Formal Recognition and Classification of Gene Transfer Agents as Viriforms

Roman Kogay[1], Sonja Koppenhöfer[2], J. Thomas Beatty[3], Jens H. Kuhn[4], Andrew S. Lang[2], and Olga Zhaxybayeva[1,#]

[1]Department of Biological Sciences, Dartmouth College, Hanover, NH, USA; [2]Department of Biology, Memorial University of Newfoundland, St. John's, NL, Canada; [3]Department of Microbiology & Immunology, University of British Columbia, Vancouver, BC, Canada; [4]Integrated Research Facility at Fort Detrick, Department of Clinical Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Frederick, MD, USA

[#]Correspondence to Olga Zhaxybayeva, olga.zhaxybayeva@dartmouth.edu

Keywords: gene transfer agent; GTA; ICTV; International Committee on Taxonomy of Viruses; viriform; virus classification; virus nomenclature; virus taxonomy

16 **ABSTRACT**

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

28 **INTRODUCTION**

29 In 2021, the International Committee on Taxonomy of Viruses (ICTV) ratified a taxonomic
30 proposal to formally accept a new operational definition of the term "virus" (Walker et al. 2021;
31 Koonin et al. 2021; Kuhn et al. 2020). Consequently, the most current version of the
32 International Code of Virus Classification and Nomenclature (ICVCN) states that viruses are
33
34 " … a type of MGEs [mobile genetic elements] that encode at least one protein that is a
35 major component of the virion encasing the nucleic acid of the respective MGE and
36 therefore the gene encoding the major virion protein itself; or MGEs that are clearly
37 demonstrable to be members of a line of evolutionary descent of such major virion
38 protein-encoding entities" (ICVCN Rule 3.3) (International Committee on Taxonomy of
39 Viruses 2022; Kuhn et al. 2020).
40
41 This definition also formalized the postulate that some MGEs, long understood by the
42 general virology community to be distinct from viruses, are indeed distinct. At the time, the
43 ICTV had already classified viroids and satellite nucleic acids in taxa separate from viral taxa (in
44 families/genera with names that end with suffixes -*viroidae*/-*viroid* and -*satellitidae*/-*satellite*,
45 respectively, as opposed to -*viridae*/-*virus*) (International Committee on Taxonomy of Viruses
46 2022), and these elements were logically placed into the perivirosphere rather than the
47 orthovirosphere (Koonin et al. 2021; Kuhn et al. 2020).
48 The adoption of the new virus definition brought into question the taxonomic standing of
49 one family: *Polydnaviridae*. Indeed, entities classified into this likely polyphyletic family
50 fundamentally deviate from MGEs fulfilling the virus definition: "polydna" particles encapsidate

51 multiple segments of circular double-stranded DNAs that, however, do not encode the entire
52 "polydna" genomes. Instead, the complete genomes are permanently endogenized into the
53 "polydna" host (i.e., parasitoid wasp) genomes and inherited vertically. The resultant non-mobile
54 nonviral entities are used by the wasps to deliver immunomodulatory genes into insects that
55 serve as prey for the wasps (Drezen et al. 2017; Herniou et al. 2013). "Polydna" entities are
56 likely evolutionarily derived from various groups of insect viruses, including nudivirids
57 (Darboux et al. 2019; Drezen et al. 2017; Gauthier et al. 2018; Petersen et al. 2022; Strand and
58 Burke 2020; Thézé et al. 2011), but, because they have lost the ability to replicate and instead
59 have been fully exapted by their wasp hosts, they have left the virosphere altogether (Koonin and
60 Krupovic 2018; Koonin et al. 2021). Consequently, in 2021, the ICTV recognized "polydna"
61 entities as a new category distinct from viruses called "viriforms" (Koonin et al. 2021; Kuhn et
62 al. 2020; Walker et al. 2021), and reclassified *Polydnaviridae* as, still likely polyphyletic,
63 *Polydnaviriformidae* (Kuhn et al. 2021; Walker et al. 2022). In the ICVCN, viriforms are defined
64 operationally as
65
66 " … a type of virus-derived MGEs that have been exapted by their organismal (cellular)
67 hosts to fulfill functions important for the host life cycle; or MGEs that are derived from
68 such entities in the course of evolution" (ICVCN Rule 3.3) (International Committee on
69 Taxonomy of Viruses 2022; Kuhn et al. 2020).
70
71 Importantly, the following comment was added to the Rule 3.3:
72

73 "Gene transfer agents (GTAs) and the MGEs previously classified in the family
74 *Polydnaviridae* are considered to be viriforms in classification and nomenclature"
75 (International Committee on Taxonomy of Viruses 2022; Kuhn et al. 2020).
76
77 GTAs have never been classified. They are analogous but unrelated to
78 polydnaviriformids in the following ways (reviewed in Lang et al. 2012; Lang et al. 2017). First,
79 GTAs are encoded by genes endogenized in bacterial and archaeal chromosomes in several
80 blocks rather than a single genomic segment. Second, GTAs do not preferentially encapsidate
81 genes encoding the structural proteins of the virion-like GTA particles but instead encapsidate
82 mostly "random" pieces of host DNA. Last, the proteins used to produce GTA particles have
83 evolved from viral ancestors. Thus, GTAs have lost the ability to replicate and have become
84 fully exapted by their cellular hosts. Under specific conditions, GTAs are produced, the cell
85 lyses, and the released particles mediate horizontal gene transfer (HGT) by delivering their
86 packaged DNA to other cells, typically those of the same species.
87 The first GTA discovered, of the alphaproteobacterium *Rhodobacter capsulatus*, was
88 described in 1974 by Barry Marrs (Marrs 1974). Since that time, distinct functional GTAs have
89 been described in other alphaproteobacteria, a sulfate-reducing deltaproteobacterium, a
90 spirochete domestic pig pathogen, and a methanogenic archaeon (Bertani 1999; Guy et al. 2013;
91 Humphrey et al. 1997; Rapp and Wall 1987). Additionally, clusters of genes homologous to
92 those encoding the *R. capsulatus* GTA are found in many alphaproteobacterial genomes,
93 suggesting a wider prevalence of GTA production than presently appreciated (Kogay et al. 2019;
94 Lang et al. 2002; Lang and Beatty 2007; Shakya et al. 2017). Indeed, some of these bacteria
95 produce functional GTAs (Biers et al. 2008; Nagao et al. 2015; Tomasch et al. 2018).

96 The recent ICTV recognition of viriforms and the formal establishment of
97 *Polydnaviriformidae* provides an opportunity to initiate a systematic classification of GTAs.
98 Here we outline initial steps to establish such a formal taxonomic scheme for GTA viriforms,
99 focusing specifically on GTAs experimentally documented as being produced by cells and
100 performing gene transfer—and for which the genetic basis of particle production has been
101 established. Simultaneously, we also officially propose this taxonomic scheme to the ICTV for
102 the 2022–2023 proposal cycle.
103
104 **NOMENCLATURE OF GENE TRANSFER AGENTS AND ASSOCIATED TAXA**
105 Per ICTV rules, virus names are written in lower case (except if a name component is a proper
106 noun), without italics in any part of the name (even if a host species name is part of the name),
107 and ending in the term "virus", which in virus name abbreviations is "V". Examples are measles
108 virus (MeV) and Ebola virus (EBOV). The nomenclature of already classified viriforms
109 (polydnaviriformids) follows these rules, with "virus" being replaced by "viriform" and the
110 abbreviation "V" being replaced with "Vf" (e.g., "Glyptapanteles liparidis bracoviriform" is
111 abbreviated "GlBVf"). We suggest applying these general rules to GTAs, but with "viriform"
112 being replaced by "gene transfer agent" due to the long-established use of this phrase and "Vf"
113 being replaced with "GTA". Therefore, the gene transfer agent produced by *Rhodobacter*
114 *capsulatus* would be called "Rhodobacter capsulatus gene transfer agent" and abbreviated as
115 "RcGTA", consistent with the established use of this abbreviation in the literature.
116
117 Rules for viriform taxon naming have been established by the ICVCN. Specifically,
118

119 “[t]he formal endings for taxon names of viriforms are the suffixes "-*viriformia*" for
120 realms, "-*viriforma*" for subrealms, "-*viriformae*" for kingdoms, "-*viriformites*" for
121 subkingdoms, "-*viriformicota*" for phyla, "-*viriformicotina*" for subphyla,
122 "-*viriformicetes*" for classes, "-*viriformicetidae*" for subclasses, "-*viriformales*" for
123 orders, "-*viriformineae*" for suborders, "-*viriformidae*" for families, "-*viriforminae*" for
124 subfamilies, and "-*viriform*" for genera and subgenera” (ICVCN Rule 3.26) (International
125 Committee on Taxonomy of Viruses 2022; Kuhn et al. 2020)
126
127 and
128
129 “[a] species name shall consist of only two distinct word components separated by a
130 space. The first word component shall begin with a capital letter and be identical in
131 spelling to the name of the genus to which the species belongs. The second word
132 component shall not contain any suffixes specific for taxa of higher ranks. The entire
133 species name (both word components) shall be italicized” (International Committee on
134 Taxonomy of Viruses 2022).
135
136 We suggest adding the infix -*gta*- prior to the taxon-specific suffixes for immediate recognition
137 of GTA-specific taxa (e.g., -*gtaviriform*).
138
139 **GENE TRANSFER AGENTS CAN BE ASSIGNED TO AT LEAST THREE MAJOR**
140 **CLADES**

141 Based on functionally and genetically characterized GTAs, at least three major GTA clades can
142 be delineated.
143
144 **Alphaproteobacterial type I GTAs**
145 The best characterized GTA of this clade is RcGTA, the gene transfer agent produced by *R.*
146 *capsulatus* (*Pseudomonadota*: *Alphaproteobacteria*: *Rhodobacterales*: *Rhodobacteraceae*) (Yen
147 et al. 1979), which we designate here as the founding member of one major GTA clade, the
148 alphaproteobacterial type I GTAs. Homologous GTAs are produced by other bacteria from the
149 order *Rhodobacterales*: *Dinoroseobacter shibae* (Dinoroseobacter shibae gene transfer agent
150 [DsGTA]) (Tomasch et al. 2018), *Ruegeria pomeroyi* (Ruegeria pomeroyi gene transfer agent
151 [RpGTA]) (Biers et al. 2008), and *Rhodovulum sulfidophilum* (Rhodovulum sulfidophilum gene
152 transfer agent [RsGTA]) (Nagao et al. 2015). Additionally, genes encoding RcGTA-like GTAs
153 are conserved in most genomes in the order *Rhodobacterales* and in many genomes of the
154 alphaproteobacterial orders *Caulobacterales*, *Sphingomonadales*, *Parvibaculales,* and
155 *Hyphomicrobiales* (formerly *Rhizobiales*) (Kogay et al. 2019; Lang et al. 2002; Lang and Beatty
156 2007; Shakya et al. 2017).
157 RcGTA and RcGTA-like GTA genes are similar in sequence to those of viruses classified
158 in the uroviricot class *Caudoviricetes* (*Duplodnaviria*: *Heunggongvirae*) (Shakya et al. 2017).
159 These GTAs are transmitted vertically from a bacterial parent to progeny during cell division
160 (Lang and Beatty 2007; Shakya et al. 2017), similar to propagation of temperate viruses
161 ("prophages"). However, in contrast to temperate virus genomes, the set of genes required for
162 production of the GTA particle (the GTA "genome") is not necessarily localized in one region of
163 the host genome. In the case of RcGTA, known structural and regulatory genes are scattered

164 across five loci in the *R. capsulatus* genome (Hynes et al. 2016), cumulatively spanning
165 approximately 20 kilobases (kb). Moreover, cellular regulatory genes are involved in controlling
166 GTA particle production (Westbye et al. 2017), adding another factor that makes the GTA
167 genome difficult to differentiate from its host's genome.

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

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

186 In a comprehensive evolutionary analysis of homologs of the large subunit of the DNA
187 packaging terminase enzyme (TerL, encoded by the *g2* gene in the RcGTA genome), RcGTA
188 and RcGTA-like GTAs form a clade closely related to, but distinct from, duplodnavirians
189 (Esterman et al. 2021). To illustrate the relationships of alphaproteobacterial type I GTAs to each
190 other and to their closest viral homologs, we reconstructed evolutionary histories of their TerL
191 proteins and the HK97-like major capsid proteins (HK97-MCP, encoded by the *g5* gene in the
192 GTA genome, is the hallmark protein that defines the virus realm *Duplodnaviria* (Koonin et al.
193 2020)). Consistent with an earlier analysis (Esterman et al. 2021), RcGTA and RcGTA-like
194 GTAs formed a clade closely related to, but distinct from, caudoviricetes (**Figure 1**), with a few
195 exceptions that are likely artefacts of phylogenetic reconstruction.
196 Specifically, in the TerL phylogeny (**Figure 1A**), all viral homologs except one
197 (Caulobacter virus Sansa) are separated from GTA proteins (with a solid bootstrap support of
198 81%). Caulobacter virus Sansa groups with one GTA sequence from a bacterium of the order
199 *Sphingomonadales* (with a low bootstrap support of 50%), whereas all other GTAs of bacteria
200 classified in *Sphingomonadales* group together (with a strong bootstrap support of 96%). We
201 hypothesize that the phylogenetic placement of the Caulobacter virus Sansa TerL is due to the
202 long-branch attraction artefact (Felsenstein 1978). We searched for a maximum-likelihood tree in
203 which caudoviricete- and GTA-derived TerLs were required to group separately from each other
204 and compared that tree to the tree depicted in **Figure 1A**. We found that the likelihoods of the
205 two trees are not significantly different (approximately unbiased [AU] test; p-value = 0.555),
206 confirming that the placement of the Caulobacter virus Sansa sequence within the GTA
207 sequences is unreliable.

208 In the HK97-MCP phylogeny (**Figure 1B**), GTAs and most caudoviricetes are separated
209 by a branch with 63% bootstrap support. Several caudoviricetes that group within GTAs are
210 located on long branches, are situated outside of well-supported groups of GTAs from several
211 alphaproteobacterial orders and have very low bootstrap support for their placements. It is
212 therefore likely that the positions of these viral homologs are unreliable. To test this hypothesis,
213 we identified a maximum-likelihood phylogeny among trees in which GTAs and caudoviricetes
214 were required to be separated by a branch. The likelihoods of this tree and the phylogeny shown
215 in **Figure 1B** are not significantly different (AU test; p-value = 0.534). Therefore, these viruses
216 are likely positioned in different places in trees reconstructed from different bootstrap replicates,
217 which would lead to their artificial (and poorly supported) basal positions with the GTA
218 homologs on the tree shown in **Figure 1B**.
219 In the **Figure 1** trees, GTA branches have shorter lengths than their caudoviricete
220 counterparts, conforming with the reported slower evolutionary rate of GTAs compared to
221 viruses (Shakya et al. 2017). Additionally, on both phylogenetic trees, GTAs from
222 alphaproteobacteria of different orders form separate groups with very high support, supporting
223 vertical inheritance of most GTA genes (Lang et al. 2002; Lang and Beatty 2007; Shakya et al.
224 2017).
225 Together, these results justify the classification of RcGTA and three RcGTA-like GTAs
226 in a common viriform taxon: family *Rhodogtaviriformidae* (from *Rhodobacterales*, infix -*gta*-,
227 and family-specific suffix -*viriformidae*). For now, we argue that at least four rhodogtaviriformid
228 genera, each for GTAs of bacteria classified in distinct genera included in *Rhodobacterales*,
229 ought to be established:

230 • *Dinogtaviriform* (named after DsGTA host genus *Dinoroseobacter*, infix -*gta*-, and genus-
231 specific suffix -*viriform*) to include one new species, *Dinogtaviriform tomaschi* (species
232 epithet to honor GTA researcher Jürgen Tomasch, who was instrumental in the discovery of
233 DsGTA) for DsGTA;
234 • *Rhodobactegtaviriform* (named after RcGTA host genus Rhodobacter, infix -*gta*-, and genus-
235 specific suffix -*viriform*) to include one new species, *Rhodobactegtaviriform marrsi* (species
236 epithet to honor GTA researcher Barry Marrs, who first discovered GTAs and coined the
237 term "gene transfer agent") for RcGTA;
238 • *Rhodovulugtaviriform* (named after RsGTA host genus *Rhodovulum*, infix -*gta*-, and genus-
239 specific suffix -*viriform*) to include one new species, *Rhodovulugtaviriform kikuchii* (species
240 epithet to honor GTA researcher Yo Kikuchi, who was instrumental in the discovery of
241 RsGTA) for RsGTA; and
242 • *Ruegerigtaviriform* (named after RpGTA host genus *Ruegeria*, infix -*gta*-, and genus-specific
243 suffix -*viriform*) to include one new species, *Ruegerigtaviriform cheni* (species epithet to
244 honor GTA researcher Feng Chen, who was instrumental in the discovery of RpGTA) for
245 RpGTA.
246

247 **Alphaproteobacterial type II GTAs**

248 There was a lag between discovery of these elements and their recognitions as bona fide GTAs.
249 Phage-like particles, originally referred to as bacteriophage-like particles (BLPs), that contained
250 heterogenous DNA from *Bartonella* host genomes were first characterized in *B. henselae*
251 (Anderson et al. 1994), and noted to be similar in structure to particles produced by *B.*
252 *bacilliformis* (Umemori et al. 1992). These *B. bacilliformi* particles were subsequently shown to

253 also contain heterogeneous genomic DNA fragments, but attempts to demonstrate their gene
254 transfer ability were not successful (Barbian and Minnick 2000). Functionality of the particles
255 produced by *Bartonella* for gene transfer (Bartonella gene transfer agent [BaGTA]) was
256 eventually demonstrated by work on *B. henselae* (*Pseudomonadota*: *Alphaproteobacteria*:
257 *Hyphomicrobiales*: *Bartonellaceae*) (Guy et al. 2013). BaGTA genes were initially proposed to
258 be located within a single cluster of 12 genes spanning approximately 14 kb (Guy et al. 2013).
259 However, a subsequent screen for genes essential for BaGTA functionality identified a total of
260 29 genes located within a larger (approximately 79-kb-long) region (Québatte et al. 2017).
261 Homologs of BaGTA genes (BaGTA-like GTAs) were found in the genomes of multiple species
262 of *Bartonella* (Berglund et al. 2009; Guy et al. 2013; Tamarit et al. 2018). BaGTA genes are
263 located near an active virus-derived origin of replication and next to genes encoding secretion
264 systems (Guy et al. 2013). As a result, the region of the genome containing BaGTA and these
265 secretion-system genes are amplified and packaged more often than other genomic regions (Guy
266 et al. 2013; Québatte et al. 2017). These findings led to the hypothesis that BaGTA and BaGTA-
267 like GTAs have been maintained due to their mediation of HGT of secretion-system and toxin
268 genes, thereby enabling *Bartonella* bacteria to adapt to diverse hosts (Guy et al. 2013). However,
269 actual GTA-mediated DNA transfer among bacterial cells has only been demonstrated for *B.*
270 *henselae* (Guy et al. 2013). There, BaGTA production is restricted to a distinct subpopulation of
271 fast-growing cells, which comprise about 6% of the total population (Québatte et al. 2017), and
272 the uptake of BaGTA-packaged DNA was proposed to be limited to cells undergoing division
273 (Québatte et al. 2017).
274 There are some discrepancies in the literature regarding the structure of BaGTA particles,
275 suggesting some bacteria might release additional phage-like particles. The *B. henselae* particles

276 were originally reported as particles without tails or with short non-contractile tails with a head
277 diameter of 40 nm (Anderson et al. 1994). The head diameter of the *B. bacilliformis* particles
278 was originally measured at 40 nm (Umemori et al. 1992) and subsequently 80 nm (Barbian and
279 Minnick 2000). Those of *B. grahamii* were reported as possessing long non-contractile tailss and
280 icosahedral heads of 50–70 nm or 80 nm and tails of 100 nm (Berglund et al. 2009). Although
281 BaGTA particles are potentially able to package the entire main structural gene cluster of 12
282 genes, they cannot package all 29 genes required for BaGTA production due to a capacity of 14
283 kb (Berglund et al. 2009; Anderson et al. 1994; Guy et al. 2013; Lang et al. 2017).
284 In the TerL phylogeny, BaGTA-like homologs are separated from almost all
285 caudoviricetes by longer branches (with 100% bootstrap support; **Figure 2A**). Two caudoviricete
286 homologs (Sulfitobacter phage pCB2047-C and Sulfitobacter phage NYA-2014a) group together
287 and are nested within the BaGTA-like group (with 84% bootstrap support). We hypothesize that
288 the *terL* gene was horizontally transferred from GTAs to these caudoviricetes, with similar HGT
289 events documented between RcGTA-like GTAs and caudoviricetes infecting bacteria of the
290 *Rhodobacterales* (Zhan et al. 2016). In the HK97-MCP phylogeny, BaGTA homologs are
291 located on shorter branches than their caudoviricete counterparts and are separated from
292 caudoviricete homologs with 100% bootstrap support (**Figure 2B**).
293 Together, these results justify the classification of BaGTA and BaGTA-like GTAs in a
294 common viriform taxon, family *Bartogtaviriformidae* (from *Bartonella*, infix -*gta*-, and family-
295 specific suffix -*viriformidae*). For now, we argue that at least one bartogtaviriformid genus ought
296 to be established: *Bartonegtaviriform* (named after BaGTA host genus *Bartonella*, infix -*gta*-,
297 and genus-specific suffix -*viriform*) including one new species, *Bartonegtaviriform andersoni*

298 (species epithet to honor GTA researcher Burt Anderson, who first discovered BaGTA particles
299 (Anderson et al. 1994)) for BaGTA.
300
301 **GTAs of spirochaetes**
302 A GTA originally called virus of *Serpulina hyodysenteriae* 1 (VSH-1) was identified in
303 *Brachyspira* (formerly *Serpulina*) *hyodysenteriae* (*Spirochaetota*: *Spirochaetia*: *Brachyspirales*:
304 *Brachyspiraceae*) (Humphrey et al. 1997). In accordance with the nomenclature rules established
305 here, we suggest renaming this GTA to Brachyspira hyodysenteriae gene transfer agent
306 (BhGTA). The structural gene cluster responsible for production of BhGTA particles—i.e., the
307 BhGTA "genome"—is 16.3 kb in length (Matson et al. 2005).
308 BhGTA particles have a head diameter of 45 nm and a flexible non-contractile tail of 65
309 nm (Humphrey et al. 1997). Like other GTAs, BhGTAs is unable to package and transfer its
310 entire genome, given the limiting capacity of 7.5 kb (Humphrey et al. 1997; Matson et al. 2005).
311 Restriction enzyme digests of the packaged DNA and the range of marker genes that can be
312 transferred by BhGTA particles suggest that they package any region of the *B. hyodysenteriae*
313 genome (Humphrey et al. 1997) without an obvious bias for the genomic region that encodes
314 BhGTA. The proportion of *B. hyodysenteriae* cells in a population that naturally produce and
315 release BhGTA particles has not been quantified. However, BhGTA particle production can be
316 induced by DNA-damaging agents, such as mitomycin C and antibiotics (Stanton et al. 2008),
317 and this induction results in large-scale lysis of cells. BhGTA particles are capable of
318 transferring antimicrobial resistance genes within the bacterial population (Stanton et al. 2008),
319 pointing at possible selective advantages of maintaining the capability of BhGTA particle
320 production.

321 Homologs of genes in the BhGTA genome were found in the genomes of other members
322 of the genus *Brachyspira*, but there is no gene synteny in their organization (Motro et al. 2009).
323 Unlike in rhodogtaviriformids and bartogtaviriformids, an endolysin-encoding gene is the only
324 gene in the BhGTA genome that has a significant sequence similarity to caudoviricete genes in
325 the National Center for Biotechnology Information (NCBI) Reference Sequence (RefSeq)
326 database (accessed in May 2022). Some genes encoding the BhGTA particle proteins were
327 experimentally validated, and the particles structurally resemble those of caudoviricetes (Matson
328 et al. 2005). Therefore, the absence of their homologs in the viral RefSeq database is likely due
329 to the limited sampling of the virosphere. Yet, the *B. hyodysenteriae* genome encodes a single
330 copy of an annotated *terL* gene, which is located outside of the currently delineated BhGTA
331 genome. Homologs of this *terL* gene are also present in a single copy in genomes of other
332 *Brachyspira* bacteria that encode BhGTA-like MCPs. These homologs are highly conserved,
333 with pairwise amino-acid identities of 81–100%. In a phylogenetic tree, the *Brachyspira* TerLs
334 are separated from all caudoviricete TerLs by a longer branch (with 100% bootstrap support;
335 **Figure 3**). Although the role of this TerL homolog in the BhGTA lifecycle has not been
336 experimentally validated, the presence of the encoding gene as the only identifiable *terL* in the
337 *Brachyspira* genomes, its high degree of conservation within the *Brachyspira* genus, and its
338 divergence from the related caudoviricete sequences support its potential involvement in the
339 packaging of DNA into the BhGTA particles.
340 Together, these results justify the classification of BhGTA and BhGTA-like GTAs in a
341 common viriform taxon, family *Brachygtaviriformidae* (form *Brachy*spira, infix -*gta*-, and
342 family-specific suffix -*viriformidae*). For now, we argue that at least one brachygtaviriformid
343 genus ought to be established: *Brachyspigtaviriform* (named after BhGTA host genus

344 *Brachyspira*, infix -*gta*-, and genus-specific suffix -*viriform*) to include one new species,
345 *Brachyspigtaviriform stantoni* (species epithet to honor GTA researcher Thaddeus Stanton, who
346 first discovered BhGTA particles (Humphrey et al. 1997)) for BhGTA.
347

348 **Independent origins of the three GTAs**

349 Genes from the genomes of these three GTAs are either not homologous or too divergent to have
350 significant sequence similarity in BLASTP searches of the encoded proteins. For example,
351 pairwise amino-acid identity of TerLs, which is one of the most conserved GTA and
352 caudoviricete proteins, is 14–20% among RcGTA, BaGTA, and BhGTA. Nevertheless, a
353 sophisticated sequence-similarity search established the homology among known TerL proteins
354 that include RcGTA, BaGTA, and putative BhGTA TerLs (Esterman et al. 2021). The
355 evolutionary history of RcGTA-like, BaGTA-like, putative BhGTA-like TerLs, and their closest
356 known caudoviricete homologs (**Figure 4**) demonstrates that GTA-like TerLs appear in three
357 distinct clades within viral TerLs. Based on this phylogenetic evidence, we propose that these
358 three GTA clades are a result of three independent exaptation events. Therefore, just like viruses
359 (which are classified in at least six unrelated realms), GTA viriforms are polyphyletic.
360

361 **DISCUSSION**

362 Based on the evolutionary differences between GTA and caudoviricete genes encoding well-
363 conserved proteins and on morphological differences of GTA particles, we propose three
364 families for these GTAs. The greatest number of functionally confirmed and putative GTAs are
365 in the alphaproteobacterial type I clade, which, for now, is proposed to be a family
366 *Rhodogtaviriformidae* that includes at least four genera. The members of this family are

367 currently restricted to a single cellular order (*Rhodobacterales*). The TerLs and MCPs of these
368 RcGTA-like GTAs and alphaproteobacterial type II GTAs (*Bartogtaviriformidae*) are clearly
369 distinguishable from each other and their caudoviricete homologs and evolve much slower
370 (**Figure 1** and **Figure 2**) (Esterman et al. 2021; Shakya et al. 2017). The spirochaete GTAs
371 (*Brachygtaviriformidae*) are more difficult to distinguish from caudoviricetes due to a lack of
372 available viral representatives in GenBank for experimentally validated BhGTA genes. However,
373 the putative BhGTA TerL and its *Brachyspira* bacteria homologs also evolve relatively more
374 slowly and form a well-supported cluster distinct from caudoviricetes lineages (**Figure 3**). As in
375 the case with the experimentally validated RcGTA, the "genome" of BhGTA is also likely
376 dispersed across multiple loci.

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

388 Based on the evolutionary history of TerL proteins (**Figure 4**), it is likely that the
389 proposed three GTA families had distinct virus progenitors. Since the exaptation events,

390 however, the three families have evolved as part of the host genomes (Esterman et al. 2021; Lang
391 et al. 2002; Lang and Beatty 2007; Shakya et al. 2017), in the case of the rhodogtaviriformids for
392 hundreds of millions of years (Shakya et al. 2017). As a result, GTAs effectively became a
393 component of cellular genomes, integrated into cellular regulatory circuits that also control
394 processes, such as, motility, quorum sensing, extracellular polysaccharide synthesis, and biofilm
395 formation (Lang et al. 2017; Shimizu et al. 2022). There is also mounting evidence that GTA
396 genes experience selective pressures to be maintained in their host genomes (Kogay et al. 2020;
397 Lang et al. 2012). Although the fitness benefits associated with GTA production remain to be
398 elucidated, the time is now ripe to have the known GTAs officially recognized and classified as
399 specific viriforms. We recognize this step as the initiation of a taxonomic framework that
400 undoubtedly will rapidly expand and change in the future.
401
402 **METHODS**
403 To identify alphaproteobacterial type I GTAs, we searched for GTA-like sequences in 1,248
404 complete alphaproteobacterial genomes extracted from the NCBI RefSeq database (accessed in
405 October 2020) using GTA-Hunter (Kogay et al. 2019). We identified 503 genomes that
406 contained at least six RcGTA homologs in the same genetic neighborhood and had both *g2*
407 (encoding TerL) and *g5* (encoding HK97-MCP) genes. To remove redundancy, we clustered
408 genomes into the operational taxonomic units (OTUs) using an average nucleotide identity
409 threshold of 95%. From all genomes within an OTU, we selected one genome with the largest
410 number of GTA genes. This strategy resulted in 290 representative GTAs selected for further
411 analysis. We identified the closest viral homologs of the TerL and HK97-MCP proteins from
412 these GTAs by conducting a BLASTP search (Altschul et al. 1997) of the RefSeq database

413 (accessed in March 2021) (O'Leary et al. 2016), using TerL and HK97-MCP proteins from
414 representative GTAs as queries, an e-value cutoff of 0.001, and query coverage of at least 50%.
415 Retrieved viral homologs with identical amino acid sequences were removed from further
416 analyses. For both proteins, we aligned amino acid sequences of GTA and virus homologs using
417 MAFFT v7.455 with *-linsi* option (Katoh and Standley 2013). We reconstructed phylogenetic
418 trees using IQ-TREE v2 (Minh et al. 2020), identifying the best substitution matrix using the
419 built-in ModelFinder (Kalyaanamoorthy et al. 2017). Branch support values were assessed using
420 1,000 ultrafast bootstrap replicates and a hill-climbing nearest-neighbor interchange search for
421 optimal trees (Hoang et al. 2018). Additionally, for both protein phylogenies, we reconstructed a
422 phylogenetic tree in IQ-TREE v2 (Minh et al. 2020) using a tree search that was constrained by
423 requiring all GTAs and all viruses to be separated by a branch. We compared the resultant trees
424 in unconstrained and constrained searches using the AU test (Shimodaira 2002), as implemented
425 in the IQ-TREE v2 program.

426 To identify alphaproteobacterial type II GTAs, we used the BaGTA TerL and HK97-
427 MCP sequences (accession numbers WP_034448260.1 and WP_011181178.1, respectively) as
428 queries in a BLASTP search against the 57 complete *Bartonella* genomes extracted from the
429 RefSeq database (accessed in May 2022) to identify their cellular homologs. We restricted our
430 search only to matches for which BaGTA TerL and HK97-MCP homologs are in the same
431 genomic neighborhood (defined as being within 5 kb of each other). In genomes with multiple
432 matches to the query protein, we retained only the homolog with the highest BLASTP bit score.
433 We clustered redundant genomes using a 95% average nucleotide identity (ANI) threshold and
434 randomly selected one representative from each cluster for phylogenetic analysis. We identified
435 caudoviricete homologs by conducting a BLASTP search (e-value cutoff of 0.001, and query

436 coverage of at least 50%) against viral RefSeq database (accessed in May 2022), and performed
437 phylogenetic reconstructions, as described above for alphaproteobacterial type I GTAs.
438 To identify GTAs of spirochaetes, we used BhGTA's MCP sequence (GenBank
439 accession number WP_012671344.1) as a query in a BLASTP search (with an e-value cutoff of
440 0.001 and query coverage of at least 50%) against the 13 complete *Brachyspira* genomes
441 extracted from the RefSeq database (accessed in May 2022). Similarly, we used TerL of *B.*
442 *hyodysenteriae* (GenBank accession number WP_012671469.1) as a query in a BLASTP search
443 (with an e-value cutoff of 0.001 and query coverage of at least 50%) against the same set of 13
444 genomes. We identified caudoviricete homologs by doing BLASTP search (e-value cutoff of
445 0.001 and query coverage of at least 50%) against viral RefSeq database (accessed in May 2022).
446 We clustered genomes using a 95% ANI threshold and randomly selected one representative
447 from each cluster for phylogenetic analysis. We performed phylogenetic reconstructions as
448 described above for the alphaproteobacterial type I GTAs.
449 To reconstruct the phylogeny that includes all three clades of GTAs, we combined all
450 TerL homologs extracted in the above described procedures into one dataset. We aligned the
451 TerL sequences using the MAFFT v7.455 with -dash option (Rozewicki et al. 2019), and
452 trimmed the obtained alignment using ClipKIT with -gappy option (Steenwyk et al. 2020). We
453 computed the phylogenetic tree using IQ-TREE v2. To reconstruct the phylogeny that includes
454 all three clades of GTAs, we combined all TerL homologs extracted in the above-described
455 procedures into one dataset. We aligned the TerL sequences using MAFFT v7.455 with -dash
456 option (Rozewicki et al. 2019), and trimmed the obtained alignment using ClipKIT with -gappy
457 option (Steenwyk et al. 2020). We computed the phylogenetic tree using IQ-TREE v2 (Minh et

458 al. 2020) as described above. We rooted the tree using a larger TerL phylogeny presented in
459 (Esterman et al. 2021).
460 We visualized all phylogenetic trees in iTOL v6 (Letunic and Bork 2021).
461

462 **ACKNOWLEDGEMENTS**

463 We thank Anya Crane and Jiro Wada (NIH/NIAID/DCR/IRF-Frederick) for editing the
464 manuscript and figures, respectively.
465

466 **FUNDING**

467 This work was supported in part through Laulima Government Solutions, LLC, prime contract
468 with the U.S. National Institute of Allergy and Infectious Diseases (NIAID) under Contract No.
469 HHSN272201800013C; J.H.K. performed this work as an employee of Tunnell Government
470 Services (TGS), a subcontractor of Laulima Government Solutions, LLC, under Contract No.
471 HHSN272201800013C. This work was also supported by the Simons Foundation Investigator in
472 Mathematical Modeling of Living Systems award #327936 (O.Z.), the Canadian Natural
473 Sciences and Engineering Research Council (NSERC) award RGPIN 2018-03898 (J.T.B.),
474 NSERC award RGPIN-2017-04636 (A.S.L.). S.K. was partially supported by funding from the
475 Memorial University of Newfoundland School of Graduate Studies.
476

477 **CONFLICTS OF INTEREST STATEMENT**

478 The authors declare no conflicts of interest.
479

480 **SUPPLEMENTARY DATA**

481 Amino acid sequence alignments in FASTA format and phylogenetic trees in Newick format for

482 trees in **Figures 1–4** are available in **Supplementary_Data.zip** file.

483 **FIGURES**

484 **Figure 1. Phylogenies of (A) large terminase (TerL) subunits and (B) HK97 major capsid**
485 **protein (HK97-MCP) sequences of rhodogtaviriformids and their closest known**
486 **caudoviricete homologs.** Alphaproteobacterial type I gene transfer agent (GTA)
487 (rhodogtoviriformid) lineages are shown in orange. Caudoviricete lineages that are nested within
488 GTA lineages are shown in dashed black lines. Other caudoviricete lineages are shown in solid
489 black lines. Bootstrap support values are shown only for selected branches. Scale bars represent
490 substitutions per site. DsGTA, Dinoroseobacter shibae gene transfer agent; GTA, gene transfer
491 agent; RcGTA, Rhodobacter capsulatus gene transfer agent; RpGTA, Ruegeria pomeroyi gene
492 transfer agent; RsGTA, Rhodovolum sulfidophilum, gene transfer agent.

493

494 **Figure 2. Phylogenies of (A) large terminase (TerL) subunits and (B) HK97 major capsid**
495 **protein (HK97-MCP) sequences of bartogtaviriformids and their closest known**
496 **caudoviricete homologs.** Alphaproteobacterial type II gene transfer agent (GTA)
497 (bartogtaviriformid) lineages are shown in blue. Caudoviricete lineages are shown in black. Two
498 nearly identical caudoviricete lineages that are nested within GTA lineages are shown in dashed
499 black lines. A bootstrap support value is shown only for the branch separating GTA and
500 caudoviricete sequences. Scale bars indicate substitutions per site. BaGTA, Bartonella gene
501 transfer agent; GTA, gene transfer agent.

502

503 **Figure 3. Phylogeny of the putative large terminase (TerL) subunits of**
504 **brachygtaviriformids and their closest known caudoviricete homologs.** *Brachyspira* gene
505 transfer agent (GTA) lineages are shown in purple. Caudoviricete lineages are shown in black. A

506 bootstrap support value is shown only for the branch separating GTA and caudoviricete
507 sequences. Scale bar indicates substitutions per site. BhGTA, Brachyspira hyodysenteriae gene
508 transfer agent; GTA, gene transfer agent.
509
510 **Figure 4. Phylogeny of the large terminase (TerL) subunits of three major clades of GTAs**
511 **and their closest known caudoviricete homologs.** This tree includes all TerL homologs from
512 Figures 1A, 2A, and 3, using their color codings. Bootstrap support values are shown only for the
513 branches separating three GTA clades and their closest caudoviricete sequences. Scale bar
514 indicates substitutions per site.. GTA, gene transfer agent.

515 **REFERENCES**

516 Altschul, S. F., et al. (1997), 'Gapped BLAST and PSI-BLAST: a new generation of protein
517 database search programs', *Nucleic Acids Res,* 25 (17), 3389-402.

518 Anderson, B., et al. (1994), 'Bacteriophage-like particle of *Rochalimaea henselae*', *Mol*
519 *Microbiol,* 13 (1), 67-73.

520 Barbian, K. D. and Minnick, M. F. (2000), 'A bacteriophage-like particle from *Bartonella*
521 *bacilliformis*', *Microbiology (Reading),* 146 (Pt 3), 599-609.

522 Bárdy, P., et al. (2020), 'Structure and mechanism of DNA delivery of a gene transfer agent', *Nat*
523 *Commun,* 11 (1), 3034.

524 Berglund, E. C., et al. (2009), 'Run-off replication of host-adaptability genes is associated with
525 gene transfer agents in the genome of mouse-infecting *Bartonella grahamii*', *PLoS Genet,*
526 5 (7), e1000546.

527 Bertani, G. (1999), 'Transduction-like gene transfer in the methanogen *Methanococcus voltae*', *J*
528 *Bacteriol,* 181 (10), 2992-3002.

529 Biers, E. J., et al. (2008), 'Occurrence and expression of gene transfer agent genes in marine
530 bacterioplankton', *Appl Environ Microbiol,* 74 (10), 2933-9.

531 Brimacombe, C. A., et al. (2013), 'Quorum-sensing regulation of a capsular polysaccharide
532 receptor for the *Rhodobacter capsulatus* gene transfer agent (RcGTA)', *Mol Microbiol,*
533 87 (4), 802-17.

534 Darboux, I., Cusson, M., and Volkoff, A. N. (2019), 'The dual life of ichnoviruses', *Curr Opin*
535 *Insect Sci,* 32, 47-53.

536 Drezen, J.-M., et al. (2017), 'Endogenous viruses of parasitic wasps: variations on a common
537 theme', *Curr Opin Virol,* 25, 41-48.

538 Esterman, E. S., et al. (2021), 'Evolution of DNA packaging in gene transfer agents', *Virus Evol,*
539 7 (1), veab015.
540 Felsenstein, Joseph (1978), 'Cases in which parsimony or compatibility methods will be
541 positively misleading', *Syst Biol,* 27 (4), 401-10.
542 Fogg, P. C. M., Westbye, A. B., and Beatty, J. T. (2012), 'One for all or all for one:
543 heterogeneous expression and host cell lysis are key to gene transfer agent activity in
544 *Rhodobacter capsulatus*', *PLoS One,* 7 (8), e43772.
545 Fu, Yunyun, et al. (2010), 'High diversity of *Rhodobacterales* in the subarctic North Atlantic
546 Ocean and gene transfer agent protein expression in isolated strains', *Aquat Microb Ecol,*
547 59, 283–93.
548 Gauthier, J., Drezen, J.-M., and Herniou, E. A. (2018), 'The recurrent domestication of viruses:
549 major evolutionary transitions in parasitic wasps', *Parasitology,* 145 (6), 713-23.
550 Guy, L., et al. (2013), 'A gene transfer agent and a dynamic repertoire of secretion systems hold
551 the keys to the explosive radiation of the emerging pathogen *Bartonella*', *PLoS Genet,* 9
552 (3), e1003393.
553 Herniou, E. A., et al. (2013), 'When parasitic wasps hijacked viruses: genomic and functional
554 evolution of polydnaviruses', *Philos Trans R Soc Lond B Biol Sci,* 368 (1626), 20130051.
555 Hoang, D. T., et al. (2018), 'UFBoot2: improving the Ultrafast Bootstrap approximation', *Mol*
556 *Biol Evol,* 35 (2), 518-22.
557 Humphrey, S. B., et al. (1997), 'Purification and characterization of VSH-1, a generalized
558 transducing bacteriophage of *Serpulina hyodysenteriae*', *J Bacteriol,* 179 (2), 323-9.

559 Hynes, A. P., et al. (2012), 'DNA packaging bias and differential expression of gene transfer
560 agent genes within a population during production and release of the *Rhodobacter*
561 *capsulatus* gene transfer agent, RcGTA', *Mol Microbiol,* 85 (2), 314-25.
562 Hynes, A. P., et al. (2016), 'Functional and evolutionary characterization of a gene transfer
563 agent's multilocus "genome"', *Mol Biol Evol,* 33 (10), 2530-43.
564 International Committee on Taxonomy of Viruses (2022), 'ICTV Code. The International Code
565 of Virus Classification and Nomenclature (ICVCN). March 2021.
566 https://talk.ictvonline.org/information/w/ictv-information/383/ictv-code'.
567 Kalyaanamoorthy, S., et al. (2017), 'ModelFinder: fast model selection for accurate phylogenetic
568 estimates', *Nat Methods,* 14 (6), 587-89.
569 Katoh, K. and Standley, D. M. (2013), 'MAFFT multiple sequence alignment software version 7:
570 improvements in performance and usability', *Mol Biol Evol,* 30 (4), 772-80.
571 Kogay, R., et al. (2020), 'Selection for reducing energy cost of protein production drives the GC
572 content and amino acid composition bias in gene transfer agents', *mBio,* 11 (4), e01206-
573 20.
574 Kogay, R., et al. (2019), 'Machine-learning classification suggests that many
575 alphaproteobacterial prophages may instead be gene transfer agents', *Genome Biol Evol,*
576 11 (10), 2941-53.
577 Koonin, E. V. and Krupovic, M. (2018), 'The depths of virus exaptation', *Curr Opin Virol,* 31, 1-
578 8.
579 Koonin, E. V., et al. (2021), 'Viruses defined by the position of the virosphere within the
580 replicator space', *Microbiol Mol Biol Rev,* 85 (4), e0019320.

581 Koonin, E. V., et al. (2020), 'Global organization and proposed megataxonomy of the virus
582 world', *Microbiol Mol Biol Rev,* 84 (2), e00061-19.
583 Kuhn, J. H., et al. (2020), 'Expand, amend, and emend the International Code of Virus
584 Classification and Nomenclature (ICVCN; "the Code") and the Statutes to clearly define
585 the remit of the ICTV. International Committee on Taxonomy of Viruses (ICTV)
586 TaxoProp 2020.005G.R.Code_and_Statute_Change.
587 https://talk.ictvonline.org/files/ictv_official_taxonomy_updates_since_the_8th_report/m/
588 general-2008/11061'.
589 Kuhn, J. H., et al. (2021), 'Rename the family *Polydnaviridae* (as *Polydnaviriformidae*), rename
590 the genus *Bracovirus* (as *Bracoviriform*) and rename all polydnaviriformid species to
591 comply with the newly ICTV-mandated binomial format. International Committee on
592 Taxonomy of Viruses (ICTV) TaxoProp
593 2021.006D.R.Polydnaviriformidae_1renfam_3rensp.
594 https://talk.ictvonline.org/files/ictv_official_taxonomy_updates_since_the_8th_report/m/
595 animal-dna-viruses-and-retroviruses/13248'.
596 Lang, A. S. and Beatty, J. T. (2007), 'Importance of widespread gene transfer agent genes in α-
597 proteobacteria', *Trends Microbiol,* 15 (2), 54-62.
598 Lang, A. S., Taylor, T. A., and Beatty, J. T. (2002), 'Evolutionary implications of phylogenetic
599 analyses of the gene transfer agent (GTA) of *Rhodobacter capsulatus*', *J Mol Evol,* 55
600 (5), 534-43.
601 Lang, A. S., Zhaxybayeva, O., and Beatty, J. T. (2012), 'Gene transfer agents: phage-like
602 elements of genetic exchange', *Nat Rev Microbiol,* 10 (7), 472-82.

603 Lang, A. S., Westbye, A. B., and Beatty, J. T. (2017), 'The distribution, evolution, and roles of
604 Ggne transfer agents in prokaryotic genetic exchange', *Annu Rev Virol,* 4 (1), 87-104.
605 Letunic, I. and Bork, P. (2021), 'Interactive Tree Of Life (iTOL) v5: an online tool for
606 phylogenetic tree display and annotation', *Nucleic Acids Res,* 49 (W1), W293-w96.
607 Marrs, B. (1974), 'Genetic recombination in *Rhodopseudomonas capsulata*', *Proc Natl Acad Sci*
608 *U S A,* 71 (3), 971-3.
609 Matson, E. G., et al. (2005), 'Identification of genes of VSH-1, a prophage-like gene transfer
610 agent of *Brachyspira hyodysenteriae*', *J Bacteriol,* 187 (17), 5885-92.
611 McDaniel, L. D., et al. (2010), 'High frequency of horizontal gene transfer in the oceans',
612 *Science,* 330 (6000), 50.
613 Minh, B. Q., et al. (2020), 'IQ-TREE 2: new models and efficient methods for phylogenetic
614 inference in the genomic era', *Mol Biol Evol,* 37 (5), 1530-34.
615 Motro, Y., et al. (2009), 'Identification of genes associated with prophage-like gene transfer
616 agents in the pathogenic intestinal spirochaetes *Brachyspira hyodysenteriae*, *Brachyspira*
617 *pilosicoli* and *Brachyspira intermedia*', *Vet Microbiol,* 134 (3-4), 340-5.
618 Nagao, N., et al. (2015), 'The gene transfer agent-like particle of the marine phototrophic
619 bacterium *Rhodovulum sulfidophilum*', *Biochem Biophys Rep,* 4, 369-74.
620 O'Leary, N. A., et al. (2016), 'Reference sequence (RefSeq) database at NCBI: current status,
621 taxonomic expansion, and functional annotation', *Nucleic Acids Res,* 44 (D1), D733-45.
622 Petersen, J. M., et al. (2022), 'The naked truth: An updated review on nudiviruses and their
623 relationship to bracoviruses and baculoviruses', *J Invertebr Pathol,* 189, 107718.
624 Québatte, M., et al. (2017), 'Gene transfer agent promotes evolvability within the fittest
625 subpopulation of a bacterial pathogen', *Cell Syst,* 4 (6), 611-21.e6.

626 Rapp, B. J. and Wall, J. D. (1987), 'Genetic transfer in *Desulfovibrio desulfuricans*', *Proc Natl*
627 *Acad Sci U S A,* 84 (24), 9128-30.
628 Rozewicki, J., et al. (2019), 'MAFFT-DASH: integrated protein sequence and structural
629 alignment', *Nucleic Acids Res,* 47 (W1), W5-w10.
630 Shakya, M., Soucy, S. M., and Zhaxybayeva, O. (2017), 'Insights into origin and evolution of α-
631 proteobacterial gene transfer agents', *Virus Evol,* 3 (2), vex036.
632 Shimizu, T., et al. (2022), 'Persulfide-responsive transcription factor SqrR regulates gene transfer
633 and biofilm formation via the metabolic modulation of cyclic di-GMP in *Rhodobacter*
634 *capsulatus*', *Microorganisms,* 10 (5).
635 Shimodaira, H. (2002), 'An approximately unbiased test of phylogenetic tree selection', *Syst Biol,*
636 51 (3), 492-508.
637 Stanton, T. B., et al. (2008), 'Collateral effects of antibiotics: carbadox and metronidazole induce
638 VSH-1 and facilitate gene transfer among *Brachyspira hyodysenteriae* strains', *Appl*
639 *Environ Microbiol,* 74 (10), 2950-6.
640 Steenwyk, J. L., et al. (2020), 'ClipKIT: a multiple sequence alignment trimming software for
641 accurate phylogenomic inference', *PLoS Biol,* 18 (12), e3001007.
642 Strand, M. R. and Burke, G. R. (2020), 'Polydnaviruses: evolution and function', *Curr Issues Mol*
643 *Biol,* 34, 163-82.
644 Tamarit, D., et al. (2018), 'Origin and evolution of the *Bartonella* gene transfer agent', *Mol Biol*
645 *Evol,* 35 (2), 451-64.
646 Thézé, J., et al. (2011), 'Paleozoic origin of insect large dsDNA viruses', *Proc Natl Acad Sci U S*
647 *A,* 108 (38), 15931-5.

648 Tomasch, J., et al. (2018), 'Packaging of *Dinoroseobacter shibae* DNA into gene transfer agent
649 particles is not random', *Genome Biol Evol,* 10 (1), 359-69.
650 Umemori, E., et al. (1992), 'A phage in *Bartonella bacilliformis*', *Microbiol Immunol,* 36 (7),
651 731-6.
652 Walker, P. J., et al. (2021), 'Changes to virus taxonomy and to the International Code of Virus
653 Classification and Nomenclature ratified by the International Committee on Taxonomy of
654 Viruses (2021)', *Arch Virol,* 166 (9), 2633-48.
655 --- (2022), 'Changes to virus taxonomy and to the International Code of Virus Classification and
656 Nomenclature ratified by the International Committee on Taxonomy of Viruses (2022)',
657 *Arch Virol*, In preparation.
658 Westbye, A. B., Beatty, J. T., and Lang, A. S. (2017), 'Guaranteeing a captive audience:
659 coordinated regulation of gene transfer agent (GTA) production and recipient capability
660 by cellular regulators', *Curr Opin Microbiol,* 38, 122-29.
661 Yen, H. C., Hu, N. T., and Marrs, B. L. (1979), 'Characterization of the gene transfer agent made
662 by an overproducer mutant of *Rhodopseudomonas capsulata*', *J Mol Biol,* 131 (2), 157-
663 68.
664 Zhan, Y., et al. (2016), 'A novel roseobacter phage possesses features of podoviruses,
665 siphoviruses, prophages and gene transfer agents', *Sci Rep,* 6, 30372.
666 Zhao, Y., et al. (2009), 'Gene transfer agent (GTA) genes reveal diverse and dynamic
667 *Roseobacter* and *Rhodobacter* populations in the Chesapeake Bay', *ISME J,* 3 (3), 364-
668 73.

Figure 1

A
Rhodobacterales GTAs
Caulobacterales GTAs
RpGTA
DsGTA
RcGTA
RsGTA
96
93
81
50
Hyphomicrobiales GTAs
Sphingomonadales GTAs
0.5
Caudoviricetes
B
Sphingomonadales GTAs
Caulobacterales GTAs
Hyphomicrobiales GTAs
Rhodobacterales GTAs
RpGTA
DsGTA
RcGTA
RsGTA
100
100
93
63
0.5
Caudoviricetes

Figure 2

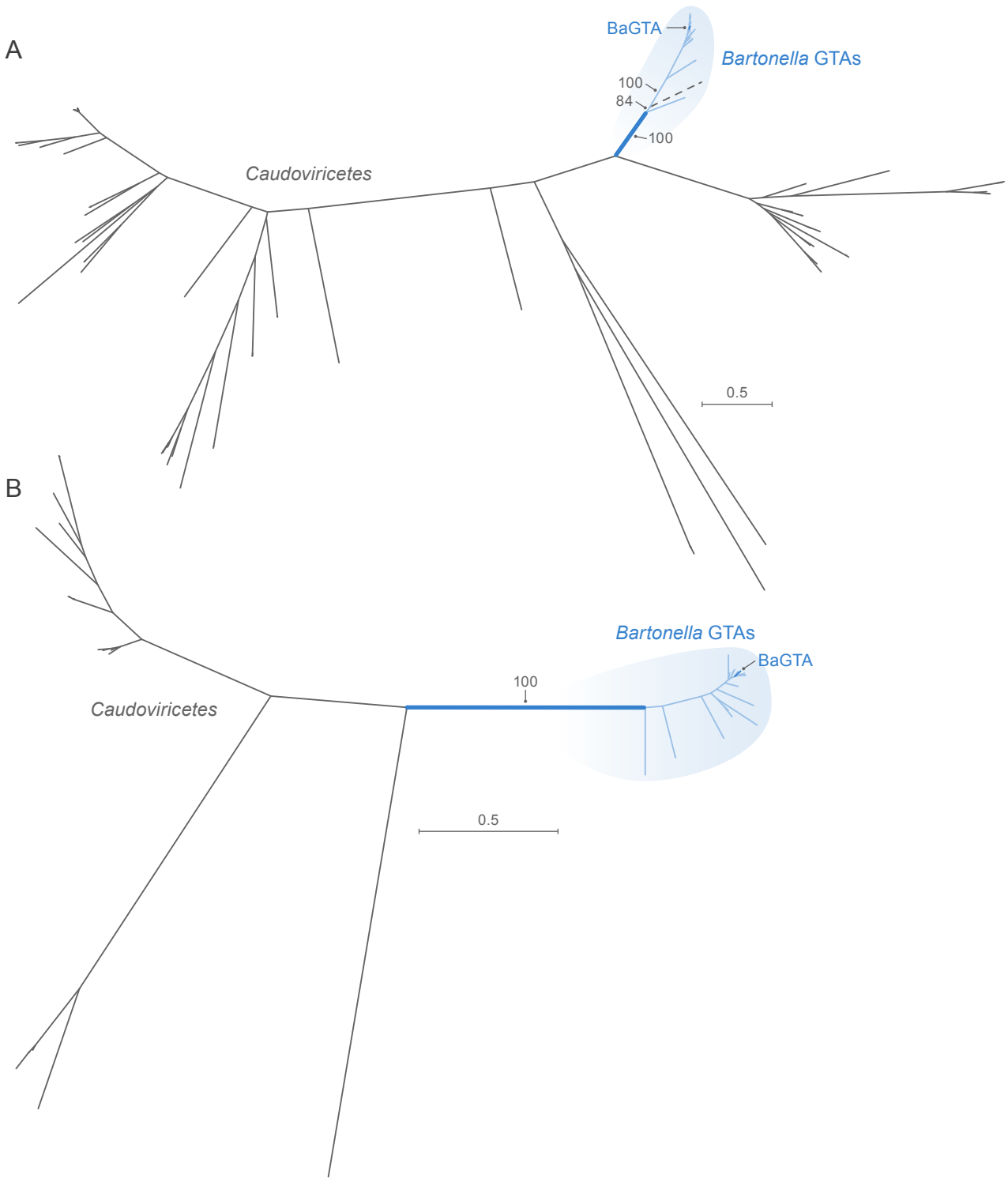

A
BaGTA
Bartonella GTAs
100
84
100
Caudoviricetes
0.5
B
Bartonella GTAs
BaGTA
100
Caudoviricetes
0.5

Figure 3

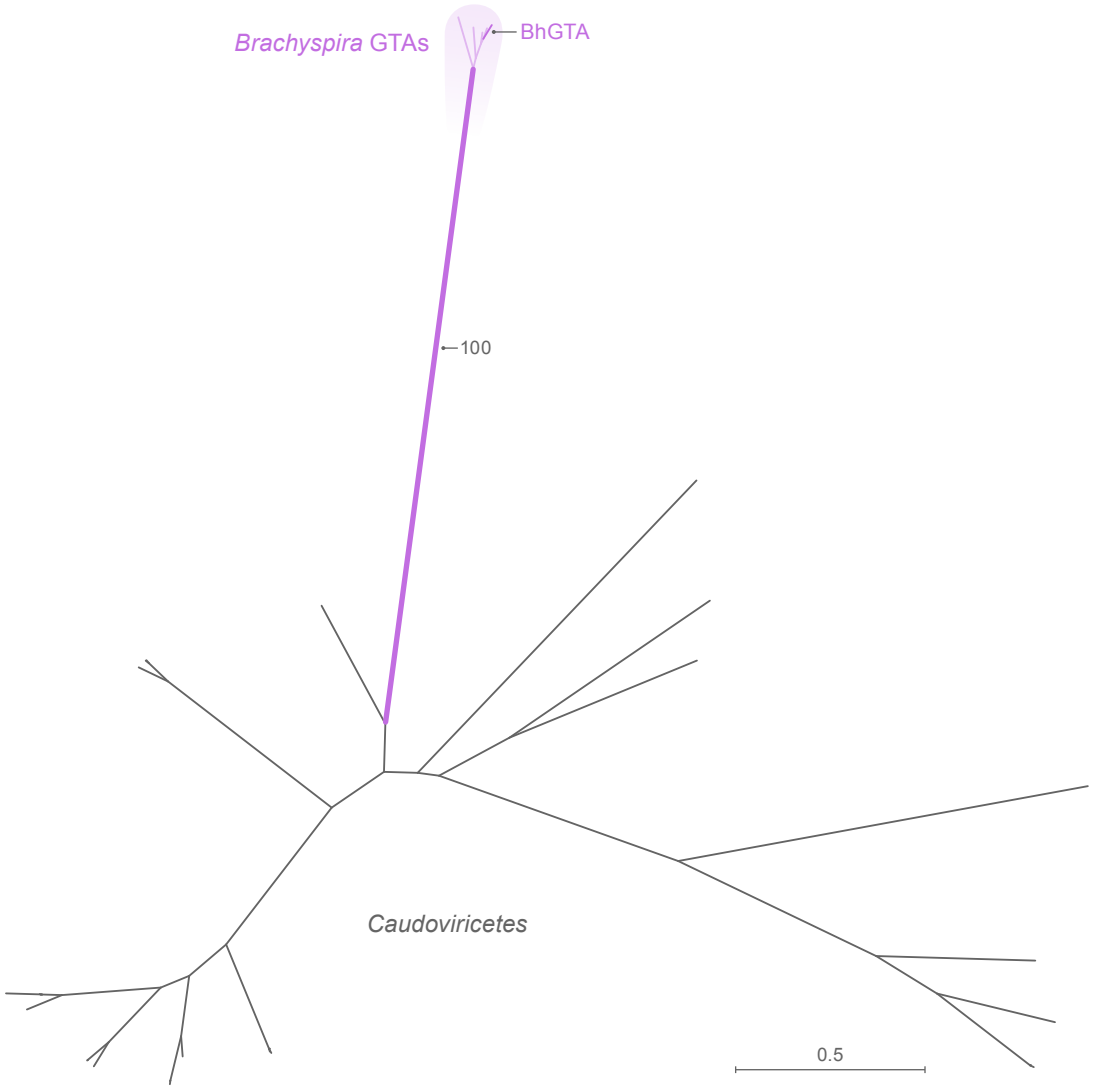
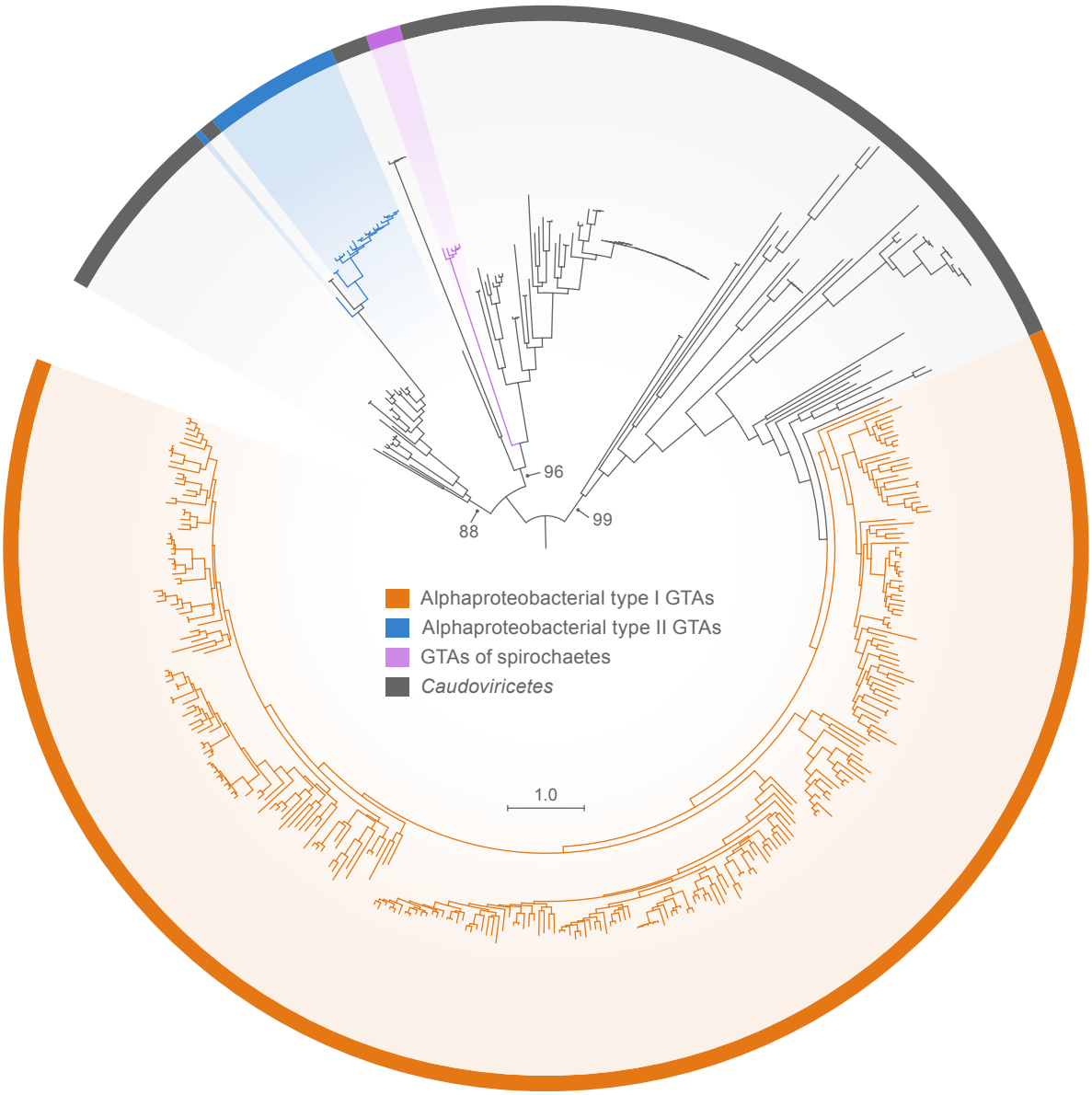


Figure 4


96
88
99
Alphaproteobacterial type I GTAs
Alphaproteobacterial type II GTAs
GTAs of spirochaetes
Caudoviricetes
1.0