetically distinct populations where novel genes can be sourced: if host cells are always infected with small clonal populations, then these effects would be quite limited.

A completely different alternative functional explanation for the distribution and selection pressures on GTAs in reduced endosymbionts could be that they have been repurposed for non-HGT functions, and one possibility is that endosymbiotic GTAs are used for eukaryotic host interactions. An example of repurposed phage structures used in bacterial interactions with eukaryotes can be found in marine bacteria that produce an array of phage tails packaged with proteins that induce the metamorphosis of a tube worm (Shikuma et al. 2014; Ericson et al. 2019). This also shows that eukaryotic-interacting proteins can be packaged inside phage-like structures, and it is possible that something other than DNA is packaged inside endosymbiotic GTA capsids. Additionally, capsid-like structures called encapsulins are used as storage compartments for iron and other elements in bacteria and archaea (McHugh et al. 2014), and the reduced GTAs could potentially serve as storage compartments. Determining between these very different functions will require direct experimental data on symbiont GTAs, including their expression in endosymbiont populations and in varying host–symbiont interactions, and direct evidence for what is packaged in GTA capsids; all intriguing but technically difficult problems in these complex systems.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We thank Vojtěch Žárský for the development of the sequence alignment script used in the dN/dS analysis, and

Vittorio Boscaro for his advice and knowledge on symbiont evolution. We also thank Thomas Beatty and Andrew Lang for their GTA expertise and discussion of GTAs in endosymbionts. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN-2014-03994) and from the Gordon and Betty Moore Foundation (https://doi.org/10.37807/GBMF9201) to P.J.K. as well as the Gordon and Betty Moore Foundation (https://doi.org/10.37807/GBMF9354) to J.L. We also acknowledge support from the Czech-BioImaging projects (LM2018129 and OPVVV 18_046/0016045) and OPVVV 16_019/0000759 to J.L., as well as support from the Czech Grant Agency (project 21-26209S) to D.T. and A.H.

Data Availability

Metatranscriptomic data produced in this study have been deposited in the NCBI Sequence Read Archive (SRA) database under the BioProject ID PRJNA813724 (BioSamples: SAMN26517683, SAMN26517684, SAMN26517685) and will be publicly released at acceptance. Previously published metatranscriptomic data are also available on the SRA database and can be accessed with SRX5472374 and SRX11449858. All other data underlying this article are available in supplementary tables S3 and S5, Supplementary Material online.

Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410. doi:10.1016/S0022-2836(05)80360-2

Bárdy P, et al. 2020. Structure and mechanism of DNA delivery of a gene transfer agent. Nat Commun. 11:303. doi:10.1038/s41467-020-16669-9

Bordenstein SR, Bordenstein SR. 2016. Eukaryotic association module in phage WO genomes from *Wolbachia*. Nat Commun. 7:13155. doi:10.1038/ncomms13155

Castelli M, et al. 2019. *Deianiraea*, an extracellular bacterium associated with the ciliate *Paramecium*, suggests an alternative scenario for the evolution of *Rickettsiales*. ISME J. 13:2280–2294. doi:10.1038/s41396-019-0433-9

Castelli M, et al. 2021. '*Candidatus* Sarmatiella mevalonica' endosymbiont of the ciliate *Paramecium* provides insights on evolutionary plasticity among *Rickettsiales*. Environ Microbiol. 23:1684–1701. doi:10.1111/1462-2920.15396

Castelli M, et al. 2022. '*Candidatus* Gromoviella agglomerans', a novel intracellular *Holosporaceae* parasite of the ciliate *Paramecium* showing marked genome reduction. Environ Microbiol Rep. 14:34–49. doi:10.1111/1758-2229.13021

Christensen S, Serbus LR. 2020. Gene transfer agents in symbiotic microbes. In: Kloc M, editor. Symbiosis: cellular, molecular, medical and evolutionary aspects. Vol. 69. Springer. p. 25–76.

Dale C, Wang B, Moran N, Ochman H. 2003. Loss of DNA recombinational repair enzymes in the initial stages of genome degeneration. Mol Biol Evol. 20:1188–1194. doi:10.1093/molbev/msg138

Dohra H, Tanaka K, Suzuki T, Fujishima M, Suzuki H. 2014. Draft genome sequences of three *Holospora* species (*Holospora obtusa*, *Holospora undulata*, and *Holospora elegans*), endonuclear symbiotic bacteria of the ciliate *Paramecium caudatum*. FEMS Microbiol Lett. 359:16–18. doi:10.1111/1574-6968.12577

Ericson CF, et al. 2019. A contractile injection system stimulates tubeworm metamorphosis by translocating a proteinaceous effector. eLife 8:46845. doi:10.7554/ELIFE.46845

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39:W29–W37. doi:10.1093/nar/gkr367

Floriano AM, et al. 2018. The genome sequence of "*Candidatus* Fokinia solitaria": insights on reductive evolution in *Rickettsiales*. Genome Biol Evol. 10:1120–1126. doi:10.1093/gbe/evy072

Fogg PCM, Westbye AB, Beatty JT. 2012. One for all or all for one: heterogeneous expression and host cell lysis are key to gene transfer agent activity in *Rhodobacter capsulatus*. PLoS One 7:e43772. doi:10.1371/JOURNAL.PONE.0043772

Garushyants SK, et al. 2018. Comparative genomic analysis of *Holospora* spp., intranuclear symbionts of paramecia. Front Microbiol. 9:738. doi:10.3389/fmicb.2018.00738

George EE, et al. 2020. Highly reduced genomes of protist endosymbionts show evolutionary convergence. Curr Biol. 30:925–933.e3. doi:10.1016/j.cub.2019.12.070

Ghequire MGK, De Mot R. 2015. The tailocin tale: peeling off phage tails. Trends Microbiol. 23:587–590. doi:10.1016/J.TIM.2015.07.011

Giannotti D, Boscaro V, Husnik F, Vannini C, Keeling PJ. 2022. The 'other' *Rickettsiales*: an overview of the family '*Candidatus* Midichloriaceae'. Appl Environ Microbiol. 88:e0243221. doi:10.1128/AEM.02432-21

Husnik F, et al. 2021. Bacterial and archaeal symbioses with protists. Curr Biol. 31:R862–R877. doi:10.1016/J.CUB.2021.05.049

Hynes AP, et al. 2016. Functional and evolutionary characterization of a gene transfer agent's multilocus "genome". Mol Biol Evol. 33:2530–2543. doi:10.1093/MOLBEV/MSW125

Hynes AP, Mercer RG, Watton DE, Buckley CB, Lang AS. 2012. DNA packaging bias and differential expression of gene transfer agent genes within a population during production and release of the Rhodobacter capsulatus gene transfer agent, RcGTA. Mol Microbiol. 85:314–325. doi:10.1111/J.1365-2958.2012.08113.X

Kaur B, et al. 2020. Gene fragmentation and RNA editing without borders: eccentric mitochondrial genomes of diplonemids. Nucleic Acids Res. 48:2694–2708. doi:10.1093/nar/gkz1215

Lang AS, Beatty JT. 2007. Importance of widespread gene transfer agent genes in α-proteobacteria. Trends Microbiol. 15:54–62. doi:10.1016/J.TIM.2006.12.001

Lang AS, Westbye AB, Beatty JT. 2017. The distribution, evolution, and roles of gene transfer agents in prokaryotic genetic exchange. Annu Rev Virol. 4:87–104. doi:10.1146/annurev-virology-101416-041624

Lang AS, Zhaxybayeva O, Beatty JT. 2012. Gene transfer agents: phage-like elements of genetic exchange. Nat Rev Microbiol. 10:472–482. doi:10.1038/nrmicro2802

Leiman PG, et al. 2009. Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. Proc Natl Acad Sci U S A. 106:4154–4159. doi:10.1073/pnas.0813360106

McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol. 10:13–26. doi:10.1038/nrmicro2670

McDaniel LD, et al. 2010. High frequency of horizontal gene transfer in the oceans. Science (80-.) 330:50. doi:10.1126/SCIENCE.1192243

McHugh CA, et al. 2014. A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. EMBO J. 33:1896–1911. doi:10.15252/embj.201488566

Midha S, Rigden DJ, Siozios S, Hurst GDD, Jackson AP. 2021. *Bodo saltans* (Kinetoplastida) is dependent on a novel *Paracaedibacter*-like endosymbiont that possesses multiple putative toxin-antitoxin systems. ISME J. 15:1680–1694. doi:10.1038/s41396-020-00879-6

Moran NA, et al. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc Natl Acad Sci U S A. 93: 2873–2878. doi:10.1073/pnas.93.7.2873

Muñoz-Gómez SA, et al. 2019. An updated phylogeny of the *Alphaproteobacteria* reveals that the parasitic *Rickettsiales* and *Holosporales* have independent origins. eLife 8:42535. doi:10.7554/eLife.42535

Naito M, Pawlowska TE. 2016. Defying Muller's ratchet: ancient heritable endobacteria escape extinction through retention of recombination and genome plasticity. mBio 7:e02057-1. doi:10.1128/mBio.02057-15

Nunan LM, Pantoja CR, Gomez-Jimenez S, Lightner DV. 2013. "*Candidatus* Hepatobacter penaei," an intracellular pathogenic enteric bacterium in the hepatopancreas of the marine shrimp *Penaeus vannamei* (Crustacea: Decapoda). Appl Environ Microbiol. 79:1407–1409. doi:10.1128/AEM.02425-12

Picelli S, et al. 2014. Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc. 9:171–181. doi:10.1038/nprot.2014.006

Pilgrim J, et al. 2017. Torix group *Rickettsia* are widespread in *Culicoides* biting midges (Diptera: Ceratopogonidae), reach high frequency and carry unique genomic features. Environ Microbiol. 19:4238–4255. doi:10.1111/1462-2920.13887

Prokopchuk G, et al. 2019. Morphological, ultrastructural, motility and evolutionary characterization of two new Hemistasiidae species. Protist 170:259–282. doi:10.1016/J.PROTIS.2019.04.001

Prokopchuk G, et al. 2022. Trophic flexibility of marine diplonemids – switching from osmotrophy to bacterivory. ISME J. 16:1409–1419. doi:10.1038/s41396-022-01192-0

Québatte M, Dehio C. 2019. *Bartonella* gene transfer agent: evolution, function, and proposed role in host adaptation. Cell Microbiol. 21: e13068. doi:10.1111/CMI.13068

Russell SL, Cavanaugh CM. 2017. Intrahost genetic diversity of bacterial symbionts exhibits evidence of mixed infections and recombinant haplotypes. Mol Biol Evol. 34:2747–2761. doi:10.1093/MOLBEV/MSX188

Sassera D, et al. 2006. "*Candidatus* Midichloria mitochondrii", an endosymbiont of the tick *Ixodes ricinus* with a unique intramitochondrial lifestyle. Int J Syst Evol Microbiol. 56:2535–2540. doi:10.1099/ijs.0.64386-0

Schulz F, et al. 2014. Life in an unusual intracellular niche: a bacterial symbiont infecting the nucleus of amoebae. ISME J. 8:1634–1644. doi:10.1038/ismej.2014.5

Schulz F, et al. 2016. A *Rickettsiales* symbiont of amoebae with ancient features. Environ Microbiol. 18:2326–2342. doi:10.1111/1462-2920.12881

Shakya M, Soucy SM, Zhaxybayeva O. 2017. Insights into origin and evolution of α-proteobacterial gene transfer agents. Virus Evol. 3:vex036. doi:10.1093/VE/VEX036

Sherlock D, Leong JX, Fogg PCM. 2019. Identification of the first gene transfer agent (GTA) small terminase in Rhodobacter capsulatus and its role in GTA production and packaging of DNA. J Virol. 93:e01328019. doi:10.1128/JVI.01328-19

Shikuma NJ, et al. 2014. Marine tubeworm metamorphosis induced by arrays of bacterial phage tail-like structures. Science 343:529–533. doi:10.1126/science.1246794

Tashyreva D, et al. 2018. Life cycle, ultrastructure, and phylogeny of new diplonemids and their endosymbiotic bacteria. mBio 9: e02447-17. doi:10.1128/mBio.02447-17

Westbye AB, et al. 2013. Phosphate concentration and the putative sensor kinase protein CckA modulate cell lysis and release of the *Rhodobacter capsulatus* gene transfer agent. J Bacteriol. 195: 5025–5040. doi:10.1128/JB.00669-13/SUPPL_FILE/ZJB999092877SO1.PDF

Westbye AB, Beatty JT, Lang AS. 2017. Guaranteeing a captive audience: coordinated regulation of gene transfer agent (GTA) production and recipient capability by cellular regulators. Curr Opin Microbiol. 38:122–129. doi:10.1016/J.MIB.2017.05.003

Yurchenko T, et al. 2018. A gene transfer event suggests a long-term partnership between eustigmatophyte algae and a novel lineage of endosymbiotic bacteria. ISME J. 12:2163–2175. doi:10.1038/s41396-018-0177-y

Zaila KE, et al. 2017. Diversity and universality of endosymbiotic *Rickettsia* in the fish parasite *Ichthyophthirius multifiliis*. Front Microbiol. 8:189. doi:10.3389/fmicb.2017.00189

Associate editor: Tal Dagan