

This is a repository copy of *Identification and characterization of a direct activator of a gene transfer agent*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/141545/>

Version: Accepted Version

Article:

Fogg, Paul Christopher Michael orcid.org/0000-0001-5324-4293 (2019) Identification and characterization of a direct activator of a gene transfer agent. *Nature Communications*. 595. p. 595. ISSN 2041-1723

<https://doi.org/10.1038/s41467-019-08526-1>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

This is a post-peer-review, pre-copyedit version of an article published in Nature Communications. The final authenticated version is available online at:
<http://dx.doi.org/10.1038/s41467-019-08526-1>

1 **Title:** Identification and characterization of a direct activator of a gene transfer agent

2 **Author:** Paul C.M. Fogg¹.

3 **Affiliations:**

4 ¹University of York, Biology Department, Wentworth Way, York, United Kingdom. YO10 5DD

5 *Correspondence to: paul.fogg@york.ac.uk

6

7 **Abstract**

8 Gene transfer agents (GTAs) are thought to be ancient bacteriophages that have been co-opted
9 into serving their host and can now transfer any gene between bacteria. Production of GTAs is
10 controlled by several global regulators through unclear mechanisms. In *Rhodobacter*
11 *capsulatus*, gene *rcc01865* encodes a putative regulatory protein that is essential for GTA
12 production. Here, I show that *rcc01865* (hereafter *gafA*) encodes a transcriptional regulator that
13 binds to the GTA promoter to initiate production of structural and DNA packaging components.
14 Expression of *gafA* is in turn controlled by the pleiotropic regulator protein CtrA and the quorum-
15 sensing regulator GtaR. GafA and CtrA work together to promote GTA maturation and eventual
16 release through cell lysis. Identification of GafA as a direct GTA regulator allows the first
17 integrated regulatory model to be proposed and paves the way for discovery of GTAs in other
18 species that possess *gafA* homologues.

19 **Main Text**

20 Rapid bacterial evolution is a fundamental process that allows bacteria to adapt to
21 changes in their environment and to explore new environmental niches. The primary
22 mechanisms for the rapid spread of genes are known collectively as Horizontal Gene Transfer
23 (HGT). In contrast to hereditary transfer, HGT allows genes to be passed directly between
24 individual bacteria at a much faster rate ^{1,2}. The genes being transferred may improve fitness or
25 resilience but can also lead to antimicrobial resistance (AMR) or increased virulence.

26 Traditionally, bacterial HGT consists of three broad mechanisms of genetic exchange –
27 conjugation, transformation and transduction. Transduction by bacteriophages is generally
28 accepted to be the most influential mechanism for the exchange of genes between bacteria, in
29 particular, the generalized transducing (GT) phages and the recently described lateral transducing
30 (LT) phages play a crucial role ³. During phage replication, host bacterial DNA is packaged into
31 a significant proportion of phage particles instead of the phage genome; the host DNA can be
32 randomly selected (GT phages) or it can be from a large hypermobile region (LT phages). The
33 packaged host DNA is then protected by the phage capsid and delivered to a new host cell, where
34 it can be integrated into the target genome by homologous recombination.

35 Gene transfer agents (GTAs) are an unusual method of HGT, which appears to be a
36 hybrid of bacteriophage transduction and natural transformation ⁴. First discovered in the 1970s,
37 GTAs are small virus-like particles that transfer random fragments of the entire genome of their
38 bacterial host between cells ⁵. Unlike the transducing phages, whose primary aim is still self-
39 preservation, GTAs have no preference for the spread of their own genes and their survival is
40 entirely dependent upon their hosts' wellbeing ^{6,7}. It is the complete lack of DNA selectivity that

41 makes GTAs particularly intriguing and raises important questions about their impact on HGT,
42 bacterial evolution and the selective pressures that allow them to persist ⁸.

43 A rough estimate of the number of viruses in the oceans alone is 4×10^{30} ⁹. Metagenomic
44 analyses of the marine virome typically reveal that >60% of the sequences are unrelated to any
45 known viruses, and there has been speculation that GTAs are a significant contributor to this
46 cloud genome ^{10,11}. A seminal study of antibiotic gene transfer by GTAs in in situ marine
47 microcosms, observed frequencies that were orders of magnitude greater than any known
48 mechanism ¹². In the model host, *Rhodobacter capsulatus*, RcGTAs are under the control of a
49 number of conserved global regulatory systems such as the cell cycle regulator CtrA ¹³⁻¹⁵, the
50 quorum sensing regulator GtaR ^{16,17} and various phosphorelay components such as DivL and
51 CckA ^{15,18}, however, all of these regulators affect RcGTA production indirectly and thus the
52 mechanism of activation is unclear.

53 In this study, I identify and characterize a transcription factor (Rcc01865 renamed GafA
54 here) that binds directly to the RcGTA promoter. The *gafA* promoter is in turn bound by both the
55 pleiotropic regulators CtrA and GtaR near the transcription start site. CtrA and GafA are both
56 required for optimal RcGTA expression, packaging of DNA and release of infective particles.
57 The data presented here indicates that GafA is the missing link that connects RcGTA production
58 with host regulatory systems and allows construction of the most comprehensive model of
59 RcGTA regulation to date.

60 **Results and Discussion**

61 **All RcGTA genes are upregulated in an RcGTA hyperproducer.** RcGTAs are usually
62 produced from a small sub-population, making in-depth analysis of RcGTA producers
63 problematic ^{6,19}. Here, we compared the transcriptome of an RcGTA hyperproducer, *R.*

64 capsulatus DE442, to the wild-type by RNAseq ¹⁹. 152 upregulated and 37 down regulated genes
65 were identified (Supplementary Tables 1 & 2). The top 29 upregulated genes had a beta value (b)
66 of 4.0 or greater (Supplementary Table 3), equivalent to a 16-fold increase in transcript
67 abundance, and contained all of the genes from the core RcGTA structural gene cluster ¹⁴, head
68 spikes ²⁰, tail fibre ²¹, lysis genes ¹⁸ and a putative RcGTA maturation protein ²². One further
69 gene, rcc01865, was previously shown to be essential for RcGTA production but its precise role
70 is unknown ²². Rcc01865 encodes a protein with a predicted helix-turn-helix (HTH) DNA
71 binding motif in the N-terminal domain that structurally resembles the DNA binding domain
72 (DBD) of the genome replication initiator protein DnaA (e.g. Mycobacterium tuberculosis
73 DnaA-DBD, 3PVV; Supplementary Figure 1), which led to the assumption that it is a regulator
74 protein ²². The C-terminus contains a region that has similarity to various sigma factors,
75 including a high HHPRED probability match to Rhodobacter sphaeroides RpoE (Supplementary
76 Figure 1). Given that rcc01865 is essential for RcGTA production ²² and encodes the only
77 putative transcription factor in the top 29 upregulated genes in the RNAseq data (Supplementary
78 Table 3), it is a strong candidate to be a specific initiator of RcGTA production. Rcc001865 will
79 hereafter be referred to as GTA Activation Factor A (gafA).

80 **GafA (rcc01865) activates production of RcGTA particles.** Deletion of gafA completely
81 prevents RcGTA gene transfer ²², even in the hyperproducer strain R. capsulatus DE442 (Figure
82 1A) where RcGTA gene expression, gene transfer frequencies and the proportion of the
83 producing RcGTAs are normally substantially increased ^{6,19}. Furthermore, in DE442, packaged
84 GTA DNA can be seen as a distinct 4 kb band in a total DNA purification. Deletion of gafA
85 prevents any detectable GTA DNA in this assay (Figure 1B), indicating that RcGTA production
86 is fundamentally undermined at or before the DNA packaging stage. Overexpression of gafA in

87 wild-type *R. capsulatus* SB1003 increased antibiotic gene transfer frequencies 57-fold (SD=7,
88 n=8), compared to 94-fold for the stable hyperproducer phenotype (SD=19, n=8) (Figure 1A)¹⁹.
89 In addition, total DNA from the *gafA* overexpressor contained large quantities of 4 kb GTA
90 DNA after 6 h (Figure 1B). After 24 h, the cells partially dampened RcGTA production,
91 although the levels observed were still far greater than WT (Figure 1B). Dampening of RcGTA
92 production is not unexpected as uniform expression in all cells is likely to be highly deleterious
93 ^{6,18,19,23}.

94 **CtrA overexpression does not lead to RcGTA overproduction.** Previous work showed that
95 the global regulator protein CtrA is also essential for RcGTA production¹⁴, however, the
96 mechanism has never been discovered. Similar to *gafA*, deletion of *ctrA* prevents any detectable
97 RcGTA gene transfer or production of the RcGTA capsid protein¹⁴. Activity of CtrA is
98 modulated by phosphorylation of an aspartic acid residue (D51), and its phosphorylation state is
99 important for RcGTA production^{15,24}. The RNAseq data showed that CtrA is upregulated (2.5-
100 fold) in DE442 (Supplementary Tables 1 & 3) along with known CtrA regulon genes for
101 chemotaxis and motility (Supplementary Table 1). If *gafA* is a simple constituent of the CtrA
102 regulon then increasing the abundance CtrA should lead to RcGTA overproduction.
103 Overexpression of WT *ctrA* or phosphomimetic *ctrA*^{D51E} led to a slight reduction in RcGTA gene
104 transfer, whereas non-phosphorylatable *ctrA*^{D51A} increased gene transfer 2-fold (Figure 2A)²⁵.
105 No GTA DNA bands were detected in total DNA for any of the *ctrA* overexpressor strains
106 (Figure 2B), consistent with no effect or a modest increase in RcGTA production. Similar to the
107 *gafA* deletion, *ctrA* knock-outs were not able to produce any detectable RcGTAs in WT^{13,14} or
108 hyperproducer strains (Figure 2).

109 **CtrA is controls GafA activity, RcGTA maturation and lysis.** Overexpression of *gafA* in cells
110 lacking *ctrA* still led to substantial intracellular GTA DNA accumulation (Figure 2B), albeit at a
111 lower level than in *ctrA* replete cells (Figure 2B), indicating that the essential role of CtrA in
112 expression of the GTA structural gene cluster is upstream of GafA. Overexpression of *gafA*,
113 however, did not rescue RcGTA gene transfer ability in the *ctrA* knock-out, DNaseI insensitive
114 DNA was not detectable in the culture supernatant and manual lysis of the cells did not release
115 any detectable infective RcGTA particles. Taken together, these data show that GafA activates
116 synthesis of the RcGTA structural genes and packaging of host DNA, whilst, CtrA is required
117 for maturation and release of infective RcGTA particles.

118 To further investigate the relationship between CtrA, GafA and RcGTA production,
119 transcription of various GTA-related genes was measured. As expected from the phenotypic
120 profiles, deletion of *ctrA* or *gafA* in DE442 eliminated the hyperproducer expression profile.
121 Expression of the RcGTA terminase, capsid and endolysin genes all reduced to basal levels
122 (Figure 3A). Deletion of *ctrA* also reduced *gafA* expression but deletion of *gafA* did not affect
123 *ctrA* expression, consistent with the hypothesis that *gafA* is part of the CtrA regulon.

124 Overexpression of *ctrA* did not lead to a substantial increase in transcription of the
125 RcGTA structural genes, lysis cassette or *gafA* (Figure 3B), but did increase the abundance of
126 native *ctrA* transcripts indicating positive autoregulation (Figure 3C). Overexpression of *gafA* in
127 WT cells led to a large increase in RcGTA gene expression (Figure 3D). After six hours, *gafA*
128 was overexpressed 34-fold leading to a large increase in terminase (78-fold), capsid (6-fold) and
129 endolysin (6-fold) transcripts, supporting the hypothesis that GafA is an activator of core RcGTA
130 gene expression and is also involved in the endgame of RcGTA release. In the *ctrA* knockout,
131 overexpression of *gafA* was even greater (198-fold) with an associated increase in terminase

132 (126-fold) and capsid (22-fold) transcription but endolysin upregulation was diminished (Figure
133 3D). Lack of lysis in the absence of *ctrA* is a likely explanation for increased transcript
134 abundance for *gafA* and the RcGTA genes. The requirement of CtrA for endolysin production is
135 presumably to allow temporal control of the different stages of RcGTA production, e.g. lysis
136 must not occur before RcGTA particles are fully mature and infective. Transcription of *gafA*
137 from the native promoter also increased 31-fold in response to ectopic *gafA* expression (Figure
138 3E). Strong positive *gafA* autoregulation could represent a hair trigger that, once initiated, locks
139 the cell into a lytic fate. In contrast, no increase in native *gafA* transcripts was detected in the
140 absence of *ctrA* (Figure 3E). These data clearly indicate that GafA induces expression of the core
141 RcGTA genes independent of CtrA, however, positive autoregulation of its own transcription is
142 CtrA dependent, providing further evidence that CtrA is required for activation of GafA.
143 Meanwhile, given that deletion of either *ctrA* or *gafA* in DE442 downregulates endolysin
144 expression and GafA only induces endolysin expression in *ctrA* replete cells, both CtrA and
145 GafA must act in concert to promote lytic release of RcGTAs.

146 **LexA and DivL are upregulated in RcGTA overproducers.** In other species such as
147 *Caulobacter crescentus*, *ctrA* is an essential cell cycle regulator^{25,26} and in *Rhodobacter*,
148 although not essential, it must control the timing of distinct phases of RcGTA production. Recent
149 work identified a phosphorelay (ChpT/CckA/DivL) that modulates CtrA phosphorylation^{15,18}
150 and dysregulation of the PAS/PAC domain protein DivL led to increased RcGTA production¹⁵.
151 DivL transcript abundance was 4 to 7-fold upregulated in DE442 (Figure 3A & Supplementary
152 Table 3) but unaffected by *gafA* overexpression and mildly increased by *ctrA* overexpression
153 (Supplementary Figure 2A). DivL was, however, significantly down regulated in *ctrA* knock-outs
154 (Supplementary Figure 2A). The SOS repressor, *lexA*, is also required for efficient RcGTA

155 production by regulating the production of CckA²⁷. GafA and ctrA overexpression both led to a
156 marginal increase (1.5 to 2-fold) in lexA transcription and, in DE442, lexA transcripts were 2 to
157 8-fold higher than WT (Figure 3A, Supplementary Figure 2B & Supplementary Table 3). It is
158 likely that a moderate increase in LexA represses CckA, which in turn shifts the CtrA
159 equilibrium toward the unphosphorylated state and thus boosts RcGTA production²⁷.

160 **CtrA binds near the gafA transcription start site.** Clearly, CtrA and GafA work together to
161 control RcGTA production. There is an obvious CtrA binding site in its own promoter (GTAAC-
162 N₆-TTAAC, Figure 4A) and the GafA promoter contains an almost identical sequence (TTAAC-
163 N₆-GTAAC, Figure 4A)^{13,28}. Alignment of the *R. capsulatus* gafA promoter with gafA promoters
164 from 14 different species (Supplementary Figure 3), revealed remarkable conservation of the
165 CtrA binding site and its distance to the start codon (usually 65-71 bases) despite otherwise
166 divergent flanking sequences. In an electrophoretic motility shift assay (EMSA), purified CtrA
167 had no detectable binding affinity for its own promoter (≤ 8000 nM Protein, Supplementary
168 Figure 4A), however, CtrA^{D51E} was able to bind to the promoter at low affinity (Supplementary
169 Figure 4B). In contrast, CtrA bound to the gafA promoter with much greater affinity than the
170 ctrA promoter (Kd 54.91 nM, SD 6.12, Figure 4B & C), in agreement with the observations that
171 CtrA is essential for gafA transcription. Furthermore, the hypothesis that CtrA regulates gafA
172 transcription was strengthened by mapping raw RNAseq transcript reads onto the gafA promoter
173 sequence, which revealed that the transcription start site is likely to be ~87 bp upstream of the
174 start codon and coincides with the CtrA binding site (Figure 4A). To test whether CtrA binding
175 to the gafA promoter is required for RcGTA production, SB1003 gafA Δ was complemented in
176 trans with plasmids containing either gafA expressed from its unaltered native promoter
177 (pCMF180) or with either of the two CtrA binding half-sites mutated by site directed

178 mutagenesis (pCMF214 and pCMF215) (Supplementary Figure 5). Complementation with the
179 wild-type promoter construct increased gene transfer frequency to 337% of WT (SD=2%, n=3,
180 ANOVA p-value=<0.001), presumably due to increased copy number of the plasmid borne *gafA*,
181 whereas, both mutated promoter constructs were significantly impaired for gene transfer (10-
182 22% of WT, n=3, ANOVA p-value=<0.001).

183 **The quorum sensing regulator GtaR binds the *gafA* promoter.** CtrA is evidently important
184 for *GafA* production, however, it is unlikely to be the only regulator acting on *gafA*. CtrA is
185 expressed throughout all growth stages, whereas RcGTA are only produced in stationary phase
186 ^{5,29}, and its expression is homogenous in wild-type cells ³⁰, whereas RcGTA are only produced
187 by <1% of the population ^{6,19}. Moreover, overexpression of *ctrA* does not lead to a substantial
188 increase in *gafA* transcription or RcGTA production (Figure 2 & 3). The GtaI/R quorum sensing
189 system is also essential for RcGTA production^{16,17,31}. Regulation by quorum sensing would
190 certainly allow *gafA* and RcGTA expression to be limited to stationary phase and heterogeneity
191 of the response to homoserine lactone inducer signal could also be responsible for RcGTA phase
192 variation ³²⁻³⁴. Band shifts were carried out using the same *gafA* promoter region that contains
193 the CtrA binding site (Figure 4A) and purified GtaR. GtaR binding was detected at
194 concentrations of 375 nM or above (Figure 5). The only known binding site for GtaR is within its
195 own promoter ¹⁶ and no analogous sequence was detected in the 50 bp promoter fragment used
196 here, which is not unexpected. Binding sites for quorum sensing proteins are thought to be highly
197 degenerate and thus difficult to predict; indeed Leung et al. (2013) reported that the best matches
198 to the model GtaR binding site in *R. capsulatus* were not bound in vitro ¹⁶. It is notable that GtaR
199 binds to its own promoter at a location spanning the predicted -10 Shine Delgarno element and

200 the transcription start site ¹⁶, and the *gafA* promoter region bound by GtaR here contains the
201 same promoter features (Figure 4A).

202 **GafA, but not CtrA, binds to the RcGTA promoter.** The data presented so far suggest that
203 GafA acts as a direct regulator of RcGTA expression and it is likely to bind to the promoter
204 region of the structural gene cluster, hereafter referred to as the RcGTA promoter. The RcGTA
205 promoter is not well characterized and no transcription factors have been identified that bind in
206 this region. An EMSA was carried out with five overlapping 50 bp probes that were designed to
207 cover the 174 bp region immediately upstream of RcGTA *g1* (Figure 6A & B). GafA binding
208 was only detected with one of the five probes (pGTA2, Figure 6C) spanning the region 76 to 125
209 bp upstream of the RcGTA *g1* start codon (Figure 6A). Titration of the GafA protein revealed
210 detectable binding to pGTA2 with low as 16 nM protein (Figure 6D). Accurate estimation of the
211 K_d was not possible because there were insufficient data points at full saturation, however, it is
212 likely to be in the high nanomolar range. The pGTA2 promoter region contains the predicted -10
213 element and the transcription start site, which was confirmed by analysis of the raw RNAseq
214 mRNA coverage (Figure 6A). Binding of GafA to the region containing the -10 and TSS,
215 together with phenotypic and qPCR data described above, strongly supports the hypothesis that
216 GafA is a direct regulator of RcGTA at the transcriptional level, possibly as an alternative sigma
217 factor. Mercer et al (2014) reported a putative partner switching signalling pathway, comprising
218 RbaV, RbaW and RbaY, that when disrupted had a moderate but significant effect on RcGTA
219 production (<3-fold) ²⁴. RbaW was predicted to be an anti-sigma factor and extensive attempts
220 were made to identify the cognate sigma factor, including deletion of all known sigma factors
221 except RpoN and RpoD, none of which were found to interact with RbaW or affect expression of

222 RcGTA. GafA had not been linked to RcGTA at that time and thus was not considered, but it is
223 possible that GafA is the target of RbaW.

224 Meanwhile, no CtrA binding was detected to the full length RcGTA promoter (Figure
225 6E), confirming that CtrA regulation is indirect. The data presented are the first evidence of a
226 transcription factor activating an RcGTA promoter and for the first time a direct link has been
227 established with core host regulatory pathways via CtrA and GtaR. Furthermore, GafA binds to
228 its own promoter region (Supplementary Figure 6A) to positively auto-regulate its own
229 expression (Figure 3E) and to the lysis cassette promoter (Supplementary Figure 6B) to induce
230 endolysin expression (Figure 3D), indicating that GafA plays a critical role in both RcGTA
231 production and subsequent release.

232 **GafA is a core component of an RcGTA regulation model.** The results presented here allow a
233 model of RcGTA regulation to be constructed (Figure 7). Rhodobacter RcGTA production
234 begins in stationary growth phase, controlled by the quorum sensing protein¹⁶. Once RcGTA
235 production begins, unphosphorylated CtrA activates *gafA* expression; GafA then enhances its
236 own expression, activates expression of the core GTA structural cluster and packaging of DNA
237 into capsids. GTAs are normally produced in a small proportion of any given population^{6,19,35},
238 however, in wild-type cells CtrA expression is more or less homogenous³⁰ and simple
239 overexpression of *ctrA* does not lead to high level expression of *gafA* (Figure 2), which suggests
240 that there are other unknown factors in play. There is no evidence that epigenetic factors, such
241 methylation or DNA inversions, influence RcGTA production but heterogeneity in the quorum
242 sensing response is a possible explanation for RcGTA phase variation. Relative fitness has been
243 implicated as a factor that induces *Bartonella* GTA (BaGTA)³⁵, i.e. the fittest subpopulation
244 spontaneously produce BaGTAs presumably to spread the most beneficial genes, but

245 contradictory data has been reported for RcGTA suggesting that it is starvation that leads to
246 production^{18,27,36}. Subsequent to induction of the RcGTA structural genes, CtrA is
247 phosphorylated by the DivL/CckA/ChpT phosphorelay¹⁵. CtrA-P activates expression of
248 maturation and secondary structural proteins required for infectivity¹⁵. Finally, GafA binds to
249 the endolysin promoter and induces CtrA-dependent cell lysis and RcGTA release.

250 Hynes et al. (2016) reported that GafA homologues are present throughout the
251 Rhodobacterales, including in each of the confirmed GTA producers, and local synteny of GafA
252 is broadly conserved i.e. it is usually flanked by lipoyl synthase (lipA) and GMP synthase (gua1)
253 genes²². Overexpression of gafA homologues from two known GTA producers (*Ruegeria*
254 *mobilis* & *Roseovarius nubinhibens*, Supplementary Figure 7) also led to increased GTA
255 production (Supplementary Figure 8), demonstrating that activation of GTAs by GafA is not
256 unique to *R. capsulatus*. Although GafA is present in various different species, its rate of
257 evolution was reported to be faster than most components of the RcGTA genome, albeit only
258 marginally so²². In general, all RcGTA genes tend to be evolving faster than core host genes and
259 slower than comparable phage genes²². Beyond the Rhodobacterales, gafA homologues can be
260 found in the Rhizobiales³⁷, a bacterial order that includes plant and animal pathogens such as
261 *Agrobacterium tumefaciens* and *Brucella abortus*. Rhizobiales gafA genes are usually share less
262 than 25% homology with their Rhodobacterales counterparts³⁷ or are split into two separate
263 ORFs, for example in *A. tumefaciens* (NZ_ASXY01000077) each ORF product is homologous
264 to the either the N-terminal DnaA DBD-like domain or C-terminal sigma factor-like domains.

265 GTAs are thought to be derived from ancient bacteriophage that have been hijacked by
266 their host²², although the lack of significant matches to GTA genes in α -proteobacterial CRISPR
267 spacer regions suggest that the hypothetical progenitor phage is extinct³⁷. Several marine

268 Roseophages, such as RDJLΦ1, contain several GTA-like structural genes as well as both GafA
269 and its neighbour, rcc01866^{22,38}, but they are separated by a single intervening gene with clear
270 homology to CtrA⁷. The phage version of CtrA lacks the N-terminus, which contains the
271 response regulator domain, but retains the transcriptional activator domain. The presence of
272 homologues of essential RcGTA regulator and structural genes in a phage suggests that the
273 relationship between these regulators and GTA production is ancient.

274 GTAs have the potential to drive bacterial evolution and genome plasticity, including the
275 spread of virulence and AMR genes. Here, GafA is identified as the first direct activator of GTA
276 expression to be reported for any species. The data allow the construction of a comprehensive
277 model of RcGTA regulation that brings together the roles of the pleiotropic regulator CtrA,
278 quorum sensing, the SOS response and a conserved phosphorelay chain. Furthermore, many
279 aspects of GTA biology make them intractable for high throughput studies, but identification of
280 direct activators of GTAs in widespread species could open up a new frontier in GTA research.

281

282 **Methods**

283 **Bacterial Strains.** Two wild-type *Rhodobacter* strains were used – rifampicin resistant SB1003
284 (ATCC BAA-309) and rifampicin sensitive B10³⁹. The RcGTA overproducer strain DE442 is of
285 uncertain provenance but has been used in a number of RcGTA publications^{19,40}. The *E. coli*
286 S17-1 strain, which contains chromosomally integrated tra genes, was used as a donor for all
287 conjugations. NEB 10-beta Competent *E. coli* (New England Biolabs, NEB) were used for
288 standard cloning and plasmid maintenance; T7 Express Competent *E. coli* (NEB) were used for
289 overexpression of proteins for purification. *Ruegeria mobilis* (DSM 23403), *Roseovarius*

290 nubinhibens (DSM 15170) and *Ruegeria pomeroyi* (DSM 15171) are reported GTA producers
291 that were all obtained from DSMZ.

292 **Cloning.** All oligonucleotides were obtained from IDT (Supplementary Table 4) and designed
293 with an optimal annealing temperature of 60°C when used with Q5 DNA Polymerase (NEB). All
294 cloning reactions were carried out with either the In-Fusion Cloning Kit (CloneTech) or
295 NEBuilder (NEB) to produce the constructs listed in Supplementary Table 5. In summary,
296 destination plasmids were linearized using a single restriction enzyme (pCM66T (BamHI),
297 pEHIS-TEV (NcoI) and pSRKBB (NdeI)), or linearized by PCR (pETFPP_2 using primers CleF
298 and CleR). Inserts were amplified using primers with 15 bp 5' overhangs that have
299 complementary sequence to the DNA with which it is to be recombined.

300 **Transformation.** Plasmids were introduced into all species except *Rhodobacter* by
301 transformation. *E. coli* was transformed by standard heat shock transformation⁴¹. For *Ruegeria*
302 and *Roseovarius*, 200 ml cultures were washed three times in ice cold 10% glycerol (100 ml then
303 50 ml then 5 ml). 100 µl aliquots were mixed with 100 ng plasmid DNA and incubated on ice for
304 30 min. Electroporation was carried out in 2 mm electroporation cuvettes (Scientific Laboratory
305 Supplies) at 2.5 kV, 25 µF and 100 Ω. 1 ml of marine broth was added and cells incubated at
306 30°C for 4 h, then plated onto MB agar + 50 µg ml⁻¹ kanamycin.

307 **Conjugation.** 1 ml aliquots of overnight cultures of the *E. coli* S17-1 donor and *Rhodobacter*
308 recipient strains were centrifuged at 5,000 x g for 1 min, washed with 1 ml SM buffer,
309 centrifuged again and resuspended in 100 µl SM buffer. 10 µl of concentrated donor and
310 recipient cells were mixed and spotted onto YPS agar or spotted individually as negative
311 controls. Plates were incubated o/n at 30°C. Spots were scraped, suspended in 100 µl YPS broth
312 and plated on YPS + 100 µg ml⁻¹ rifampicin (counter-selection against *E. coli*) + 10 µg ml⁻¹

313 kanamycin (plasmid selection). Plates were incubated o/n at 30°C then restreaked onto fresh agar
314 to obtain single colonies.

315 **Nucleic Acid Purification.** 1 ml samples of relevant bacterial cultures were taken for each
316 nucleic acid purification replicate. Generally, sampling occurred during stationary phase but for
317 overexpression experiments samples were taken 6 h and 24 h after transition to anaerobic
318 growth. Total DNA was purified according to the Purification of Nucleic Acids by Extraction
319 with Phenol:Chloroform protocol ⁴¹. In brief, cell pellets were resuspended in 567 μ l TE buffer
320 then 30 μ l of 10% SDS and 3 μ l of 10 mg ml⁻¹ proteinase K were added. Cells were incubated at
321 37°C for 1 h to allow complete lysis. 100 μ l of 5 M NaCl was added to each tube and mixed
322 thoroughly, before addition of 80 μ l of 1% CTAB in 100 mM NaCl. The cell lysates were
323 incubated at 65°C for 10 minutes. Nucleic acids were purified by addition of an equal volume of
324 Phenol:Chloroform:Isoamyl Alcohol (25:24:1, pH 8.0), vigorous mixing by inversion and
325 centrifugation for 5 min at 14,000 x g. The upper aqueous layer containing DNA was carefully
326 pipetted into a fresh tube and the phenol:chloroform:isoamyl alcohol step was repeated a further
327 two times. Traces of phenol were removed by addition of an equal volume of chloroform,
328 vigorous mixing by inversion and centrifugation for 5 min at 14,000 x g. The aqueous fraction
329 was transferred to a fresh tube and nucleic acids were precipitated by addition of 0.6 volume of
330 ice cold isopropanol, incubation at -20°C for 1 h and centrifugation at 14,000 x g for 20 min.
331 DNA pellets were washed with 70% ethanol, air dried for ~10 min and resuspended in 50-100 μ l
332 of TE buffer. RNA was removed by addition of 1 μ l of 10 mg ml⁻¹ RNase and incubation at 37°C
333 for 1 h. Total RNA was purified using the NucleoSpin RNA Kit (Macherey-Nagel) and DNaseI
334 treated on column according to the recommended protocol. RNA was quantified using a

335 Nanodrop spectrophotometer. 1 μ g of total RNA was converted to cDNA using the LunaScript
336 RT SuperMix Kit (NEB).

337 **RNAseq.** Production of GTAs is thought to lead to cell death through packaging of host cell's
338 entire genome followed by lysis from within ^{18,19,42}. To inhibit lysis cultures were grown in a
339 high phosphate medium, RCV, to stationary phase where total RNA was isolated ¹⁸. RNA yield
340 was quantified and quality checked using a Nanodrop spectrophotometer and Aglient
341 bioanalyser. Ribosomal RNA was removed from 1 μ g good quality total RNA using the Ribo-
342 Zero rRNA Removal Kit (Bacteria; Illumina). Libraries were then prepared from rRNA-depleted
343 samples using the NEBNext RNA Ultra II Directional Library preparation kit for Illumina, with
344 single 6 bp indices, according the manufacturer's guidelines for insert sizes of approximately 200
345 - 350 bp. Libraries were pooled at equimolar ratios, and the pool was sent for 2 x 150 base paired
346 end sequencing on a HiSeq 3000 at the University of Leeds Next Generation Sequencing
347 Facility.

348 Abundance of transcripts were compared between the wild-type *R. capsulatus* strain
349 SB1003 (n=4), a GTA hyperproducer DE442 (n=4) and a DE442 culture that had been passaged
350 three times (n=4). Reads were quality checked and trimmed using FastQC version 11.0.5 ⁴³ and
351 Cutadapt version 1.8.3 ⁴⁴, respectively. Kallisto version 0.43.1 ⁴⁵ was used to pseudo-align reads
352 to the *R. capsulatus* SB1003 reference transcriptome, and to quantify gene expression.

353 Differential expression analysis was performed using Sleuth version 0.29.0 ⁴⁶. A full linear
354 model containing strain, passage and sequencing batch was fit to the data. In order to look at the
355 effect of strain, the full model was compared to a reduced model based only on passage and
356 batch. The effect size of the test variable, i.e. strain DE442 vs SB1003, was calculated using the
357 Wald test to give the beta value (b), based on fitting a linear model to the data, in log₂ units. The

358 se_b value is the standard error. The q-value (qval) is the p-value adjusted by false discovery
359 rate, where the p-value was calculated using the likelihood ratio test (LRT) in Sleuth. RNAseq
360 data was submitted to the GEO database with the record ID GSE118116 - Comparison of the
361 expression profiles of wild-type *Rhodobacter capsulatus* and a GTA hyperproducer (DE442) by
362 RNAseq.

363 **Gene Knock-Outs.** Knock-outs were created by RcGTA transfer. pCM66T plasmid constructs
364 were created with a gentamicin resistance cassette flanked by 500-1000 bp of DNA from either
365 side of the target gene. Assembly was achieved by a one-step, four component NEBuilder (NEB)
366 reaction and transformation into NEB 10-beta cells. Deletion constructs were introduced into the
367 RcGTA hyperproducer strain by conjugation and a standard GTA bio-assay was carried out to
368 replace the intact chromosomal gene with the deleted version.

369 **GafA Overexpression in Rhodobacter.** Gene overexpression in *Rhodobacter* was achieved by a
370 transcriptional fusion of the genes of interest to the *puf* photosynthesis promoter¹⁹. Growth and
371 general strain maintenance of *Rhodobacter* strains containing overexpression plasmids was
372 carried out at 30°C under aerobic, chemotrophic growth conditions where transcription from the
373 *puf* promoter is strongly repressed. To produce overexpression conditions 12 ml cultures were
374 grown to stationary phase aerobically, mixed 1:1 with fresh media and immediately transferred
375 to 23 ml sealed tubes. Cultures were then incubated at 30°C with illumination to induce *puf*
376 promoter activity.

377 **Rhodobacter Gene Transfer Assays.** In *Rhodobacter*, the assays were carried out essentially as
378 defined by Leung and Beatty (2013)⁴⁷. RcGTA donor cultures were grown anaerobically with
379 illumination in YPS for ~72 h and recipient cultures were grown aerobically in RCV for ~24 h.
380 For overexpression experiments, donor cultures were first grown aerobically to stationary phase

381 then anaerobically for 24 h. Cells were cleared from donor cultures by centrifugation and the
382 supernatant filtered through a 0.45 μm syringe filter. Recipient cells were concentrated 3-fold by
383 centrifugation at 5,000 x g for 5 min and resuspension in 1/3 volume G-Buffer (10 mM Tris-HCl
384 (pH 7.8), 1 mM MgCl_2 , 1 mM CaCl_2 , 1 mM NaCl, 0.5 mg ml^{-1} BSA). Reactions were carried out in
385 polystyrene culture tubes (Starlab) containing 400 μl G-Buffer, 100 μl recipient cells and 100 μl
386 filter donor supernatant, then incubated at 30°C for 1 h. 900 μl YPS was added to each tube and
387 incubated for a further 3 h. Cells were harvested by centrifugation at 5,000 x g and plated on
388 YPS + 100 $\mu\text{g ml}^{-1}$ rifampicin (for standard GTA assays) or 3 $\mu\text{g ml}^{-1}$ gentamicin (for gene
389 knock-outs).

390 **Quantitative Reverse Transcriptase PCR.** 1 in 50 dilutions of the cDNA template were
391 prepared and 1 μl used per reaction. Reactions contained Fast Sybr Green Mastermix (Applied
392 Biosystems), cDNA and primers (500 nM). Standard conditions were used with an annealing
393 temperature of 60°C. All primer efficiencies were calculated as between 90 and 110%. Relative
394 gene expression was determined using the $\Delta\Delta\text{Ct}$ method⁴⁸. For each sample, variance was
395 calculated for three independent biological replicates, which were each the mean of three
396 technical replicates. QuantStudio 3 Real-Time PCR System was used for all experiments
397 (Applied Biosystems).

398 **Protein Purification.** For His6-tagged proteins, 500 ml cultures of E. coli containing the
399 relevant expression plasmid were induced at mid-exponential growth phase with 0.2 mM IPTG
400 overnight at 20°C. Concentrated cells were lysed in 20 ml binding buffer (1 M NaCl, 75 mM
401 Tris; pH 7.75) plus 0.2 mg ml^{-1} lysozyme and 500 U Basemuncher Endonuclease (Expedeon
402 Ltd.) for 30 min on ice and then sonicated. Cleared supernatant was applied to a 5 ml HisTrap FF
403 crude column (GE Healthcare) and the bound, his-tagged protein was eluted with 125 mM

404 imidazole. Eluted protein was desalted on a HiPrep 26/10 desalting column (GE Healthcare) and
405 then further separated by size exclusion chromatography on a HiLoad 16/60 Superdex 200
406 preparative grade gel filtration column. All chromatography steps were carried out on an AKTA
407 Prime instrument (GE Healthcare). Purified proteins were concentrated in a Spin-X UF
408 Centrifugal Concentrator (Corning) and quantified by the nanodrop extinction co-efficient
409 method (Thermo Scientific). Samples were stored at -80 °C in binding buffer plus 50% glycerol.
410 MBP-tagged proteins were purified as above except the cells were induced with 1 mM IPTG,
411 MBP binding buffer was used (200 mM NaCl, 20 mM Tris, 1 mM EDTA; pH 7.4), the lysate
412 was applied to a 5 ml MBPTrap FF column (GE Healthcare) and purified protein was eluted with
413 10 mM maltose in binding buffer.

414 **Electrophoretic motility shift assays (EMSA).** For all 50 bp binding substrates, 50 base Cy5
415 5'-labelled oligos (IDT) were annealed to unlabelled complimentary oligos (IDT). Both oligos
416 were mixed to a final concentration of 40 μ M in annealing buffer (1 M Potassium Acetate, 300
417 mM HEPES; pH 7.5) and heated to 98°C for 5 min then allowed to cool to room temperature. 10
418 μ l EMSA mixtures contained 80 nM annealed Cy5-dsDNA, standard binding buffer (25 mM
419 HEPES, 50 mM K-glutamate, 50 mM MgSO₄, 1 mM dithiothreitol, 0.1 mM EDTA, 0.05%
420 Triton X-100; pH 8.0)⁴⁹ for all assays except those testing GtaR for which a modification of the
421 published buffer was used (10 mM HEPES, 40 mM NaCl; pH8)¹⁶, 1 μ g poly dI:dC, 4% glycerol
422 and the specified concentrations of purified protein⁵⁰. 500-fold excess of competitor DNA was
423 added to control mixtures – specific competitor was unlabelled but otherwise identical to the
424 binding substrate and the non-specific competitor was an unlabelled 50 bp annealed oligo
425 matching an arbitrary location elsewhere in the *R. capsulatus* genome. All assays except GtaR
426 were incubated for 15 min at 30°C then immediately loaded onto a 5 % Acrylamide gel (1 x

427 TBE) without loading dye. GtaR assays were incubated at 37°C for 30 minutes¹⁶. Gels were run
428 at 100 V for 1 h at room temperature in 1 x TBE. Fluorescence was imaged using a Typhoon
429 Biomolecular Imager (Amersham) and analysed using ImageQuant (Amersham) and FIJI⁵¹
430 software. For the RcGTA promoter (pGTA), a 5' Cy5 labelled oligo was used to create a 633 bp
431 PCR product. The pGTA DNA was used under the same conditions as the annealed oligos,
432 except the concentration was 2 ng μl^{-1} , reactions were run at 100 V for 4 h. Non-fluorescent
433 reactions used 100 ng of unlabelled PCR products as binding substrates and were run on 1% high
434 resolution MicroSieve 3:1 Agarose (Cambridge Reagents) in 1 x TBE at 100 V for 2 h. Gels
435 were stained with Sybr Safe (Invitrogen) and imaged on a GelDoc transilluminator (BioRad).

436 **Ruegeria/Roseovarius Gene Transfer Assays.** Assays were carried out as originally reported in
437 Biers et al. (2008)⁵². In brief, spontaneous rifampicin or streptomycin resistant colonies were
438 isolated by plating onto selective agar. Cultures were grown in $\frac{1}{2}$ YTSS medium for 5 days, static
439 and without illumination. For co-culture experiments, a rifampicin resistant strain was grown
440 together with a streptomycin resistant strain then plated on marine broth agar with both
441 antibiotics to assess transfer of resistance. For in vitro assays, resistant strains were grown
442 separately for 5 days and filtered through a 0.45 μm syringe filter. The filtered supernatant was
443 then added to antibiotic sensitive cells, shaken at 200 rpm for 1h in the dark and plated on marine
444 broth agar containing the relevant antibiotics. The gafA homologues were cloned into pSRKBB
445 to produce pCMF195 & 6 (Supplementary Table 5); gafA expression was induced from the lac'
446 promoter by addition of 1 mM IPTG when growth had reached late logarithmic phase (OD_{600} :
447 ~0.8-1.0).

448 **Bioinformatics.** Helix turn helix predictions were carried out using NPS@^{53,54} and Gym2.0⁵⁵
449 using the default settings. HHPRED^{56,57} analysis of GafA was carried out using the

450 pdb_mmcif70_5_oct database and the default parameters i.e. HHBlits uniprot20_2016_02 MSA
451 generation method, maximal generation steps = 3 and an E-value threshold of 1e-3. Minimum
452 coverage was 20%, minimum sequence identity was 0%. Secondary structure scoring was done
453 during alignment (local). Initial full length protein query was refined and resubmitted according
454 to the automatic suggestions provided by the software for the two respective domains. The NCBI
455 BlastP search for GafA homologues was performed with the default parameters - expect
456 threshold=10, word size=6, blosum62 similarity matrix, gap costs of existence=11 and
457 extension=1. No taxonomic constraints were applied but sequences from
458 uncultured/environmental samples. The top ten hits belonging to different species were
459 arbitrarily selected for analysis irrespective of alignment score, the most distant match used
460 (*Sulfitobacter* spp.) produced a score of 377 and an E-value of 6e-126 from 100% coverage and
461 55% sequence identity. Promoter sequences for each protein were then identified in the
462 nucleotide database for each sequence. Promoter -10/-35 elements were predicted with BPROM
463 ⁵⁸. FIJI software ⁵¹ was used to measure band intensities in EMSA experiments with the Gel
464 Analyzer plug in, ClustalW2 ⁵⁹ and ClustalΩ⁶⁰ were used for DNA/protein alignments as
465 indicated in the figure legends, Jalview⁶¹ was used to visualize alignments. Transcript abundance
466 was visualized using the Broad Institute's IGV viewer⁶². Statistical analysis was carried out
467 using Sigmaplot software version 13 (Systat Software Inc., www.systatsoftware.com.) and, for
468 each use, the test parameters are indicated in the text and/or figure legends. The Sigmaplot
469 Ligand Binding macro was also used to calculate dissociation constants (kD) from EMSA band
470 intensities.

471 **Data and materials availability:** All data needed to evaluate the conclusions of the paper are
472 present in the paper and the supplementary information file. Source data for all graphs and gel

473 images are provided as a Source Data file. The complete RNAseq data was submitted to the
474 NCBI Gene Expression Omnibus (GEO) Database, accession number GSE118116
475 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE118116>]. All bacterial strains or
476 genetic constructs are securely stored locally and are available on request

477 **References:**

- 478 1. Soucy, S. M., Huang, J. & Gogarten, J. P. Horizontal gene transfer: building the web of life.
479 *Nat Rev Genet* **16**, 472–482 (2015).
- 480 2. Koonin, E. V. & Wolf, Y. I. Genomics of bacteria and archaea: the emerging dynamic view
481 of the prokaryotic world. *Nucleic Acids Res* **36**, 6688–6719 (2008).
- 482 3. Chen, J. et al. Genome hypermobility by lateral transduction. *Science* **362**, 207–212 (2018).
- 483 4. Lang, A. S., Zhaxybayeva, O. & Beatty, J. T. Gene transfer agents: phage-like elements of
484 genetic exchange. *Nat Rev Microbiol* **10**, 472–482 (2012).
- 485 5. Solioz, M. & Marrs, B. The gene transfer agent of *Rhodospseudomonas capsulata*.
486 Purification and characterization of its nucleic acid. *Arch Biochem Biophys* **181**, 300–307
487 (1977).
- 488 6. Hynes, A. P., Mercer, R. G., Watton, D. E., Buckley, C. B. & Lang, A. S. DNA packaging
489 bias and differential expression of gene transfer agent genes within a population during
490 production and release of the *Rhodobacter capsulatus* gene transfer agent, RcGTA. *Mol*
491 *Microbiol* **85**, 314–325 (2012).
- 492 7. Lang, A. S., Westbye, A. B. & Beatty, J. T. The distribution, evolution, and roles of gene
493 transfer agents in prokaryotic genetic exchange. *Annual review of virology* **4**, 87–104
494 (2017).
- 495 8. Redfield, R. J. & Soucy, S. M. Evolution of bacterial gene transfer agents. *Front Microbiol*
496 **9**, 2527 (2018).
- 497 9. Suttle, C. A. Viruses in the sea. *Nature* **437**, 356–361 (2005).
- 498 10. Kristensen, D. M., Mushegian, A. R., Dolja, V. V. & Koonin, E. V. New dimensions of the
499 virus world discovered through metagenomics. *Trends Microbiol* **18**, 11–19 (2010).
- 500 11. Angly, F. E. et al. The marine viromes of four oceanic regions. *PLoS Biol* **4**, e368 (2006).
- 501 12. McDaniel, L. D. et al. High frequency of horizontal gene transfer in the oceans. *Science*
502 **330**, 50 (2010).
- 503 13. Mercer, R. G. et al. Loss of the response regulator CtrA causes pleiotropic effects on gene
504 expression but does not affect growth phase regulation in *Rhodobacter capsulatus*. *J*
505 *Bacteriol* **192**, 2701–2710 (2010).
- 506 14. Lang, A. S. & Beatty, J. T. Genetic analysis of a bacterial genetic exchange element: the
507 gene transfer agent of *Rhodobacter capsulatus*. *Proc Natl Acad Sci U S A* **97**, 859–864
508 (2000).
- 509 15. Westbye, A. B. et al. The protease ClpXP and the PAS-domain protein DivL regulate CtrA
510 and gene transfer agent production in *Rhodobacter capsulatus*. *Appl Environ Microbiol*
511 (2018). doi:10.1128/AEM.00275-18
- 512 16. Leung, M. M., Brimacombe, C. A., Spiegelman, G. B. & Beatty, J. T. The GtaR protein
513 negatively regulates transcription of the gtaRI operon and modulates gene transfer agent
514 (RcGTA) expression in *Rhodobacter capsulatus*. *Mol Microbiol* **83**, 759–774 (2012).

- 515 17. Brimacombe, C. A. et al. Quorum-sensing regulation of a capsular polysaccharide receptor
516 for the *Rhodobacter capsulatus* gene transfer agent (RcGTA). *Mol Microbiol* **87**, 802–817
517 (2013).
- 518 18. Westbye, A. B. et al. Phosphate concentration and the putative sensor kinase protein CckA
519 modulate cell lysis and release of the *Rhodobacter capsulatus* gene transfer agent. *J*
520 *Bacteriol* **195**, 5025–5040 (2013).
- 521 19. Fogg, P. C. M., Westbye, A. B. & Beatty, J. T. One for all or all for one: heterogeneous
522 expression and host cell lysis are key to gene transfer agent activity in *Rhodobacter*
523 *capsulatus*. *PLoS ONE* **7**, e43772 (2012).
- 524 20. Westbye, A. B., Kuchinski, K., Yip, C. K. & Beatty, J. T. The Gene Transfer Agent RcGTA
525 Contains Head Spikes Needed for Binding to the *Rhodobacter capsulatus* Polysaccharide
526 Cell Capsule. *J Mol Biol* **428**, 477–491 (2016).
- 527 21. Chen, F. et al. Proteomic analysis and identification of the structural and regulatory proteins
528 of the *Rhodobacter capsulatus* gene transfer agent. *J Proteome Res* **8**, 967–973 (2009).
- 529 22. Hynes, A. P. et al. Functional and evolutionary characterization of a gene transfer agent’s
530 multilocus “genome”. *Mol Biol Evol* **33**, 2530–2543 (2016).
- 531 23. Matson, E. G., Thompson, M. G., Humphrey, S. B., Zuerner, R. L. & Stanton, T. B.
532 Identification of genes of VSH-1, a prophage-like gene transfer agent of *Brachyspira*
533 *hyodysenteriae*. *J Bacteriol* **187**, 5885–5892 (2005).
- 534 24. Mercer, R. G. & Lang, A. S. Identification of a predicted partner-switching system that
535 affects production of the gene transfer agent RcGTA and stationary phase viability in
536 *Rhodobacter capsulatus*. *BMC Microbiol* **14**, 71 (2014).
- 537 25. Reisenauer, A., Quon, K. & Shapiro, L. The CtrA response regulator mediates temporal
538 control of gene expression during the *Caulobacter* cell cycle. *J Bacteriol* **181**, 2430–2439
539 (1999).
- 540 26. Pini, F. et al. Cell Cycle Control by the Master Regulator CtrA in *Sinorhizobium meliloti*.
541 *PLoS Genet* **11**, e1005232 (2015).
- 542 27. Kuchinski, K. S., Brimacombe, C. A., Westbye, A. B., Ding, H. & Beatty, J. T. The SOS
543 Response Master Regulator LexA Regulates the Gene Transfer Agent of *Rhodobacter*
544 *capsulatus* and Represses Transcription of the Signal Transduction Protein CckA. *J*
545 *Bacteriol* **198**, 1137–1148 (2016).
- 546 28. Laub, M. T., Chen, S. L., Shapiro, L. & McAdams, H. H. Genes directly controlled by CtrA,
547 a master regulator of the *Caulobacter* cell cycle. *Proc Natl Acad Sci U S A* **99**, 4632–4637
548 (2002).
- 549 29. Marrs, B. Genetic recombination in *Rhodopseudomonas capsulata*. *Proc Natl Acad Sci U S*
550 *A* **71**, 971–973 (1974).
- 551 30. Brimacombe, C. A., Ding, H., Johnson, J. A. & Beatty, J. T. Homologues of Genetic
552 Transformation DNA Import Genes Are Required for *Rhodobacter capsulatus* Gene
553 Transfer Agent Recipient Capability Regulated by the Response Regulator CtrA. *J Bacteriol*
554 **197**, 2653–2663 (2015).

- 555 31. Schaefer, A. L., Taylor, T. A., Beatty, J. T. & Greenberg, E. P. Long-chain acyl-homoserine
556 lactone quorum-sensing regulation of *Rhodobacter capsulatus* gene transfer agent
557 production. *J Bacteriol* **184**, 6515–6521 (2002).
- 558 32. Grote, J., Krysciak, D. & Streit, W. R. Phenotypic heterogeneity, a phenomenon that may
559 explain why quorum sensing does not always result in truly homogenous cell behavior. *Appl*
560 *Environ Microbiol* **81**, 5280–5289 (2015).
- 561 33. Pérez, P. D. & Hagen, S. J. Heterogeneous response to a quorum-sensing signal in the
562 luminescence of individual *Vibrio fischeri*. *PLoS ONE* **5**, e15473 (2010).
- 563 34. Anetzberger, C., Pirsch, T. & Jung, K. Heterogeneity in quorum sensing-regulated
564 bioluminescence of *Vibrio harveyi*. *Mol Microbiol* **73**, 267–277 (2009).
- 565 35. Québatte, M. et al. Gene Transfer Agent Promotes Evolvability within the Fittest
566 Subpopulation of a Bacterial Pathogen. *Cell systems* **4**, 611–621.e6 (2017).
- 567 36. Westbye, A. B., O’Neill, Z., Schellenberg-Beaver, T. & Beatty, J. T. The *Rhodobacter*
568 *capsulatus* gene transfer agent is induced by nutrient depletion and the RNAP omega
569 subunit. *Microbiology (Reading, Engl)* **163**, 1355–1363 (2017).
- 570 37. Shakya, M., Soucy, S. M. & Zhaxybayeva, O. Insights into origin and evolution of α -
571 proteobacterial gene transfer agents. *Virus evolution* **3**, vex036 (2017).
- 572 38. Huang, S., Zhang, Y., Chen, F. & Jiao, N. Complete genome sequence of a marine
573 roseophage provides evidence into the evolution of gene transfer agents in
574 alphaproteobacteria. *Virol J* **8**, 124 (2011).
- 575 39. Wall, J. D., Weaver, P. F. & Gest, H. Gene transfer agents, bacteriophages, and bacteriocins
576 of *Rhodopseudomonas capsulata*. *Arch Microbiol* **105**, 217–224 (1975).
- 577 40. Ding, H., Moksa, M. M., Hirst, M. & Beatty, J. T. Draft Genome Sequences of Six
578 *Rhodobacter capsulatus* Strains, YW1, YW2, B6, Y262, R121, and DE442. *Genome*
579 *Announc* **2**, (2014).
- 580 41. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual*. Cold
581 Spring Harbor laboratory press. 931–957 (1982).
- 582 42. Matson, E. G., Zuerner, R. L. & Stanton, T. B. Induction and transcription of VSH-1, a
583 prophage-like gene transfer agent of *Brachyspira hyodysenteriae*. *Anaerobe* **13**, 89–97
584 (2007).
- 585 43. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence
586 Data. at <<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>>
- 587 44. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
588 *EMBnet.journal* **17**, 10 (2011).
- 589 45. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq
590 quantification. *Nat Biotechnol* **34**, 525–527 (2016).
- 591 46. Pimentel, H., Bray, N. L., Puente, S., Melsted, P. & Pachter, L. Differential analysis of
592 RNA-seq incorporating quantification uncertainty. *Nat Methods* **14**, 687–690 (2017).
- 593 47. Leung, M. & Beatty, J. *Rhodobacter capsulatus* Gene Transfer Agent Transduction Assay.
594 *Bio-protocol* **3**, (2013).

- 595 48. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time
596 quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408 (2001).
- 597 49. Wiethaus, J., Schubert, B., Pfänder, Y., Narberhaus, F. & Masepohl, B. The GntR-like
598 regulator TauR activates expression of taurine utilization genes in *Rhodobacter capsulatus*. *J*
599 *Bacteriol* **190**, 487–493 (2008).
- 600 50. Wiethaus, J., Wirsing, A., Narberhaus, F. & Masepohl, B. Overlapping and specialized
601 functions of the molybdenum-dependent regulators MopA and MopB in *Rhodobacter*
602 *capsulatus*. *J Bacteriol* **188**, 8441–8451 (2006).
- 603 51. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods*
604 **9**, 676–682 (2012).
- 605 52. Biers, E. J. et al. Occurrence and expression of gene transfer agent genes in marine
606 bacterioplankton. *Appl Environ Microbiol* **74**, 2933–2939 (2008).
- 607 53. Dodd, I. B. & Egan, J. B. Improved detection of helix-turn-helix DNA-binding motifs in
608 protein sequences. *Nucleic Acids Res* **18**, 5019–5026 (1990).
- 609 54. Combet, C., Blanchet, C., Geourjon, C. & Deléage, G. NPS@: network protein sequence
610 analysis. *Trends Biochem Sci* **25**, 147–150 (2000).
- 611 55. Narasimhan, G. et al. Mining protein sequences for motifs. *J Comput Biol* **9**, 707–720
612 (2002).
- 613 56. Hildebrand, A., Remmert, M., Biegert, A. & Söding, J. Fast and accurate automatic structure
614 prediction with HHpred. *Proteins* **77 Suppl 9**, 128–132 (2009).
- 615 57. Zimmermann, L. et al. A Completely Reimplemented MPI Bioinformatics Toolkit with a
616 New HHpred Server at its Core. *J Mol Biol* **430**, 2237–2243 (2018).
- 617 58. Solovyev, V. & Salamov, A. Automatic Annotation of Microbial Genomes and
618 Metagenomic Sequences. In *Metagenomics and its Applications in Agriculture, Biomedicine*
619 *and Environmental Studies* (Ed. R.W. Li), Nova Science Publishers, p.61-78. 61–78 (2011).
- 620 59. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948
621 (2007).
- 622 60. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence
623 alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).
- 624 61. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview
625 Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**,
626 1189–1191 (2009).
- 627 62. Robinson, J. T. et al. Integrative genomics viewer. *Nat Biotechnol* **29**, 24–26 (2011).
- 628
- 629

630 **Supplementary Information:**

631 Supplementary Tables 1-5

632 Supplementary Figures 1-8

633 Supplementary References (1-10)

634 **Acknowledgments:** I would like to thank the University of York Technology Facility for
635 providing access to equipment and expert technical assistance when required; in particular I
636 would like to acknowledge Dr Katherine Newling for RNAseq quality control and statistical
637 analysis. I also thank Dr Jelena Kusakina for critical reading of the manuscript. **Funding:** This
638 work was wholly supported by a Wellcome Trust/Royal Society Sir Henry Dale Fellowship
639 Grant (109363/Z/15/Z).

640

641 **Authors:** Paul C.M. Fogg.

642 **Affiliations:** University of York, Biology Department, Wentworth Way, York, United Kingdom.
643 YO10 5DD

644 **Correspondence to:** paul.fogg@york.ac.uk

645 **Author contributions:** P.C.M.F conceived, designed and implemented this study and prepared
646 the manuscript. **Competing interests:** The author declares no competing interests.

647 **Figure 1. Confirmation of the RcGTA Activator, GafA.** **A**, GTA gene transfer assays for *R.*
648 *capsulatus* SB1003 (WT), SB1003 *gafA* overexpressor (*gafA* OX), RcGTA hyperproducer strain *R.*
649 *capsulatus* DE442 (DE442) and DE442 with *gafA* deleted (DE442 *gafA* Δ). Individual replicates
650 are shown as diamonds. All conditions were significantly different; One Way ANOVA
651 significance is indicated above the bars ($n=8$, three asterisks = $p<0.001$). **B**, Agarose gels of total
652 DNA isolated from the annotated *R. capsulatus* strains - RcGTA hyperproducer strain *R.*
653 *capsulatus* DE442, *ctrA* (*ctrA* Δ) and *gafA* (*gafA* Δ) knock-outs in DE442, wild-type *R. capsulatus*
654 SB1003 compared to *gafA* overexpressor (OX) derivatives of SB1003. Time post induction of
655 *gafA* is noted in hours, GTA and genomic DNA (gDNA) are indicated by labelled arrows. NEB 1
656 kb Extend DNA Ladder (M1) or Bioline HyperLadder 1 kb DNA ladder were used (M2); the 4 kb
657 band is annotated with a white arrow head. Source data are provided as a Source Data file.

658 **Figure 2. The role of CtrA in RcGTA production.** **A**, GTA gene transfer assays for *R.*
659 *capsulatus* SB1003, *ctrA* overexpressor (*ctrA* OX), non-phosphorylatable *ctrA* overexpressor
660 (D51A) and phosphomimetic *ctrA* overexpressor (D51E OX). Individual replicates are shown as
661 diamonds ($n=3$), One Way ANOVA significance versus the control (SB1003) is indicated above
662 the chart (n.s. = not significant i.e. $p>0.05$, three asterisks = $p<0.001$). **B**, Agarose gels of total
663 DNA isolated from *R. capsulatus* SB1003 and the annotated derivatives - wild-type *R.*
664 *capsulatus* SB1003, *ctrA* knock-out (*ctrA* Δ), *ctrA* overexpressor (*ctrA* OX), phosphomimetic
665 *ctrA* overexpressor (D51E OX), non-phosphorylatable *ctrA* overexpressor (D51A OX) and a
666 *gafA* overexpressor in a *ctrA* knock-out background (*ctrA* Δ , *gafA* OX). GTA and genomic DNA
667 (gDNA) are indicated by labelled arrows. Bioline HyperLadder 1 kb DNA ladder was used
668 (M2); the 4 kb band is annotated with a white arrow head. Source data are provided as a Source
669 Data file.

670 **Figure 3. Relative Transcription of RcGTA-Related Genes.** The *R. capsulatus* strains and
671 gene targets assessed are annotated on each graph. OX indicates a gene overexpressor and Δ is a
672 gene knock-out. All Y-axis fold expression changes are normalized using *uvrD* as an endogenous
673 reference gene (Δ Ct) and relative to the wild-type SB1003 strain ($\Delta\Delta$ Ct). Dot plots of individual
674 replicates are overlaid onto each bar (biological replicates, $n\geq 3$ for all samples). Statistical
675 significance was determined using a two-tail t-test (one asterisk = $p<0.05$, two asterisks =
676 $p<0.01$, three asterisks = $p<0.001$, hash = transcript not detected in knock-out lines, n.s. = not
677 significant i.e. $p>0.05$). Total transcripts were measured in **A**, **B** & **D** and transcripts originating
678 from the native promoter only in **C** & **E**. Source data are provided as a Source Data file.

679 **Figure 4. CtrA binding to the *gafA* promoter.** **A**, Alignment of the DNA probe sequences
680 containing CtrA binding sites that were used for EMSAs (double headed arrow). CtrA half sites
681 are represented by solid lines and the spacer sequence as a dashed line and the predicted Shine
682 Delgarno -10 site is annotated. mRNA transcript coverage for the *gafA* promoter, obtained from
683 RNAseq data, is shown as a histogram above the alignment. **B**, EMSA band shift of Cy5-labelled
684 *gafA* promoter DNA incubated with the protein concentrations specified. The lane labelled N
685 contained 500-fold excess of an unlabelled non-specific competitor and S contained 500-fold
686 excess of an unlabelled specific competitor. **C**, Quantification of two independent band shifts of
687 CtrA vs the *gafA* promoter. Error bars are standard deviation, $n=2$. Source data are provided as a
688 Source Data file.

689 **Figure 5. Binding of the GtaR quorum sensing protein to the *gafA* promoter.** EMSA band
690 shift of Cy5-labelled *gafA* promoter DNA (see Figure 4A) incubated with the protein
691 concentrations specified. The lane labelled N contained 500-fold excess of an unlabelled non-

692 specific competitor and S contained 500-fold excess of an unlabelled specific competitor. Source
693 data are provided as a Source Data file.

694 **Figure 6. GafA binding to the RcGTA cluster promoter.** **A**, Map of the RcGTA structural
695 gene cluster promoter indicating the predicted locations of the Shine Delgarno -10 and -35 sites,
696 the ribosome binding site (RBS), RcGTA g1 start codon and transcription start site (TSS).
697 mRNA transcript coverage, obtained from RNAseq data, is shown as a histogram. **B**, Map of the
698 overlapping 50 bp regions of the RcGTA promoter used as EMSA probes (pGTA1-5). **C**, EMSA
699 band shifts of Cy5-labelled pGTA1-5 versus 2 μ M GafA protein. **D**, EMSA band shift of titrated
700 GafA protein at the concentrations indicated versus Cy5-pGTA2. The lane labelled N contained
701 500-fold excess of an unlabelled non-specific competitor and S contained 500-fold excess of an
702 unlabelled specific competitor. **E**, Unshifted Cy5 labelled, 633bp RcGTA promoter DNA after
703 incubation with up to 4 μ M of either CtrA or GafA. Source data are provided as a Source Data
704 file.

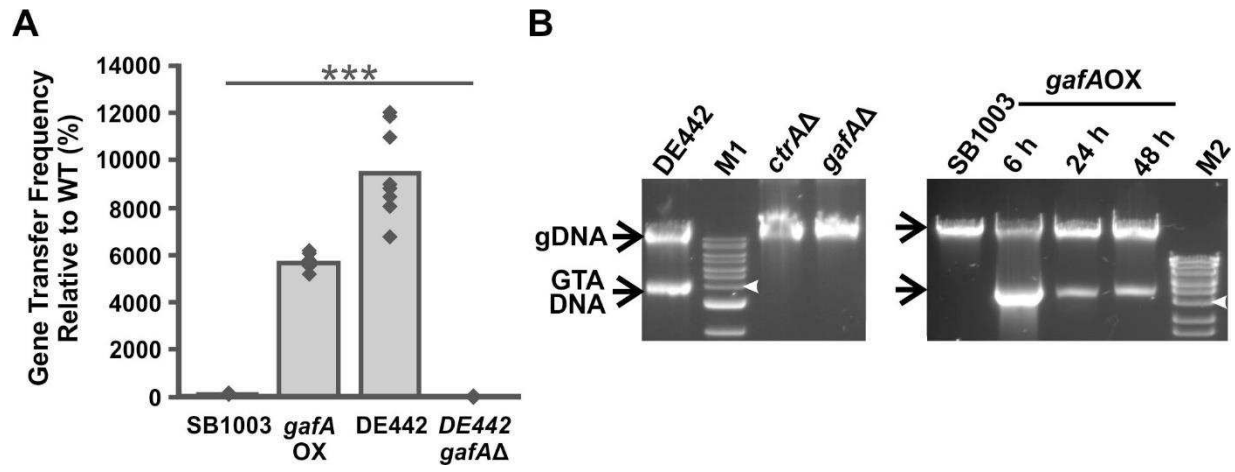
705 **Figure 7. Model of RcGTA regulation.** The interactions depicted are inferred from the data in
706 this study, raw microarray data ¹⁶ and published results ^{18,19,25,30}. Bent, perpendicular arrows
707 represent promoters and are annotated with the preceding gene name. CtrA (*) or GtaR (^)
708 binding sites are labelled where present. Proteins are depicted as coloured ellipses with
709 phosphate groups (P) in orange circles. Solid arrows indicate direct regulation, dashed arrows
710 indicate indirect or unknown route of regulation and emboldened arrows indicate that the
711 regulator is essential for target expression.

712

713 Figure 1

714

715

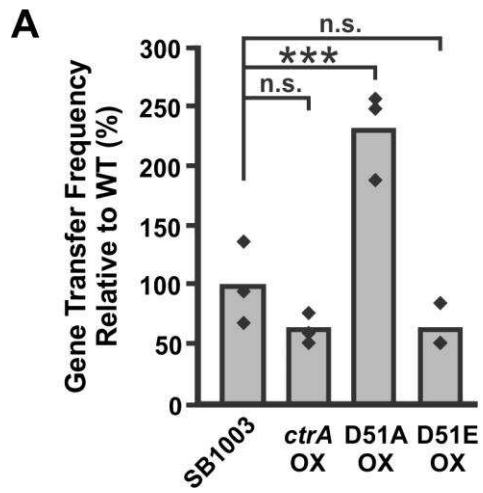


716

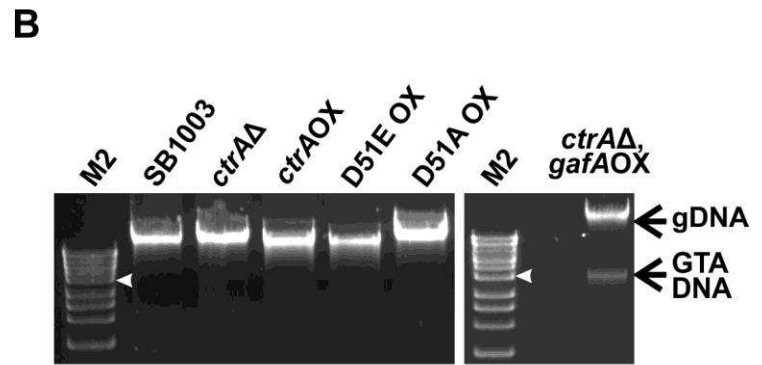
717 Figure 2

718

719



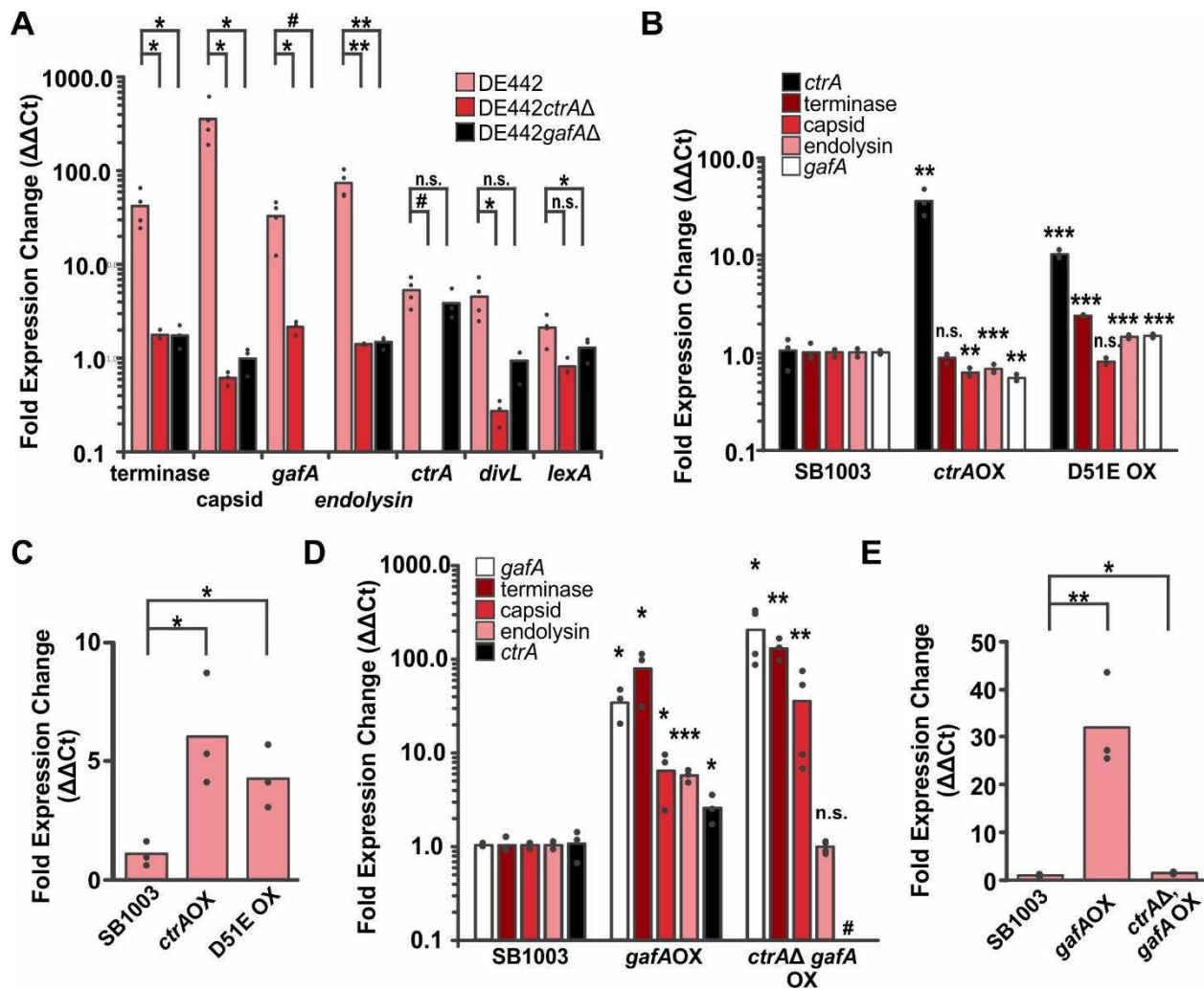
720



721 Figure 3

722

723

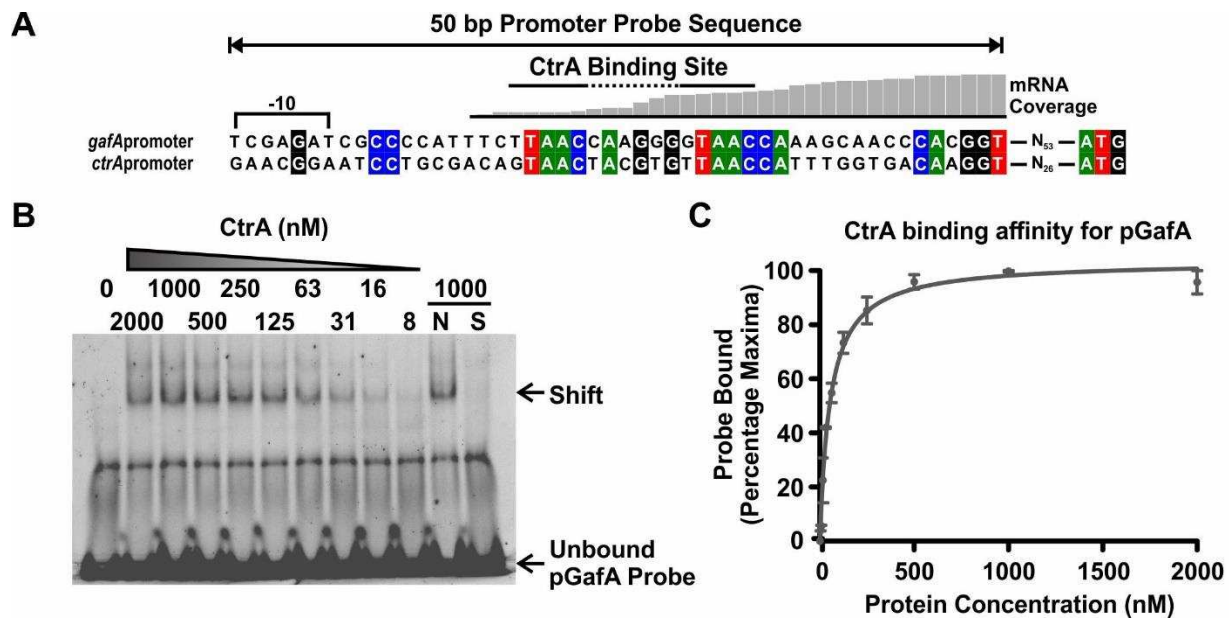


724

725 Figure 4

726

727

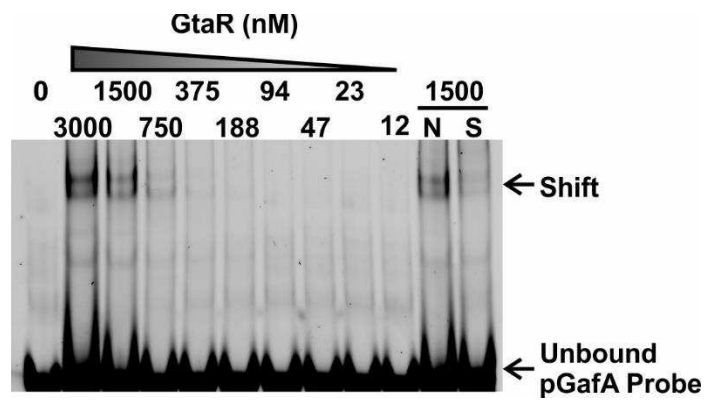


728

729 Figure 5

730

731

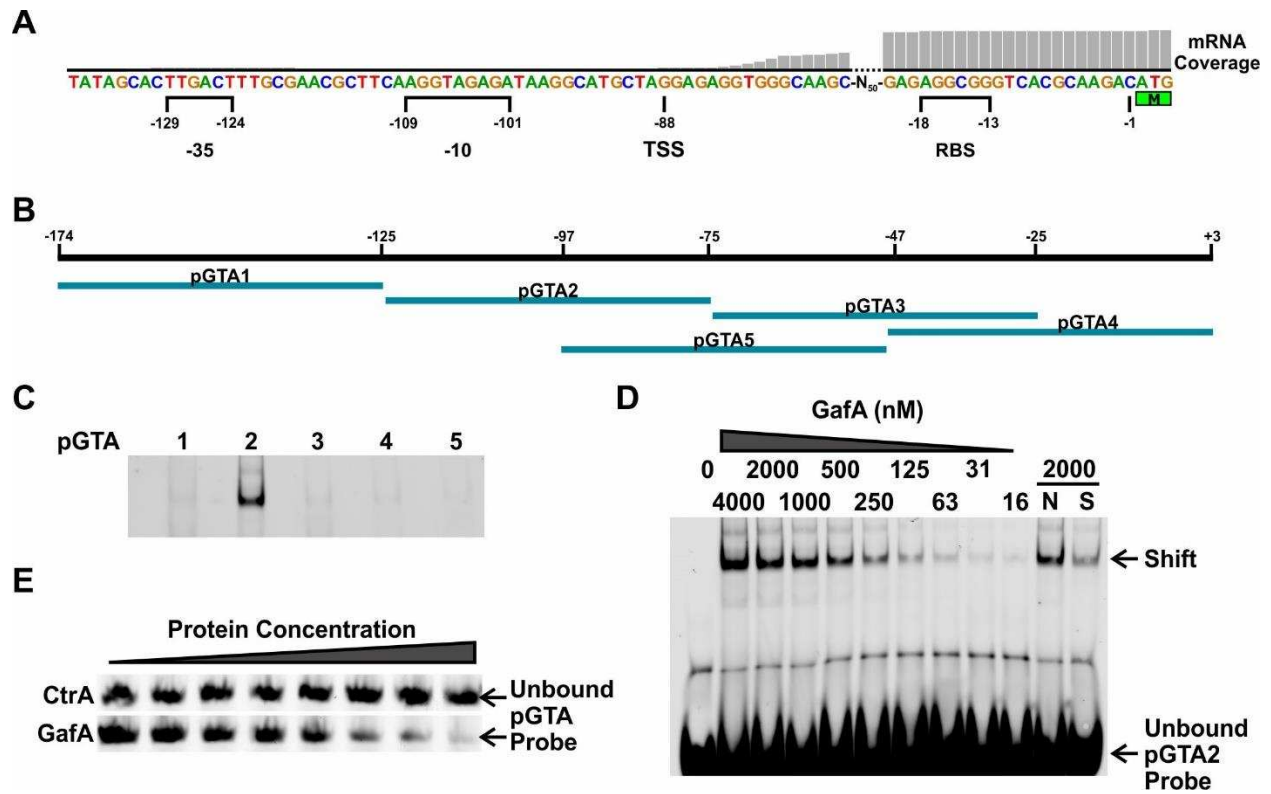


732

733 Figure 6

734

735

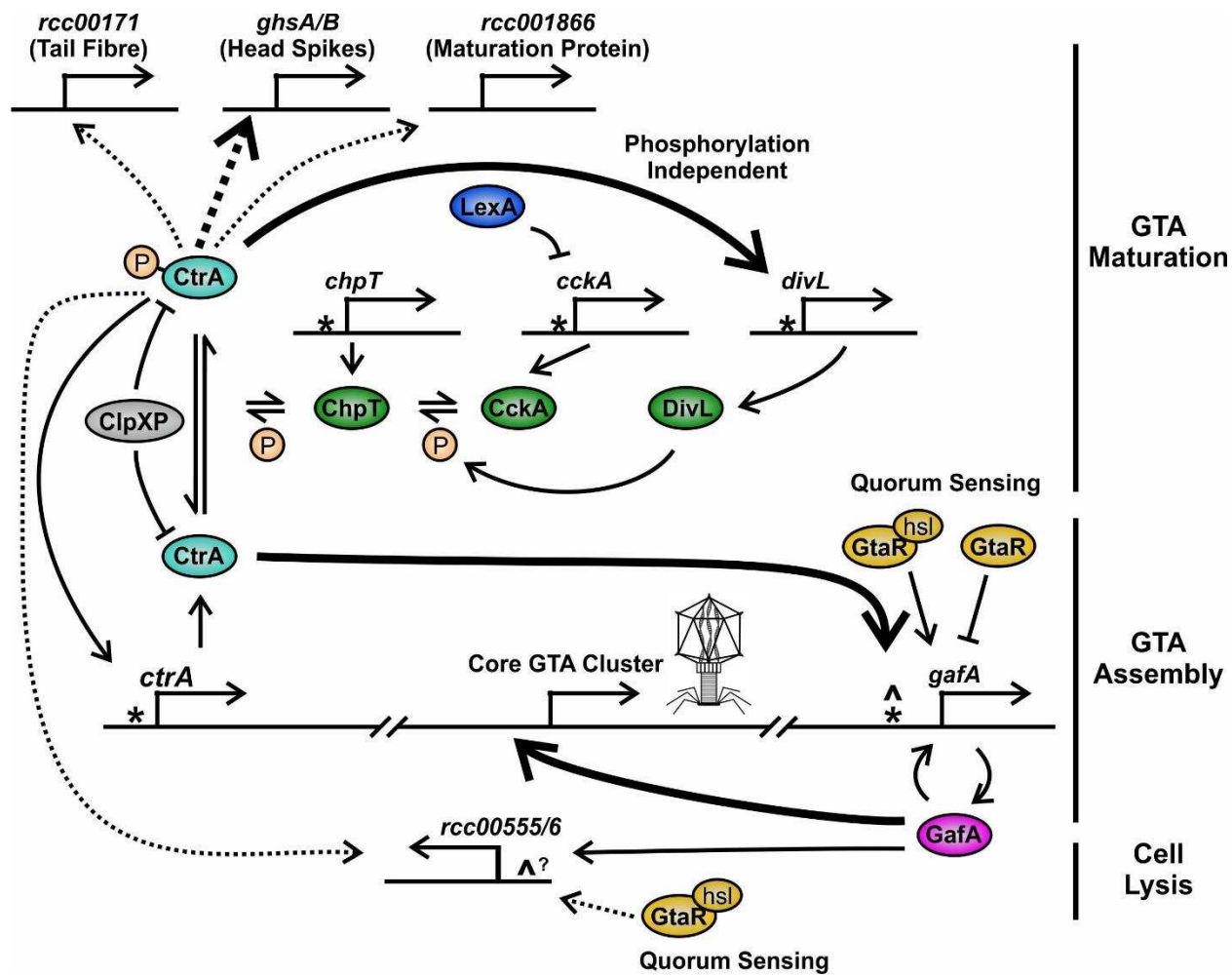


736

737 Figure 7

738

739



740