Supplementary data

Supplementary Table 1.

Figure S1. A, cumulative relative abundance (fraction of Illumina reads aligned) of different IAS-virus like phage contigs in several faecal virome metagenomic sequencing datasets[1–3]; **B**, relative abundance of φAPC-LOC110 in the same datasets.

Figure S2. A, one-step growth curve of φcrAss001 in early log-phase culture *B. intestinalis* 919/174 (OD_{600} = 0.2 at inoculation, MOI = 0.7); LP, latency period; BS, burst size (see text for details); **B**, adsorption curve of φcrAss001 on early-log phase *B. intestinalis* 919/174 cells in FAB medium (MOI = 0.7); **C**, growth curves of *B. intestinalis* 919/174 infected at the early-log growth phase (OD_{600} = 0.2 at inoculation) with different MOI levels of φcrAss001 (MOI ~ 0.00025 – 25).

Supplementary methods

Faecal filtrates, bacterial strains and phage enrichment

Faeces were collected from 20 adult consenting Irish volunteers according to study protocol APC055, approved by the Cork Research Ethics Committee (CREC). Samples were collected (without fixative or preservative) in the volunteer's home and transported to the research facility at ambient temperature, avoiding exposure to heat. They were generally stored at -80°C until processed, unless indicated otherwise.

Faecal filtrates were prepared by homogenizing 0.5 g of thawed faeces in 10 ml SM buffer using vigorous vortexing for 5 min. Tubes were then chilled on ice for 5 minutes prior to centrifugation at 5,000 rpm in a swing bucket rotor for 10 min at +4°C. Supernatants were transferred to new tubes and centrifugation was repeated once again. Supernatant were subsequently filtered twice through a 0.45 µm pore PES syringe-mounted membrane filters. Pooled faecal filtrates were prepared by combining equal volumes of filtrates from 20 donors.

A total of 53 bacterial strains were used for the enrichment experiment (Table S1). Strains were grown on YCFA broth[4] supplemented with a mixture of carbohydrates (D-glucose, soluble potato starch, D-cellobiose and D-maltose at concentrations of 2 g/L of each. Hungate tubes were filled with 9 ml of pre-reduced YCFA broth in an anaerobic chamber and sterilized by autoclaving. Tubes were inoculated with 1ml of overnight cultures using 23G hypodermic needles and incubated at 37°C until the OD_{600} = 0.3 (CO8000 Cell Density Meter, Biochrom WPA). One millilitre of

culture was then removed from the Hungate tube and replaced with the pooled faecal filtrate, maintaining a volume of 10 ml. Incubation was continued at 37°C overnight. Cultures were centrifuged twice as described above and supernatants were filtered through 0.45 µm pore PES filters. One millilitre of this filtrate was then added to an exponentially growing culture of the same strain as described above. The enrichment process was repeated a total of 3 times.

Extraction of VLP-associated DNA and shotgun sequencing

The VLP DNA fraction was extracted from 10 ml bacterial cultures after 3rd round of phage enrichment. Briefly, cultures were centrifuged twice and filtered as described above. NaCl and PEG-8000 powders were then added to filtrates to give a final concentration of 0.5 M and 10% w/v, respectively. After complete dissolving samples were incubated overnight (16 hours) at +4°C. The remaining DNA extraction steps were performed as described before[5]. One microliter of DNA sample was then amplified using MDA technology with Illustra GenomiPhi V2 kit (GE Healthcare). The latter step was done in triplicate for each sample. Products from all three MDA reactions were pooled together and subjected to additional round of purification using DNeasy Blood & Tissue Kit.

Amplified DNA was quantified using Qubit dsDNA HS Assay Kit (Invitrogen/ThermoFisher Scientific) and subjected to random shotgun library preparation using Nextera XT DNA Library Preparation Kit (Illumina) and bead-based normalisation following the standard manufacturer's protocol. Ready-to-load libraries were sequenced using a proprietary modified protocol using 2 × 300 bp paired-end chemistry on an Illumina MiSeq platform (Illumina, San Diego, California) at GATC Biotech AG, Germany.

Analysis of shotgun sequencing data and annotation of φAPC-LOC110 genome

Raw Illumina Miseq paired-end reads were quality-checked using FastQC v0.11.5, trimmed and filtered using Trimmomatic v0.36[6] using sliding window approach with window size of 4 nt and minimum allowed Phred score of 20. In addition, all reads were cropped to a length of 230 nt with first 10 nt removed. Reads shorter than 60 nt were discarded. Filtered reads were assembled on per-sample basis using meta-spAdes v3.10.0[7] with standard parameters. Contigs shorter than 1000 nt were discarded. Contigs from all enrichment samples were then pooled together and demultiplexed by picking the longest representatives for each group of contigs with >90% sequence identity and >90% of sequence overlap (as determined by BLASTn v2.2.28+[8]). To quantify presence of various contigs in the enrichment samples, filtered reads were aligned back to the common demultiplexed database of contigs using Bowtie v2.1.0[9] in the end-to-end mode. Counts of aligned reads were extracted from alignment data using Samtools v0.1.19.

Annotation of genomic contig representing phage φAPC-LOC110 was done using VIGA v0.10.3[10] with BLASTp searches against NCBI nr database (snapshot of 2018-01-15), and HMM searches against UniProt/Swiss-Prot database (snapshot of 2018-01-24). Additionally, amino acid sequences of the encoded proteins were annotated using HHpred Web-server[11] (https://toolkit.tuebingen.mpg.de/#/tools/hhpred) with HMM profile-profile searches against the following databases: PDB mm_CIF70_25_Feb, Pfam-A v31.0, NCBI CD v3.16, TIGRFAMs v15.0. Circular map of φAPC-LOC110 was visualized using GView v1.7.

The PCR primers φAPC-LOC110-F1 (5'-AATAAGGTGGAAGATGCTGAC-3') and φAPC-LOC110-R1 (5'-TTATCCATTTGGTCAACAGCTC-3'), specific towards gene 20 of φAPC-LOC110 were developed for detection and quantification of the phage in cultures of *Bacteroides intestinalis* 919/174.

Propagation of φAPC-LOC110, one step-growth curve, adsorption, and efficiency of lysogeny.

The host strain *B. intestinalis* 919/174 is routinely maintained in Fastidious Anaerobe Broth (FAB, Oxoid) anaerobically at 37°C. Cultures were infected with φAPC-LOC110 at various multiplicities of infection (MOI = 1 for optimal phage yield) in early logarithmic phase of growth (OD_{600} = 0.2, corresponding to ~2 x 10^8 cfu/ml) with or without addition of $CaCl_2$ and $MgCl_2$ to final concentration of 1 mM of each. Infected culture were collected after overnight incubation, centrifuged at 5,000 rpm, 4°C for 15 min to remove cells and then filtered through 0.45 µm pore PES syringe-mounted membrane filters. Phage cultures could be stored at 4°C without any further treatment for a period of up to 3 months without significant loss of titre. Plaque and spot assays were performed in a standard manner using 3 ml of 0.4% Bacto agar (Becton Dickinson) for overlays on 100 mm diameter plates with FAA solid medium. Two hundred µl of overnight *B. intestinalis* 919/174 culture in FAB and 100 µl of phage sample were added to molten overlay agar tubes kept at 45°C, followed by vortexing and pouring on pre-made FAA plates. Plates were incubated anaerobically at 37°C for 24 hours before plaque counting.

For one-step growth experiment early logarithmic phase culture of *B. intestinalis* 919/174 was infected at an MOI of 0.7 for 5 min at room temperature, followed by centrifugation at 5,000 rpm, 4°C for 10 min, removal of supernatant and re-suspending of the infected cells in fresh FAB medium. Incubation was continued anaerobically at 37°C for further 225 min with removal of 1 ml samples every 15 min. Samples were filtered through 0.45 µm pore PES filters and subjected to standard plaque assays with appropriate dilutions.

Adsorption experiments were performed in a similar manner. Early logarithmic phase culture *B. intestinalis* 919/174 were mixed with φAPC-LOC110 at an MOI of 1 and incubated at room temperature for 1 hour with removal of aliquots every 5 min followed by immediate filtering through 0.45 µm pore PES filters and subsequent standard plaque assay.

Efficiency of lysogeny experiments were conducted by spreading 200 µl of phage lysates (10^9 – 10^{10} pfu/ml) onto 100 mm diameter FAA agar plates. After fully dried, both phage-covered and negative control plates were inoculated with serial 10-fold dilutions of *B. intestinalis* 919/174 overnight culture. Efficiency of lysogeny was determined as a percentage of colonies observed on phage-coated plates relative to the total counts on negative control plates after 48 hours of anaerobic incubation at 37°C.

Transmission electron microscopy (TEM) and proteomic analysis of φAPC-LOC110

Sixty ml of phage lysate obtained as described above was ultra-centrifuged at 120,000g for 3 hours using a F65L-6x13.5 rotor (ThermoScientific). The resulting pellets were re-suspended in 5 ml SM buffer. The viral suspensions were further purified by overlaying them onto a caesium chloride (CsCl) step gradient of 5M and 3M, followed by centrifugation at 105,000g for 2.5 hours. A band of viral particles visible under side illumination was collected and buffer-exchanged using 3 sequential rounds of 10-fold diluting and concentrating to the original volume by ultra-filtration using Amicon Centifugal Filter Units 10,000 MWCO (Merck). Following this, 5µl aliquots of the viral fraction were applied to Formvar/Carbon 200 Mesh, Cu grids (Electron Microscopy Sciences) with subsequent removal of excess sample by blotting. Grids were then negatively contrasted with 0.5% (w/v) uranyl acetate and examined at UCD Conway Imaging Core Facility (University College Dublin, Dublin, Ireland) by transmission electron microscope. The same CsCl gradient fraction was further concentrated 10-fold using Amicon Ultra-0.5 Centrifugal Filter Unit with 3 kDa MWCO membrane (Merck, Ireland). The obtained sample (25 µl) was loaded onto a pre-made Bolt 4-12% Bis-Tris Plus reducing SDS-PAGE gel (Invitrogen) and separated at 200 V for 30 minutes using 1X NuPAGE MOPS SDS Running Buffer. Thirteen detectable bands with approximate molecular weights of 15, 35, 54, 58, 70, 77, 80, 85, 100, 150 and 270 kDa were excised and subjected to MALDI-TOF/TOF (Bruker Ultraflex III) protein identification following in-gel trypsinization, at Metabolomics & Proteomics Technology Facility (University of York, York, UK).

References

1. Manrique, P. *et al.* Healthy human gut phageome. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 10400–10405 (2016).
2. Norman, J. M. *et al.* Disease-specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. *Cell* **160,** 447–460 (2015).
3. 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).
4. Duncan, S. H., Hold, G. L., Harmsen, H. J. M., Stewart, C. S. & Flint, H. J. Growth requirements and fermentation products of Fusobacterium prausnitzii, and a proposal to reclassify it as Faecalibacterium prausnitzii gen. nov., comb. nov. *Int. J. Syst. Evol. Microbiol.* **52,** 2141–2146 (2002).
5. Shkoporov, A. N. *et al.* Reproducible protocols for metagenomic analysis of human faecal phageomes. *Microbiome* **6,** 68 (2018).
6. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).
7. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27,** 824–834 (2017).
8. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
9. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).
10. González-Tortuero, E. *et al.* VIGA: a sensitive, precise and automatic de novo VIral Genome Annotator. *bioRxiv* 277509 (2018). doi:10.1101/277509
11. 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).