Formal recognition and classification of gene transfer agents as viriforms

Roman Kogay,1 Sonja Koppenhöfer,2∗ J. Thomas Beatty3†, Jens H. Kuhn1§, Andrew S. Lang2‡ and Olga Zhaxybayeva1∗††

1Department of Biological Sciences, Dartmouth College, 78 College Street, Hanover, NH, USA; 2Department of Biology, Memorial University of Newfoundland, 45 Arctic Ave., St. John’s, NL A1C 5S7, Canada; 3Department of Microbiology and Immunology, University of British Columbia, 2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada and *Integrated Research Facility at Fort Detrick, Division of Clinical Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, B-8200 Research Plaza, Fort Detrick, Frederick, MD 21702, USA

Abstract

Morphological and genetic features strongly suggest that gene transfer agents (GTAs) are caudoviricete-derived entities that have evolved in concert with cellular genomes to such a degree that they should not be considered viruses. Indeed, GTA particles resemble caudoviricete virions, but, in contrast to caudoviricetes (or any viruses), GTAs can encapsidate at best only part of their own genomes, are induced solely in small subpopulations of prokaryotic host cells, and are transmitted vertically as part of cellular genomes during replication and division. Therefore, the lifecycles of GTAs are analogous to virus-derived entities found in the parasitoid wasps, which have recently been recognized as non-virus entities and therefore reclassified as viriforms. We evaluated three distinct, independently exapted GTA groups, for which the genetic basis for GTA particle production has been established. Based on the evidence, we outline a classification scheme for these viriforms.

Key words: GTA; polydnaviriformid; Brachyspira; Bartonella; Rhodobacterales; nomenclature.

Introduction

In 2021, the International Committee on Taxonomy of Viruses (ICTV) ratified a taxonomic proposal to formally accept a new operational definition of the term ‘virus’ (Kuhn et al. 2020; Koonin et al. 2021; Walker et al. 2021). Consequently, the most current version of the International Code of Virus Classification and Nomenclature (ICVCN) states that viruses are

‘... a type of MGEs [mobile genetic elements] that encode at least one protein that is a major component of the virion encasing the nucleic acid of the respective MGE and therefore the gene encoding the major virion protein itself; or MGEs that are clearly demonstrable to be members of a line of evolutionary descent of such major virion protein-encoding entities’ (ICVCN Rule 3.3) (Kuhn et al. 2020; International Committee on Taxonomy of Viruses 2022).

This definition also formalized the postulate that some MGEs, long understood by the general virology community to be distinct from viruses, are indeed distinct. At the time, the ICTV had already classified viroids and satellite nucleic acids in taxa separated from viral taxa (in families/genera with names that end with suffixes -viroidae/-viroid and -satellitidae/-satellite, respectively, as opposed to -viridae/-virus) (International Committee on Taxonomy of Viruses 2022), and these elements were logically placed into the perivirosphere rather than the orthovirosphere (Kuhn et al. 2020; Koonin et al. 2021).

The adoption of the new virus definition brought into question the taxonomic standing of one official virus family, Polydnaviridae. Indeed, entities classified into this polyphyletic family fundamentally deviate from MGEs fulfilling the virus definition because ‘polydna’ particles encapsidate multiple segments of circular double-stranded DNA that, however, do not encode the entire ‘polydna’ genomes. Instead, the genomes are permanently endogenized into the ‘polydna’ host (i.e. parasitoid wasp) genomes and inherited vertically. The resultant non-mobile non-viral entities are used by the wasps to deliver immunomodulatory genes into insects that serve as prey for the wasps (Herniou et al. 2013; Drezen et al. 2017). ‘Polydna’ entities are likely evolutionarily derived from various groups of insect viruses, including nudivirids (Thézé et al. 2011; Drezen et al. 2017; Gauthier, Drezen, and Herniou 2018; Darboux, Cusson, and Volkoff 2019; Strand and...
Burke 2020; Petersen et al. 2022), but, because they have lost the ability to replicate and instead have been fully exapted by their wasp hosts, they have left the virophage altogether (Koonin and Krupovic 2018; Koonin et al. 2021). Consequently, in 2021, the ICTV recognized ‘polydna’ entities as representatives of a new MGE category distinct from viruses called ‘viriforms’ (Kuhn et al. 2020; Koonin et al. 2021; Walker et al. 2021) and reclassified Polydnaviridae as (still polyphyletic) Polydnaviriformidae (Kuhn et al. 2021; Walker et al. 2022). In the ICVCN, viriforms are defined operationally as

‘... a type of virus-derived MGEs that have been exapted by their organismal (cellular) hosts to fulfill functions important for the host life cycle, or MGEs that are derived from such entities in the course of evolution’ (ICVCN Rule 3.3) (Kuhn et al. 2020; International Committee on Taxonomy of Viruses 2022).

Importantly, the following comment was added to the Rule 3.3:

‘Gene transfer agents (GTAs) and the MGEs previously classified in the family Polydnaviridae are considered to be viriforms in classification and nomenclature’ (Kuhn et al. 2020; International Committee on Taxonomy of Viruses 2022).

Notably, there are no discernible evolutionary relationships between GTAs and polydnaviriformids. The term ‘viriform’, similar to the term ‘virus’, is an umbrella term for certain MGEs with comparable lifecycles and properties; it is currently applied to six realms of MGEs that are not evolutionary related to each other.

Based on the properties of entities referred to as ‘GTAs’ in the literature (reviewed in Lang, Zhaxybayeva, and Beatty 2012; Lang, Westbye, and Beatty 2017), we define GTAs as viriforms with the following features:

1. GTAs use caudoviricete ancestor-derived proteins (established either via significant similarity of at least some GTA proteins to caudoviricete proteins or by image-based evidence of caudoviricete-like particles) to form caudoviricete-like particles;
2. GTAs mostly encapsidate random pieces of host DNA (established experimentally);
3. GTA genomes are fully endogenized in host genomes, often across multiple loci (established experimentally and via genomic sequence examination);
4. GTA genomes are not/cannot be fully packaged into particles due to limited particle head size (established via comparison of the packaged DNA length and size of GTA loci);
5. GTA genomes are mostly vertically inherited, and GTAs co-diversify with their hosts (established via congruence between phylogenies of host and GTA genes); and
6. DNA encapsidated in GTA particles is delivered to other cells (established experimentally).

Having these attributes, GTAs have lost the ability to replicate and have become fully exapted by their cellular hosts. They are produced under specific conditions (e.g. nutrient depletion [Westbye et al. 2017b]) and mediate horizontal gene transfer (HGT), typically among cells of the same species.

The first GTA discovered, of the alphaproteobacterium Rhodobacter capsulatus, was described in 1974 by Barry Marrs (Marrs 1974). Since that time, distinct functional GTAs have been described in other alphaproteobacteria, a sulfate-reducing deltaproteobacterium, a methanogenic archaeon, and a spirochete that infects domestic pigs (Rapp and Wall 1987; Humphrey et al. 1997; Bertani 1999; Guy et al. 2013). Additionally, clusters of genes homologous to those encoding the R. capsulatus GTA (RcGTA) are found in many alphaproteobacterial genomes, suggesting a wider prevalence of GTA production than presently appreciated (Lang, Taylor, and Beatty 2002; Lang and Beatty 2007; Shakya, Soucy, and Zhaxybayeva 2017; Kogay et al. 2019). Indeed, some of these bacteria produce functional GTAs (Biers et al. 2008; Nagao et al. 2015; Tomash et al. 2018).

The recent ICTV recognition of viriforms and the formal establishment of Polydnaviriformidae provide an opportunity to initiate a systematic classification of GTAs. Here, we outline initial steps to establish such a formal taxonomic scheme for GTA viriforms—focusing specifically on GTAs experimentally documented as being produced by cells and performing gene transfer—and for which the genetic basis of particle production has been established. Simultaneously, we have also officially proposed this taxonomic scheme to the ICTV for the 2022–3 proposal cycle.

**Nomenclature of GTAs and associated taxa**

Per ICTV rules, virus names are written in lower case (except if a name component is a proper noun), without italics in any part of the name (even if a host species name is part of the name), and ending in the term ‘virus’, which in virus name abbreviation is ‘V’. Examples are measles virus (MeV) and Ebola virus (EBOV). The nomenclature of already classified viriforms (polydnaviriformids) follows these rules, with ‘virus’ being replaced by ‘viriform’ and the abbreviation ‘V’ being replaced with ‘Vf’ (e.g. ‘Glyptapanteles liparidis bracoviriform’ is abbreviated as ‘GlBVf’). We suggest applying these general rules to GTAs, but with ‘viriform’ being replaced by ‘gene transfer agent’ due to the long-established use of this phrase and ‘Vf’ being replaced with ‘GTA’. Therefore, the GTA produced by Rhodobacter capsulatus would be called ‘Rhodobacter capsulatus gene transfer agent’ and abbreviated as ‘RcGTA’, consistent with the established use of this abbreviation in the literature.

Rules for viriform taxon naming have been established by the ICVCN. Specifically,

‘[t]he formal endings for taxon names of viriforms are the suffixes “-viriforma” for realms, “-viriform” for subrealms, “-viriformae” for kingdoms, “-viriformites” for subkingdoms, “-viriformicota” for phyla, “-viriformicotina” for subphyla, “-viriformicetes” for classes, “-viriformicetidae” for subclasses, “-viriformales” for orders, “-viriforminae” for suborders, “-viriformidae” for families, “-viriformae” for subfamilies, and “-viriform” for genera and subgenera’ (ICVCN Rule 3.26) (Kuhn et al. 2020; International Committee on Taxonomy of Viruses 2022)

and

‘[a] species name shall consist of only two distinct word components separated by a space. The first word component shall begin with a capital letter and be identical in spelling to the name of the genus to which the species belongs. The second word component shall not contain any suffixes specific for taxa of higher ranks. The entire species name (both word components) shall be italicized’ (International Committee on Taxonomy of Viruses 2022).
GTAs can be assigned to at least three major clades

Based on functionally and genetically characterized GTAs, at least three major GTA clades can be delineated.

**Alphaproteobacterial type I GTAs**

The best characterized GTA of this clade is RcGTA, produced by *Rhodobacter capsulatus* (Pseudomonadota: Alphaproteobacteria: Rhodobacterales: Rhodobacteraceae). We designate RcGTA here as the founding member of one major GTA clade, the alphaproteobacterial type I GTAs. For many years after its discovery (Marrs 1974), RcGTA was the only known GTA. Now, we know that homologous GTAs are produced by other bacteria from the order Rhodobacterales: Dinoroseobacter shibae (Dinoroseobacter shibae gene transfer agent [DsGTA]) (Tomasch et al. 2018), Ruegeria pomeroyi (Ruegeria pomeroyi gene transfer agent [RpGTA]) (Biers et al. 2008), and Rhodovulum sulfidophilum (Rhodovulum sulfidophilum gene transfer agent [RsGTA]) (Nagao et al. 2015). Additionally, genes encoding RcGTA-like GTAs are conserved in most genomes in the order Rhodobacterales and in many genomes of the alphaproteobacterial orders Caulobacterales, Spingomonadales, Parvibaculales, and Hyphomicrobiiales (formerly Rhizobiiales) (Lang, Taylor, and Beatty 2002; Lang and Beatty 2007; Shakya, Soucy, and Zhaxybayeva 2017; Kogay et al. 2019).

RcGTA and RcGTA-like GTA genes are similar in sequence to those of viruses classified in the uroviricote class Caudoviricetes (Duplodnaviria: Heunggongviirae) (Shakya, Soucy, and Zhaxybayeva 2017). These GTAs are transmitted vertically from a bacterial parent to progeny during cell division (Lang and Beatty 2007; Shakya, Soucy, and Zhaxybayeva 2017), similar to propagation of temperate viruses (‘prophages’). However, in contrast to temperate virus genomes, the set of genes required for the production of GTA particle (the GTA ‘genome’) is not necessarily localized in one region of the host genome. In the case of RcGTA, known structural and regulatory genes are scattered across five loci in the *R. capsulatus* genome (Hynes et al. 2016), cumulatively spanning approximately 20 kb (Fig. 1 and Supplementary Table S1). Moreover, cellular regulatory genes are involved in controlling GTA particle production (Westbye, Beatty, and Lang 2017a), adding another factor that makes the GTA genome difficult to differentiate from its host’s genome.

RcGTA particles resemble virions of caulimoviricetes (Yen, Hu, and Marrs 1979) and have been structurally characterized at high resolution (Bárdy et al. 2020). RcGTA particles have head diameters of 38 nm and tail lengths of 49 nm. A small percentage of RcGTA particles have T = 3 quasi-icosahedral heads, but the capsid shape of most particles is oblate, as they lack the five hexamers of capsid protein needed to form genuine icosahedral heads. Because of the small head size, RcGTA particles can only package double-stranded DNA of approximately 4 kb in length (Yen, Hu, and Marrs 1979). The DNA is also encapsidated at 10–25 per cent lower density than typical caulimoviricetes (Bárdy et al. 2020). Both RcGTA particle production and acquisition of the GTA-packaged DNA by other host cells in the population are controlled by the same cellular regulatory systems (Westbye, Beatty, and Lang 2017a). Only 0.1–3.0 per cent of cells produce GTA particles (Fogg, Westbye, and Beatty 2012; Hynes et al. 2012), whereas the remaining cells produce a GTA receptor (Brimacombe et al. 2013).

Compositionally, structural proteins encoded by RcGTA and RcGTA-like GTAs are biased toward amino acids that are energetically cheaper to produce (Kogay et al. 2020). To date, such a bias has not yet been associated with viruses. Based on this difference in amino-acid composition, GTA proteins can be distinguished from their viral homologs using a machine–learning approach, which is implemented in the publicly available GTA-Hunter program (Kogay et al. 2019).

In a comprehensive evolutionary analysis of homologs of the large subunit of the DNA packaging terminase enzyme (large terminase [TerL], encoded by the g2 gene in the RcGTA genome), RcGTA and RcGTA-like GTAs form a clade closely related to, but distinct from, duplodnavirians (Esterman et al. 2021). To illustrate the relationships of alphaproteobacterial type I GTAs to each other and to their closest viral homologs, we reconstructed evolutionary histories of their TerL proteins and the HK97-like major capsid protein (HK97-MCP, encoded by the g5 gene in the GTA genome, is the hallmark protein that defines the virus realm Duplodnaviria [Koonin et al. 2020]). Consistent with an earlier analysis (Esterman et al. 2021), RcGTA and RcGTA-like GTAs formed a clade closely related to, but distinct from, caulimoviricetes (Fig. 2), with a few exceptions that are likely artifacts of phylogenetic reconstruction.

Specifically, in the TerL phylogeny (Fig. 2A), all viral homologs except one (Caulobacter virus Sansa) are separated from GTA proteins (with a solid bootstrap support of 81 per cent). Caulobacter virus Sansa groups with one GTA sequence from a bacterium of the order Spingomonadales (with a low bootstrap support of 50 per cent), whereas all other GTAs of Spingomonadales bacteria group together (with a strong bootstrap support of 96 per cent). We hypothesize that the phylogenetic placement of the Caulobacter virus Sansa TerL is due to the long-branch attraction artifact (Felsenstein 1978). We searched for a maximum likelihood tree in which caulimoviricete- and GTA-derived TerLs were required to group separately from each other and compared that tree to the tree depicted in Fig. 2A. We found that the likelihoods of the
Virus Evolution

Figure 2. Maximum likelihood phylogenies of (A) TerL subunits and (B) HK97-MCP sequences of rhodogtviriformids and their closest known caudoviricete homologs. Alphaproteobacterial type I GTA (rhodogtviriformid) lineages are shown in the shaded areas within the top portion of the figure. Caudoviricete lineages that are nested within GTA lineages are shown in dashed lines. Other caudoviricete lineages are shown in solid lines in the bottom portion of the figure. Bootstrap support values are shown only for selected branches. Scale bars represent substitutions per site. DsGTA, Dinoroseobacter shibae gene transfer agent; GTA, gene transfer agent; RcGTA, Rhodobacter capsulatus gene transfer agent; RpGTA, Ruegeria pomeroyi gene transfer agent; RsGTA, Rhodovulum sulfidophilum gene transfer agent.

The two trees are not significantly different (approximately unbiased [AU] test, \( P \)-value = 0.555), confirming that the placement of the Caulobacter virus Sansa sequence within the GTA sequences is unreliable. In the HK97-MCP phylogeny (Fig. 2B), GTAs and most caudoviricete are separated by a branch with 63 per cent bootstrap support. Several caudoviricete that group within GTAs are located on long branches, situated outside of well-supported groups of GTAs from several alphaproteobacterial orders, and have very low bootstrap support for their placements. It is therefore likely that the positions of these viral homologs are unreliable. To test this hypothesis, we identified a maximum likelihood phylogeny among trees in which GTAs and caudoviricetes were required to be separated by a branch. The likelihoods of this tree and the phylogeny shown in Fig. 2B are not significantly different (AU test; \( P \)-value = 0.534). Therefore, these viruses are likely positioned in different places in trees reconstructed from different bootstrap replicates, which would lead to their artificial (and poorly supported) basal positions with the GTA homologs on the tree shown in Fig. 2B.

In the Fig. 2 trees, GTA branches have shorter lengths than their caudoviricete counterparts, conforming to the reported slower
evolutionary rate of GTAs compared to viruses (Shakya, Soucy, and Zhaxybayeva 2017). Additionally, on both phylogenetic trees, GTAs from alphaproteobacteria of different orders form separate groups with very high support, corroborating vertical inheritance of most GTA genes (Lang, Taylor, and Beatty 2002; Lang and Beatty 2007; Shakya, Soucy, and Zhaxybayeva 2017).

Together, these results justify the classification of RcGTA and three RCGTA-like GTAs in a common viriron taxon: family Rhodogtaviriformidae (from Rhodobacterales, infix -gta-, and family-specific suffix -viriformidae). Given the limited dataset size (i.e. just four GTAs), it is challenging to establish quantifiable criteria for demarcating taxonomic relationships among the four GTAs. In the future, when more GTA sequences become available for analyses, a criterion based on percent sequence similarity among shared genes should be considered. For now, based on the evidence of co-evolution of these GTAs and their specific hosts, we argue that at least four rhodogtaviriformid genera, each for GTAs of bacteria classified in distinct genera included in Rhodobacterales, ought to be established:

- Dinogtaviriform (named after DsGTA host genus Dimorosobacter, infix -gta-, and genus-specific suffix -viriform) to include one new species, Dinogtaviriform tomashchi (species epithet to honor GTA researcher Jurgen Tomashch, who was instrumental in the discovery of DsGTA) for DsGTA (Supplementary Table S2);
- Rhodobacte_gtaviriform (named after RcGTA host genus Rhodobacter, infix -gta-, and genus-specific suffix -viriform) to include one new species, Rhodobacte_gtaviriform marri (species epithet to honor GTA researcher Barry Marrs, who first discovered GTAs and coined the term ‘gene transfer agent’) for RcGTA (Supplementary Table S1);
- Rhodovulugtaviriform (named after RsGTA host genus Rhodovulum, infix -gta-, and genus-specific suffix -viriform) to include one new species, Rhodovulugtaviriform kikuchii (species epithet to honor GTA researcher Yo Kikuchi, who was instrumental in the discovery of RsGTA) for RsGTA (Supplementary Table S3); and
- Ruegerigtaviriform (named after RpGTA host genus Ruegeria, infix -gta-, and genus-specific suffix -viriform) to include one new species, Ruegerigtaviriform cheni (species epithet to honor GTA researcher Feng Chen, who was instrumental in the discovery of RpGTA) for RpGTA (Supplementary Table S4).

Alphaproteobacterial type II GTAs

There was a lag between discovery of these elements and their recognitions as bona fide GTAs. Phage-like particles, originally referred to as bacteriophage-like particles, that contained heterogeneous DNA from Bartonella host genomes, were first characterized in Bartonella henselae (Anderson et al. 1994) and noted to be similar in structure to the particles produced by Bartonella bacilliformis (Umemori et al. 1992). These B. bacilliformis particles were subsequently shown to also contain heterogeneous genomic DNA fragments, but attempts to demonstrate their gene transfer ability were not successful (Barbian and Minnick 2000). Functionality of the particles produced by Bartonella for gene transfer (Bartonella gene transfer agent [BaGTA]) was eventually demonstrated by work on B. henselae (Pseudomonadota: Alphaproteobacteria: Hyphomicrobiales: Bartonellaceae) (Guy et al. 2013). BaGTA genes were initially proposed to be located within a single cluster of 11–13 genes spanning approximately 14 kb (Guy et al. 2013). However, a subsequent screen for genes essential for BaGTA functionality identified a total of twenty-nine genes located within a larger (approximately 79-kb-long) region (Québatte et al. 2017) (Fig 3 and Supplementary Table S5). Homologs of BaGTA genes (BaGTA-like GTAs) were found in the genomes of multiple species of Bartonella (Berglund et al. 2009; Guy et al. 2013; Tamarit et al. 2018). BaGTA genes are located near an active virus-derived origin of replication and next to genes encoding secretion systems (Guy et al. 2013). As a result, the region of the genome containing BaGTA and these secretion-system genes are amplified and packaged more often than other genomic regions (Guy et al. 2013; Québatte et al. 2017). These findings led to the hypothesis that BaGTA and BaGTA-like GTAs have been maintained due to their mediation of HGT of secretion-system and toxin genes, thereby enabling Bartonella bacteria to adapt to diverse hosts (Guy et al. 2013). However, actual GTA-mediated DNA transfer among bacterial cells has only been demonstrated for B. henselae (Guy et al. 2013). There, BaGTA production is restricted to a distinct subpopulation of fast-growing cells, which comprise about 6 per cent of the total population (Québatte et al. 2017), and the uptake of BaGTA-packaged DNA was proposed to be limited to cells undergoing division (Québatte et al. 2017).

There are some discrepancies in the literature regarding the structure of BaGTA particles, suggesting some bacteria might release additional phage-like particles. The B. henselae particles were originally reported as particles without tails or with short non-contractile tails with a head diameter of 40 nm (Anderson et al. 1994). The head diameter of the B. bacilliformis particles was originally measured at 40 nm (Umemori et al. 1992) and subsequently 80 nm (Barbain and Minnick 2000). Those of Bartonella grahamii were reported as possessing long non-contractile tails and icosahedral heads of 50–70 nm or 80 nm and tails of 100 nm (Berglund et al. 2009). Although BaGTA particles are potentially able to package the entire main structural gene cluster of 11–3 genes, they cannot package all twenty-nine genes required for BaGTA production due to a capacity of 14 kb (Anderson et al. 1994; Guy et al. 2013; Lang, Westbye, and Beatty 2017).

In the TerL phylogeny, BaGTA-like homologs are separated from almost all caudoviricetes by longer branches (with 100 per cent bootstrap support; Fig. 4A). Two caudoviricete homologs (Sulfitobacter phage pCB2047-C and Sulfitobacter phage NYA-2014a) group together and are nested within the BaGTA-like group (with 84 per cent bootstrap support). We hypothesize that the terL gene was horizontally transferred from GTAs to these caudoviricetes, with similar HGT events documented between RCGTA-like GTAs and caudoviricetes infecting bacteria of the Rhodobacterales (Zhan et al. 2016). In the HK97-MCP phylogeny, BaGTA homologs are located on shorter branches than their caudoviricete counterparts and are separated from caudoviricete homologs with 100 per cent bootstrap support (Fig 4B). Phylogenomic analyses suggest that Bartonella GTAs have co-evolved with their hosts (Tamarit et al. 2018).

Together, these results justify the classification of BaGTA and BaGTA-like GTAs in a common viriron taxon, family Bartogtaviriformidae (from Bartonella, infix -gta-, and family-specific suffix -viriformidae). For now, we argue that at least one bartogtaviriformid genus ought to be established: Bartongtaviriform (named after BaGTA host genus Bartonella, infix -gta-, and genus-specific suffix -viriform) including one new species, Bartongtaviriform andersoni (species epithet to honor GTA researcher Burt Anderson, who first discovered BaGTA particles [Anderson et al. 1994]) for BaGTA.
GTAs of spirochetes

A GTA, originally called virus of *Serpulina hyodysenteriae* 1 (VSH-1) was identified in Brachyspira (formerly *Serpulina*) hyodysenteriae (Spirochaetota: Spirochaetia: Brachyspirales: Brachyspiraceae) (Humphrey et al. 1997). In accordance with the nomenclature rules established here, we suggest renaming this GTA to *Brachyspira hyodysenteriae* gene transfer agent (BhGTA). The structural gene cluster responsible for the production of BhGTA particles—i.e. the BhGTA ‘genome’—is 16.3 kb in length (Matson et al. 2005) (Fig. 5 and Supplementary Table S6).

BhGTA particles have a head diameter of 45 nm and a flexible non-contractile tail of 65 nm (Humphrey et al. 1997). Like other GTAs, BhGTA is unable to package and transfer its entire genome, given the limiting capacity of 7.5 kb (Humphrey et al. 1997; Matson et al. 2005). Restriction enzyme digests of the packaged DNA and the range of marker genes that can be transferred by BhGTA particles suggest that they package any region of the *B. hyodysenteriae* genome (Humphrey et al. 1997) without an obvious bias for the genomic region that encodes BhGTA. The induction of BhGTA particle production by DNA-damaging agents, such as mitomycin C, results in large-scale lysis of cells (Stanton et al. 2008). However, the proportion of *B. hyodysenteriae* cells in a population that naturally produce and release BhGTA particles has not been quantified. BhGTA particles are capable of transferring antimicrobial resistance genes within the bacterial population (Stanton et al. 2008), pointing at possible selective advantages of maintaining the capability of BhGTA particle production.

Homologs of genes in the BhGTA genome were found in the genomes of these three GTAs are either not homologous or too divergent to have significant sequence similarity in those of caudoviricetes (Matson et al. 2005). Therefore, the absence of their homologs in the viral RefSeq database is likely due to the limited sampling of the virosphere.

In the endolysin phylogeny, the *Brachyspira* homologs group together and are separated from all caudoviricetes by a long branch (with 100 per cent bootstrap support; Fig. 6A). Additionally, the *B. hyodysenteriae* genome encodes a single copy of an identifiable terL gene, which is located outside of the currently delineated BhGTA genome. Homologs of this terL gene are also present in a single copy in genomes of other Brachyspira bacteria that encode BhGTA-like MCPs. These homologs are highly conserved, with pairwise amino-acid identities of 81–100 per cent. In a phylogenetic tree, the *Brachyspira* TerLs are separated from all caudoviricete TerLs by a longer branch (with 100 per cent bootstrap support; Fig. 6B). Although the role of this TerL homolog in the BhGTA lifecycle has not been experimentally validated, the presence of the encoding gene as the only identifiable terL in the Brachyspira genomes, its high degree of conservation within the Brachyspira genus, and its divergence from the related caudoviricete sequences support its potential involvement in the packaging of DNA into the BhGTA particles. Based on comparison of Brachyspira GTA and host genes, GTAs have co-diversified with Brachyspira (Motro et al. 2009).

Together, these results justify the classification of BhGTA and BhGTA-like GTAs in a common viriform taxon, family *Brachytagtaviriformidae* (form *Brachyspira*, infix -gta-, and family-specific suffix -viriformidae). For now, we argue that at least one brachytaviriformid genus ought to be established: *Brachyspigtagtaviriform* (named after BhGTA host genus *Brachyspira*, infix -gta-, and genus-specific suffix -viriform) to include one new species, *Brachyspigtagtaviriform* stantonii (species epithet to honor GTA researcher Thaddeus Stanton, who first discovered BhGTA particles [Humphrey et al. 1997]) for BhGTA.

Independent origins of the three GTAs

Genes from the genomes of these three GTAs are either not homologous or too divergent to have significant sequence similarity in
Figure 4. Maximum likelihood phylogenies of (A) TerL subunits and (B) HK97-MCP sequences of bartogtaviniformids and their closest known caudoviricete homologs. Alphaproteobacterial type II GTA (bartogtaviniformid) lineages are shown in the shaded areas of the figure. Caudoviricete lineages are shown in the unshaded areas of the figure. Two nearly identical caudoviricete lineages that are nested within GTA lineages are shown in dashed lines. A bootstrap support value is shown only for the branch separating GTA and caudoviricete sequences. Scale bars indicate substitutions per site. BaGTA, Bartonella gene transfer agent; GTA, gene transfer agent.

BLASTP searches of the encoded proteins. For example, pairwise amino-acid identities of TerLs, which is one of the most conserved GTA and caudoviricete proteins, is 14–20 per cent among RcGTA, BaGTA, and BhGTA. Nevertheless, an iterative clustering-alignment-phylogeny procedure (Wolf et al. 2018) established the homology among known TerL proteins that include RcGTA, BaGTA,
and putative BhGTA TerLs (Esterman et al. 2021). The evolutionary history of RcGTA-like, BaGTA-like, and putative BhGTA-like TerLs and their closest known caudoviricete homologs (Fig. 7) demonstrates that GTA-like TerLs appear in three distinct clades within viral TerLs. Based on this phylogenetic evidence, we propose that these three GTA clades are a result of three independent exaptation events. Therefore, just like viruses (which are classified in at least six unrelated realms), GTA viriforms are polyphyletic.

**Discussion**

Based on the evolutionary differences between GTA and caudoviricete genes encoding well-conserved proteins and on morphological differences of GTA particles, we propose three families for these GTAs. The greatest number of functionally confirmed and putative GTAs is in the alphaproteobacterial type I clade, which, for now, is proposed to be a family Rhodogtaviriformidae that includes at least four genera. The members of this family are currently restricted to a single cellular order (Rhodobacterales). The TerLs and MCPs of these RcGTA-like GTAs and alphaproteobacterial type II GTAs (Bartogtaviriformidae) are clearly distinguishable from each other and their caudoviricete homologs and evolve at a slower rate (Figs 2 and 4) (Shakya, Soucy, and Zhaxybayeva 2017; Esterman et al. 2021). The spirochete GTAs (Brachygtaviriformidae) are more difficult to distinguish from caudoviricetes due to a lack of available viral representatives in GenBank for all but one experimentally validated BhGTA gene. Nevertheless, both the experimentally validated BhGTA endolysin and the putative BhGTA TerL and their Brachyspira homologs also form a well-supported cluster distinct from caudoviricete lineages; moreover, brachygtaviriformid TerLs evolve at a slower rate than their spirochete homologs (Fig. 6B). As in the case with the experimentally validated RcGTA, the ‘genome’ of BhGTA is also likely dispersed across multiple loci.

Analyses of environmental samples and genome sequences suggest the existence of a large number of GTAs, especially those related to the rhodogtaviriformids (Biers et al. 2008; Zhao et al. 2009; Fu et al. 2010; McDaniel et al. 2010). In a genome-wide screen of 1,423 alphaproteobacterial genomes, 57.5 per cent were found to encode RcGTA-like ‘genomes’, which are often annotated as either intact or incomplete prophages (Kogay et al. 2019). The great majority of RcGTA-like genes in alphaproteobacterial genomes are associated with bacteria for which a GTA-based gene-transfer activity has not been documented, and it is possible that some of these RcGTA-like genes may not be expressed to produce functional particles. Therefore, we have restricted our proposal to those GTAs that have been shown to be functional. However, we speculate that at least some (and perhaps many) of these GTA-like gene clusters will be shown to produce functional GTAs that will need to be classified.

Based on the evolutionary history of TerL proteins (Fig. 7), it is likely that the proposed three GTA families had distinct caudoviricete progenitors. Eventual deduction of the relatives of these progenitors may make it possible (or necessary) to include these GTA families in the virus class Caudoviricetes, thereby creating an overarching taxon for distinct MGEs (viruses and viriforms). Since the exaptation events, however, the three families have evolved as part of the host genomes (Lang, Taylor, and Beatty 2002; Lang and Beatty 2007; Shakya, Soucy, and Zhaxybayeva 2017; Esterman et al. 2021), in the case of the rhodogtaviriformids for hundreds of millions of years (Shakya, Soucy, and Zhaxybayeva 2017). As a result, GTAs effectively became a component of cellular genomes, integrated into cellular regulatory circuits that also control processes such as motility, quorum sensing, extracellular polysaccharide synthesis, and biofilm formation (Lang, Westbye, and Beatty 2017; Pailegar et al. 2020; Shimizu et al. 2022). There is also mounting evidence that GTA genes experience selective pressures to be maintained in their host genomes (Lang, Zhaxybayeva, and Beatty 2012; Kogay et al. 2020). Although the fitness benefits associated with GTA production remain to be elucidated, the time is now ripe to have the known GTAs officially recognized and classified as specific viriforms. We recognize this step as the initiation of a taxonomic framework that undoubtedly will rapidly expand and change in the future.

**Methods**

To identify alphaproteobacterial type I GTAs, we searched for RcGTA-like sequences in 1,248 complete alphaproteobacterial genomes extracted from the NCBI RefSeq database (accessed in October 2020) using GTA-Hunter (Kogay et al. 2019). We identified 503 genomes that contained at least six RcGTA homologs in the same genetic neighborhood and had both $g_2$ (encoding TerL) and $g_5$ (encoding HK97-MCP) genes. To remove redundancy, we clustered genomes into the operational taxonomic units (OTUs) using an average nucleotide identity threshold of 95 per cent. From all genomes within an OTU, we selected one genome with the largest number of GTA genes. This strategy resulted in 290 representative GTAs selected for further analysis. We identified the closest viral homologs of the TerL and HK97-MCP proteins from these GTAs by conducting a BLASTP search (Altschul et al. 1997) of the RefSeq database (accessed in March 2021) (O’Leary et al. 2016), using TerL and HK97-MCP proteins from representative GTAs as queries, an e-value cut-off of 0.001, and query coverage of at least 50 per cent. Retrieved viral homologs with identical amino-acid sequences were removed from further analyses. For both proteins, we aligned amino-acid sequences of GTA and virus homologs using MAFFT v7.455 with -insi option (Katoh and Standley 2013). We reconstructed phylogenetic trees using
Figure 6. Maximum likelihood phylogenies of (A) endolysin and (B) the putative TerL subunits of brachygtaviriformids and their closest known caudoviricete homologs. Brachyspira GTA lineages are shown in the shaded areas of the figure. Caudoviricete lineages are shown in the unshaded areas of the figure. A bootstrap support value is shown only for the branch separating GTA and caudoviricete sequences. Scale bar indicates substitutions per site. BhGTA, Brachyspira hyodysenteriae gene transfer agent; GTA, gene transfer agent.

IQ-TREE v2 (Minh et al. 2020), identifying the best substitution models using the built-in ModelFinder (Kalyaanamoorthy et al. 2017). The selected models were LG + F + R9 and LG + F + R7 for TerL and HK97-MCP datasets, respectively. Branch support values were assessed using 1,000 ultrafast bootstrap replicates and a hill-climbing nearest-neighbor interchange search for optimal trees (Hoang et al. 2018). Additionally, for both protein phylogenies, we reconstructed a phylogenetic tree in IQ-TREE v2 (Minh et al. 2020).
using a tree search that was constrained by requiring all GTAs and all viruses to be separated by a branch. We compared the resultant trees in unconstrained and constrained searches using the AU test (Shimodaira 2002), as implemented in the IQ-TREE v2 program.

To identify alphaproteobacterial type II GTAs, we used the BaGTA TerL and HK97-MCP sequences (accession numbers WP_034448260.1 and WP_011181178.1, respectively) as queries in a BLASTP search against the fifty-seven complete Bartonella genomes extracted from the RefSeq database (accessed in May 2022). We restricted our search only to matches for which BaGTA TerL and HK97-MCP homologs are in the same genomic neighborhood (defined as being within 5 kb of each other). In genomes with multiple matches to the query protein, we retained only the homolog with the highest BLASTP bit score. We clustered fifty-seven genomes using a 95 per cent average nucleotide identity (ANI) threshold and randomly selected one TerL and HK97-MCP representative from each cluster for phylogenetic analysis. We identified caudoviricete homologs by conducting a BLASTP search (e-value cut-off of 0.001 and query coverage of at least 50 per cent) against viral RefSeq database (accessed in May 2022). We performed phylogenetic reconstructions as described above for alphaproteobacterial type I GTAs. The selected best substitution models were LG + R6 and LG + G4 for TerL and HK97-MCP datasets, respectively.

To identify GTAs of spirochetes, we used BhGTA’s MCP sequence (GenBank accession number WP_012671344.1) as a query in a BLASTP search (with an e-value cut-off of 0.001 and query precision as described above for alphaproteobacterial type I GTAs.

Figure 7. Maximum Likelihood phylogeny of the TerL subunits of three major clades of GTAs and their closest known caudoviricete homologs. This tree includes all TerL homologs from Figs 2A, 4A, and 6A. Bootstrap support values are shown only for the branches separating three GTA clades and their closest caudoviricete sequences. Scale bar indicates substitutions per site. GTA, gene transfer agent.
coverage of at least 50 per cent) against the thirteen complete Brachyspira genomes extracted from the RefSeq database (accessed in May 2022). We used TerL of B. hyodysenteriae (GenBank accession number WP_012671469.1) and endolysin protein of B. hyodysenteriae (GenBank accession number WP_012671356.1) as queries in a BLASTP search (with an e-value cut-off of 0.001 and query coverage of at least 50 per cent) against the same set of thirteen genomes. For endolysins, we only retained matches that co-localized within the BhGTA region on the chromosome. We clustered thirteen genomes using a 95 per cent ANI threshold and randomly chose one TerL and endolysin representative from each cluster for phylogenetic analyses. We identified caudoviricete homologs by doing BLASTP searches (e-value cut-off of 0.001 and query coverage of at least 50 per cent) against the viral RefSeq database (accessed in May 2022). We performed phylogenetic reconstructions as described above for the alphaproteobacterial type I GTAs. The selected best substitution models were VT + F + R3 and WAG + R6 for TerL and endolysin datasets, respectively.

To reconstruct the phylogeny that includes all three clades of GTAs, we combined all TerL homologs extracted in the above-described procedures into one dataset. We aligned the TerL sequences using MAFFT v7.455 with -dash option (Rozewicki et al. 2019) and trimmed the obtained alignment using ClipKIT with -gappy option (Steenwyk et al. 2020). We computed the phylogenetic tree using IQ-TREE v2 (Minh et al. 2020) as described above with the LG + F + R10 substitution model selected by ModelFinder. We rooted the tree using a larger TerL phylogeny presented in Esterman et al. (2021).

We visualized all phylogenetic trees in iTOL v6 (Letunic and Bork 2021).

**Data availability**

All data used in this manuscript were retrieved from publicly available GenBank databases, as described in the Methods. The accession numbers of database records used in the phylogenetic analyses can be found in alignments that are included in the Supplementary data. Phylogenetic trees shown in Figures 2, 4, 6, and 7 are provided in Newick format in the Supplementary Data.

**Supplementary data**

Supplementary data are available at Virus Evolution online.

**Acknowledgements**

We thank Anya Crane and Jiro Wada (Integrated Research Facility at Fort Detrick, Division of Clinical Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health) for editing the manuscript and figures, respectively.

**Funding**

This work was supported in part through Laulima Government Solutions, LLC, prime contract with the U.S. National Institute of Allergy and Infectious Diseases under Contract No. HHSN27201800013C; J.H.K. performed this work as an employee of Tunnell Government Services, a subcontractor of Laulima Government Solutions, LLC, under Contract No. HHSN27201800013C. This work was also supported by the Simons Foundation Investigator in Mathematical Modeling of Living Systems award #327936 (O.Z.), the Canadian Natural Sciences and Engineering Research Council (NSERC) award RGPIN-2018-03898 (J.T.B.), and NSERC award RGPIN-2017-04636 (A.S.L.). S.K. was partially supported by funding from the Memorial University of Newfoundland School of Graduate Studies.

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Health and Human Services or of the institutions and companies affiliated with the authors, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

**Conflict of interest:** The authors declare no conflicts of interest.

**References**

Altschul, S. F. et al. (1997) ‘Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs’, Nucleic Acids Research, 25: 3389–402.

Anderson, B. et al. (1994) ‘Bacteriophage-like Particle of Rochalimaea henselae’, Molecular Microbiology, 13: 67–73.

Barbrian, K. D. and Minnick, M. F. (2000) ‘A Bacteriophage-like Particle from Bartonella bacilliformis’, Microbiology (Reading), 146: 599–609.

Bárdy, P. et al. (2020) ‘Structure and Mechanism of DNA Delivery of a Gene Transfer Agent’, Nature Communications, 11: 3034.

Berglund, E. C. et al. (2009) ‘Run-off Replication of Host-adaptability Genes Is Associated with Gene Transfer Agents in the Genome of Mouse-infecting Bartonella grahamii’, PLoS Genetics, 5: e1000564.

Bertani, G. (1999) ‘Transduction-Like Gene Transfer in the Methanogen Methanococcus voltae’, Journal of Bacteriology, 181: 2992–3002.

Biers, E. J. et al. (2008) ‘Occurrence and Expression of Gene Transfer Agent Genes in Marine Bacterioplankton’, Applied and Environmental Microbiology, 74: 2933–9.

Brimacombe, C. A. et al. (2013) ‘Quorum-sensing Regulation of a Capsular Polysaccharide Receptor for the Rhodobacter capsulatus Gene Transfer Agent (RcGTA)’, Molecular Microbiology, 87: 802–17.

Darboux, I., Cusson, M., and Volkoff, A. N. (2019) ‘The Dual Life of Ichnoviruses’, Current Opinion in Insect Science, 32: 47–53.

Drezen, J.-M. et al. (2017) ‘Endogenous Viruses of Parasitic Wasps: Variations on a Common Theme’, Current Opinion in Virology, 25: 41–8.

Esterman, E. S. et al. (2021) ‘Evolution of DNA Packaging in Gene Transfer Agents’, Virus Evolution, 7: veab015.

Felsenstein, J. (1978) ‘Cases in Which Parsimony or Compatibility Methods Will Be Positively Misleading’, Systematic Biology, 27: 401–10.

Fogg, P. C. M., Westbye, A. B., and Beatty, J. T. (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.

Fu, Y. et al. (2010) ‘High Diversity of Rhodobacterales in the Subarctic North Atlantic Ocean and Gene Transfer Agent Protein Expression in Isolated Strains’, Aquatic Microbial Ecology, 59: 283–93.

Gauthier, J., Drezen, J.-M., and Herniou, E. A. (2018) ‘The Recurrent Domestication of Viruses: Major Evolutionary Transitions in Parasitic Wasps’, Parasitology, 145: 713–23.

Guy, L. et al. (2013) ‘A Gene Transfer Agent and a Dynamic Repertoire of Secretion Systems Hold the Keys to the Explosive Radiation of the Emerging Pathogen Bartonella’, PLoS Genetics, 9: e1003393.

Herniou, E. A. et al. (2013) ‘When Parasitic Wasps Hijacked Viruses: Genomic and Functional Evolution of Polydnaviruses’, Philosophical Transactions of the Royal Society B: Biological Sciences, 368: 20130051.
Hoang, D. T. et al. (2018) 'UFBoot2: Improving the Ultrafast Bootstrap Approximation', Molecular Biology and Evolution, 35: 518–22.

Humphrey, S. B. et al. (1997) ‘Purification and Characterization of VSH-1, a Generalized Transducing Bacteriophage of Serpulina hydysenteriae’, Journal of Bacteriology, 179: 323–9.

Hynes, A. P. et al. (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, RvGTA’, Molecular Microbiology, 85: 314–25.

——— et al. (2016) ‘Functional and Evolutionary Characterization of a Gene Transfer Agent’s Multilocus “Genome”’, Molecular Biology and Evolution, 33: 2530–43.

International Committee on Taxonomy of Viruses (2022), ICTV Code: The International Code of Virus Classification and Nomenclature (ICVCN). March 2021. <https://ictv.global/about/code>.

Kalyaanamoorthy, S. et al. (2016) 'Functional and Evolutionary Characterization of a Gene Transfer Agent Particles Is Not Random', Microbiology and Molecular Biology Reviews, 84: e00061–19.

Kogay, R. et al. (2019) ‘Machine-learning Classification Suggests that Many Alphaproteobacterial Prophages May Instead Be Gene Transfer Agents’, Genome Biology and Evolution, 11: 2941–53.

——— et al. (2020) ‘Selection for Reducing Energy Cost of Protein Production Drives the GC Content and Amino Acid Composition Bias in Gene Transfer Agents’, mBio, 11: e01206–20.

Koonin, E. V. et al. (2020) ‘Global Organization and Proposed Megatransfer in the Virophere among the Replicator Space’, Microbiology and Molecular Biology Reviews, 85: e00061–19.

Koonin, E. V., and Krupovic, M. (2018) ‘The Depths of Virus Exaptation’, Current Opinion in Virology, 31: 1–8.

Kuhn, J. H. et al. (2020), ‘Expand, Amend, and Emend the International Code of Virus Classification and Nomenclature (ICVCN; “The Code”) and the Statutes to Clearly Define the Remit of the ICTV’. International Committee on Taxonomy of Viruses (ICTV) TaxoProp 2020.005G.R.Code_and_Statute_Change. <https://ictv.global/filebrowser/download/7146> accessed 20 Oct 2022.

——— et al. (2021), ‘Rename the Family Polynavirusidae (As Polynaviriformae), Rename the Genus Brachovirus (As Brachoviriform) and Rename All Polynaviriformid Species to Comply with the Newly ICTV-mandated Binomial Format’, International Committee on Taxonomy of Viruses (ICTV) TaxoProp 2021.006D.R.Polynaviriformae_1renfam_Srensp. <https://ictv.global/filebrowser/download/4986>.

Lang, A. S., and Beatty, J. T. (2007) ‘Importance of Widespread Gene Transfer Agent Genes in α-proteobacteria’, Trends in Microbiology, 15: 54–62.

Lang, A. S., Taylor, T. A., and Beatty, J. T. (2002) ‘Evolutionary Implications of Phylogenetic Analyses of the Gene Transfer Agent (GTA) of Rhodobacter capsulatus’, Journal of Molecular Evolution, 55: 534–43.

Lang, A. S., Westbye, A. B., and Beatty, J. T. (2017) ‘The Distribution, Evolution, and Roles of Gene Transfer Agents in Prokaryotic Genetic Exchange’, Annual Review of Virology, 4: 87–104.

Lang, A. S., Zhaxybayeva, O., and Beatty, J. T. (2012) ‘Gene Transfer Agents: Phage-like Elements of Genetic Exchange’, Nature Reviews Microbiology, 10: 472–82.

Letunic, I., and Bork, P. (2021) ‘Interactive Tree of Life (Itol) V5: An Online Tool for Phylogenetic Tree Display and Annotation’, Nucleic Acids Research, 49: W293–w96.

Marrs, B. (1974) ‘Genetic Recombination in Rhodopseudomonas capsulata’, Proceedings of the National Academy of Sciences of the United States of America, 71: 971–3.

Matson, E. G. et al. (2005) ‘Identification of Genes of VSH-1, a Prophage-Like Gene Transfer Agent of Brachyspira hydysenteriae’, Journal of Bacteriology, 187: 5885–92.

McDaniel, L. D. et al. (2010) ‘High Frequency of Horizontal Gene Transfer in the Oceans’, Science, 330: 50.

Minh, B. Q. et al. (2020) ‘IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era’, Molecular Biology and Evolution, 37: 1530–4.

Motro, Y. et al. (2009) ‘Identification of Genes Associated with Prophage-Like Gene Transfer Agents in the Pathogenic Intestinal Spirochaetes Brachyspira hydysenteriae, Brachyspira pilosicoli and Brachyspira intermedia’, Veterinary Microbiology, 134: 340–5.

Nagao, N. et al. (2015) ‘The Gene Transfer Agent-like Particle of the Marine Phototrophic Bacterium Rhodovulum sulfadophilum’, Biochemistry and Biophysics Reports, 4: 369–74.

O’Leary, N. A. et al. (2016) ‘Reference Sequence (Refseq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation’, Nucleic Acids Research, 44: D733–45.

Paliard, P. et al. (2020) ‘Cyclic di-GMP-mediated Regulation of Gene Transfer and Motility in Rhodobacter capsulatus’, Journal of Bacteriology, 202: e00554–19.

Petersen, J. M. et al. (2022) ‘The Naked Truth: An Updated Review on Nudiviruses and Their Relationship to Bracoviruses and Baculoviruses’, Journal of Invertebrate Pathology, 189: 1107718.

Québatte, M. et al. (2017) ‘Gene Transfer Agent Promotes Evolvability within the Fittest Subpopulation of a Bacterial Pathogen’, Cell Systems, 4: 611–21.e6.

Rapp, B. J., and Wall, J. D. (1987) ‘Genetic Transfer in Brucella melitensis and Brucella abortus’, Journal of Bacteriology, 179: 323–9.

Rozewicki, J. et al. (2019) ‘MAFFT-DASH: Integrated Protein Sequence and Structural Alignment’, Nucleic Acids Research, 47: W5–w10.

Shakya, M., Soucy, S. M., and Zhaxybayeva, O. (2017) ‘Insights into Origin and Evolution of α-proteobacterial Gene Transfer Agents’, Virus Evolution, 3: vex036.

Shimizu, T. et al. (2022) ‘Persulfide-responsive Transcription Factor SqrR Regulates Gene Transfer and Biofilm Formation via the Metabolic Modulation of Cyclic di-GMP in Rhodobacter capsulatus’, Microorganisms, 10: 908.

Shimodaira, H. (2002) ‘An Approximately Unbiased Test of Phylogenetic Tree Selection’, Systematic Biology, 51: 492–508.

Stanton, T. B. et al. (2008) ‘Collateral Effects of Antibiotics: Carbachol and Metronidazole Induce VSH-1 and Facilitate Gene Transfer among Brachyspira hydysenteriae Strains’, Applied and Environmental Microbiology, 74: 2950–6.

Steenwyk, J. L. et al. (2020) ‘ClipKIT: A Multiple Sequence Alignment Trimming Software for Accurate Phylogenomic Inference’, PLOS Biology, 18: e3001007.

Strand, M. R., and Burke, G. R. (2020) ‘Polydnaviruses: Evolution and Function’, Current Issues in Molecular Biology, 34: 163–82.

Tamarin, D. et al. (2018) ‘Origin and Evolution of the Bartonella Gene Transfer Agent’, Molecular Biology and Evolution, 35: 451–64.

Thézé, J. et al. (2011) ‘Paleozoic Origin of Insect Large dsDNA Viruses’, Proceedings of the National Academy of Sciences of the United States of America, 108: 15931–5.

Tomasch, J. et al. (2018) ‘Packaging of Dinoroseobacter shibae DNA into Gene Transfer Agent Particles Is Not Random’, Genome Biology and Evolution, 10: 359–69.

Umemori, E. et al. (1992) ‘A Phage in Bartonella bacilliformis’,Microbiology and Immunology, 36: 731–6.
Walker, P. J. et al. (2021) 'Changes to Virus Taxonomy and to the International Code of Virus Classification and Nomenclature Ratified by the International Committee on Taxonomy of Viruses (2021)', Archives of Virology, 166: 2633–48.

——— et al. (2022) 'Changes to Virus Taxonomy and to the International Code of Virus Classification and Nomenclature Ratified by the International Committee on Taxonomy of Viruses (2022)', Archives of Virology 167: 2429–40.

Westbye, A. B. et al. (2017b) 'The Rhodobacter capsulatus Gene Transfer Agent Is Induced by Nutrient Depletion and the RNAP Omega Subunit', Microbiology (Reading), 163: 1355–63.

Westbye, A. B., Beatty, J. T., and Lang, A. S. (2017a) 'Guaranteeing a Captive Audience: Coordinated Regulation of Gene Transfer Agent (GTA) Production and Recipient Capability by Cellular Regulators', Current Opinion in Microbiology, 38: 122–9.

Wolf, Y. I. et al. (2018) 'Origins and Evolution of the Global RNA Virome', mBio, 9: e02329–18.

Yen, H. C., Hu, N. T., and Marrs, B. L. (1979) 'Characterization of the Gene Transfer Agent Made by an Overproducer Mutant of Rhodopseudomonas capsulata', Journal of Molecular Biology, 131: 157–68.

Zhan, Y. et al. (2016) 'A Novel Roseobacter Phage Possesses Features of Podoviruses, Siphoviruses, Prophages and Gene Transfer Agents', Scientific Reports, 6: 30372.

Zhao, Y. et al. (2009) 'Gene Transfer Agent (GTA) Genes Reveal Diverse and Dynamic Roseobacter and Rhodobacter Populations in the Chesapeake Bay', The ISME Journal, 3: 364–73.