

**Part 1:** **TITLE, AUTHORS, APPROVALS, etc**

|  |  |  |
| --- | --- | --- |
| **Code assigned:** | **2022.001G** |  |
| **Short title:** Classification of Gene Transfer Agents (GTAs) as Viriforms | | |
|  | | |

**Author(s) and email address(es)**

|  |  |
| --- | --- |
| Kogay R, Koppenhöfer S, Beatty JT, Kuhn JH, Lang AS, Zhaxybayeva O | [Roman.Kogay.GR@dartmouth.edu](mailto:Roman.Kogay.GR@dartmouth.edu); [skoppenhofer@mun.ca](mailto:skoppenhofer@mun.ca); [j.beatty@ubc.ca](mailto:j.beatty@ubc.ca); [kuhnjens@mail.nih.gov](mailto:kuhnjens@mail.nih.gov); [aslang@mun.ca](mailto:aslang@mun.ca); [Olga.Zhaxybayeva@dartmouth.edu](mailto:Olga.Zhaxybayeva@dartmouth.edu) |

**Author(s) institutional address(es) (optional)**

|  |
| --- |
| * Dartmouth College, Hanover, NH, USA (RK, OZ) * Memorial University of Newfoundland, St. John’s, NL, Canada (SK, ASL) * University of British Columbia, Vancouver, BC, Canada (JTB) * Integrated Research Facility at Fort Detrick, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Frederick, MD, USA (JHK) |

**Corresponding author**

|  |
| --- |
| Zhaxybayeva O |

**List the ICTV Study Group(s) that have seen this proposal**

|  |
| --- |
| There are currently no Study Groups for GTAs; hence this proposal was sent, per advice from the ICTV President, to the ICTV President. |

**ICTV Study Group comments and response of proposer**

|  |
| --- |
| N/A |

**ICTV Study Group votes on proposal**

|  |  |  |  |
| --- | --- | --- | --- |
| **Study Group** | **Number of members** | | |
| **Votes support** | **Votes against** | **No vote** |
|  |  |  |  |

**Authority to use the name of a living person**

|  |  |
| --- | --- |
| **Is any taxon name used here derived from that of a living person (Y/N)** | Y |

|  |  |  |
| --- | --- | --- |
| **Taxon name** | **Person from whom the name is derived** | **Permission attached (Y/N)** |
| *Bartonegtaviriform andersoni* | Burt Anderson | Y |
| *Brachyspigtaviriform stantoni* | Thad Stanton | Y |
| *Dinogtaviriform tomaschi* | Jürgen Tomasch | Y |
| *Rhodobactegtaviriform marrsi* | Barry Marrs | Y |
| *Rhodovulugtaviriform kikuchii* | Yo Kikuchi | Y |
| *Ruegerigtaviriform cheni* | Feng Chen | Y |

**Submission dates**

|  |  |
| --- | --- |
| Date first submitted to SC Chair | May 27, 2022 |
| Date of this revision (if different to above) | November 3, 2022 |

**ICTV-EC comments and response of the proposer**

|  |
| --- |
| The EC voiced the following concerns:   1. There may not be a monophyletic distinction between caudoviricetes and GTAs   Our response: We agree with that assessment – all discussed GTAs indeed appear to have evolved from caudoviricetes, although a very long time ago likely from ancestors that are no longer extant. However, based on current ICVCN definitions, GTAs are no longer viruses, but are now viriforms – hence we proposed classifying them as such. We now explicitly acknowledge this issue in the revised proposal:  “Based on the evolutionary history of TerL proteins, 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 [[9](#_ENREF_9), [28](#_ENREF_28), [29](#_ENREF_29), [43](#_ENREF_43)], in the case of the rhodogtaviriformids for hundreds of millions of years [[43](#_ENREF_43)]. 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 [[31](#_ENREF_31), [40](#_ENREF_40), [44](#_ENREF_44)]. There is also mounting evidence that GTA genes experience selective pressures to be maintained in their host genomes [[23](#_ENREF_23), [30](#_ENREF_30)].”   1. Accession numbers should be provided for the GTAs that are proposed to be classified.   Our response: We apologize for the oversight. Accession numbers are now included in this revision via new Supplementary Tables S1–S6 (separate file) also detailing GTA genes, their locus tags, and their functional annotations in the host bacterium genomes. The Excel module has been updated accordingly.   1. Since GTA genomes are, by definition, partial rather than complete (or since it is difficult to impossible to define what “complete” means), minimal criteria for GTA classification ought to be outlined.   Our response: We agree that the unclear boundaries of GTA genomes are a complication for GTA taxonomy compared to viruses. We added the following paragraph:  “Here we outline initial steps to establish such a formal taxonomic scheme for well-characterized 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. We propose such data as minimal requirements for GTA classification.”   1. GTA genome cartoons should be added   Our response: genome cartoons have been added in the form of new Figures 1, 3, and 5.   1. GTA name changes should be deleted; instead, the TaxoProp should focus on taxa.   Our response: We deleted the GTA name change section.  Additional author comments:  An article detailing this proposal and providing additional information was recently published open access by the proposal authors in *Virus Evolution* at [https://doi.org/10.1093/ve/veac100](https://gcc02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdoi.org%2F10.1093%2Fve%2Fveac100&data=05%7C01%7Ckuhnjens%40niaid.nih.gov%7Caf0f836010084fb235f908dabe769491%7C14b77578977342d58507251ca2dc2b06%7C0%7C0%7C638031713757726307%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=bM%2BnWREbu2dh4BV93P9aR13TiL7%2BP4gPSbruYI8O024%3D&reserved=0). We replaced figures and other material in this revised TaxoProp with updated and more detailed text and material from that publication (Figures 1, 2, and 3; now Figures 2, 4, and 6) and added additional material (new Figure 7). |

**Part 2:** **NON-TAXONOMIC PROPOSAL**

**Text of proposal**

|  |
| --- |
| N/A |

**Part 3:** **TAXONOMIC PROPOSAL**

**Name of accompanying Excel module**

|  |
| --- |
| 2022.001G.Uc.v2.GTA\_viriforms.xlxs |

**Abstract**

|  |
| --- |
| Gene transfer agents (GTAs) are caudoviricete-derived entities that have been exapted by host bacteria and therefore have lost their mobility. GTAs are therefore analogous in their lifecycles to virus-derived entities found in parasitoid wasps that have recently been removed from the virosphere and reclassified as so-called viriforms. Here we propose an initial taxonomic framework for GTAs comprising three initial families for well-characterized GTAs. |

**Text of proposal**

|  |  |
| --- | --- |
| |  | | --- | | **Introduction**  In 2021, the ICTV officially recognized a new type of classifiable entities, viriforms. Viriforms are now defined in the ICVCN 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) [[19](#_ENREF_19), [26](#_ENREF_26)]. They are therefore classifiable into taxa with viriform-specific name suffixes.  The first replicators that have been classified as viriforms were those of the virus family *Polydnaviridae*, which consequently was renamed *Polydnaviriformidae* [[25-27](#_ENREF_25), [51](#_ENREF_51), [52](#_ENREF_52)]. Polydnaviriformid particles encapsidate multiple segments of circular double-stranded DNAs that, however, do not encode the entire polydnaviriformid genomes. Instead, the complete genomes are permanently endogenized into the hosts (*i.e.*, parasitoid wasp) genomes and inherited vertically. The result are non-mobile nonviral entities that are used by wasps in a mutualist fashion to deliver immunomodulatory genes into insects that serve as prey for the wasps [[8](#_ENREF_8), [14](#_ENREF_14)].  The ICVCN already recognized that there is at least one other group of viriforms, gene transfer agents (GTAs), that could be classified (ICVCN Rule 3.3) [[19](#_ENREF_19), [26](#_ENREF_26)].  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 [[30](#_ENREF_30), [31](#_ENREF_31)] 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 caudeviricete proteins or by image-based evidence of caudovirion-like particles) to form caudovirion-like particles; 2. GTAs encapsidate mostly 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 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 [[54](#_ENREF_54)]) and mediate horizontal gene transfer (HGT), typically among cells of the same species.  Here we outline initial steps to establish such a formal taxonomic scheme for well-characterized 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. We propose such data as minimal requirements for GTA classification.  **Proposed nomenclature for GTA taxa**  GTA taxon naming is proposed to follow the current ICVCN stipulations for viriforms:  “[t]he formal endings for taxon names of viriforms are the suffixes "‑*viriformia*" for realms, "‑*viriforma*" for subrealms, "‑*viriformae*" for kingdoms, "‑*viriformites*" for subkingdoms, "‑*viriformicota*" for phyla, "‑*viriformicotina*" for subphyla, "‑*viriformicetes*" for classes, "‑*viriformicetidae*" for subclasses, "‑*viriformales*" for orders, "‑*viriformineae*" for suborders, "‑*viriformidae*" for families, "‑*viriforminae*" for subfamilies, and "‑*viriform*" for genera and subgenera” (ICVCN Rule 3.26) [[19](#_ENREF_19), [26](#_ENREF_26)]  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.” [[19](#_ENREF_19)]  We propose adding the infix -*gta*- prior to the taxon-specific suffixes for immediate recognition of GTA-specific taxa (e.g., -*gtaviriform*).  **Proposed taxonomic framework for GTAs**  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 *R. capsulatu*s (*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 since its discovery [[33](#_ENREF_33)], RcGTA was the only known GTA. Now we know that homologous GTAs are produced by other bacteria fromtheorder *Rhodobacterales*: *Dinoroseobacter shibae* (Dinoroseobacter shibae gene transfer agent [DsGTA]) [[49](#_ENREF_49)]*, Ruegeria* *pomeroyi* (Ruegeria pomeroyi gene transfer agent [RpGTA]) [[6](#_ENREF_6)], and *Rhodovulum* *sulfidophilum* (Rhodovulum sulfidophilum gene transfer agent [RsGTA]) [[38](#_ENREF_38)]. Additionally, genes encoding RcGTA-like GTAs are conserved in most genomes in the order *Rhodobacterales* and in many genomes of the alphaproteobacterial orders *Caulobacterales*, *Sphingomonadales*, *Parvibaculales,* and *Hyphomicrobiales* (formerly *Rhizobiales*) [[22](#_ENREF_22), [28](#_ENREF_28), [29](#_ENREF_29), [43](#_ENREF_43)].  RcGTA and RcGTA-like GTA genes are similar in sequence to those of viruses classified in the uroviricot class *Caudoviricetes* (*Duplodnaviria*: *Heunggongvirae*) [[43](#_ENREF_43)]. These GTAs are transmitted vertically from a bacterial parent to progeny during cell division [[29](#_ENREF_29), [43](#_ENREF_43)], similar to propagation of temperate viruses (“prophages”). However, in contrast to temperate virus genomes, the set of genes required for production of the 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 [[18](#_ENREF_18)], cumulatively spanning approximately 20 kilobases (kb) (**Figure 1** and **Supplementary Table S1**). Moreover, cellular regulatory genes are involved in controlling GTA particle production [[53](#_ENREF_53)], adding another factor that makes the GTA genome difficult to differentiate from its host’s genome.  RcGTA particles resemble virions of caudoviricetes [[56](#_ENREF_56)] and have been structurally characterized at high resolution [[4](#_ENREF_4)]. 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 [[56](#_ENREF_56)]. The DNA is also encapsidated at 10–25% lower density than typical caudoviricetes [[4](#_ENREF_4)]. 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 [[53](#_ENREF_53)]. Only 0.1–3.0% of cells produce GTA particles [[11](#_ENREF_11), [17](#_ENREF_17)], whereas the remaining cells produce a GTA receptor [[7](#_ENREF_7)].  Compositionally, structural proteins encoded by RcGTA and RcGTA-like GTAs are biased towards amino acids that are energetically cheaper to produce [[23](#_ENREF_23)]. 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 [[22](#_ENREF_22)].  In a comprehensive evolutionary analysis of homologs of the large subunit of the DNA packaging terminase enzyme (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 [[9](#_ENREF_9)]. 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 proteins (HK97-MCP, encoded by the *g5* gene in the GTA genome, is the hallmark protein that defines the virus realm *Duplodnaviria* [[24](#_ENREF_24)]). Consistent with an earlier analysis [[9](#_ENREF_9)], RcGTA and RcGTA-like GTAs formed a clade closely related to, but distinct from, caudoviricetes (**Figure 2**), with a few exceptions that are likely artefacts of phylogenetic reconstruction.  Specifically, in the TerL phylogeny (**Figure 2A**), all viral homologs except one (Caulobactervirus Sansa) are separated from GTA proteins (with a solid bootstrap support of 81%). Caulobactervirus Sansa groups with one GTA sequence from a bacterium of the order *Sphingomonadales* (with a low bootstrap support of 50%), whereas all other GTAs of *Sphingomonadales* bacteria group together (with a strong bootstrap support of 96%). We hypothesize that the phylogenetic placement of the Caulobactervirus Sansa TerL is due to the long-branch attraction artefact [[10](#_ENREF_10)]. We searched for a maximum-likelihood tree in which caudoviricete- and GTA-derived TerLs were required to group separately from each other and compared that tree to the tree depicted in **Figure 2A**. We found that the likelihoods of the two trees are not significantly different (approximately unbiased [AU] test; p-value = 0.555), confirming that the placement of the Caulobactervirus Sansa sequence within the GTA sequences is unreliable.  In the HK97-MCP phylogeny (**Figure 2B**), GTAs and most caudoviricetes are separated by a branch with 63% bootstrap support. Several caudoviricetes that group within GTAs are located on long branches, are 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 **Figure 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 **Figure 2B**.  In the **Figure 2** trees, GTA branches have shorter lengths than their caudoviricete counterparts, conforming with the reported slower evolutionary rate of GTAs compared to viruses [[43](#_ENREF_43)]. 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 [[28](#_ENREF_28), [29](#_ENREF_29), [43](#_ENREF_43)].  Together, these results justify the classification of RcGTA and three RcGTA-like GTAs in a common viriform taxon: family *Rhodogtaviriformidae* (from *Rhodobacterales*, infix -*gta*-, and family-specific suffix -*viriformidae*). Given 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 *Dinoroseobacter*, infix -*gta*-, and genus-specific suffix -*viriform*) to include one new species, *Dinogtaviriform tomaschi* (species epithet to honor GTA researcher Jürgen Tomasch, who was instrumental in the discovery of DsGTA) for DsGTA (**Supplementary Table S2**); * *Rhodobactegtaviriform* (named after RcGTA host genus Rhodobacter, infix -*gta*-, and genus-specific suffix -*viriform*) to include one new species, *Rhodobactegtaviriform marrsi* (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 (BLPs), that contained heterogenous DNA from *Bartonella* host genomes were first characterized in *B. henselae* [[2](#_ENREF_2)], and noted to be similar in structure to particles produced by *B. bacilliformis* [[50](#_ENREF_50)]. 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 [[3](#_ENREF_3)]. 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*) [[13](#_ENREF_13)]. BaGTA genes were initially proposed to be located within a single cluster of 11–13 genes spanning approximately 14 kb [[13](#_ENREF_13)]. However, a subsequent screen for genes essential for BaGTA functionality identified a total of 29 genes located within a larger (approximately 79-kb-long) region [[41](#_ENREF_41)] (**Figure 3** and **Supplementary Table S5**). Homologs of BaGTA genes (BaGTA-like GTAs) were found in the genomes of multiple species of *Bartonella* [[5](#_ENREF_5), [13](#_ENREF_13), [48](#_ENREF_48)]. BaGTA genes are located near an active virus-derived origin of replication and next to genes encoding secretion systems [[13](#_ENREF_13)]. 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 [[13](#_ENREF_13), [41](#_ENREF_41)]. 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 [[13](#_ENREF_13)]. However, actual GTA-mediated DNA transfer among bacterial cells has only been demonstrated for *B. henselae* [[13](#_ENREF_13)]. There, BaGTA production is restricted to a distinct subpopulation of fast-growing cells, which comprise about 6% of the total population [[41](#_ENREF_41)], and the uptake of BaGTA-packaged DNA was proposed to be limited to cells undergoing division [[41](#_ENREF_41)].  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 [[2](#_ENREF_2)]. The head diameter of the *B. bacilliformis* particles was originally measured at 40 nm [[50](#_ENREF_50)] and subsequently 80 nm [[3](#_ENREF_3)]. Those of *B. grahamii* were reported as possessing long non-contractile tails and icosahedral heads of 50–70 nm or 80 nm and tails of 100 nm [[5](#_ENREF_5)]. Although BaGTA particles are potentially able to package the entire main structural gene cluster of 11–13 genes, they cannot package all 29 genes required for BaGTA production due to a capacity of 14 kb [[2](#_ENREF_2), [13](#_ENREF_13), [31](#_ENREF_31)].  In the TerL phylogeny, BaGTA-like homologs are separated from almost all caudoviricetes by longer branches (with 100% bootstrap support; **Figure 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% 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* [[57](#_ENREF_57)]. In the HK97-MCP phylogeny, BaGTA homologs are located on shorter branches than their caudoviricete counterparts and are separated from caudoviricete homologs with 100% bootstrap support (**Figure 4B**). Phylogenomic analyses suggest that *Bartonella* GTAs have co-evolved with their hosts [[48](#_ENREF_48)].  Together, these results justify the classification of BaGTA and BaGTA-like GTAs in a common viriform 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: *Bartonegtaviriform* (named after BaGTA host genus *Bartonella*, infix -*gta*-, and genus-specific suffix -*viriform*) including one new species, *Bartonegtaviriform andersoni* (species epithet to honor GTA researcher Burt Anderson, who first discovered BaGTA particles [[2](#_ENREF_2)]) for BaGTA. GTAs of spirochaetes A GTA originally called virus of *Serpulina hyodysenteriae* 1 (VSH-1) was identified in *Brachyspira* (formerly *Serpulina*) *hyodysenteriae* (*Spirochaetota*: *Spirochaetia*: *Brachyspirales*: *Brachyspiraceae*) [[16](#_ENREF_16)]. 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 production of BhGTA particles—i.e., the BhGTA “genome”—is 16.3 kb in length [[34](#_ENREF_34)] (**Figure 5** and **Supplementary Table S6**).  BhGTA particles have a head diameter of 45 nm and a flexible non-contractile tail of 65 nm [[16](#_ENREF_16)]. Like other GTAs, BhGTA is unable to package and transfer its entire genome, given the limiting capacity of 7.5 kb [[16](#_ENREF_16), [34](#_ENREF_34)]. 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 [[16](#_ENREF_16)] 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 and antibiotics, results in large-scale lysis of cells [[46](#_ENREF_46)]. 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 [[46](#_ENREF_46)], 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 other members of the genus *Brachyspira*, but there is no gene synteny in their organization [[37](#_ENREF_37)]. Unlike in rhodogtaviriformids and bartogtaviriformids, an endolysin-encoding gene is the only gene in the BhGTA genome that has a significant sequence similarity to caudoviricete genes in the National Center for Biotechnology Information (NCBI) Reference Sequence (RefSeq) database (accessed in May 2022). Some genes encoding the BhGTA particle proteins were experimentally validated (including endolysin), and the particles structurally resemble those of caudoviricetes [[34](#_ENREF_34)]. 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% bootstrap support; **Figure 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%. In a phylogenetic tree, the *Brachyspira* TerLs are separated from all caudoviricete TerLs by a longer branch (with 100% bootstrap support; **Figure 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* [[37](#_ENREF_37)].  Together, these results justify the classification of BhGTA and BhGTA-like GTAs in a common viriform taxon, family *Brachygtaviriformidae* (form *Brachyspira*, infix -*gta*-, and family-specific suffix -*viriformidae*). For now, we argue that at least one brachygtaviriformid genus ought to be established: *Brachyspigtaviriform* (named after BhGTA host genus *Brachyspira*, infix -*gta*-, and genus-specific suffix -*viriform*) to include one new species, *Brachyspigtaviriform stantoni* (species epithet to honor GTA researcher Thaddeus Stanton, who first discovered BhGTA particles [[16](#_ENREF_16)]) for BhGTA. Independent origins of the three major GTA clades Genes from the genomes of these three GTAs are either not homologous or too divergent to have significant sequence similarity in BLASTP searches of the encoded proteins. For example, pairwise amino-acid identity of TerLs, which is one of the most conserved GTA and caudoviricete proteins, is 14–20% among RcGTA, BaGTA, and BhGTA. Nevertheless, an iterative clustering-alignment-phylogeny procedure [[55](#_ENREF_55)] established the homology among known TerL proteins that include RcGTA, BaGTA, and putative BhGTA TerLs [[9](#_ENREF_9)]. The evolutionary history of RcGTA-like, BaGTA-like, putative BhGTA-like TerLs, and their closest known caudoviricete homologs (**Figure 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.  **Conclusions**  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 are 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 (**Figure 2** and **Figure 4**) [[9](#_ENREF_9), [43](#_ENREF_43)]. The spirochaete 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 (**Figure 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 [[6](#_ENREF_6), [12](#_ENREF_12), [35](#_ENREF_35), [58](#_ENREF_58)]. In a genome-wide screen of 1,423 alphaproteobacterial genomes, 57.5% were found to encode RcGTA-like “genomes”, which are often annotated as either intact or incomplete prophages [[22](#_ENREF_22)]. 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 (**Figure 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 [[9](#_ENREF_9), [28](#_ENREF_28), [29](#_ENREF_29), [43](#_ENREF_43)], in the case of the rhodogtaviriformids for hundreds of millions of years [[43](#_ENREF_43)]. 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 [[31](#_ENREF_31), [40](#_ENREF_40), [44](#_ENREF_44)]. There is also mounting evidence that GTA genes experience selective pressures to be maintained in their host genomes [[23](#_ENREF_23), [30](#_ENREF_30)]. 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. | |

**Supporting evidence**

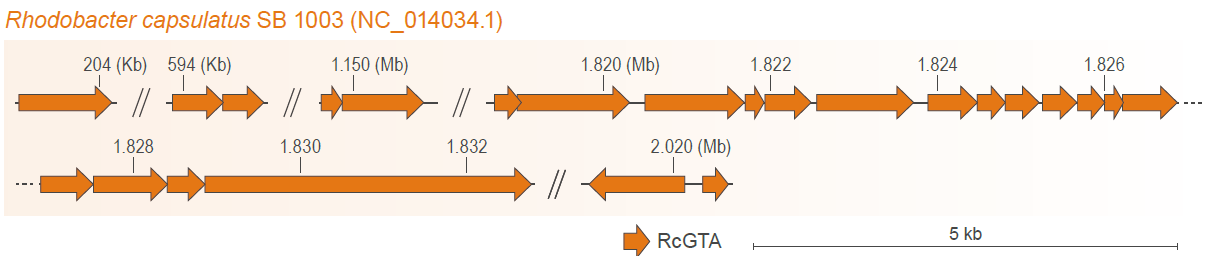
To identify alphaproteobacterial type I GTAs, we searched for RcGTA-like sequences in 1,248 complete alphaproteobacterial genomes extracted from the NCBI RefSeq database (accessed in October 2020) using GTA-Hunter [[22](#_ENREF_22)]. We identified 503 genomes that contained at least six RcGTA homologs in the same genetic neighborhood and had both *g2* (encoding TerL) and *g5* (encoding HK97-MCP) genes. To remove redundancy, we clustered genomes into the operational taxonomic units (OTUs) using an average nucleotide identity threshold of 95%. 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 [[1](#_ENREF_1)] of the RefSeq database (accessed in March 2021) [[39](#_ENREF_39)], using TerL and HK97-MCP proteins from representative GTAs as queries, an e-value cutoff of 0.001, and query coverage of at least 50%. 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 -linsi option [[21](#_ENREF_21)]. We reconstructed phylogenetic trees using IQ-TREE v2 [[36](#_ENREF_36)], identifying the best substitution models using the built-in ModelFinder [[20](#_ENREF_20)]. 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 [[15](#_ENREF_15)]. Additionally, for both protein phylogenies, we reconstructed a phylogenetic tree in IQ-TREE v2 [[36](#_ENREF_36)] 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 [[45](#_ENREF_45)], 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 57 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 57 genomes using a 95% 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 cutoff of 0.001, and query coverage of at least 50%) 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.

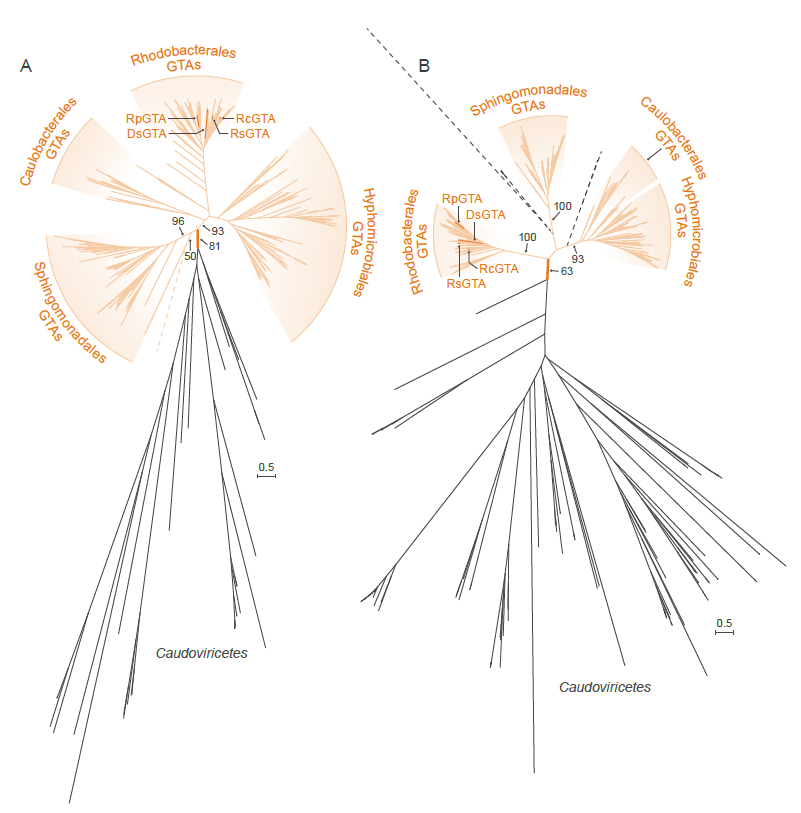
To identify GTAs of spirochaetes, we used BhGTA’s MCP sequence (GenBank accession number WP\_012671344.1) as a query in a BLASTP search (with an e-value cutoff of 0.001 and query coverage of at least 50%) against the 13 complete *Brachyspira* genomes extracted from the RefSeq database (accessed in May 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 cutoff of 0.001 and query coverage of at least 50%) against the same set of 13 genomes. For endolysins, we only retained matches that co-localized within the BhGTA region on the chromosome. We clustered 13 genomes using a 95% 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 cutoff of 0.001 and query coverage of at least 50%) 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 [[42](#_ENREF_42)] and trimmed the obtained alignment using ClipKIT with -gappy option [[47](#_ENREF_47)]. We computed the phylogenetic tree using IQ-TREE v2 [[36](#_ENREF_36)] as described above with the LG+F+R10 substitution model selected by ModelFinder. We rooted the tree using a larger TerL phylogeny presented in [[9](#_ENREF_9)].

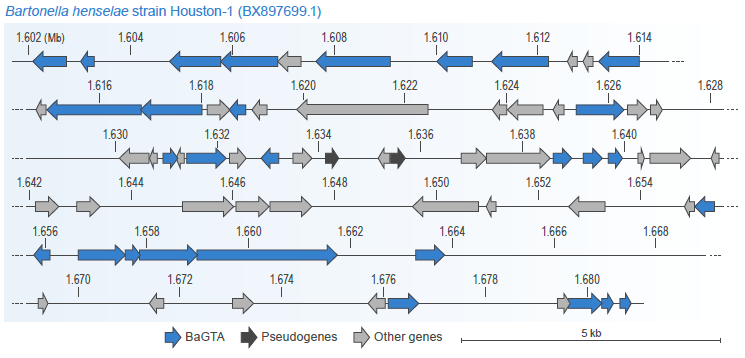
We visualized all phylogenetic trees in iTOL v6 [[32](#_ENREF_32)].

****

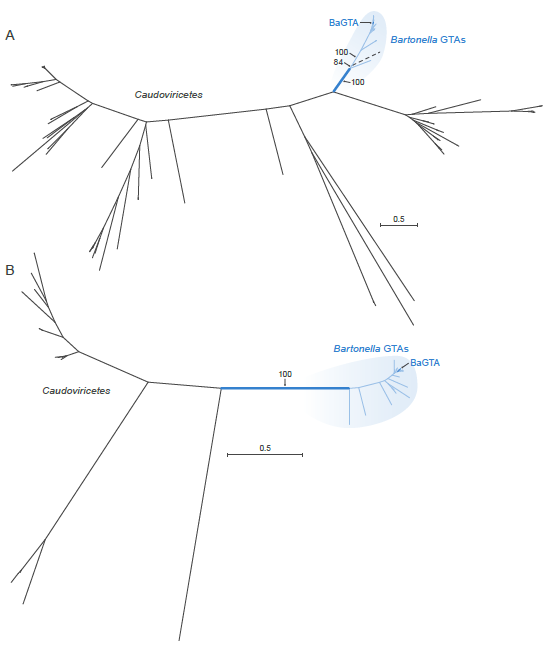
**Figure 1. Genome of Rhodobacter capsulatus gene transfer agent (RcGTA).** Genes(arrows) are depicted to scale, in their locations in the host genome (*R. capsulatus*). Exact coordinates of the RcGTA genes, their locus tags, and their functional annotations are listed in **Supplementary Table S1**.

****

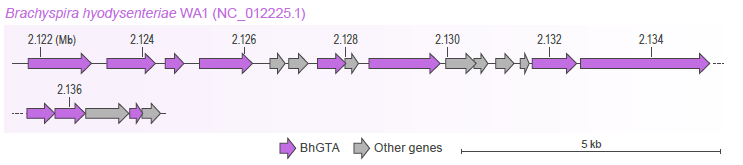
**Figure 2. Maximum Likelihood phylogenies of (A) large terminase (TerL) subunits and (B) HK97 major capsid protein (HK97-MCP) sequences of rhodogtaviriformidsandtheir closest known caudoviricete homologs.** Alphaproteobacterial type I gene transfer agent (GTA) (rhodogtaviriformid) lineages are shown in orange. Caudoviricete lineages that are nested within GTA lineages are shown in dashed black lines. Other caudoviricete lineages are shown in solid black lines. 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, Rhodovolum sulfidophilum, gene transfer agent.



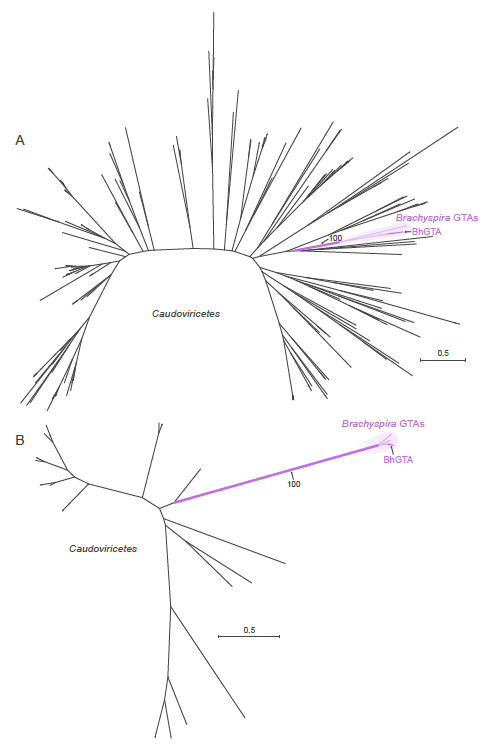
**Figure 3. Genome of Bartonella gene transfer agent (BaGTA).** Genes(arrows) are depicted to scale, in their locations in the host genome (*B. henselae*). Exact coordinates of the BaGTA genes, their locus tags, and their functional annotations are listed in **Supplementary Table S5**.



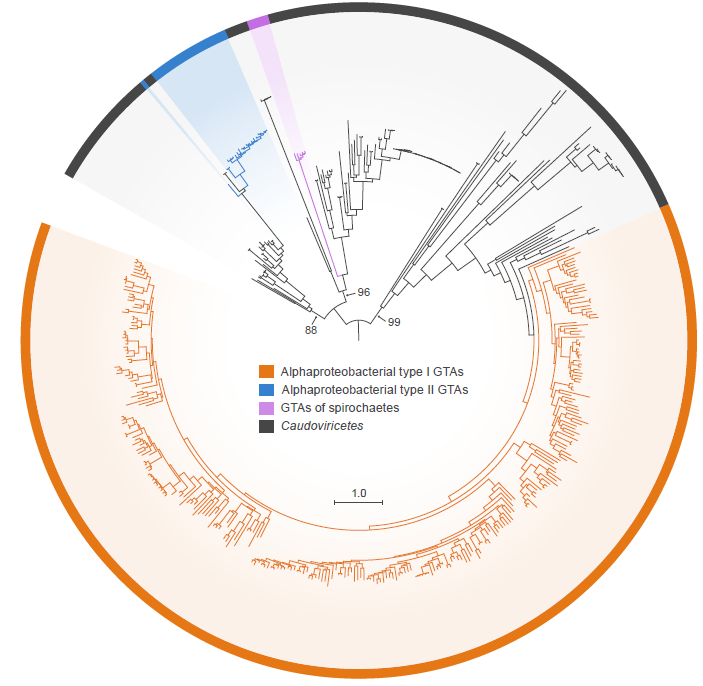
**Figure 4. Maximum Likelihood phylogenies of (A) large terminase (TerL) subunits and (B) HK97 major capsid protein (HK97-MCP) sequences of bartogtaviriformidsand their closest known caudoviricete homologs.** Alphaproteobacterial type II gene transfer agent (GTA) (bartogtaviriformid) lineages are shown in blue. Caudoviricete lineages are shown in black. Two nearly identical caudoviricete lineages that are nested within GTA lineages are shown in dashed black 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.



**Figure 5. Genome of Brachyspira hyodysenteriae gene transfer agent (BhGTA).** Genes(arrows) are depicted to scale, in their locations in the host genome (*B. hyodysenteriae*). Exact coordinates of the BhGTA genes, their locus tags, and their functional annotations are listed in **Supplementary Table S6**.

****

**Figure 6. Maximum Likelihood phylogenies of (A) endolysin and (B) the putative large terminase (TerL) subunits of brachygtaviriformidsand their closest known caudoviricete homologs.** *Brachyspira* gene transfer agent (GTA) lineages are shown in purple. Caudoviricete lineages are shown in black. 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.



**Figure 7. Maximum Likelihood phylogeny of the large terminase (TerL) subunits of three major clades of GTAs and their closest known caudoviricete homologs.** This tree includes all TerL homologs from Figures 2A, 4A, and 6A, using their color coding. 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.

**References**

1. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-402. PMID: 9254694. PMCID: PMC146917. doi: 10.1093/nar/25.17.3389.

2. Anderson B, Goldsmith C, Johnson A, Padmalayam I, Baumstark B (1994) Bacteriophage-like particle of *Rochalimaea henselae*. Mol Microbiol 13:67-73. PMID: 7527118. doi: 10.1111/j.1365-2958.1994.tb00402.x.

3. Barbian KD, Minnick MF (2000) A bacteriophage-like particle from *Bartonella bacilliformis*. Microbiology (Reading) 146 (Pt 3):599-609. PMID: 10746763. doi: 10.1099/00221287-146-3-599.

4. Bárdy P, Füzik T, Hrebík D, Pantůček R, Thomas Beatty J, Plevka P (2020) Structure and mechanism of DNA delivery of a gene transfer agent. Nat Commun 11:3034. PMID: 32541663. PMCID: PMC7296036. doi: 10.1038/s41467-020-16669-9.

5. Berglund EC, Frank AC, Calteau A, Vinnere Pettersson O, Granberg F, Eriksson A-S, Näslund K, Holmberg M, Lindroos H, Andersson SGE (2009) Run-off replication of host-adaptability genes is associated with gene transfer agents in the genome of mouse-infecting *Bartonella grahamii*. PLoS Genet 5:e1000546. PMID: 19578403. PMCID: PMC2697382. doi: 10.1371/journal.pgen.1000546.

6. Biers EJ, Wang K, Pennington C, Belas R, Chen F, Moran MA (2008) Occurrence and expression of gene transfer agent genes in marine bacterioplankton. Appl Environ Microbiol 74:2933-9. PMID: 18359833. PMCID: PMC2394915. doi: 10.1128/aem.02129-07.

7. Brimacombe CA, Stevens A, Jun D, Mercer R, Lang AS, Beatty JT (2013) Quorum-sensing regulation of a capsular polysaccharide receptor for the *Rhodobacter capsulatus* gene transfer agent (RcGTA). Mol Microbiol 87:802-17. PMID: 23279213. PMCID: PMC3641046. doi: 10.1111/mmi.12132.

8. Drezen JM, Leobold M, Bezier A, Huguet E, Volkoff AN, Herniou EA (2017) Endogenous viruses of parasitic wasps: variations on a common theme. Curr Opin Virol 25:41-48. PMID: 28728099. doi: 10.1016/j.coviro.2017.07.002.

9. Esterman ES, Wolf YI, Kogay R, Koonin EV, Zhaxybayeva O (2021) Evolution of DNA packaging in gene transfer agents. Virus Evol 7:veab015. PMID: 33732503. PMCID: PMC7947584. doi: 10.1093/ve/veab015.

10. Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. Syst Biol 27:401-410. doi: 10.1093/sysbio/27.4.401.

11. 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. PMID: 22916305. PMCID: PMC3423380. doi: 10.1371/journal.pone.0043772.

12. Fu Y, MacLeod DM, Rivkin RB, Chen F, Buchan A, Lang AS (2010) High diversity of *Rhodobacterales* in the subarctic North Atlantic Ocean and gene transfer agent protein expression in isolated strains. Aquat Microb Ecol 59:283–293.

13. Guy L, Nystedt B, Toft C, Zaremba-Niedzwiedzka K, Berglund EC, Granberg F, Näslund K, Eriksson A-S, Andersson SG (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 Genet 9:e1003393. PMID: 23555299. PMCID: PMC3610622. doi: 10.1371/journal.pgen.1003393.

14. Herniou EA, Huguet E, Theze J, Bezier A, Periquet G, Drezen JM (2013) When parasitic wasps hijacked viruses: genomic and functional evolution of polydnaviruses. Philos Trans R Soc Lond B Biol Sci 368:20130051. PMID: 23938758. PMCID: PMC3758193. doi: 10.1098/rstb.2013.0051.

15. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018) UFBoot2: improving the Ultrafast Bootstrap approximation. Mol Biol Evol 35:518-522. PMID: 29077904. PMCID: PMC5850222. doi: 10.1093/molbev/msx281.

16. Humphrey SB, Stanton TB, Jensen NS, Zuerner RL (1997) Purification and characterization of VSH-1, a generalized transducing bacteriophage of *Serpulina hyodysenteriae*. J Bacteriol 179:323-9. PMID: 8990282. PMCID: PMC178700. doi: 10.1128/jb.179.2.323-329.1997.

17. 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 85:314-25. PMID: 22640804. doi: 10.1111/j.1365-2958.2012.08113.x.

18. Hynes AP, Shakya M, Mercer RG, Grüll MP, Bown L, Davidson F, Steffen E, Matchem H, Peach ME, Berger T, Grebe K, Zhaxybayeva O, Lang AS (2016) Functional and evolutionary characterization of a gene transfer agent's multilocus "genome". Mol Biol Evol 33:2530-43. PMID: 27343288. PMCID: PMC5026251. doi: 10.1093/molbev/msw125.

19. International Committee on Taxonomy of Viruses (2021) The International Code of Virus Classification and Nomenclature (ICVCN). March 2021. <https://talk.ictvonline.org/information/w/ictv-information/383/ictv-code>.

20. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587-589. PMID: 28481363. PMCID: PMC5453245. doi: 10.1038/nmeth.4285.

21. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772-80. PMID: 23329690. PMCID: PMC3603318. doi: 10.1093/molbev/mst010.

22. Kogay R, Neely TB, Birnbaum DP, Hankel CR, Shakya M, Zhaxybayeva O (2019) Machine-learning classification suggests that many alphaproteobacterial prophages may instead be gene transfer agents. Genome Biol Evol 11:2941-2953. PMID: 31560374. PMCID: PMC6821227. doi: 10.1093/gbe/evz206.

23. Kogay R, Wolf YI, Koonin EV, Zhaxybayeva O (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. PMID: 32665274. PMCID: PMC7360931. doi: 10.1128/mBio.01206-20.

24. Koonin EV, Dolja VV, Krupovic M, Varsani A, Wolf YI, Yutin N, Zerbini FM, Kuhn JH (2020) Global organization and proposed megataxonomy of the virus world. Microbiol Mol Biol Rev 84:e00061-19. PMID: 32132243. PMCID: PMC7062200. doi: 10.1128/mmbr.00061-19.

25. Koonin EV, Dolja VV, Krupovic M, Kuhn JH (2021) Viruses defined by the position of the virosphere within the replicator space. Microbiol Mol Biol Rev 85:e0019320. PMID: 34468181. PMCID: PMC8483706. doi: 10.1128/mmbr.00193-20.

26. Kuhn J, Dolja VV, Krupovic M, Adriaenssens EM, Di Serio F, Dutilh BE, Flores R, Harrach B, Mushegian A, Owens B, Randles J, Rubino L, Sabanadzovic S, Simmonds P, Varsani A, Zerbini M, Koonin EV (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://talk.ictvonline.org/files/ictv_official_taxonomy_updates_since_the_8th_report/m/general-2008/11061>.

27. Kuhn J, Postler T, Dolja V, Krupovic M, Adriaenssens E, Di Serio F, Dutilh B, Flores R, Harrach B, Mushegian A, Owens B, Randles J, Rubino L, Sabanadzovic S, Simmonds P, Varsani A, Zerbini M, Koonin E (2021) Rename the family Polydnaviridae (as Polydnaviriformidae), rename the genus Bracovirus (as Bracoviriform) and rename all polydnaviriformid species to comply with the newly ICTV-mandated binomial format. International Committee on Taxonomy of Viruses (ICTV) TaxoProp 2021.006D.R.Polydnaviriformidae\_1renfam\_3rensp. <https://talk.ictvonline.org/files/ictv_official_taxonomy_updates_since_the_8th_report/m/animal-dna-viruses-and-retroviruses/13248>.

28. Lang AS, Taylor TA, Beatty JT (2002) Evolutionary implications of phylogenetic analyses of the gene transfer agent (GTA) of *Rhodobacter capsulatus*. J Mol Evol 55:534-43. PMID: 12399927. doi: 10.1007/s00239-002-2348-7.

29. Lang AS, Beatty JT (2007) Importance of widespread gene transfer agent genes in α-proteobacteria. Trends Microbiol 15:54-62. PMID: 17184993. doi: 10.1016/j.tim.2006.12.001.

30. Lang AS, Zhaxybayeva O, Beatty JT (2012) Gene transfer agents: phage-like elements of genetic exchange. Nat Rev Microbiol 10:472-82. PMID: 22683880. PMCID: PMC3626599. doi: 10.1038/nrmicro2802.

31. Lang AS, Westbye AB, Beatty JT (2017) The distribution, evolution, and roles of Ggne transfer agents in prokaryotic genetic exchange. Annu Rev Virol 4:87-104. PMID: 28784044. doi: 10.1146/annurev-virology-101416-041624.

32. Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293-w296. PMID: 33885785. PMCID: PMC8265157. doi: 10.1093/nar/gkab301.

33. Marrs B (1974) Genetic recombination in *Rhodopseudomonas capsulata*. Proc Natl Acad Sci U S A 71:971-3. PMID: 4522805. PMCID: PMC388139. doi: 10.1073/pnas.71.3.971.

34. Matson EG, Thompson MG, Humphrey SB, Zuerner RL, Stanton TB (2005) Identification of genes of VSH-1, a prophage-like gene transfer agent of *Brachyspira hyodysenteriae*. J Bacteriol 187:5885-92. PMID: 16109929. PMCID: PMC1196157. doi: 10.1128/jb.187.17.5885-5892.2005.

35. McDaniel LD, Young E, Delaney J, Ruhnau F, Ritchie KB, Paul JH (2010) High frequency of horizontal gene transfer in the oceans. Science 330:50. PMID: 20929803. doi: 10.1126/science.1192243.

36. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37:1530-1534. PMID: 32011700. PMCID: PMC7182206. doi: 10.1093/molbev/msaa015.

37. Motro Y, La T, Bellgard MI, Dunn DS, Phillips ND, Hampson DJ (2009) Identification of genes associated with prophage-like gene transfer agents in the pathogenic intestinal spirochaetes *Brachyspira hyodysenteriae*, *Brachyspira pilosicoli* and *Brachyspira intermedia*. Vet Microbiol 134:340-5. PMID: 18950961. doi: 10.1016/j.vetmic.2008.09.051.

38. Nagao N, Yamamoto J, Komatsu H, Suzuki H, Hirose Y, Umekage S, Ohyama T, Kikuchi Y (2015) The gene transfer agent-like particle of the marine phototrophic bacterium *Rhodovulum sulfidophilum*. Biochem Biophys Rep 4:369-374. PMID: 29124226. PMCID: PMC5669443. doi: 10.1016/j.bbrep.2015.11.002.

39. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44:D733-45. PMID: 26553804. PMCID: PMC4702849. doi: 10.1093/nar/gkv1189.

40. Pallegar P, Peña-Castillo L, Langille E, Gomelsky M, Lang AS (2020) Cyclic di-GMP-mediated regulation of gene transfer and motility in *Rhodobacter capsulatus*. J Bacteriol 202. PMID: 31659012. PMCID: PMC6941535. doi: 10.1128/JB.00554-19.

41. Québatte M, Christen M, Harms A, Körner J, Christen B, Dehio C (2017) Gene transfer agent promotes evolvability within the fittest subpopulation of a bacterial pathogen. Cell Syst 4:611-621.e6. PMID: 28624614. PMCID: PMC5496983. doi: 10.1016/j.cels.2017.05.011.

42. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K (2019) MAFFT-DASH: integrated protein sequence and structural alignment. Nucleic Acids Res 47:W5-w10. PMID: 31062021. PMCID: PMC6602451. doi: 10.1093/nar/gkz342.

43. Shakya M, Soucy SM, Zhaxybayeva O (2017) Insights into origin and evolution of α-proteobacterial gene transfer agents. Virus Evol 3:vex036. PMID: 29250433. PMCID: PMC5721377. doi: 10.1093/ve/vex036.

44. Shimizu T, Aritoshi T, Beatty JT, Masuda T (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. PMID: 35630353. PMCID: PMC9143464. doi: 10.3390/microorganisms10050908.

45. Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. Syst Biol 51:492-508. PMID: 12079646. doi: 10.1080/10635150290069913.

46. Stanton TB, Humphrey SB, Sharma VK, Zuerner RL (2008) Collateral effects of antibiotics: carbadox and metronidazole induce VSH-1 and facilitate gene transfer among *Brachyspira hyodysenteriae* strains. Appl Environ Microbiol 74:2950-6. PMID: 18359835. PMCID: PMC2394957. doi: 10.1128/aem.00189-08.

47. Steenwyk JL, Buida TJ, III, Li Y, Shen X-X, Rokas A (2020) ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. PLoS Biol 18:e3001007. PMID: 33264284. PMCID: PMC7735675. doi: 10.1371/journal.pbio.3001007.

48. Tamarit D, Neuvonen M-M, Engel P, Guy L, Andersson SGE (2018) Origin and evolution of the *Bartonella* gene transfer agent. Mol Biol Evol 35:451-464. PMID: 29161442. doi: 10.1093/molbev/msx299.

49. Tomasch J, Wang H, Hall ATK, Patzelt D, Preusse M, Petersen J, Brinkmann H, Bunk B, Bhuju S, Jarek M, Geffers R, Lang AS, Wagner-Döbler I (2018) Packaging of *Dinoroseobacter shibae* DNA into gene transfer agent particles is not random. Genome Biol Evol 10:359-369. PMID: 29325123. PMCID: PMC5786225. doi: 10.1093/gbe/evy005.

50. Umemori E, Sasaki Y, Amano K, Amano Y (1992) A phage in *Bartonella bacilliformis*. Microbiol Immunol 36:731-6. PMID: 1406375. doi: 10.1111/j.1348-0421.1992.tb02075.x.

51. Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM, Alfenas-Zerbini P, Davison AJ, Dempsey DM, Dutilh BE, García ML, Harrach B, Harrison RL, Hendrickson RC, Junglen S, Knowles NJ, Krupovic M, Kuhn JH, Lambert AJ, Łobocka M, Nibert ML, Oksanen HM, Orton RJ, Robertson DL, Rubino L, Sabanadzovic S, Simmonds P, Smith DB, Suzuki N, Van Dooerslaer K, Vandamme AM, Varsani A, Zerbini FM (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). Arch Virol 166:2633-2648. PMID: 34231026. doi: 10.1007/s00705-021-05156-1.

52. Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM, Alfenas-Zerbini P, Davison AJ, Dempsey DM, Dutilh BE, García ML, Harrach B, Harrison RL, Hendrickson RC, Junglen S, Knowles NJ, Krupovic M, Kuhn JH, Lambert AJ, Łobocka M, Nibert ML, Oksanen HM, Orton RJ, Robertson DL, Rubino L, Sabanadzovic S, Simmonds P, Smith DB, Suzuki N, Van Dooerslaer K, Vandamme AM, Varsani A, Zerbini FM (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). Arch Virol: In preparation.

53. 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. PMID: 28599143. doi: 10.1016/j.mib.2017.05.003.

54. Westbye AB, O'Neill Z, Schellenberg-Beaver T, Beatty JT (2017) The *Rhodobacter capsulatus* gene transfer agent is induced by nutrient depletion and the RNAP omega subunit. Microbiology (Reading) 163:1355-1363. PMID: 28869409. doi: 10.1099/mic.0.000519.

55. Wolf YI, Kazlauskas D, Iranzo J, Lucía-Sanz A, Kuhn JH, Krupovic M, Dolja VV, Koonin EV (2018) Origins and evolution of the global RNA virome. mBio 9:e02329-18. PMID: 30482837. PMCID: PMC6282212. doi: 10.1128/mBio.02329-18.

56. Yen HC, Hu NT, Marrs BL (1979) Characterization of the gene transfer agent made by an overproducer mutant of *Rhodopseudomonas capsulata*. J Mol Biol 131:157-68. PMID: 490646. doi: 10.1016/0022-2836(79)90071-8.

57. Zhan Y, Huang S, Voget S, Simon M, Chen F (2016) A novel roseobacter phage possesses features of podoviruses, siphoviruses, prophages and gene transfer agents. Sci Rep 6:30372. PMID: 27460944. PMCID: PMC4961962. doi: 10.1038/srep30372.

58. Zhao Y, Wang K, Budinoff C, Buchan A, Lang A, Jiao N, Chen F (2009) Gene transfer agent (GTA) genes reveal diverse and dynamic *Roseobacter* and *Rhodobacter* populations in the Chesapeake Bay. Isme j 3:364-73. PMID: 19020557. doi: 10.1038/ismej.2008.115.