

ΦCrAss001, a member of the most abundant bacteriophage family in the human gut, infects *Bacteroides*

Andrey N. Shkoporov, Ekaterina V. Khokhlova, C. Brian Fitzgerald, Stephen R. Stockdale,
Lorraine A. Draper, R. Paul Ross, Colin Hill
APC Microbiome Ireland, University College Cork, Cork, Ireland

Abstract

ΦCrAss001, isolated from human faecal material, is the first member of the extensive crAssphage family to be grown in pure culture. The bacteriophage infects the human gut symbiont *Bacteroides intestinalis*, confirming *in silico* predictions of the likely host. Genome analysis demonstrated that the phage DNA is 102 kb in size, has an unusual genome organisation and does not possess any obvious genes for lysogeny. In addition, electron microscopy confirms that φcrAss001 has a podovirus-like morphology. Despite the absence of lysogeny genes, φcrAss001 replicates in a way that does not disrupt proliferation of the host bacterium and is able to maintain itself in continuous host culture.

Keywords: crAssphage, crAss-like phages, human gut virome, bacteriophage, *Bacteroides*.

The human gut virome contains a vast number of bacterial, mammalian, plant, fungal and archaeal viruses¹⁻³. There is a growing body of evidence supporting specific and consistent alterations of the gut virome in a number of human diseases and conditions, including inflammatory bowel disease (IBD), malnutrition, and AIDS⁴⁻⁶. Currently, our understanding of the physiological significance of the human gut virome is limited by the fact that the vast majority of these viruses

cannot be taxonomically classified or linked to any particular hosts (so-called “viral dark matter”⁷). Several recent studies identified a highly abundant (representing up to 90% of gut viruses in some individuals), unique uncultured group termed crAss-like phages or crAssphage^{8–11}. Here we report on the first successful isolation of a crAssphage on a single host and describe its key biological properties. Based on electron microscopy of propagated crAssphage001 morphology we confirm that crAss-like phages possess podovirus-like morphology. We also demonstrate the ability of the phage to stably co-replicate with its *Bacteroides intestinalis* host in equilibrium for many generations *in vitro*, which mimics earlier observed ability of crAss-like phages to maintain stable colonization of the mammalian gut^{11,12}.

In 2014, a study by Dutilh *et al.*⁸ demonstrated the presence in ~50% of human samples of highly abundant sequences, which when cross-assembled from multiple sources, indicated the presence of a unique bacteriophage they named crAssphage (cross Assembly). The fully assembled 97 kb DNA genome showed no homology to any known virus, even though in some subjects it completely dominated the gut virome (up to 90% of sequencing reads). Indirect evidence suggested that *Bacteroides* species is a likely host⁸. A later study, based on detailed sequence analysis of proteins encoded by this phage, predicted *Podoviridae*-like morphology and placed it into the novel diverse and expansive family-level phylogenetic group of loosely related bacteriophages termed crAss-like phages⁹. Genomes of members of this group are highly abundant in the human gut and are also present in other diverse habitats including termite gut, terrestrial/groundwater, and oceans^{9,13,14}. However, in the absence of a known host, no member of this family has been isolated and nothing is known of the biological properties of these crAss-like phages from the human gut.

In a recent analysis¹¹, we *de novo* assembled 244 genomes of crAss-like phages from the human gut and classified them into several genus- and subfamily-level taxonomic groups based on percentage of shared orthologous genes. As a result of this study, we can state that 98-100% of

51 healthy adults from Western cohorts carry at least one or more types of crAss-like phages, albeit
52 with widely varying relative abundance. Here, we report the first documented isolation in culture of
53 a crAss-like phage from the human gut virome and describe its key biological properties.

54 The replication strategy of crAss-like bacteriophages is unknown and all attempts to use
55 standard plaque assays on semi-solid agar have failed. We therefore attempted to detect crAss-like
56 phage replication using a broth enrichment strategy. Phage-enriched filtrates of faecal samples were
57 collected from 20 healthy adult Irish volunteers, pooled and used to infect pure cultures of 53
58 bacterial strains representing the commensal human gut microbiota (Supplementary Table 1). After
59 three successive rounds of enrichment, cell free supernatants were subjected to shotgun
60 metagenomic sequencing. Analysis of the assembled sequencing reads demonstrated that the
61 supernatant from strain *Bacteroides intestinalis* 919/174 was dominated by a single 102.7 kb contig
62 (~98% of reads) related to a known but previously uncultured crAss-like bacteriophage, IAS virus,
63 isolated from the human gut^{9,15}. All other species of *Bacteroides* tested were insensitive to
64 ϕ crAss001.

65 The genome of bacteriophage ϕ crAss001 is 102,679 bp and is circular or circularly permuted.
66 We identified 105 protein coding genes (ORFs of length 90-7,239 bp) and 25 tRNA genes specific
67 for 17 different amino acids. Cumulative G+C content of the phage genome is 34.7%, which is
68 significantly lower than that of published *B. intestinalis* genomes (42.8-43.5%). An interesting
69 structural feature of the genome is its apparent division into two parts of roughly equal size with
70 strictly opposite gene orientation and inverted GC skew, possibly reflecting the direction of
71 transcription and/or replication (Figure 1). Functional gene annotation was performed using a
72 comprehensive approach, which included BLASTp amino acid sequence homology searches against
73 NCBI nt database, hidden Markov model (HMM) searches against UniProtKB/TrEMBL database¹⁶
74 and profile-profile HMM searches with HHpred against PDB, PFAM, NCBI-CDD and TIGRFAM
75 databases^{17,18}. This allowed for the functional annotation of 57 genes and assignment of further 11

76 genes to conserved protein families with unknown functions. Ten identifiable structural protein
 77 genes (phage head, tail, appendages) as well as three genes responsible for lytic functions were
 78 clustered on the right-hand side of the genome, suggesting that the remaining un-annotated genes in
 79 this part of the genome may also be responsible for structure and assembly of phage particles, as
 80 well as cell lysis. By contrast, the left-hand side predominantly harboured genes involved in
 81 replication, recombination, transcription and nucleotide metabolism. Putative DNA-binding and
 82 transcriptional regulation proteins were located in the proximal portions of the two oppositely
 83 oriented genome halves, suggesting their role in governing the transcription of gene modules
 84 located downstream on both sides of the genome (Fig. 1). Two genes (3, β -fructosidase and 5,
 85 ferredoxin-thioredoxin reductase) were predicted to be involved in auxiliary metabolic processes, in
 86 that they are unrelated with phage replication and virion assembly. No identifiable lysogeny
 87 module, or integrase or recombinase genes were identified. MALDI-TOF analysis of virion proteins
 88 separated by SDS-PAGE identified presence of most of the predicted structural proteins.
 89 Unexpectedly, three high molecular weight subunits of phage RNA polymerase⁹ (gene products 47,
 90 49 and 50) were also detected as part of virion structure (Fig. 2a).

91 TEM (transmission electron microscopy) of ϕ crAss001 virions revealed a podoviral
 92 morphology (Fig. 2b). Phage heads are isometric with a diameter of 77.2 ± 3.3 nm (mean \pm SD). Tails
 93 are 36.1 ± 3.6 nm long with elaborate structural features and several side appendages of variable
 94 length. This is in agreement with our previous TEM observations of podoviruses of similar
 95 dimensions (~ 76.5 nm in diameter) from a faecal sample rich in a mixture of several highly
 96 prevalent crAss-like bacteriophages¹¹.

97 According to the recently proposed classification scheme based on functional gene repertoire
 98 and protein sequence homology⁹, ϕ crAss001 fits into the IAS virus subgroup¹⁵ of crAss-like
 99 bacteriophages (Fig. 2c). Our recent analysis identified a common presence of similar uncultured
 100 bacteriophages in the gut microbiotas of healthy Irish and US adults, Irish elderly people, as well as

101 in healthy Irish infants and healthy and malnourished Malawian infants¹¹, where they were
102 detectable in 62%, 44%, 50%, 10%, 14% and 16% of cases, respectively, and in some case
103 represented up to 61% of virome reads (Fig. S1).

104 When genomes of ϕ crAss001, IAS virus and other related putative phages were compared,
105 most of the sequence variability (including variation in the number and size of ORFs) was
106 concentrated in a region putatively coding for tail spike/tail fiber subunits (gene products 22, 23, 25,
107 26, Fig. 2c). This suggests high level of variability of receptor-binding proteins and potentially high
108 level of host specialization of crAss-like phages.

109 Φ crAss001 could be effectively propagated *in vitro* using standard techniques using an EPS-
110 producing *B. intestinalis* 919/174 as its host. It was able to form readily visible plaques in agar
111 overlays (Fig. 2d) and reached 10^{10} pfu/mL when propagated in broth culture. In a one-step growth
112 experiment with multiplicity of infection (MOI) of ~ 1 the phage demonstrated a long latent period
113 of 120 min that was followed with a very small burst of progeny (2.5 pfu per infected cell). A
114 second burst of roughly the same size occurred 90 min later (Fig. S2a). An adsorption curve shows
115 that $\sim 74\%$ of phage bound to cells in first 5 min, and $>90\%$ were bound by 20 min (Fig. S2b).
116 Efficiency of lysogeny/mutation rate tests demonstrated $2\pm 1\%$ of cells of the strain 919/174 are
117 resistant to ϕ crAss001 on initial contact. These clones (potentially lysogens) were initially PCR
118 positive for presence of ϕ crAss001 (using gene 20 as a target) and were resistant to the phage in
119 plaque/spot assays. Successive rounds of propagation resulted in loss of the ϕ crAss001 PCR
120 positive outcome, but the phage-resistance phenotype was retained in all clones.

121 Infection of exponentially growing cells of *B. intestinalis* 919/174 with ϕ crAss001 at different
122 multiplicity of infections (MOI) did not result in complete culture lysis, but caused a delay in
123 stationary phase onset time and final density at stationary phase (Fig. S2c). A brief lysis period
124 occurred in the first few hours after infection, with timing dependent on the MOI, followed by
125 recovery of bacterial growth. In order to investigate the fate of phage in co-culture, bacterial cells

126 infected with phage at high MOI were allowed to reach stationary phase and then passaged daily or
127 bi-daily for a period of 23 days (Fig. 2e). Phage titre, which on the first two passages reached $\sim 10^{10}$,
128 was reduced and remained steady between 10^6 - 10^8 for the duration of the experiment.

129 Collectively, this suggests that ϕ crAss001 uses an unusual infection strategy to replicate *in*
130 *vitro* on its *B. intestinalis* host very efficiently on semi-solid agar (giving rise to large clear plaques)
131 and yet propagate in liquid culture without causing lysis of the host bacterium. Despite not being
132 able to form true lysogens or pseudolysogens, the phage seem to be able to co-exist with its host in
133 an equilibrium which is likely to confer an ecological advantage to both partners (similar to the
134 recently described carrier state life cycle¹⁹). Taken together, these results are puzzling. Given that in
135 liquid culture we used an MOI of 1 and there is a high adsorption rate, we cannot explain why two
136 bursts seemed to occur and why the broth culture did not clear. We can conclude that the phage
137 probably causes a successful lytic infection in a subset of infected cells (giving rise to a ‘false’
138 overall burst size of 2.5), but must enter into an alternative interaction with some or all of the
139 remaining cells. Overall, this allows both phage and host to co-exist in a stable interaction over
140 prolonged passages. This may explain why the *Bacteroides* host and crAssphage can both remain in
141 high numbers for prolonged period in the gut. The nature of this interaction warrants further
142 investigation.

143 Published studies of gut viromes in humans and germ-free mice with transplanted human
144 viromes^{12,20}, as well as our own unpublished longitudinal observations of the human gut virome
145 support the hypothesis that crAssphage use an unusual strategy to establish themselves at high
146 levels in the gut and to then persist stably within the microbial communities for several weeks to as
147 long as several months or even years. The absence of any detectable integrase gene and our inability
148 to isolate stable lysogens, together with the lack of any evidence that this or any other crAss-like
149 bacteriophages can occur in the form of prophages, suggest that pseudolysogeny and other
150 mechanisms, such as physical binding of phage to the intestinal mucous gel²¹ might be responsible

151 for long term persistence of crAss-like phages *in vivo*. Further studies will be required to fully
152 understand the replication cycle of crAss-like bacteriophages, their peculiar ability to persist at high
153 titres in the human gut microbiota, as well as their significance for human intestinal physiology and
154 disease. The first report of a phage-host pair should accelerate our understanding of these highly
155 abundant and unusual phages.

156

157

158 **Figure legends**

159 **Figure 1. Circular map of ϕ crAss001 genome.** Inner circle (green and blue), GC skew;
160 Middle circle (black), G+C content; Outer circle (Pink and purple fill), protein-coding genes
161 (ORFs); Outermost circle (orange fill), tRNA genes. Stroke colour on ORFs and fill colour on gene
162 numbers corresponds with the general predicted function (see colour legend for details); black
163 stroke, introns; genes with no functional annotations (coding for hypothetical proteins) are left
164 unlabelled.

165

166 **Figure 2. Morphology, growth and adsorption of ϕ crAss001.** **A**, SDS-PAGE analysis of
167 protein content in ϕ crAss001 virions and identification of selected polypeptides using MALDI-TOF
168 (see methods for details); **B**, TEM image of uranyl acetate negatively contrasted ϕ crAss001 virions
169 (62,000x magnification, accelerating voltage of 120kV); **C**, BLASTn comparisons between the
170 genome of the uncultured human faecal IAS virus, ϕ crAss001, and uncultured human phage
171 genomic contigs ERR843986_ms_1, ERR844058_ms_2, SRR4295175_s_4 (for contigs description
172 see Guerin *et al.*, 2018¹), highly variable region is marked with a red asterisk; **D**, plaque
173 morphology of ϕ crAss001 after 48 h incubation in a 0.3% FAA agar overlay with *B. intestinalis*
174 919/174 as host strain; **E**, persistence of ϕ crAss001 in a periodic culture of phage infected *B.*

175 *intestinalis* 919/174 with daily (or bi-daily) transfers for 23 days (see supplementary methods for
176 details).

177

178 **References**

1. Ogilvie, L. A. & Jones, B. V. The human gut virome: a multifaceted majority. *Front. Microbiol.* **6**, (2015).
2. Carding, S. R., Davis, N. & Hoyles, L. Review article: the human intestinal virome in health and disease. *Aliment. Pharmacol. Ther.* **46**, 800–815 (2017).
3. Manrique, P. *et al.* Healthy human gut phageome. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 10400–10405 (2016).
4. Norman, J. M. *et al.* Disease-specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. *Cell* **160**, 447–460 (2015).
5. Reyes, A. *et al.* Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11941–11946 (2015).
6. Monaco, C. L. *et al.* Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. *Cell Host Microbe* **19**, 311–322 (2016).
7. Aggarwala, V., Liang, G. & Bushman, F. D. Viral communities of the human gut: metagenomic analysis of composition and dynamics. *Mob. DNA* **8**, 12 (2017).
8. Dutilh, B. E. *et al.* A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* **5**, ncomms5498 (2014).
9. Yutin, N. *et al.* Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. *Nat. Microbiol.* **3**, 38–46 (2018).
10. Liang, Y. Y., Zhang, W., Tong, Y. G. & Chen, S. P. crAssphage is not associated with diarrhoea and has high genetic diversity. *Epidemiol. Amp Infect.* **144**, 3549–3553 (2016).
11. Guerin, E. *et al.* Biology and taxonomy of crAss-like bacteriophages, the most abundant virus in the human gut. *bioRxiv* 295642 (2018). doi:10.1101/295642
12. Yarygin, K. *et al.* Abundance profiling of specific gene groups using precomputed gut metagenomes yields novel biological hypotheses. *PLOS ONE* **12**, e0176154 (2017).
13. Cinek, O. *et al.* Quantitative CrAssphage real-time PCR assay derived from data of multiple geographically distant populations. *J. Med. Virol.* **90**, 767–771 (2018).
14. Holmfeldt, K. *et al.* Twelve previously unknown phage genera are ubiquitous in global oceans. *Proc. Natl. Acad. Sci.* **110**, 12798–12803 (2013).

15. Oude Munnink, B. B. *et al.* Unexplained diarrhoea in HIV-1 infected individuals. *BMC Infect. Dis.* **14**, 22 (2014).
16. González-Tortuero, E. *et al.* VIGA: a sensitive, precise and automatic de novo Viral Genome Annotator. *bioRxiv* 277509 (2018). doi:10.1101/277509
17. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005).
18. Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* (2017). doi:10.1016/j.jmb.2017.12.007
19. Siringan, P., Connerton, P. L., Cummings, N. J. & Connerton, I. F. Alternative bacteriophage life cycles: the carrier state of *Campylobacter jejuni*. *Open Biol.* **4**, 130200 (2014).
20. Reyes, A., Wu, M., McNulty, N. P., Rohwer, F. L. & Gordon, J. I. Gnotobiotic mouse model of phage–bacterial host dynamics in the human gut. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20236–20241 (2013).
21. Barr, J. J. *et al.* Bacteriophage adhering to mucus provide a non–host-derived immunity. *Proc. Natl. Acad. Sci.* **110**, 10771–10776 (2013).

179

180 **End notes**

181 **Acknowledgements:** This research was conducted with the financial support of Science
 182 Foundation Ireland (SFI) under Grant Number SFI/12/RC/2273 a Science Foundation Ireland's
 183 Spokes Programme which is co-funded under the European Regional Development Fund under
 184 Grant Number SFI/14/SP APC/B3032, and a research grant from Janssen Biotech, Inc.

185 **Supplementary Information** is available in the online version of the paper (Supplementary
 186 table 1, Supplementary figure 1, Supplementary figure 2, Supplementary methods)

187 **Author contributions:** ANS, EVK and CBF performed the experiments; ANS, SRS, LAD,
 188 RPR and CH analysed the data; RPR and CH supervised the project.

189 **Author information:** Reprints and permissions information is available at
 190 www.nature.com/reprints. The authors declare no competing financial interests. Correspondence
 191 and requests for materials should be addressed to CH (c.hill@ucc.ie) or ANS
 192 (andrey.shkoporov@ucc.ie).



