

Title	Mining for novel human gut bacteriophages against Bacteroidales
Authors	Guerin, Emma
Publication date	2020-02
Original Citation	Guerin, E. 2020. Mining for novel human gut bacteriophages against Bacteroidales. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2020, Emma Guerin. - https://creativecommons.org/licenses/by-nc-nd/4.0/
Download date	2024-04-19 05:02:08
Item downloaded from	https://hdl.handle.net/10468/9987



Mining for novel human gut bacteriophages against Bacteroidales

A Thesis Presented to the National University of Ireland
for the Degree of
Doctor of Philosophy
by

Emma Guerin B.Sc.

Student ID No. 110327635

February 2020

Supervisors: Prof. Colin Hill and Prof. Paul Ross

School of Microbiology

University College Cork

Head of School: Professor Paul O'Toole

Contents

Declaration.....	5
Acronyms	6
Thesis abstract.....	8
CHAPTER I Shining light on the viral dark matter of the human gut microbiome	11
Abstract.....	12
Introduction	14
1. The human gut phageome	19
1.1 Phage lifecycles	19
1.2 Taxonomy.....	21
1.3 Composition and structure.....	22
2. Phage-host interactions.....	28
2.1 Host defense mechanisms and phage counter defense strategies	28
2.2 Ecological models and the GIT	33
2.3 Other influencing factors and alternative interactions.....	35
2.4 Interactions with the mammalian host.....	40
3. Current methodologies, challenges and potential solutions	42
3.1 Viral metagenomics, sequencing and <i>in silico</i> methods.....	42
3.2 <i>In vitro</i> , <i>ex vivo</i> and <i>in vivo</i> methods.....	51
3.3 Case Study: The crAssphage family story: from <i>de novo</i> assembly and <i>in silico</i> characterisation to an <i>in vitro</i> reality	62
4. The merits of studying bacteriophages and future prospects	71
5. Conclusions	74
6. Figures.....	76
7. References	84
CHAPTER II Biology and taxonomy of crAss-like bacteriophages, the most abundant virus in the human gut	120
Author contributions	121
Graphical abstract and highlights.....	122

2.1 Summary	123
2.2 Introduction	124
2.3 Experimental procedures.....	127
2.4 Results	138
2.5 Discussion	146
2.6 Figures.....	154
2.7 Tables	169
2.8 References	170
 CHAPTER III Isolation and characterisation of a <i>Bacteroides xylanisolvens</i> infecting crAss-like phage isolated from the human gut, ΦcrAss002	181
3.1 Summary	182
3.2 Introduction	184
3.3 Experimental Procedure	187
3.4 Results	199
3.5 Discussion	210
3.6 Figures.....	219
3.7 Tables	236
3.8 References	247
 CHAPTER IV Isolation and characterization of a novel <i>Parabacteroides distasonis</i> bacteriophage isolated from the human gut.....	261
4.1 Summary	262
4.2 Introduction	263
4.3 Experimental procedures.....	265
4.4. Results	275
4.5. Discussion	281
4.6 Figures.....	287

4.7 Tables.....	301
4.8 References	307
Thesis summary and future work.....	320
Acknowledgments	323

This thesis has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under grant numbers SFI/12/RC/2273 and SFI/14/SP APC/B3032, and a research grant from Janssen Biotech.

Declaration

This is to certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism.

Signed: _____
Emma Guerin

Date: _____

Acronyms

BAM = bacteriophage adherence to mucus

Bp = base pair

CBA = Columbia blood agar

CD = Crohn's disease

CFU = colony forming units

dsDNA = double stranded DNA

EM = electron microscopy

FAB = fastidious anaerobe broth

FACS = fluorescence activated cell sorting

FFT = faecal filtrate transfer

FMT = faecal microbiota transfer

FSI = frozen standard inoculum

GIT = gastrointestinal tract

HMM = hidden Markov model

IBD = inflammatory bowel disease

ICTV = International Committee on Taxonomy of Viruses

MDA = multiple displacement amplification

MOI = multiplicity of infection

OMVs = outer membrane vesicles

P-crAssphage = prototypical crAssphage

PFU = plaque forming units

PPV = personal persistent virome

PV = phase variation

qPCR = quantitative PCR

QS = quorum sensing

RBP = receptor binding protein

SAGs = single amplified genomes

SSB = single stranded binding protein

ssDNA = single stranded DNA

SVGs = single viral genomics

TDV = transiently detected virome

TEM = transmission electron microscopy

UC = ulcerative colitis

VLPs = virus-like particles

VMR = virus-to-microbe ratio

VT = viral tagging

YCFA = yeast extract, casitone and fatty acid agar

Thesis abstract

The importance of the gut microbiome in human health and disease has become apparent in recent years. The bacterial component is well characterised in comparison to the viral fraction, that is predominantly composed of bacteria infecting viruses called bacteriophages (phages). Phages are largely understudied in the context of the human gut despite their potential as driving forces of microbiome composition and function. Changes in the composition of the microbiome have been linked to disease states such as IBD, colorectal cancer and diabetes. However, the root causes of these changes remain to be elucidated. While gradual improvements in virome metagenomic analyses have provided us with important insights, we are only just beginning to understand the complex roles of gut phages. If we are to truly understand how phages may shape our microbiome and influence our health, it will be necessary to determine the mechanisms behind phage-bacterium host interactions as well as how phages affect the human host such as through interactions with the immune system. As metagenomics provides little insight into the biological properties of phages, the isolation of novel phage-host pairs from the human gut is required in conjunction with characterisation using *in vitro*, *ex vivo* and *in vivo* models with naturally relevant conditions. However, this is not without challenges.

To gain some scope into the human gut phageome, this thesis focuses on one of the most interesting and abundant phages identified within the “dark matter” of viral sequences, crAssphage. This phage was predicted to infect hosts of the Bacteroidales order, one of the most important bacterial groups in the healthy human gut. We performed bioinformatic analyses of crAss-like phage relatives using faecal phageome datasets originating from multiple human cohorts. This led to the *de novo* assembly of

244 novel crAss-like phages and their taxonomic classification at the genus-level. We demonstrated the first evidence of *ex vivo* crAss-like phage propagation and generated preliminary electron micrographs using faeces from a crAss-rich donor, confirming a podovirus morphology.

Following this, the objective was to isolate members of the crAss-like phage family *in vitro* using a targeted screening approach. Both phage and host enrichments were performed from the same crAssphage-rich faeces which was aided by the use of antibiotics to selectively promote the growth of Bacteroidales with the parallel expansion of associated crAss-like phages. This led the isolation of a novel member of the crAss-like family in pure culture, Φ crAss002, the characterisation of which provided interesting insights into gut phage-host dynamics.

A similar methodology was applied with the goal of isolating further Bacteroidales associated phages, however, a more traditional screening approach was implemented. This resulted in the isolation and first detailed characterisation of a virulent *Parabacteroides distasonis* phage to be reported as of 2020. Given that no other virulent phages targeting this important bacterial genus are deposited on NCBI Taxonomy, the need for *in vitro* isolation of further phage-host pairs is emphasised.

These novel gut phages show that virulent phages do not necessarily present with traditional biological properties such as plaque formation or complete lysing of host cultures. Interestingly, both phage isolates were able to co-exist with their host over time. This is consistent with the observed persistence of virulent phages in the human gut over extensive periods of time. Consequently, our view of phage-host interactions in the human gut is likely to change and expand in years to come. This will be necessary if we are to fully understand the extent to which these bacteria

infecting viruses shape the microbiome and mediate homeostasis of the gut. This will be crucial if we are to manipulate them for therapeutic applications.

Chapter I

Shining light on the viral dark matter of the human gut microbiome

Emma Guerin, Stephen Stockdale, Paul Ross and Colin Hill

Abstract

The human gut is a complex environment that contains a multitude of microorganisms that are collectively termed the microbiome. Multiple factors have a role to play in driving the composition of human gut bacterial communities either towards homeostasis or the instability that is associated with many disease states. One of the most important forces are likely to be bacteriophages, bacteria infecting viruses that constitute by far the largest portion of the human gut virome. Despite this, bacteriophages (phages) are the one of the least studied residents of the gut. This is largely due to the challenges associated with studying these difficult to culture entities. Modern sequencing technologies have played an important role in improving our understanding of the human gut phageome but much of the generated sequencing data remains uncharacterised. Overcoming this requires database-independent bioinformatic pipelines and even those phages that are successfully characterised, limited insight into associated biological properties is provided, and thus the majority of viral sequences exist only *in silico* as “dark matter”. Fundamental to understanding the role of phages in shaping the human gut microbiome, and in turn perhaps influencing human health, is how they interact with their bacterial hosts. An essential aspect is the isolation of novel phage-bacterium host pairs by direct isolation through various screening methods, which can transform *in silico* phages into a biological reality. However, this is also beset with multiple challenges including culturing difficulties and the use of traditional methods, such as plaquing, which may bias the phage-host pairs that can be successfully isolated. Phage-bacterium interactions may be influenced by many aspects of complex human gut biology which can be difficult to reproduce under laboratory conditions. Here we discuss some of the main findings associated with the human gut phageome to date including composition, our

understanding of phage-host interactions, particularly the observed persistence of virulent phages and their hosts, as well as factors that may influence these highly intricate relationships, and current methodologies and bottlenecks hindering progression in this field. We also discuss potential steps that may be useful in overcoming these hurdles in future studies.

Introduction

The human microbiome is collectively comprised of trillions of microbial cells originating from all domains of cellular life (Bacteria, Archaea, Eukarya) as well as viruses (Ursell et al., 2012). The human microbiome has received significant attention in the past two decades, and can even be considered as an organ in its own right due to its important role in multiple aspects of human health (O'Hara and Shanahan, 2006; Baquero and Nombela, 2012; Lloyd-Price et al., 2016). Microbial numbers equate to $\sim 10^{14}$ cells which is comparable to the total number of human cells (Sender et al., 2016; Cani, 2018). The vast majority of these microorganisms reside in various locations within the human gastrointestinal tract (GIT) where they partake in essential metabolic and physiological processes that impact human biology (Turnbaugh et al., 2007; Sender et al., 2016). Thus, the GIT forms a complex ecosystem where biotic factors and abiotic factors, such as immune components and physiochemical parameters, are intertwined (Mirzaei and Maurice, 2017; Shkoporov and Hill, 2019). Bacteria constitute the majority of biomass in the GIT where they have key roles in maintaining human homeostasis, immunity, and health (O'Hara and Shanahan, 2006; Qin et al., 2010; Yatsunenko et al., 2012; Belkaid and Hand, 2014; Sender et al., 2016). A healthy human gut, with no obvious signs of disease, is populated by more than 1000 bacterial species with greater than 90% resolving into two phyla, the Firmicutes and Bacteroidetes (Turnbaugh et al., 2007; Lozupone et al., 2012). In the healthy human gut, these bacterial species will be generally stable over years and perhaps even decades (Faith et al., 2013). Perturbations and shifts in the composition, abundance, and diversity, of bacterial communities has been associated with multiple disease states (Lozupone et al., 2012). These include inflammatory bowel disease, obesity, diabetes, eczema, cancer and neurological disorders (Karlsson et al., 2011; Giongo et

al., 2011; Gkouskou et al., 2014; Goodrich et al., 2014; Halfvarson et al., 2017; Casén et al., 2015; Hartstra et al., 2015; Vogt et al., 2017; Helmink et al., 2019). The driving forces that cause these shifts in the gut bacterial community remain a vital question in the field of human gut microbiome research. One of the most understudied aspects of the microbiome are viruses of bacteria, bacteriophages (phages). It is well known that phages have an important role as drivers in shaping bacterial communities in different environments both by predation or horizontal gene transfer by transduction and can be regarded as powerful evolutionary forces (Koskella and Meaden, 2013). To date, the phageome has been largely studied in the context of ecosystems such as the marine environment (Rodriguez-Brito et al., 2010; Hurwitz et al., 2014). However, in recent years, it has become apparent that phages are also highly abundant in the human gut.

Together bacteria and their viruses are the most abundant components of the human gut (Cani, 2018). Despite this, and their role in influencing bacterial communities in the GIT ecosystem, the viral fraction remains one of the least understood components of the human gut microbiome. The virome includes both bacteriophages and eukaryotic viruses; however, phages (phageome) are significantly more abundant (Gregory et al., 2019). Furthermore, recent studies have shown that microbiome composition is driven by phages and not by the eukaryotic viruses which are generally not associated with health (Moreno-Gallego et al., 2019). Despite being independently discovered over a century ago by Twort and d’Herelle, it is only in the past decade that the potential role of phages in the gut microbiome has been considered (Twort, 1915; d’Hérelle, 1917). Phage-driven alterations of bacterial composition by direct interactions or potentially via the human immune system and changes in the gut phageome composition itself have been linked to a number of disease states including inflammatory bowel disease (IBD), diabetes, malnutrition, Parkinson’s disease

(Norman et al., 2015; Reyes et al., 2015; Manrique et al., 2016; Nguyen et al., 2017; Zhao et al., 2017; Belleghem et al., 2017; Ma et al., 2018; Kieser et al., 2018; Tetz et al., 2018; Gogokhia et al., 2019; Zuo et al., 2019; Clooney et al., 2019). Faecal filtrate transfer (FFT) studies further indicate that the phageome has the potential to shape the microbiome. For example, the administration of a bacterial cell-free faecal filtrate from healthy donors to patients with *Clostridioides difficile* infection (CDI) was demonstrated to prevent disease recurrence (Ott et al., 2017). The engraftment of the donor phageome has also been observed in another study involving faecal microbiota transplantation (FMT) in CDI patients (Draper et al., 2018). Furthermore, some abundant gut phages, such as the crAss-like phages, appear to interact with their bacterial hosts in dynamics that have not been typically observed in other ecosystems. They have been found to persist in the gut over extensive periods of time without major impacts on their bacterial hosts (Maura et al., 2012a; Guerin et al., 2018; Shkoporov et al., 2018a). These intriguing interactions merit further investigation. Findings to date highlight the importance and significance of the human gut phageome but the factors and mechanisms by which phages impact on bacterial communities and overall microbiome homeostasis or instability are still unclear. Key to understanding this is gaining insight into how phages interact with their bacterial host in the human gut.

Despite a resurgence in interest in gut phages, current methodologies are limited and require improvements. Modern sequencing technology has been essential in providing insights that would have not been possible with culture-based methods (Shkoporov et al., 2019). However, this methodology is also limited as a consequence of the fact that most viral sequences lack homology with known phages in current databases, with up to 86-99% of reads remaining uncharacterised, sometimes referred

to as viral “dark matter” (Aggarwala et al., 2017). Many bioinformatic pipelines are database-dependent and only focus on the characterizable fraction. This results in the exclusion of the majority of viral sequences and skewed overall findings (Clooney et al., 2019; Sutton and Hill, 2019; Gregory et al., 2019). Furthermore, novel phages identified through viral metagenomics often only exist *in silico* and it can be difficult to predict potential bacterial hosts at a species or strain level (de Jonge et al., 2019a). Overall, current pipelines can result in incomplete and skewed analyses.

In terms of laboratory-based research there are also multiple challenges to overcome. Many gut bacteria are difficult to culture, thus making isolation of their associated phages extremely difficult (Browne et al., 2016). Furthermore, many phages are strain specific, mimicking gut conditions is problematic, and host factors such as phase variable expression of phage receptors add another level of difficulty (Porter et al., 2020). Many current studies only look at the phage host-pair under reductionist *in vitro* conditions. Although laboratory characterisation is essential and provides important insights, analysis of novel phage-host pairs should also be coupled where possible with examination under more complex, realistic conditions such as fermenter systems or mouse models. Mouse models also provide insight into the radial and distal variability along the GIT (Zhao et al., 2019; Lourenço et al., 2019). The importance of coupling *in vitro* work with other models was highlighted in a study which found that a single point mutation in the tail fibre gene of an *E. coli* infecting phage was sufficient to allow the phage to “jump” host at a strain level when faced with resistance. However, this phenomenon was overlooked under *in vitro* conditions and only detected in a conventional mouse model (De Sordi et al., 2017).

This review will discuss the current understandings of the human gut phageome and phage-host interactions. We will also discuss the importance of isolating and

characterising novel phage-host pairs from the gut to attain new insights and improve our knowledge on phage-bacteria dynamics. We will review the current methodologies implemented in the laboratory and in bioinformatics, and the challenges associated with each that need to be overcome if both *in silico* and *in vitro/in vivo* findings are to provide a better understanding of the reality of the gut phageome. The ultimate goal is to fully understand how phages shape the microbiome and thus how this can impact on human health and disease.

1. The human gut phageome

The human gut is a densely populated ecosystem with significant variability in physical conditions as well as abiotic and biotic factors including pH, oxygen, nutrient and water availability, immunoregulators and bile acids (Mirzaei and Maurice, 2017). This creates a gradient of conditions ranging from the mouth to the stomach and to the small and large intestine, which results in niche specificity of the microbial residents of the gut (Pereira and Berry, 2017; Bauer et al., 2018). In addition to longitudinal variation, there is radial variation along the GIT due to features such as mucosa, villi and crypts (Donaldson et al., 2016; Zhao et al., 2019). Within this complex ecosystem bacteriophages, although structurally simple, can partake in intricate dynamics with bacteria and can influence microbiome homeostasis.

1.1 Phage lifecycles

Traditionally, phages have been classified as lytic/virulent or lysogenic/temperate based on the method of infection they employ when targeting a permissive host (Hobbs and Abedon, 2016). Both life cycles initiate with adsorption of the phage to its specific bacterial host where binding occurs between the phage receptor binding protein and a host cell surface feature which acts as the corresponding receptor. Following this, the phage injects its genetic material, which can be either single-stranded (ss) or double-stranded (ds) DNA or RNA, from its capsid into the host cell. It is at this point that the life cycles diverge. Lytic phages hijack host replication machinery to replicate their genomes, assemble phage components and produce progeny (Figure 1A). This ultimately ends in lysis of the host cell and release of progeny phages to commence subsequent rounds of host infection. In contrast,

lysogenic phages can integrate into the host chromosome forming a prophage and replicates in tandem with the bacterial genome (Figure 1B). In the case of pseudolysogeny, the genetic material persists in the host cytoplasm as an episome independent of the host genome (Figure 1C). In both cases, the phage genome replicates with the host and is carried in daughter cells without leading to lysis. The direct integration of a temperate phage into the bacterial genome, or the horizontal gene transfer that they are capable of mediating by transduction, can confer the host with beneficial genes and functions including virulence factors and stress tolerance. This is reviewed in detail by Obeng and colleagues (Obeng et al., 2016). Certain stimuli can cause the induction and excision of prophages which revert to a lytic lifestyle. These stimuli can include multiple intrinsic and extrinsic factors, including antibiotics and inflammation of the gut (Goerke et al., 2006; Nanda et al., 2015; Howard-Varona et al., 2017; Clooney et al., 2019). Pseudolysogeny is associated with conditions of stress such as host cell starvation and allows the phage to avoid resistance strategies. Thus, it is thought to provide phages with a means of survival. Depending on the favourability of conditions, pseudolysogenic phages can then commit to a lysogenic or lytic infection cycle (Łoś and Węgrzyn, 2012). A carrier state life cycle has been described for *Campylobacter jejuni* phages (Figure 1D). This alternative life cycle is pseudolysogenic-like in that the phage can enter an episome state in a permissive host, but it can also remain “carried” on the surface of non-permissive host cells without replication. The phage will then commit to lysis when it comes in contact with a sensitive host and conditions are favourable for progeny production. This is largely dictated by whether the host is phage permissive, thus creating an equilibrium that is favourable to both phage and host survival (Siringan et al., 2014). Chronic phage infections can be likened to the lytic cycle in that new

progeny is produced by hijacking the host, but newly produced virions are exported by alternative mechanisms without lysis thus the host cell is maintained intact (Figure 1E). This method of replication has been associated with filamentous phages such as M13 and CTX Φ (Hobbs and Abedon, 2016; Smeal et al., 2017a, 2017b). Although pseudolysogeny, the carrier state lifestyle and chronic phage infections are less characterised in comparison to the lytic and lysogenic lifecycles, the prevalence of each of the lifecycles is poorly understood in the context of the human gut. To avoid ambiguity, when discussing the phageome we generally refer to phages present within bacterial genomes as temperate, while free phage virions are referred to as virulent phages or virus-like particles (VLPs).

1.2 Taxonomy

The International Committee on Taxonomy of Viruses (ICTV) is responsible for the taxonomic classification of prokaryotic viruses which has been traditionally based on virion morphology (Lefkowitz et al., 2018). The best characterised of the phage families are the *Siphoviridae*, *Podoviridae* and *Myoviridae* that make up the Caudovirales order, all of which are tailed with dsDNA genomes. Members of these phage families have distinct head and tail morphologies: *Siphoviridae* have long non-contractile tails, *Podoviridae* have short non-contractile tails and *Myoviridae* have long contractile tails (Ackermann, 1998). Electron microscopy (EM) of human faecal filtrates shows a dominance of these phage morphologies (Hoyles et al., 2014). However, many studies focused solely on VLPs and observations can be skewed due to the ease at which phages of the Caudovirales order can be recognised. An interesting argument was put forward by Sutton and Hill regarding the current discrepancies in phage taxonomy, where two phages that shared similar genomic and functional

characteristics were classified into different phage families solely based on virion morphology (Sutton and Hill, 2019). This highlights the need for a move towards the use of sequenced-based taxonomic classification. Currently, efforts are being made to develop a genome-based taxonomic scheme (Bolduc et al., 2017; Eloie-Fadrosh, 2019; Jang et al., 2019). This will be important for future virome studies and for the classification of novel phages. However, this is not without difficulties and will require integration with the current taxonomic scheme.

1.3 Composition and structure

In the human gut, phages are estimated at $\sim 10^{10} \text{ g}^{-1}$ while the bacteria they infect equate to $\sim 10^{11} \text{ g}^{-1}$ (Shkoporov and Hill, 2019). It has been long postulated that phages greatly outnumber their bacterial hosts at 10:1 virus-to-microbe ratio (VMR) (Bergh et al., 1989; Wommack and Colwell, 2000; Chibani-Chennoufi et al., 2004). VMR has been associated with influencing lifestyle switches of phages (Howard-Varona et al., 2017). With shifts to a high VMR (high phage to low host densities), it was thought that phages would enter into a lysogenic life cycle to ensure persistence in the community. However, more recently it has been proposed that with increased bacterial abundances, phages will transition from lysis to lysogeny allowing them to take advantage of the success of their host in “Piggy-Back-The-Winner” dynamics (Knowles et al., 2016). It is important to note however, that this is largely based on marine and aquatic environments and may not necessarily be representative of the human gut. The VMR in the human gut is significantly lower compared to other ecosystems, estimated at $\sim 0.1:1$ (Shkoporov and Hill, 2019).

The ratio of virulent and temperate phages in the healthy human gut has been a question of interest. In recent years, the main consensus based on both microscopic and sequenced-based virome studies is that the human gut is dominated by temperate dsDNA phages of the Caudovirales order and ssDNA phages of the *Microviridae* family (Reyes et al., 2010; Kim et al., 2011; Minot et al., 2011; Hoyles et al., 2014; Moreno-Gallego et al., 2019). These phages, in particular Caudovirales, have been described as the core of a healthy human adult phageome with a high occurrence of prototypical crAssphage and related crAss-like phages (Dutilh et al., 2014; Manrique et al., 2016; Moreno-Gallego et al., 2019). In addition, this core was found to be highly stable over time which is consistent with reports on crAss-like phages and the slow evolution rates associated with temperate phages compared to virulent phages (Minot et al., 2011; Dutilh et al., 2014; Shkoporov et al., 2018a; Guerin et al., 2018; Shkoporov et al., 2019). Studies have found that 95% of viral genotypes are maintained after one year and 80% after two years (Reyes et al., 2010; Minot et al., 2013). This was linked to the dominance of temperate phages which have low mutation rates compared to virulent phages, such as *Microviridae*, which have mutation rates as high as 10^{-5} substitutions per day (Minot et al., 2013). Changes in the composition and the richness of the temperate core have also been associated with diseases states such as IBD (Norman et al., 2015). However, the existence of this phage core in the human gut shared across individuals has also been disputed (Gregory et al., 2019). There have also been suggestions of a healthy “core microbiota” in relation to bacterial communities in the gut and that deviations from this core are associated with disease. For defining this, it was found that specifying the core was most appropriate based on shared functionalities as opposed to taxonomy (Turnbaugh et al., 2009).

A key issue with many virome studies is that the analytical pipelines focus solely on the known, annotated component (~1-14% of sequences generated) and exclude viral dark matter which currently constitutes most of the sequences in viral datasets (Aggarwala et al., 2017). If only a fraction of the reads are examined, this obviously leads to an incomplete and potentially biased analysis. More recent studies have attempted to take this into account by implementing stricter criteria and limiting the use of database-dependent methods.

A study examining the correlation between virome and microbiome similarity among 21 adult monozygotic twin pairs with concordant or discordant microbiomes confirmed that phages indeed drive microbiome diversity (Moreno-Gallego et al., 2019). This fits well with observations made for FMT recipients whose phage and bacterial profiles reflected that of the donor (Draper et al., 2018). In addition, 18 contigs were found to be shared across all individuals tested, the majority of which were crAss-like phage family associated (Moreno-Gallego et al., 2019). These findings support the presence of a potential core virome; however, additional longitudinal studies of healthy and disease cohorts are necessary to examine the stability, composition and inter-personal variation of this supposed core over time.

A longitudinal study of the healthy human gut virome of ten individuals over a one-year period, which implemented a database-independent approach, identified a personal persistent virome (PPV) (Shkoporov et al., 2019) (Figure 2A). A total of 22 abundant and temporally stable phage clusters were identified; however, none of these were present in all individuals at one time and the presence and abundance of each cluster was individual specific. These 22 clusters were classified as virulent crAss-like phages, other virulent Caudovirales, and *Microviridae* with only three being identified as temperate. A transiently detected virome (TDV) was also observed but in contrast

to the PPV, it was less abundant and less stable but surprisingly more shared among individuals. The TDV is comprised of *Siphoviridae* and phages such as *Inoviridae* (Figure 2B). Interestingly, CRISPR-spacer analyses found that the PPV was associated with abundant and stable bacterial genera specifically adapted to a healthy gut environment, whereas more transient genera such as *Escherichia* and *Streptococci* were associated with the TDV (Shkoporov et al., 2019). Overall this study clarifies multiple discrepancies in virome studies and is in agreement with a previous study that was unable to support the idea that a defined set of specific phages are shared among all individual (Gregory et al., 2019). These findings also show that the gut is dominated by virulent phages as opposed to temperate phages which only form a subset of the virome (Reyes et al., 2010; Minot et al., 2011; Moreno-Gallego et al., 2019). In summary, there is a predominantly virulent core of phages clusters in healthy individuals that forms an individual specific PPV that results in high inter-personal variation across phageomes. The gut is indeed dominated by dsDNA Caudovirales and ssDNA *Microviridae*, but the most abundant phages are virulent. If temperate phages are not dominant as previously thought, this begs the question of how virulent phages, such as crAss-like phages, maintain themselves in the human gut at high levels over extensive periods of time? *In vitro* studies of Φ crAss001 in serial co-culture with *Bacteroides intestinalis* 919/174 showed that despite propagation at high titres, both phage and host were able to co-exist over 23 days (Shkoporov et al., 2018a). Chapter 3 of this thesis also demonstrates similar dynamics and co-existence for Φ crAss002 and *Bacteroides xylanisolvens* APCS1/XY. These findings are consistent with the observed ability of lytic phages to propagate on their hosts for multiple weeks in the murine intestine with minimal impact on host counts (Weiss et al., 2009; Maura et al., 2012a). The co-existence of virulent phages and their bacterial hosts was recently

examined in a murine model with a defined community. Findings suggests that the spatial variation in the GIT allows co-existence in that phage propagation occurs in the lumen while a subset of the target host population remains in the mucosa or crypts where they are inaccessible to the phage (Lourenço et al., 2019).

With high inter-individual variation across viromes, the identification of viral signals or biomarkers that could differentiate and group individuals based on disease states is a challenge. Clooney *et al.* examined the potential of the virome to distinguish cohorts using a combination of a previously published and an in-house dataset of healthy and IBD viromes (Norman et al., 2015; Clooney et al., 2019). This was performed using a database-independent method based on the whole virome, i.e. both the known viral sequences and dark matter were examined. Protein-based clustering was performed to identify compositional patterns that potentially separated the two cohorts using vConTACT2, a protein clustering program and the same pipeline implemented by Shkoporov *et al.* (Jang et al., 2019; Shkoporov et al., 2019). Interestingly, this protocol also revealed a person-specific virulent core of crAss-like phages and *Microviridae* in health which was absent in IBD patients. IBD viromes were found to have an increased abundance of induced temperate phages and this was reflected by a reduction in bacterial alpha-diversity (Clooney et al., 2019). The characteristic inflammation of this disease may be due to the induction of a proinflammatory immune response as a result of increased bacterial lysis. These findings support the observed expansion of specific phage populations in other IBD virome studies (Norman et al., 2015; Duerkop et al., 2018). Duerkop and colleagues proposed that members of the *Spounaviridae* subfamily could serve a biomarker of colitis (Duerkop et al., 2018). Clooney *et al.* observed an expansion of temperate

Myoviridae, as well as *Siphoviridae* (Clooney et al., 2019). As *Spounaviridae* resolve into the *Myoviridae* family, further investigation of this biomarker is merited.

In summary, this work shows that loss of, or deviation from, the healthy virulent core to a virome with increased temperate phage induction is indicative of IBD. It would be worthwhile expanding on the findings in these studies to identify genus- or species-specific changes and apply the same experimental protocol to other disease states of the GIT to examine if similar disease-specific shifts occur.

2. Phage-host interactions

The maintenance of a phage in an ecosystem is dependent on its ability to infect a suitable bacterial host. However, bacterial hosts can employ an arsenal of defence mechanisms. This leads to antagonism between the two populations, as phages have also evolved strategies to counter bacterial resistance mechanisms. If selective pressures occur, co-evolution can result in genotypic variants due to mutations at specific sites within both entities, such as phage tail fibres and bacterial surface receptors. These complex interactions are believed to be an important shaping force of multiple ecosystems. Several ecological models have been developed in attempt to explain these interactions. In recent years, a number of interesting mechanisms have been observed in this relationship of resistance and counter-resistance. To add another level of complexity, phages are also thought to interact with the mammalian host, both directly and indirectly, via the immune system. For clarity, when discussing phage-host interactions, this refers solely to the bacterial host unless stated otherwise.

2.1 Host defense mechanisms and phage counter defense strategies

Both bacterial defence and phage counter-defence strategies have been reviewed extensively (Labrie et al., 2010; Samson et al., 2013; Seed, 2015; Rostøl and Marraffini, 2019). Here we will summarise the key features and significance of these antagonistic interactions (Figure 3A). Bacteria can employ an array of defence mechanisms when faced with phage predation. These include preventing phage entry, blocking injection of or elimination of nucleic acids, inhibiting the hijacking of replication machinery and programmed cell death. Bacteria can prevent phage adsorption by modifying, masking or differentially expressing surface features such

as lipopolysaccharide and polysaccharide capsules (Bertozzi Silva et al., 2016). Some bacteria produce outer membrane vesicles (OMVs) that extend from the cell surface, irreversibly bind phage and pinch off thus acting as a decoy (Manning and Kuehn, 2011; Schwechheimer and Kuehn, 2015). Interestingly, OMVs were also shown to transfer phage receptors to resistant host cells making them transiently susceptible to phage infection (Tzipilevich et al., 2017). How bacteria use cell surface features to control phage interactions will be discussed in more detail later.

Should the phage overcome these initial barriers the host can employ other forms of defence to interfere with other stages of the phage infection cycle. Bacteria can inhibit the injection of phage genetic material in multiple ways, some of which are mediated by altering the conformation of the injection site or the inner membrane (Lu et al., 1993; Cumby et al., 2015). If the phage nucleic acids enter the cytoplasm, they can be eliminated by restriction modifications systems which cut the invading phage DNA (Roberts et al., 2003; Tock and Dryden, 2005). CRISPR-Cas systems provide bacteria with a means to form an adaptive immunity against phage infection, by cleaving a 20 – 40bp segment of the phage genome and incorporating it into the bacterial genome. This short segment is called a CRISPR-spacer and allows rapid elimination on repeated infection by the same phage by spacer sequence directed cleavage of phage DNA (Horvath and Barrangou, 2010; Wiedenheft et al., 2012). Some bacteria encode phage-inducible chromosomal islands that are excised on phage infection. These interfere with the progeny assembly stage of the infection cycle and can act by interfering with DNA packing for example. Ultimately, this leads to cell death but limits phage dissemination to neighbouring cells (Ram et al., 2012; Seed, 2015). Bacteria can also possess toxin-antitoxin systems which can interfere with different levels of the phage infection cycle. These systems are thought to have

multiple roles in addition to phage resistance (Rostøl and Marraffini, 2019). Three novel anti-phage defence systems, the BREX (bacteriophage exclusion) system, the DISARM system and the cyclic-oligonucleotide-based antiphage signalling system (CBASS), were detected only recently in 2015, 2018 and 2019 respectively (Goldfarb et al., 2015; Ofir et al., 2018; Cohen et al., 2019). BREX inhibits DNA replication and DISARM is a restriction modification type system (Goldfarb et al., 2015; Ofir et al., 2018). CBASS, analogous to the mammalian cGAS-STING innate immune response, is a signalling pathway activated on phage detection that leads to cyclic GMP-AMP production and effector activation resulting in cell death prior to phage progeny release (Cohen et al., 2019). These findings suggest that many novel systems have yet to be discovered. The final resort for a bacterial cell is abortive infection in which the cell shuts itself down to prevent phage replication and release of progeny. This protects sister cells from a similar fate (Dy et al., 2014). Although bacteria have been reported to possess multiple defence mechanisms, none have been reported to possess all possible options, probably due to the associated fitness costs (Bernheim and Sorek, 2019; De Sordi et al., 2019). Indeed, a “pan-immune system” model has been presented suggesting that bacteria can “pick and choose” defence systems from related strains by horizontal gene transfer to overcome the fitness burden of possessing specific defences (Bernheim and Sorek, 2019). When a prophage has integrated and overcome host defences it confers the lysogen with resistance to further phage infection. This phenomenon is called superinfection exclusion (Cumby et al., 2012). However, poly-lysogeny is also possible in which the bacterial genome carries several prophages but the mechanisms mediating this are poorly understood (Argov et al., 2019).

Despite the efforts of bacteria to ward off phage attack, their predators employ mechanisms to fight back. In the case of masked receptors, phages can encode enzymes which degrade capsular polysaccharides (Cornelissen et al., 2012; Leiman et al., 2007). Phages can modify their genomes to avoid restriction modification systems or they can encode anti-CRISPR proteins (Samson et al., 2013; Wiedenheft, 2013). Phages can also mimic host antitoxins to neutralise host produced toxins in host toxin-antitoxin defences allowing the phage to proceed with propagation (Blower et al., 2012). Interestingly, phage encoded CRISPR-Cas systems have also been detected which allow the phage to hijack host defences (Seed et al., 2013). When presented with a resistant host with modified receptors, phages can modify their receptor binding protein (RBP) to continued replication on the same host. For example, lambda phage originally absorbed to its host, *Escherichia coli* B, at cognate receptor LamB but with the introduction of four point mutations into the gene encoding its RBP, the phage was able to bind at a new host receptor, OmpF, while maintaining its ability to bind at the original receptor (Chatterjee and Rothenberg, 2012; Meyer et al., 2012). This bacterial adaptation during resistance and phage counter resistance has given rise to co-evolution which in turn has led to diversity and expansion of host range and strain diversity in ecosystems. Phages are also capable of switching hosts at the strain level and in rarer cases they can infect across species. It is proposed that this may occur when a host becomes spatially inaccessible or reduced in abundance. This was identified in a conventional murine model in which it was discovered that a single point mutation in the tail fibre gene of an *E. coli* LF82 infecting phage, P10, was sufficient to allow the phage to “jump host”. Subsequently, the phage could infect an originally non-permissive strain, *E. coli* MG1655, in the presence of an intermediate host within the murine microbiome. (De Sordi et al., 2017).

It has been proposed that antagonistic co-evolution plays an important role in the diversification of bacteria and phages in the human gut and merits further investigation. However, this is challenging due to the complexity of this ecosystem and the multiple influencing factors to consider (De Sordi et al., 2019). Longitudinal studies have provided some insight into the genetic variation among phages in the human gut. In the previously discussed longitudinal study performed by Shkoporov *et al.*, the accumulation of single nucleotide polymorphisms (SNPs) was monitored over a one-year time period. It was found that novel genotypes replaced the original phage strain in some cases while in other instances multiple genotypes of the same phage co-existed for some time (Shkoporov et al., 2019). These findings are indeed suggestive of arm-race dynamics; however, findings *in vivo* suggest that these dynamics may not be as prevalent in the GIT as previously thought, in particular for persistent virulent phages. It is proposed that the spatial spread of virulent phages and bacterial hosts in the GIT limits the need for selective pressures and allows co-existence without significant emergence of resistant variants (Lourenço et al., 2019). This is indicated by the low emergence of resistant bacterial clones after phage exposure in a murine model and maintenance of phage populations. These findings support observations that have been made in previous studies, including work on crAss-like phages (Weiss et al., 2009; Maura et al., 2012a; Shkoporov et al., 2018a). Conversely, a number of studies have detected the emergence of phage-resistant clones at high levels (Hsu et al., 2019). Although, the percentage of resistant mutants selected may be influenced by a specific phage-host pair or environmental conditions. Significant variation in phage-host dynamics have been observed. For example, T7 phage was found to select for only 20% resistant *E. coli* clones versus 68% resistant *E. faecalis* clones selected for in the presence of VD13 phage *in vivo* (Weiss et al., 2009; Hsu et al., 2019).

2.2 Ecological models and the GIT

A number of ecological models have been developed that describe phage-host interactions in various ecosystems. One of the earliest defined ecological models describing these interactions is the classical kill-the winner model which is derived from the Lotka-Volterra equation (Thingstad, 2000). This model describes the predation of the most abundant bacteria in an ecosystem by their associated phages in a manner that leads to an abrupt reduction in their population. This in turn results in the expansion of another “winner” and a subsequent round of predation leading to fluctuations of the dominant phages and bacteria in an ecosystem (Maslov and Sneppen, 2017). This pattern of predator-prey dynamics is generally not associated with the GIT, at least not at the genus or species level (Reyes et al., 2010; Minot et al., 2011; Shkoporov et al., 2019). Furthermore, it doesn’t take lysogenic conversion into account. This is considered in the piggyback-the-winner model which was previously discussed in relation to VMR. Traditionally, this model proposed the lysogenic conversion of phages in a high VMR (high phage: low bacterial densities) to ensure survival until conditions favoured reversion to lysis. However, more recently lysogenic conversion was also described in the case of a low VMR (low phage: high bacterial densities) in which phages take advantage of and “piggyback” on host prosperity (Knowles et al., 2016; Silveira and Rohwer, 2016).

In relation to the preceding discussion, the constant antagonistic battle that leads to the co-evolution of phages and their hosts has been described as the arms-race dynamics. This describes the constant cycling between infection, resistance, adaptation and counter resistance in phage-host interactions that is responsible for evolutionary pressures (Hall et al., 2011). It has also been proposed that these dynamics can shift towards what is referred to as the fluctuating selection dynamics

model (Gandon et al., 2008). In the context of phages, this model considers the fitness cost that comes with selection for genetic resistance. For example, when a phage initially infects its host they successfully propagate, but with time they select for resistance mutants. This reduces the number of infectible hosts, and in turn phage numbers will decrease which allows sensitive host populations to recover. (Hall et al., 2011; Scanlan et al., 2015).

It is important to consider that these ecological models have been largely discussed in the context of environmental ecosystems such as oceans and may not be directly applicable to the human gut ecosystem. A recently proposed ecological model, although discussed in the context of the marine environment, fits well with our current understanding of virulent phage-host interactions in the human gut. These interactions occur between a stable and persistence dominant core of lytic phages that can co-exist with their hosts. The “royal family” model proposes that kill-the-winner dynamics occur at a host strain or sub-strain level, which is below detectable levels of techniques such as 16S rRNA gene sequencing, and as a result population dynamics appear stable at both the genus and species level over time (Gandon et al., 2008; Breitbart et al., 2018). This is also supported by the phenomenon of “host jumping” which occurs at the strain level (De Sordi et al., 2017). Quorum sensing has also been proposed as a factor that allows phage-host co-existence (Høyland-Kroghsbo et al., 2013).

The ability to come into contact in the gut is also thought to influence phage-host interactions; for example, motility and diffusion can be affected due to variations in mucus concentrations. A “continuous time random walk” model was developed to describe how this can influence the rate at which phage-host interact (Joiner et al., 2019). Further supporting this and the role of spatial heterogeneity in the co-existence and maintenance of virulent phages with their hosts, is the source-sink dynamic model.

This accounts for the radial variation of the gut due to anatomical features and the influence that this has on the spatial dissemination of phages and their hosts (Figure 3B). The mucosa is described as the source, a phage inaccessible refuge for bacteria that can gradually disseminate into the lumen. The lumen acts as the sink where phage propagation occurs on permissive hosts (Holt, 1985; Lourenço et al., 2019). As a result, this creates a gradient of low virulent phage density in mucosal crypts which increases towards the lumen. Interestingly, this is consistent with the idea that populations of intestinal bacteria, particularly commensals, reside in microhabitats such as crypts (Figure 3B). This allows them to form reservoirs and manage the population by re-seeding of the lumen after perturbations (Donaldson et al., 2016). A number of prior studies support the idea of spatial refuges for bacteria to avoid phage predation (Schrage and Mittler, 1996; Bruttin et al., 1997; Weiss et al., 2009). If virulent phage-host interactions do indeed occur in the lumen, stability may result here due to the constant “arms-race” of resistance and counter-resistance between the two antagonistic populations at the strain (De Sordi et al., 2019). This may also account for the genotypic heterogeneity that can develop among certain virulent phages in the personal persistent virome over time (Shkoporov et al., 2019). However, it has also been suggested that “arms-race” dynamics and extension of host range do not have a role in this persistence (Lourenço et al., 2019).

2.3 Other influencing factors and alternative interactions

The ability of a phage to infect its host is not solely dependent on resistance and counter resistance. The lifestyle and physiological state of the target host is thought to be a significant influencing factor on the success of a phage infection. The environmental conditions of the GIT have an important role in this due to the impact

they can have on the metabolic state and stress levels in bacteria (Figure 3B, 3C). These conditions include the physical and anatomical variation along radial and longitudinal axes of the GIT, mucous, bile acids, pH, oxygen, water levels, nutrient levels, peristalsis and dietary components (Koziolek et al., 2015; Vandeputte et al., 2016; Shkoporov and Hill, 2019; Zhao et al., 2019). Furthermore, *in vitro* and *in vivo* laboratory conditions can also influence physiology which may have an important role in experimental outcomes. The role of host physiology in phage-host interactions has been comprehensively reviewed by Lourenço *et al.* (Lourenço et al., 2018).

The physiological state of a bacterial host can influence whether it is phage permissive. This can also generate phenotypic heterogeneity within an isogenic population with part phage sensitive and others resistant (Bull et al., 2014) (Figure 3C). A phenomenon called phase variation (PV) is thought to be an important mediator of transient resistance that results in phenotypic variation in response to environmental conditions. The effects of PV in terms of phage-host interactions were first described in detail in *Haemophilus influenzae* (Zaleski et al., 2005). Recently, the significance of PV in dictating phage-host interactions in the gut has become of interest (Jiang et al., 2019; Turkington et al., 2019; Porter et al., 2020). PV is a mechanism by which bacteria undergo transient phenotypic switching. This occurs by variable expression of gene loci switches due to hypermutation or methylation. This results in genes being reversibly switched “on or off” (Bayliss, 2009). This switching is often mediated by recombinases, integrases or invertases. These enzymes act on invertible regions containing promoters of gene loci often associated with expression of surface features such as *Sus*-like systems, capsule polysaccharides, *S*-layer and Ton-dependent transporters, which can act as phage receptors, outer membrane vesicles, flagella or restriction modification systems (Zaleski et al., 2005; Coward et al., 2006; Nakayama-

Imaohji et al., 2009; Zitomersky et al., 2011; Nakayama-Imaohji et al., 2016; Porter et al., 2020). These phenotypic switches allow the host to adapt to environmental stresses, which are abundant in the human gut, while minimising the fitness impact on the total population (Moxon et al., 1994). A recent study demonstrated that phase variable invertons are more dominant in the human gut compared to other environments, particularly among Bacteroidetes (Jiang et al., 2019). This indicates the importance of this phenomenon in the survival of bacteria in the human gut. An interesting example of how hosts can mediate phage interactions through phenotypic modifications was observed in *Campylobacter jejuni*. It is thought that PV is the most common phage resistance mechanism used by these bacteria (Gencay et al., 2018). Although not a native of the human gut, it was observed that *C. jejuni* was able to co-exist in equilibrium with its infecting phage. Electron microscopy revealed attachment of phage to host cells, noted as flagella absent, without infection resulting in a carrier state. A sub-population of cells with functional flagella were phage permissive and able to support phage replication thus resulting in phage titre remaining almost equal to host counts (Siringan et al., 2014; Brathwaite et al., 2015). Recently, the role of PV and its influence on capsular polysaccharide expression in the evasion of phage predation was effectively demonstrated in *Bacteroides thetaiotaomicron* (Porter et al., 2020). This study identified as many as nineteen host cellular functions could dictate the ability of the phage to infect its host. The phenotypic heterogeneity that PV generates is believed to act as a form of herd immunity equivalent to that seen in human disease immunity. By limiting the viral load, the impact on sensitive variants is reduced thus preventing a drastic reduction in the overall population (Turkington et al., 2019). The use of herd immunity style dynamics to limit phage predation was further supported by the persistent propagation of virulent phages on a microcolony

without its elimination (Eriksen et al., 2018). The variability of conditions in the GIT likely generates significant PV mediated heterogeneity within bacterial niches. With the presence of spatially distributed phage-permissive and non-permissive hosts variants in parallel, it is possible to stably maintain both virulent phages and their associated hosts. Incorporation of transcriptomic profiling into experimental design may provide further insight into the role of PV in phage-host interactions and its implications for phage therapy.

Interestingly, temperate phages can hijack host quorum sensing to monitor metabolic conditions and population densities, thus granting them the ability to make informed lytic-lysogenic switch decision (Laganenka et al., 2019). For example, this was observed to occur for *E. coli* phage T1 which encodes a transcriptional regulator, Pir. In the maintenance of lysogeny, *pir* expression is inhibited by host produced cAMP-CRP. In conditions of efficient host metabolism, repression of *pir* is lifted by increased influx of sugar levels. If cell densities are sufficient, indicated by high levels of autoinducer, the phage commits to lysis as conditions are optimal for progeny production (Laganenka et al., 2019).

Quorum sensing (QS) is a phenomenon employed by bacteria that allows controlled expression of specific genes for processes such as virulence and biofilm formation. This occurs through cell-cell communication mediated by bacterial produced extracellular signalling molecules, autoinducers. These molecules allow bacteria to monitor population density and make an informed collective decision on whether expression of specific genes is a fitness cost or beneficial to the overall population (Whitehead et al., 2001; Turovskiy et al., 2007; Ng and Bassler, 2009; Papenfort and Bassler, 2016). Bacteria can use QS to assess the risk of phage predation and act accordingly. For example, *E. coli* can reduce expression of λ phage receptors

in response to an autoinducer thus limiting phage adsorption and as a result limits the effect on population levels (Høyland-Kroghsbo et al., 2013). However, QS homologs have also been identified in phage genomes deposited in NCBI databases. The presence of a QS cassette has been reported in a temperate *Clostridium difficile* infecting phage, phiCDHM1 (Hargreaves et al., 2014). More recently, it was shown that *Vibrio cholerae* infecting phage VP882 encodes a QS receptor (VqmA_{Phage}) homologous to that of the host allowing it to “listen in” on its host via host produced autoinducer (DPO). In turn, the phage can monitor population densities thus allowing an informed lytic/lysogenic lifestyle switch. In high host cell density, lysis occurs resulting in optimal propagation (Silpe and Bassler, 2019). Interference with QS regulation of virulence factors in *P. aeruginosa* have been shown to cause repression of CRISPR-Cas defences. This in turn makes the host more susceptible to phage attack and may have useful application in phage therapy (Høyland-Kroghsbo et al., 2017). Phages are believed to interfere with host QS to their advantage. Indeed, this was recently shown to be true for the lytic *Pseudomonas* phage LUZ19 (Hendrix et al., 2019).

The above demonstrates that there are a multitude of factors that influence the interactions and dynamics between phages and their bacterial hosts. These factors are important for the success of both in the complex gut environment and are summarised in Figure 3.

2.4 Interactions with the mammalian host

In addition to phage-bacterium interactions, phages have also been demonstrated to interact with their mammalian host adding another level of complexity when studying the human gut phageome.

Ig-like domains have been identified in certain phages which help them to reside in the outer mucosal layer in the GIT. This led to the proposition of the “bacteriophage adherence to mucus” (BAM) model which describes the idea that such phages protect the mammalian host from bacterial infections (Barr et al., 2013). However, it has also been demonstrated, *in vitro*, that phages can migrate across the mucus layer, cross the gut epithelial barrier or enter systemic circulation either by transcytosis or by peptide-guided transport and subsequently interact directly with cells of both the innate and adaptive immune system (Bellegheem et al., 2017; Duerr et al., 2004; Moustafa et al., 2017; Nguyen et al., 2017; Van Bellegheem et al., 2019). This was further supported by *in vivo* studies in a germ-free murine model that examined the use of endotoxin-free lytic phages in the alleviation of bacterial driven colon cancer. The administered phages, specifically Caudovirales phage with DNA and not empty capsids, interacted directly with the murine immune system. This was indicated by interferon production (virus-specific immunity) which led to the exacerbation of colon inflammation (Gogokhia et al., 2019). Although it has been shown that phages can induce an inflammatory response, it is thought that their interactions with the immune system are generally anti-inflammatory; however, this is still debated (Van Bellegheem et al., 2017; Zhang et al., 2018b; Van Bellegheem et al., 2019). These interesting tripartite interactions were extensively reviewed by Van Bellegheem *et al.* (Van Bellegheem et al., 2019). Findings thus far suggest that phages can directly influence mammalian health and indirectly shape the gut microbiome via

interactions with immune system; however, we still have little understanding of the mechanistic details and further *in vivo* studies are required. Considering the various phage interactions above, there are multiple layers of conditions and factors to take into account when studying the human gut phageome.

3. Current methodologies, challenges and potential solutions

In the past decade there have been many significant advances in our understanding of the human gut virome. Findings to date show that the healthy gut virome is individual specific and temporally stable with a lytic core that is personal specific persistent and dominated by crAss-like phages, other Caudovirales and *Microviridae*, but which are not shared among all individuals. However, many important questions remain unanswered and our overall understanding of this elusive component of our gut microbiome still remains poor in comparison to our understanding of the bacterial fraction. This is largely due to challenges with *in silico*, *in vitro* and *in vivo* methodologies. Efforts are being made to highlight current shortcomings, to develop and standardise protocols, and improve overall reproducibility and comparability. This is essential to truly understand the importance of the phageome in health and disease and to take advantage of phages for therapeutic applications.

3.1 Viral metagenomics, sequencing and *in silico* methods

Next generation sequencing technologies, bioinformatic tools and metagenomic studies have provided important insights into the human gut phageome such as profiling of community composition and *in silico* identification of novel phages. However, the protocols implemented in these studies are often not optimised and don't lead to a true representation of the phageome. The majority of these studies have used faeces as the sample source of phage genetic material due to the practical difficulties with using other sampling sites of the GIT (Reyes et al., 2010; Minot et al., 2011; Dutilh et al., 2014; Norman et al., 2015; Guerin et al., 2018; Devoto et al., 2019;

Shkoporov et al., 2019; Clooney et al., 2019). In brief, the key protocol steps involved in the majority of viral metagenomics studies includes VLP enrichment, nucleic acid extraction, sequencing library preparation, followed by *in silico* characterisation and annotation of generated viral sequences using bioinformatic pipelines. However, due to the lack of universally reproducible methods being implemented in this field, cross-study comparisons of findings are difficult, and disparities can occur between different studies.

The first step in viral metagenomic studies is the removal of non-target contaminants from faeces (or other biological samples) and the enrichment of VLPs to ensure optimal yield and quality of nucleic acids for library preparations (Kleiner et al., 2015). This involves the elimination of prokaryotic and eukaryotic DNA and RNA as well as cellular and dietary debris using physical and chemical process such as homogenisation, filtration, chloroform treatment and enzymatic action (Conceição-Neto et al., 2015). There are multiple phage nucleic acid isolation protocols described and while CsCl gradient centrifugation can yield highly purified samples, this protocol is time consuming, laborious and can introduce bias (Castro-Mejía et al., 2015; Kleiner et al., 2015). It is important to consider that the phageome is comprised of virulent and temperate phages that vary in nucleic acid content (ssDNA, dsDNA, ssRNA, and dsRNA). The challenge in developing a protocol to consider all these components is difficult and can affect our overall picture of human gut phage profile. Due to the shared physio-chemical properties among temperate and virulent Caudovirales phages, many techniques have been developed to allow their co-purification. However, it is important to note that using the incorrect protocol can lead to the generation of a biased virome composition (Gregory et al., 2019). The majority of metagenomic studies have focused on enrichment of VLPs followed by extraction of

nucleic acids, although many of the protocols promote a biased focus on free-phage and the enrichment of DNA phages (Kleiner et al., 2015). Whole community metagenomic sequencing is thought to be more informative in providing data with into phage-host interactions in the gut than that derived from VLP metagenomics which may exclude certain phage types including RNA phages and integrated phages (Ma et al., 2018).

Protocols may also generate low nucleic acid yields and to overcome this, amplification methods are performed such as multiple-displacement amplification (MDA) (Kim et al., 2011; Marine et al., 2014). However, this leads to the biased amplification of ssDNA viruses such as *Microviridae* by as much as 10-fold, which in turn, skews phage community composition and is thought to be the confounding factor behind why certain viral datasets have an extremely high relative abundances of *Microviridae* (Kim et al., 2011; Roux et al., 2016; McCann et al., 2018; Garmaeva et al., 2019; Shkoporov and Hill, 2019; Gregory et al., 2019; d’Humières et al., 2019). Many virome studies interpret findings based on relative abundance and thus results may be skewed where MDA has been implemented (Sutton and Hill, 2019). MDA has also been shown to affect phage diversity and method reproducibility (d’Humières et al., 2019). MDA can introduce GC bias and coverage extremes which can have implications for *de novo* genome assemblies (Chen et al., 2013; Sutton et al., 2019). Having identified the effects of amplification on viral nucleic acids, there is a call to eliminate this step from metagenomics protocols. However, this is challenging due to the variable concentration yields of phage genetic material that can be acquired from samples due to the differences in the viral loads of samples and the strict input requirements for sequencing library preparation (Shkoporov and Hill, 2019). For example, one study found that DNA yields, despite enrichment, could range from 4 –

500 ng of DNA per gram of faeces (Shkoporov et al., 2018b). Attempts have been made to improve current sequencing library preparations, such as the adaptase-linker amplification method, by excluding the need for MDA with the goal of eliminating downstream issues and providing a more accurate representation of phage composition in a community (Roux et al., 2016).

The elimination of bacterial contamination has also been a challenge both in terms of viral nucleic acid extraction and for *in silico* analyses. Although it is difficult to completely avoid, efforts should be made to limit levels as it is thought to impact on the validity of viral sequences deposited in database (Roux et al., 2013). In the enrichment of VLPs and extraction of viral nucleic acids, the choice of filter pore size is important to avoid contaminants. The use of a larger pore size filter such as 0.8 μm is less biased as it allows larger viruses to be retained but this also leads to high levels of bacterial contamination. Studies indicate that a pore size of 0.45 is optimum (Conceição-Neto et al., 2015; Shkoporov et al., 2018b). How bacterial contamination among viral sequences can impact on conclusions drawn from findings was highlighted in a study that demonstrated the over-representation of antibiotic resistant genes of bacterial origin among phage genomes (Enault et al., 2017). The issue of contamination was further highlight by Zolfo and colleagues who examined metagenomic virome sequences from 35 studies and detected an abundance of bacterial, archaeal and fungal contaminants irrespective of VLP enrichment technique (Zolfo et al., 2019). Bioinformatic pipeline development and standardisation of criteria for the detection of bacterial contamination among viral sequences is important in ensuring the quality of viral databases and the *de novo* detection of phages. Shkoporov *et al.*, recently developed an effective pipeline to eliminate bacterial contamination by aligning reads against a database of the conserved bacterial genes, *cpn60* and setting

strict cut-offs to identify true viral contigs (Shkoporov et al., 2018b). ViromeQC also has significant potential in aiding the detection of contaminants. This pipeline was developed to allow stringent quality control of viral sequences using 31 universal bacterial genes in addition to the 16S/18S rRNA genes and 23S/28S rRNA genes (Zolfo et al., 2019). In addition to contamination, Shkoporov *et al.*, also considered the effects of sample storage conditions and operator bias in virome analyses and has proposed the use of an exogenous phage standard to spike faecal samples to allow absolute quantification of total viral loads across samples (Shkoporov et al., 2018b). The disparity between relative and absolute quantification in microbiome profiling can be significant (Stämmler et al., 2016; Vandeputte et al., 2017). This issue was recently highlighted by Shanahan and Hill. They discussed the issue of microbiome misrepresentation due to relative abundance and how it can mask variations between microbiomes that would be highlighted by absolute abundances (Shanahan and Hill, 2019). Thus, the inclusion of a standard to allow absolute quantification of sequences and viral load should form part of future virome studies to give a more accurate representation of composition. It has also suggested that viral load may have potential as a biomarker (Sutton and Hill, 2019).

Many protocols in phageome studies have been developed with a particular focus on DNA phages. This results in RNA phages being underrepresented in viral databases and highlights the need for developing of protocols specific to the isolation of these phages (Callanan et al., 2018). This will not be without difficulties due to the sensitive nature of RNA. The potential array of human gut RNA phages was recently highlighted with the *in silico* discovery of over one thousand near complete RNA phage genomes and over fifteen thousand non-redundant genomes by screening 82 publicly available metatranscriptomic datasets, generated from activated sludge and

aquatic environments studies, using profile hidden Markov models to detect conserved proteins. Overall this represents a 60-fold increase in identified ssRNA phages (Callanan et al., 2020).

The choice of sequencing platform for virome analysis can influence the read length output. Following quality control, reads are assembled into contigs generally by referenced-based or *de novo* assembly methods (Garmaeva et al., 2019). The challenges associated with former method will be discussed below. Short read platforms allow deep sequencing with low error rates and require low DNA concentrations. However, *de novo* assembly of short reads is testing due to the modular nature of phage genomes, strain heterogeneity, high incidences of hypervariable and repeat regions and MDA influenced variations in coverage and GC content (Lima-Mendez et al., 2011; Minot et al., 2012, 2013; Chen et al., 2013; Sutton et al., 2019). This leads to fragmented assemblies and hampers downstream *in silico* analyses (Sutton et al., 2019). It has been suggested that the use of long-read sequencing platforms, such as Oxford Nanopore, which can generate reads that are representative of near complete genomes may overcome the issues associated with short-read sequencing (Somerville et al., 2019; Warwick-Dugdale et al., 2019). The major drawback with this method is that high DNA concentrations are required which, as discussed previously, can be difficult to attain. Therefore, when performing assemblies from short reads, the choice of assembly software and stringent criteria to recruit true viral sequences is essential (Roux et al., 2017; Sutton et al., 2019).

In recent years, sequence-based methods have played an important role in providing insights into the poorly understood human gut phageome. Nevertheless, there is a pressing challenge for metagenomics and *in silico* analyses in current phage research. This is the dependency on insufficiently developed viral databases. The

majority of phage sequences in these databases are not annotated due to the lack of homology with known phages. When newly generated sequence reads are aligned to these databases as much as 99% of the reads from a dataset can fail to align to known phage genomes or homologs and thus these reads remain as dark matter (Aggarwala et al., 2017). As a consequence, multiple studies have based their findings solely on the identifiable fraction of reads and have excluded the abundant unexplored dark matter. Although these studies have identified cohort variation, it is unknown how the inclusion of viral dark matter would influence and validate the overall findings (Norman et al., 2015; Lim et al., 2015; Monaco et al., 2016; Zuo et al., 2019). To overcome this issue, there has been a move towards the use of database- independent methods such as open-reference and clustering approaches to include both the identifiable and unknown fraction of reads in a dataset (Shkoporov et al., 2018b; Clooney et al., 2019; Moreno-Gallego et al., 2019; Shkoporov et al., 2019). Protein-clustering programs, such as vContact2, offer an effective solution and allow the implementation of a whole virome analysis (Bolduc et al., 2017). These work by building a gene-sharing network based on shared protein families across genomes, an approach similar to that employed in the development of the crAss-like phage family taxonomy (Yutin et al., 2018; Guerin et al., 2018; Jang et al., 2019). In addition, this approach overcomes the challenge of detecting cohort variation within datasets due to the high inter-individual variation associated with the human gut phageome (Clooney et al., 2019; Shkoporov et al., 2019; Sutton and Hill, 2019). The power of this approach, in the context of the human gut, was recently demonstrated in a study by Shkoporov *et al.* This study discovered a predominantly lytic core in healthy adults that is persistent, temporally stable and individual specific (Shkoporov et al., 2019). This is contradictory to the temperate core identified in other studies (Reyes et al.,

2010; Minot et al., 2011; Moreno-Gallego et al., 2019). These findings were further supported in a study examining the IBD virome by Clooney *et al.*, which implemented the same database-independent methodology. A healthy lytic core was once again identified that was absent in IBD patients whose viromes were dominated by induced temperate phages (Clooney et al., 2019). This analysis included both an in-house and a previously published IBD dataset that was analysed using database-dependent methodology (Norman et al., 2015). The whole virome analysis applied in the replicate study identified a number of findings that were contradictory in relation to the initial study (Clooney et al., 2019). This highlights the importance of looking at the virome as a whole. The use of these methods will also gradually lead to more complete viral databases and reduce the incidence of incomplete analyses. Also, where annotation is possible, it is important to consider that some phages use alternative genetic codes such as Lak phages and crAss-like phages of specific candidate genera (Guerin et al., 2018; Devoto et al., 2019).

Although the number of phage genomes being generated is increasing rapidly due to sequencing technology and *in silico* tools, many fail to be taxonomically assigned (Korf et al., 2019). Many of the newly deposited phages are *de novo* assembled without *in vitro* characterisation and therefore do not necessarily slot into a morphology based taxonomic scheme (Ackermann, 2012). As a result, there is a significant push towards a genome-based taxonomic classification to allow universal and accurate taxonomic assignment of phages that have not yet been culture *in vitro* (Meier-Kolthoff and Göker, 2017; Simmonds et al., 2017; Aiewsakun et al., 2018; Eloë-Fadrosch, 2019). This is not without challenges due to the current incompleteness of viral databases, the mosaic structure of phages and the lack of a universal taxonomic marker shared across phages (Lima-Mendez et al., 2011; Shapiro and Putonti, 2018).

The application of gene-sharing networks, protein clustering, and whole virome analysis provides possible methodologies to aid the development of such a scheme (Clooney et al., 2019; Jang et al., 2019). This can provide insights into evolution and shared functions across phages (Shkoporov and Hill, 2019). In summary, the above highlights the need for the development, optimisation and standardisation of protocols in many aspects of viral metagenomics and *in silico* tools. CRISPR-spacer analysis of phages identified *in silico* allows for host predictions (Zhang et al., 2018a). For example, prototypical crAssphage host predictions were done in this manner in conjunction with BACON domain analysis (Dutilh et al., 2014). These predictions were later confirmed with the isolation of Φ crAss001 and its associated host, *Bacteroides intestinalis* (Shkoporov et al., 2018a). However, without *in vitro* isolation of these *in silico* phages, insights into biological characteristics remain limited. CRISPR-spacer analyses can aid the phage-host pair screening process, by predicting associated hosts at a species level (Paez-Espino et al., 2019). With time, it is hoped that strain level predictions will be possible that will greatly aid *in vitro* and *in vivo* studies.

Sequence-based analyses have provided significant insights into the composition of the human phageome which would not be possible *in vitro*. Despite this, *in silico* analysis alone can provide little insights into phage-host interactions and biological properties of *de novo* phages as is possible with *in vitro* and *in vivo* methods. However, both methodologies in unison will be essential in fully understanding the interactions between phages and the microbiome in the gut and how these relationships ultimately impact on health.

3.2 *In vitro*, *ex vivo* and *in vivo* methods

With recent findings, a number of questions have arisen regarding the human gut phageome. The observed ability of virulent phages to stably co-exist with their host over time remains poorly understood and requires further investigation (Shkoporov et al., 2018a, 2019). The isolation of dominant gut phages and their hosts, such as the crAss-like phages and *Microviridae* which form the virulent core of the phageome in healthy individuals, would provide insights into the mechanisms behind these interactions. Isolation of transient gut phages may also help us understand why certain phages are not persistent. CRISPR-spacer analyses have provided some directions for targeting potential hosts (Shkoporov et al., 2019). One of the main challenges with the isolation of novel phage-host pairs, their biological characterisation, and examining their interactions under laboratory conditions, is the ability to mimic the complex conditions of the GIT from where they originate.

Many gut bacteria are difficult to culture as they are often strict anaerobes and the conditions necessary for their growth *in vivo* is unknown. Metagenomic studies have demonstrated the diversity of bacterial species in the human gut, however, this provides limited value when an isolate is required *in vitro* (Gill et al., 2006; Turnbaugh et al., 2007; Huttenhower et al., 2012). Until recently, as many as ~80% of the bacteria identified by these studies were uncultured or unculturable (Eckburg et al., 2005). Nevertheless, significant efforts are being made to culture the “unculturable”. As a result, there has been a push towards the isolation and culturing of bacteria using novel techniques with the complement of genome sequencing. Culturing of GIT bacteria is not only necessary to improve our understanding of their role in the microbiome, it is also key for investigating phage-host dynamics and the validation and expansion of *in silico* findings.

An example of recent advancements is The Human Gastrointestinal Bacteria Culture Collection which was compiled with cultured human GIT isolates and their sequences with the goal of improving the accuracy of metagenomic analyses and archiving novel isolates (Forster et al., 2019). Improvements in culturing methods has played an important role in cultivating bacterial species of the human gut. Media developments have aided the isolation of novel obligate anaerobes from human faeces. For example, yeast extract, casitone and fatty acid (YCFA) medium has been shown to support the growth of these bacteria to high levels and can be modified with antibiotics, carbohydrates or other components to select for less abundant species or phenotypes of interest (Duncan et al., 2002; Browne et al., 2016; Das et al., 2018; Forster et al., 2019). YCFA modified with carbohydrates and antibiotics was applied in the successful isolation of bacterial hosts in Chapters 3 and 4 of this thesis.

Culturomics has also proved to be an important tool in expanding the repertoire of bacterial species isolated from the human gut (Bilen et al., 2018). This involves coupling high-throughput culture-dependent and culture-independent methods. Samples from which bacteria are to be isolated from are tested in multiple optimised culture conditions and rapid identification of isolates is performed using MALDI-TOF MS or 16S rRNA gene sequencing (Lagier et al., 2012, 2016). In the context of the human gut, culturomics has been successfully applied in the isolation of bacteria from faeces, small intestine and colonic samples (Lagier et al., 2016). The potential of this method was reviewed by Lagier and colleagues (Lagier et al., 2018). It is also important to consider screening conditions in relation to phenotypic features associated with intestinal bacteria. For example, the human gut contains a significant population of spore-forming bacterial species which are troublesome to culture (Abecasis et al., 2013; Rajilić-Stojanović and de Vos, 2014; Browne et al., 2016).

Browne and colleagues successfully cultured an array of spore-formers from faecal samples using ethanol shock enrichment to distinguish them from vegetative cells (Riley et al., 1987; Browne et al., 2016). This led to the isolation of 137 bacterial species from six healthy individual, 90 of which were on the Human Microbiome Project's "most wanted" list of species that had yet to be cultured and sequenced (Browne et al., 2016). This emphasises that even the most stubborn gut residents can be cultured *in vitro* when the correct conditions are applied; however, identifying these conditions can be fastidious.

In silico analyses, such a co-occurrence networking of metagenomic data, can also help guide the identification of suitable combinations of bacterial species in co-culture (Das et al., 2018). This is an important consideration to limit significant antagonisms between bacteria when examining phages in a defined microbial community and may have implications for the validity of findings.

Faecal samples have been widely used when screening for novel gut phages. Traditional methods of plaque and spot assays are still widely used in this process. This involves screening a phage rich suspension, such as faecal filtrate, from which bacteria and debris have been removed, against a lawn of pure bacterial culture in overlay agar. A successful phage-host pair is indicated by host lysis through spot or plaque formation, which are picked, enriched, and purified on the specific host (Furuse et al., 1978; Kai et al., 1985). Due to the variable abundance of different phages in faecal samples, enrichment is often performed to allow phage-host interactions establish and increase titres to detectable levels which is often indicated by clearing of culture (Salem et al., 2015). This is generally followed by sequencing of the phage genome. Isolating phages on a single strain has been the general practice, although it has been shown that enriching phages on strain variants of a specific host can increase

host range of a phage which may have useful applications in phage therapy (Hyman, 2019). Plaque assays may fail to detect temperate phages unless an induction treatment, such as mitomycin C, is performed to induce the phage to a lytic state (Castellazzi et al., 1972). It has also become apparent that not all lytic phages form clear plaques or spots on a suitable host nor clear liquid cultures (Porter et al., 2020) (Chapters 3 and 4 of this thesis). This can be due to phenotypic heterogeneity within an isogenic population, resulting in both phage permissive and non-phage permissive variants within a single culture. This can be mediated by phenomena such a phase variation which was discussed above. This is an important consideration when screening for phage-host pairs and highlights the need to avoid focusing on traditional signs of phage-host pairing such a plaque assays, zones of clearing and culture lysis.

Enrichment of phages on intestinal bacteria, either acquired from culture collection or through selective enrichment from faeces, followed by shotgun sequencing is a useful approach for screening of novel gut phages. This approach led to the isolation of Φ crAss001, the first member of the elusive crAss-like phage family to be isolated in pure culture with its host *Bacteroides intestinalis* (Shkoporov et al., 2018a). To achieve this, 20 healthy faecal donors were recruited, faecal filtrates were prepared, pooled and screened against 53 bacterial strains in modified YCFA broth under anaerobic conditions over three days, with each filtered lysate used to inoculate the subsequent round of enrichment. The enriched lysates were then sequenced to identify expansion on a specific host (Shkoporov et al., 2018a).

Chemostats are convenient for phage-host enrichment from faeces. They allow more controlled conditions than can be provided *in vitro* and are useful for examining phage-host interactions (Santiago-Rodriguez et al., 2015a). Chapters 3 and 4 discuss the use of antibiotics to selectively promote the growth of Bacteroidales *ex vivo*. This

was performed in a chemostat using frozen standard inoculum prepared from faeces (O'Donnell et al., 2016). This in turn allowed the forced expansion of Bacteroidales and associated phages and the generation of a phage-rich fermentate. The fermentate was then screened against individual bacterial species by serial co-culture followed by plaque assay attempts and sequencing. The bacterial species that were tested were selectively enriched *in vitro* from the same faecal sample used to prepare the fermentate, in the presence of the same antibiotics. This ultimately led to the isolation of two novel phage-host pairs, Φ crAss002 that infects *Bacteroides xylanisolvens* and Φ PDS1 that infects *Parabacteroides distasonis*. A similar protocol could be applied in the isolation of further novel phage-host pairs from the GIT.

Culture-independent methodologies have also been applied in the isolation of novel phage-host pairs. Within samples used for screening there are constantly a percentage of bacterial cells that either have free virions attached to their surface or contain phages internally either due to replication or as prophages. A number of protocols take advantage of this to match phages and their hosts.

Viral tagging (VT) has proved to be a useful tool in linking phages to their host. This method was originally used to examine marine viruses. A sample of interest is obtained, and virions are randomly tagged with a generic fluorescent marker that binds to nucleic acids. The tagged phages are then applied to potential bacterial hosts. If interactions occur, fluorescence-activated cell sorting (FACS) can isolate and discriminate between tagged phage-host pairs and free virions. The isolated phage can then be sequenced. This methodology can also provide insights into phage-host interactions (Deng et al., 2012, 2014). Single bacterial cells can be isolated from samples using FACS. This allows complete sequencing of the bacterial genome to be performed and in parallel can also capture the genomes of phages which were infecting

or attached to the cell. This method of identifying phage-host pairs is called single amplified genomes (SAGs) (Swan et al., 2013; Labonté et al., 2015). It is also possible to directly isolated single uncultured viruses from environmental samples using flow cytometry. This allows for single viral genomics (SVGs) which involves sequencing isolated viruses individually which overcomes certain assembly limitations associated with metagenomics and provided insights into strain variation and genetic diversity (Allen et al., 2011; Martinez-Hernandez et al., 2017). In one case, this method allowed the recovery of genetic information from 5,000 individual viruses sorted from a marine sample (Martínez et al., 2014). SVGs, however, excludes the host so provides little insight into phage-host pairs or interactions. FACs, although very useful in its applications, is biased toward the isolation of more abundant viruses (Martinez-Hernandez et al., 2017; Lawrence et al., 2019).

VT proves to be a particularly powerful tool for examining phage-host interaction. It has been largely applied with known panels of bacteria but could have potential in identifying pairing between both unknown bacteria and novel phages from a specific sample. A workflow incorporating single-cell VT in parallel with metagenomics and SAGs has been developed using human faecal samples. This led to the identification of 363 novel phage-host pairs (Džunková et al., 2019). Isolating a large panel of phage-host pair in this manner highlights the current advancements in this field. However, once isolated, it is still necessary to carry out characterisation of biological properties and mechanisms involved in the phage-host pairing.

As mentioned previously, many studies use faeces as the sampling source for examining the virome due sampling difficulties associated with other sites of the GIT. It has been shown that consistency and transit time of faeces can influence bacterial composition and in turn influence variation in clinical studies, suggesting that the

importance of recording this information (Vandeputte et al., 2016). This can also have a knock-on effect for virome studies. Faeces is the most easily accessible and ethical sample source for examining the GIT; however, it is important to consider that phage-host interactions vary radially and longitudinally along the GIT (Maura et al., 2012b; Galtier et al., 2017). Biogeographical sectioning of the microbiota was extensively reviewed by Donaldson and colleagues. The importance of bacterial microhabitats, such as in crypts, along the GIT is thought to play an important role in providing bacterial reservoirs which maintain homeostasis and allow rapid recovery of bacterial populations after perturbations (Donaldson et al., 2016). (Figure 3B). This is likely to have an important impact on phage-host interactions and co-existence dynamics. Martinez-Guryn and colleagues also described region to region variation of the GIT, from oral cavity to distal colon, physical, chemical and biological factors such as cell type and surface variation, mucus composition, pH, oxygen, bile acids and immune factors. This creates functional heterogeneity within the GIT and influences the regional variation in bacterial composition (Martinez-Guryn et al., 2019) (Figure 3B, 3C). This heterogeneity along with bacterial microhabitats plays an important role in influencing phage composition and interactions along the GIT. A recent study examined the virome along different sites of the GIT in nonhuman primates. The virome of the large intestine and rectum were found to be similar but distinct from the ileum. These findings indicate that faeces provide a good representation of the colon specifically rather than the GIT as a whole (Zhao et al., 2019). Microbial signatures associated with Crohn's disease (CD) were examined in CD children and were observed in mucosa; however, these signatures went undetected during the assessment of stool samples (Gevers et al., 2014). Therefore, focusing solely on stool samples would have had significant implications for the findings of this study. Faeces doesn't

necessarily represent the virome of each region of the GIT, but it is a useful proxy for providing insights into the gut phageome. However, considering the above, future studies will need to extend beyond the scope of stool samples if we are to truly understand the human gut phageome. Naturally this is not without challenges and highlights the need for developing *in vivo* models.

The examination of phages at multiple experimental levels (*in vitro* to *ex vivo* to *in vivo*), the importance of incorporating relevant conditions to experimental models and considering the influence of host metabolic state in phage-host interactions was recently discussed by Lourenço and colleagues (Lourenço et al., 2018). These are important factors in developing an experimental pipeline. The advantages and disadvantages of using defined bacterial communities in *in vitro* and *in vivo* models for studies of the human gut was recently reviewed by Elzinga and colleagues (Elzinga et al., 2019). Several studies have demonstrated the usefulness of *in vivo* models in gaining insights into phage-host interactions (Maura et al., 2012a, 2012b; Reyes et al., 2013; De Sordi et al., 2017; Galtier et al., 2017; Hsu et al., 2019). While *in vitro* studies can be of great importance, they are not necessarily always good representatives of realistic conditions. Here, we will discuss two examples where *in vivo* models proved particularly useful compared to *in vitro* analyses. De Sordi *et al.* examined the phenomenon of phage host strain jumping *in vitro* and *in vivo* in a conventional murine model; however, host jumping was only detected under the latter conditions (De Sordi et al., 2017). This emphasizes the importance of testing under multiple experimental conditions. In another study, a gnotobiotic mouse was colonized with ten human intestinal strains followed by introduction of four lytic phages. The phages were shown to knock down levels of their target hosts but without complete elimination resulting in co-existence of virulent phages and their hosts. In parallel, a knock-on

effect on non-target community members was observed which in turn has an influence on the gut metabolome and mammalian host (Hsu et al., 2019). However, perturbations of off-target bacterial communities have not been observed in other murine studies (Galtier et al., 2016; Cieplak et al., 2018; Lourenço et al., 2019). Therefore, this topic requires further investigation. Murine models also overcome the issue of acquiring samples that would be ethically difficult to access in the case of humans, such as mucosa or biopsies, and they also allow spatial heterogeneity in the GIT to be examined (De Sordi et al., 2019). Interestingly, recent work found that the phageome in healthy and colitis murine models shared some overlap with the phageome of healthy and IBD human patients. This suggests that the murine model is conducive to human-like GIT conditions making it suitable for studying phage-host interactions and the virome in diseases such as colitis (Duerkop et al., 2018).

There are a number of other techniques that can be incorporated into human gut phageome studies. Transcriptome profiling and metatranscriptomics can provide insights into gene expression during individual or community phage-host interactions. Few studies have examined this in the context of the human gut. Transcriptome profiling of phage communities in the oral cavity was recently carried out to compare gene expression in health versus periodontal disease (Santiago-Rodriguez et al., 2015b). Such studies should also be performed to compare gene expression of phages in a healthy compared to diseased gut. This technique is also useful for examining changes in bacterial host gene expression in the presence of a phage. Transcriptomics recently demonstrated the significance of phase variation in *B. theta* during phage infection (Porter et al., 2020). It can also determine the effect of an experimental model on bacterial host gene expression (Denou et al., 2007). This can aid the development of experimental pipelines and avoid inaccurate conclusions.

Bioelectronics has provided many useful biological and biomedical applications in recent years (Strakosas et al., 2015; Pitsalidis et al., 2018). With this, it is possible to mimic bacterial and mammalian membranes, called supported lipid bilayers, using bioelectronic tools and membrane biosensors which can detect specific interactions. For example, this method was used to examine membrane interactions with antibiotics to determine antibiotic targets that were discriminatory between bacterial and mammalian membranes (Su et al., 2019). Similar tools were used to monitor *Salmonella typhimurium* infection of epithelial cells (Tria et al., 2014). Bioelectronics could also have useful applications for detailed examination of phage-host interactions.

In summary, it is necessary to make improvements to bioinformatic pipelines and experimental protocols so that there is greater reproducibility allowing more accurate cross-study comparisons. Methodology can have a greater influence on the virome than the examined disease state (Gregory et al., 2019). The influence virome analysis protocols have on findings and the pros and cons of specific techniques used in analyses have been highlighted (Sutton and Hill, 2019). It is also important to use relevant conditions when examining phage-host interactions. Transcriptomic profiling should also be incorporated into studies of the human gut phageome. Overall, the progression in metagenomics is more rapid than in the isolation and characterisation of novel phage-host pairs from the human gut. Using a top-down approach that links metagenomic findings to the screening protocol may allow a more targeted approach. For example, metagenomics can aid the primer development for qPCR detection and monitoring of phage propagation *in vitro* or *in vivo*. Novel phages that have been identified *in silico* need to be isolated and characterised *in vitro* and *in vivo* to understand the mechanisms they employ when interacting with their host, and the

health and disease implications that this may have for the human host. Experimental pipelines that bring the above together will be key to this progress (Figure 4). Furthermore, we need to consider the most recent findings with regard to the human phageome, in particular the individual-specific healthy core phageome. This needs to be further elucidated and it will also be necessary to reconsider our current application of ecological models to the human gut. As many of these models were developed in the context of other ecosystems, it may be necessary to develop new models to explain the low VMR, the temporal stability, and the persistence of virulent phages and their bacterial hosts that is observed in the human gut.

3.3 Case Study: The crAssphage family story: from *de novo* assembly and *in silico* characterisation to an *in vitro* reality

CrAssphage provides an important example of novel phage discovery *in silico* using methods that overcame the limits of database-dependency. For years, following this discovery, crAssphages largely existed only *in silico* due to the challenges faced in achieving isolation on a suitable host. Here, we will discuss the crAssphage timeline from an *in silico* discovery to an *in vitro* reality, and highlight the role that this phage family has played in pushing human gut phage research towards looking beyond the known (Figure 5).

3.3.1 Discovery and importance

Despite over one hundred years of phage research, the most abundant phage in the human gut remained undetected until 2014. The database independent approach implemented proved essential to its discovery among the viral dark matter. Prior to this, the phage remained undetected due the majority of its encoded proteins having no homology with phage proteins in public viral databases (Dutilh et al., 2014). In this sense, crAssphage is an ideal example of the current status of human gut phage research in that it highlights key bottlenecks and ways in which these can be overcome. In the few years since its discovery, this phage has also played an important role in providing essential insights into the human gut phageome.

For clarity, the initial crAssphage detected in 2014, will be referred to as prototypical-crAssphage (p-crAssphage). P-crAssphage was discovered through mining of a previously published human metagenomic dataset (Reyes et al., 2010). This dataset comprised of viromes sequenced from the faeces of 12 human subjects,

of which included 4 pairs of monozygotic twins and their mothers. The metagenomes were analyzed for co-occurring contigs with similar depth-profiles thus ensuring the contigs originated from the same individual phage genome. Using this information, the p-crAssphage genome was *de novo* assembled (Dutilh et al., 2014). This was performed using Cross-Assembly (crAss) software, a reference-independent tool, from where the phage received its name (Dutilh et al., 2012). The novel dsDNA phage was found to have a circular genome ~97 kbp in size. The genome was further analyzed against public metagenomes sequenced from human faeces in Europe, Korea and the USA. P-crAssphage was detected in 73% of the faecal metagenomes examined, contributed to as much as 90% of faecal VLP-derived sequencing reads and as much as 30% of whole metagenome reads. This analysis indicated that the phage was present in ~50% of human populations (Dutilh et al., 2014). To highlight the significance of this discovery, p-crAssphage was found to be six times more abundant in public metagenomes than all other known phages combined, and it showed little to no homology with any of these known phages (Dutilh et al., 2014). Despite the abundance at which this phage occurs, it remained undetected prior to 2014, due to the database-dependency of many bioinformatic pipelines. This emphasizes the potential for novel phage discovery within the viral dark matter.

3.3.2 Taxonomic classification and identification of an expansive family

When originally discovered, p-crAssphage was considered as a single entity. Variants were detected in a Chinese dataset indicated by complete deletion of an open reading frame (orf00039) and low similarity of DNA polymerase to the equivalent gene in p-crAssphage (Liang et al., 2016). Other studies also saw indication of

crAssphage variants (Manrique et al., 2016; García-Aljaro et al., 2017). This was further developed in 2018 by Yutin and colleagues by performing the first detailed sequence-based analysis of the proteins encoded by p-crAssphage. This work was the first to propose p-crAssphage as the original member of an expansive family, the crAss-like phage family, which includes the IAS-virus (Oude Munnink et al., 2014; Yutin et al., 2018). These phages were predicted to have a podovirus-like morphology, resolve into the Caudovirales order, and were initially predicted to be predominately temperate (Yutin et al., 2018). They were identified from reads generated from sources including the human gut, termite gut, animals, marine, and soil samples, thus indicating their presence in an array of environments. Due the ubiquitous nature of these phages among humans, qPCR assays have been developed to implement them as markers in faecal source tracking and tracing the origin of antibiotic resistance genes released into the environment due to human faecal pollution (Stachler et al., 2017; Karkman et al., 2019). A number of qPCR assays have either detected p-crAssphage at low levels or not at all in non-human mammals, among other environmental sources, but not at the same abundance as seen among humans (García-Aljaro et al., 2017; Stachler et al., 2017; Ahmed et al., 2018, 2019; Karkman et al., 2019; Ahmed et al., 2019). Hidden Markov model (HMM) analysis was also applied to crAss-like phages against public virome database and all sequences detected were human faecal virome associated (Moreno-Gallego et al., 2019). This indicates these phages are particularly abundant and specific to the human gut.

In 2018, Guerin et al., proposed a taxonomic classification scheme for crAss-like phages of human gut origin, with four sub-families (Alphacrassvirinae, Betacrassvirinae, Gammacrassvirinae and Deltacrassvirinae) and ten candidate genera. This was developed using 249 gut associated crAss-like phages, 244 of which

were *de novo* assembled in this study (Guerin et al., 2018). Members of the same subfamily shared 20-40% similarity between orthologous proteins and crAss-like phages that clustered into each genus shared >40% protein similarity. This was performed using a protein clustering approach that allowed similarity to be identified across the phages despite sharing little to no similarity at the nucleotide level. This was further supported by phylogenetic analyses of four conserved genes (capsid, primase, portal protein and terminase) which is consistent with earlier studies (Yutin et al., 2018; Guerin et al., 2018). This work also presented the *ex vivo* propagation of crAss-like phages from five candidate genera including p-crAssphage, but without host identification. Electron micrographs generated from crAssphage-rich faecal filtrate also supported previous predictions that these phages have a podovirus-like morphology (Guerin et al., 2018).

3.3.3 CrAssphage host predictions and *in vitro* isolation

Through CRISPR-spacer profiling and bacterial co-abundance analysis, it was predicted that this phage infects bacteria of the *Bacteroides* genus or other members of the Bacteroidetes phylum. Further supporting this, Bacteroidetes-associated carbohydrate-binding (BACON) domains were identified within the p-crAssphage genome (Dutilh et al., 2014). It is likely that this interaction is evolutionary ancient, based on the conservation of a DNA primase among crAss-like phages that is homologous to the primase genes of Bacteroidetes. This is one of the few highly conserved crAss-like phage family replication genes that is linked to Bacteroidetes (Yutin et al., 2018). Despite these *in silico* analyses, without a phage-host pair isolated *in vitro* we would gain little insight into the biological properties of this phage.

Due to the difficulty in isolating crAss-like phages and their associated hosts, little biological characterizations have been performed until recently. In 2018, the first crAssphage-host pair was isolated in pure culture using a phage enrichment protocol from a pool of human faeces from 20 individuals against a panel of 53 strains, which included 18 members of the Bacteroidales order (Shkoporov et al., 2018a). The isolated IAS-like crAssphage, Φ crAss001, of candidate genus VI, sub-family Betacrassvirinae, was found to infect *Bacteroides intestinalis* thus confirming *in silico* predictions. Electron micrographs also confirmed the podovirus-like morphology predicted for these phages. Intriguingly, the phage was found to co-exist with its host in co-culture over 3 weeks without significant impact on host growth despite efficient propagation. The phage genome was also absent of lysogeny genes (Shkoporov et al., 2018a). Similarly, Φ crAss002, was found to stably co-exist with its host *B. xylanisolvens* (Chapter 3 of this thesis). This is consistent with the observed persistence and stability of these phages in metagenomics datasets over time (Guerin et al., 2018; Edwards et al., 2019; Shkoporov et al., 2019). This was further supported by the stable engraftment of these phages in certain individuals for up to a year following FMT (Draper et al., 2018; Siranosian et al., 2020). More recently, the isolation of two crAss-like phages which infect *Bacteroides thetaiotaomicron* and are closely related to Φ crAss001 have been reported, DAC15 and DAC17 (Hryckowian et al., 2020). Detailed biological characterization of these phage has yet to be performed but will shine further light on the traits shared across this phage family.

3.3.4 Our current understanding of crAss-like phages

Thus far we know that crAss-like phages are generally absent from the neonate gut but are acquired in infancy, however, the influence of birth mode on the transmission of these phages has been debated (McCann et al., 2018; Siranosian et al., 2020). It was initially thought that crAss-like phages were absent from the infant gut, however, now we know that these phages can be vertically transferred from mother to infant (Lim et al., 2015; Siranosian et al., 2020). The early colonization of crAss-like phages and their universal ubiquity is not surprising considering the predominance of Bacteroidetes in the healthy human gut from infancy (Rodríguez et al., 2015). These phages have also been detected in the elderly virome (Stockdale et al., 2018). With crAss-like phage variants detected in non-human primate faecal metagenomes, it is likely that this phage family has coevolved with humans over millions of years (Edwards et al., 2019). CrAss-BACONs have also provided insights into the evolution and diversity of this phage family and allowed the detection of two novel lineages (de Jonge et al., 2019b). Furthermore, the crAssphage family is globally distributed, with one or more types being detected in 77% of individuals with geographic variability of strains and genera (Cinek et al., 2018; Guerin et al., 2018; Edwards et al., 2019). For example, p-crAssphage (candidate genus I), is largely absent from the gut of hunter-gatherer populations and is more abundant among industrialized populations (Honap et al., 2020). CrAss-like phages of candidate genus I are predominant in Western infants, among other less abundant genera, compared to malnourished and healthy Malawian infants which are dominated by crAss-like phages of the candidate genera VI, VIII, and IX (Guerin et al., 2018). CrAss-like phages have not been linked to factors such as disease, age, gender, or body mass index, but a weak link to diet has been detected as well as potential links to ethnicity and geography (Guerin et al., 2018;

Edwards et al., 2019; Honap et al., 2020). Diet may be one factor driving the crAssphage family variation observed between Western and more rural non-industrialized populations. The Western diet generally drives a microbiota dominated in *Bacteroides/Clostridia*, whereas a non-Western fibre rich diet is associated with high *Prevotella*/low *Bacteroides*. Therefore, this diet driven variation in dominant bacterial species in the gut may also influence crAss-like phage composition (Gorvitovskaia et al., 2016; Guerin et al., 2018). Despite the genetic divergence, the organisation of crAss-like phage genomes are colinear in humans and primates thus suggesting the genome structure of these phages has been conserved throughout evolution due to the stability of the gut ecosystem (Edwards et al., 2019).

One of the most intriguing characteristics of these phages is their ability to persist in the human gut. Following *in silico* and *in vitro* analysis, it was observed that these phages have the remarkable ability to maintain themselves in the gut and engage in an unusual interaction with their host that allows co-existence of two antagonistic populations over extensive periods of time (Guerin et al., 2018; Shkoporov et al., 2018a). The concept of a stable healthy core gut phageome that is dominated by temperate phages and abundant in crAss-like phages has been debated (Manrique et al., 2016; Moreno-Gallego et al., 2019). This was further expanded on by Shkoporov and colleagues in a longitudinal study that monitored the gut virome of ten healthy individuals over a one-year period. This study identified a small but abundant, predominately virulent core virome of 22 phage clusters using a database-independent clustering approach. CrAss-like phages represented approximately one third of these clusters. This confirmed that temperate phages are not as dominant in the healthy human gut as previously thought. Therefore, mechanisms other than lysogeny must be at play to maintain virulent crAss-like phages in the human gut over time (Shkoporov

et al., 2019). This is supported by the absence of a lysogeny gene module in Φ crAss001 and crAss-like prophages going undetected in Bacteroidales genomes deposited in databases such as NCBI RefSeq (Shkoporov et al., 2018a). Transient phenotypic resistance mediated by phase variable expression of host encoded phage receptors, such as capsular polysaccharide, is thought to play an important role in the persistence of virulent phages and their hosts. This phenomenon creates heterogeneity within an isogenic population with sub-populations oscillating between phage-susceptible and phage-resistant phenotypes. As a result, the phage can efficiently propagate but viral load is limited by the co-occurrence of a phage-resistant sub-population which continues to grow, thus creating a form of herd immunity that protects sensitive variants by limiting phage accessibility (Turkington et al., 2019). The importance of phase variation among Bacteroidales has been recently highlighted (Jiang et al., 2019; Porter et al., 2020). The influence of this phenomenon on crAss-like phage-host interactions has been indicated by the inability of Φ crAss001 to clear liquid cultures of its host at high titres and the isolation of ~46% resistant clones and ~54% completely or partially phage-susceptible clones following co-culture experiments (Shkoporov et al., 2018a). Φ crAss002 was also unable to clear liquid cultures of *B. xylanisolvens* despite achieving high titres. On initial exposure the phage selects for resistant host variants followed by a greater selection for phage-susceptible variants (Chapter 3 of this thesis). The inability of a virulent phage to clear liquid culture of *B. thetaiotaomicron* was also observed due to population heterogeneity mediated by phase variable expression of CPS (Porter et al., 2020). In the case of Φ crAss002, it appears that the presence of the phage ultimately leads to the selection of phage-susceptible variants (Chapter 3). This may suggest that co-existing with these phages confers the host with some benefit. However, most examples of phage

conferred benefit identified to date have been observed among temperate phages (Obeng et al., 2016). Transcriptomic profiling may shine light on this and will be necessary to confirm the role of phase variation in phage-host interactions.

“Royal-family” model dynamics may also provide an explanation for the observed stability. If kill-the-winner dynamics occur at a strain or sub-strain level, stability is maintained at the species or genus levels over time (Breitbart et al., 2018). The spatial heterogeneity of the GIT could provide bacterial hosts with phage-inaccessible reservoirs in microhabitats, such as crypts and mucosa, creating source-sink dynamics with predation predominating in the lumen thus allowing maintenance of both phage and host (Lourenço et al., 2019). Overall, it is likely that the persistence of crAss-like phages in the human gut and the co-existence with their host is due to an inter-play of multiple mechanisms.

The isolation of further representatives from the crAss-like phage family from different genera will be necessary to identify shared and unique biological properties. The precise mechanisms mediating the stable and persistent colonization of these virulent phages at high levels needs to be further examined. Ideally, future studies should examine these phages in *in vivo* models that simulate realistic conditions as best as possible. This will provide important insights in the healthy gut phageome. The ultimate goal of this will be to understand how these phages shape our microbiome and influence health. Although we still have a lot to learn about these enigmatic phages, they have played a significant role in highlighting the potential of viral dark matter and in improving our understanding of the human gut phageome both in terms of composition and interactions (Figure 5).

4. The merits of studying bacteriophages and future prospects

Despite the hurdles that need be overcome in phage research and the challenges faced in studying the human gut phageome, phages have many applications which could be potentially of benefit to humans, in addition to forming an important component and driving force of the microbiome.

Interest in phages has resurged in recent years due to current and projected severity of the antibiotic resistance crises. It is estimated that the inability to treat bacterial diseases due to multi-drug resistance may lead to 10 million additional deaths globally per annum by 2050 (Sugden et al., 2016). Phage therapy is being examined as an alternative or as an aid to antibiotic treatment; however, this is not without difficulties due to possible undesirable immune responses, host resistance development, identifying suitable phage(s), and regulations (Oechslin, 2018; Brüssow, 2019). Several human diseases, such as IBD, have a bacterial component that worsens disease status due to over-stimulation of the immune system (Gevers et al., 2014). Phage-driven selective elimination of pathogens from the gut to shape microbiome composition towards homeostasis is also being examined. The ability of phages to shape the microbiome is supported by outcomes of FMT studies (Ott et al., 2017; Draper et al., 2018). Phages produce lytic enzymes, endolysins, which degrade the cell wall of their associated bacterial host, and also have therapeutic potential as alternatives to antibiotics (Love et al., 2018). While phage therapy is not within the scope of this review, we will mention a few cases of its application. The current status of phage therapy has been extensively reviewed and assessed (Czaplewski et al., 2016; Cisek et al., 2017; Lin et al., 2017; Lourenço et al., 2018; Hyman, 2019; Brüssow, 2019). Phages have also been proposed as a means of reducing the use of preservatives or antibiotics in food production (O'Sullivan et al.,

2019; Lewis and Hill, 2020). However, here we will focus on therapeutic applications. We have seen the antimicrobial capacity of phages in selectively eradicating infection and alleviating diseases with a microbial component. Recently, phages were successfully administered in eliminating cytolytic *E. faecalis* in humanized mice, a pathogen responsible for increased disease severity and mortality in patients with alcoholic hepatitis (Duan et al., 2019). Phage-guided nanoparticles were used in the eradication of pro-tumoral *Fusobacterium nucleatum* and allowed the site-specific delivery of chemotherapeutic drugs in parallel. This resulted in colorectal cancer tumour reduction in a mouse model with limited impact on healthy tissue (Zheng et al., 2019). This technology was also applied in a piglet model without induction of the immune system or adverse effects (Zheng et al., 2019). However, in another study, phage treatment of carcinogenic bacteria to alleviate colorectal cancer led to immune activation and exacerbation of colitis in a murine model, despite reducing the target bacteria in the mouse intestine (Gogokhia et al., 2019). This study also showed that high levels of Caudovirales phages in IBD patients can dictate the efficacy of FMT in treating the disease (Gogokhia et al., 2019). Phage therapy has also led to the eradication of multi-drug resistant *Klebsiella pneumoniae* by oral and rectal administration of one lytic phage in a human patient without any adverse effects (Corbellino et al., 2019). One report showed that phages can act in synergy with innate immune cells to eliminate pathogens. This was observed in a murine lung model of drug resistant *Pseudomonas aeruginosa* in which phages could only eradicate the pneumonia-causing pathogen in synergy with neutrophils which aided the elimination of resistance clones that emerge on phage exposure (Roach et al., 2017). The above highlights the potential of phage therapy in selectively eliminating pathogens, but also the need to fully decipher phage gut dynamics with both the bacterial and human host,

before we apply phages for therapeutic use. Considering the impending outcome of the antibiotic resistance crisis and the importance of microbiome homeostasis, we need to expand our knowledge rapidly.

As we begin to understand phage-host dynamics and how they drive the microbiome, specific phages or phage compositions may be identified as biomarkers of disease or health. Database-independent methods and examination of the virome as a whole will be essential for discovering disease-specific alterations. For example, whole virome analysis of the IBD and healthy virome identified a shift from a lytic core of a few but abundant clusters of *Microviridae* and crAss-like phages in health to a virome with increased induction of temperate phages such as *Myoviridae* and *Siphoviridae* in the disease state (Clooney et al., 2019). Further development of these findings and the identification of compositional patterns in the context of other diseases states may lead to the use of virome biomarkers in the future.

The recognition of a temporally stable, personal persistent virulent core in healthy humans has changed our view of the gut phageome (Shkoporov et al., 2019). Formerly it was thought that the healthy gut phageome was dominated by temperate phages with a minority of virulent phages. However, now it is known that the converse is true. Having observed the geographical variation of crAss-like phages among individuals, how location, diet and ethnicity influence the individual composition of the virulent core should be explored. Gregory *et al.* observed significant differences in the dominant viral populations in Western and non-Western gut viromes, further suggesting that lifestyle and diet differences associated with geographical location can influence both phage and bacterial composition in the gut (Gregory et al., 2019). Additionally, *in vitro* and *in vivo* isolation and characterization of representatives of the lytic core will be essential in providing insights into phage-host interactions, how

virulent phages can persistent in the gut, and drive homeostasis. The biogeography of these phages, radially and longitudinally along the GIT also requires probing. Elucidation of the personal persistent virome will serve as an important window into the gut phageome and provide us with insights into ecology, composition and dynamics including the persistence of virulent phages.

5. Conclusions

In summary, we need to expand our understanding of the mechanisms phages use to drive the compositions of the gut microbiome. One of the most critical shortcomings of human gut phage research is the incomplete analysis of metagenomic datasets due to dependency on a poorly curated viral database. Furthermore, little progress has been made in the isolation and characterisation of novel gut phage-host pairs. In recent years, efforts are being made to overcome these bottlenecks. This requires the development of universal and easily reproducible methods that limit bias during viral enrichment, nucleic acid extraction, and sequencing library preparation. *In silico* analyses need to be performed using database-independent methods that allow for a complete *de novo* virome analysis using benchmarked criteria. There is also the need to develop a sequence-based taxonomic scheme to facilitate the rapid expansion of identified phage sequences due to modern sequencing technology. The current contradictions among different gut virome studies need to be rectified and clarified. Overall progression in metagenomic analyses of the gut phageome has been more rapid than in the isolation of gut phages, which is ultimately the key for expanding our scope of the mechanisms mediating phage-host relationships. If we are to better understand the driving force of phages, more *in vitro* and *in vivo* studies are

necessary. Recent findings have provided important insights, but they also highlight how little we know about this important and enigmatic component of our gut.

6. Figures

1.

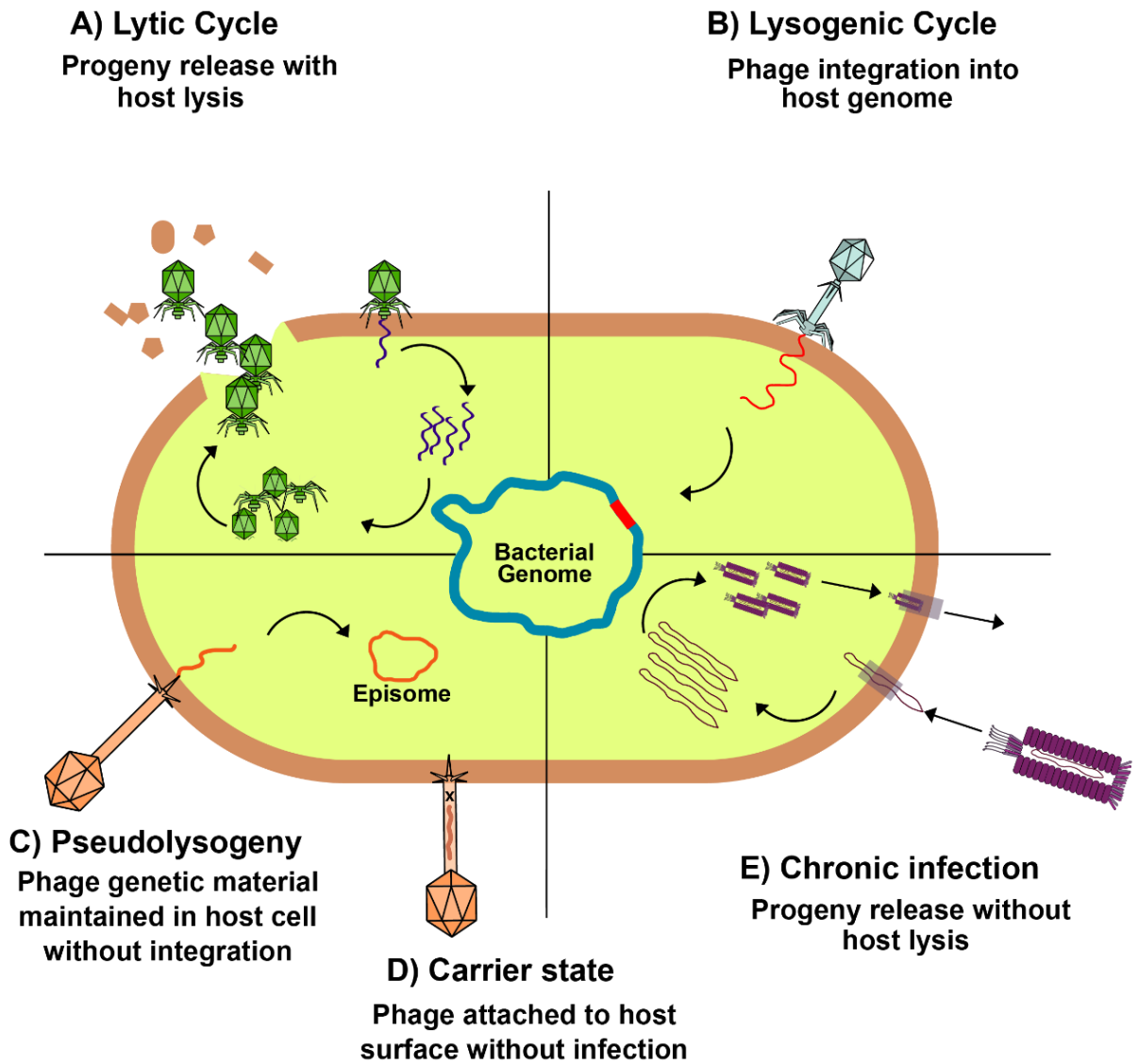


Figure 1. Overview of phage lifecycles. (A) Lytic phages hijack host cell machinery to replicate, assemble and produce progeny which are released from the cytoplasm on host cell lysis to initiate further rounds of infection. This phage lifecycle is thought to be the most prevalent in the healthy human gut. (B) Lysogenic phages integrate their genome into the bacterial host genome with which they passively replicate until stress signals trigger their induction and switch to the lytic cycle. (C) Phages that follow a pseudolysogenic lifecycle also passively replicate with the host but their genome remains independent from that of the host and is maintained in the cytoplasm as an episome. (D). When phages are in a carrier state they can remain attached to the surface of a non-permissive host without infection. (E) During a chronic infection phages produce progeny similarly to that of lytic phages but without host cell lysis. Adapted from (Lawrence et al., 2019).

2.

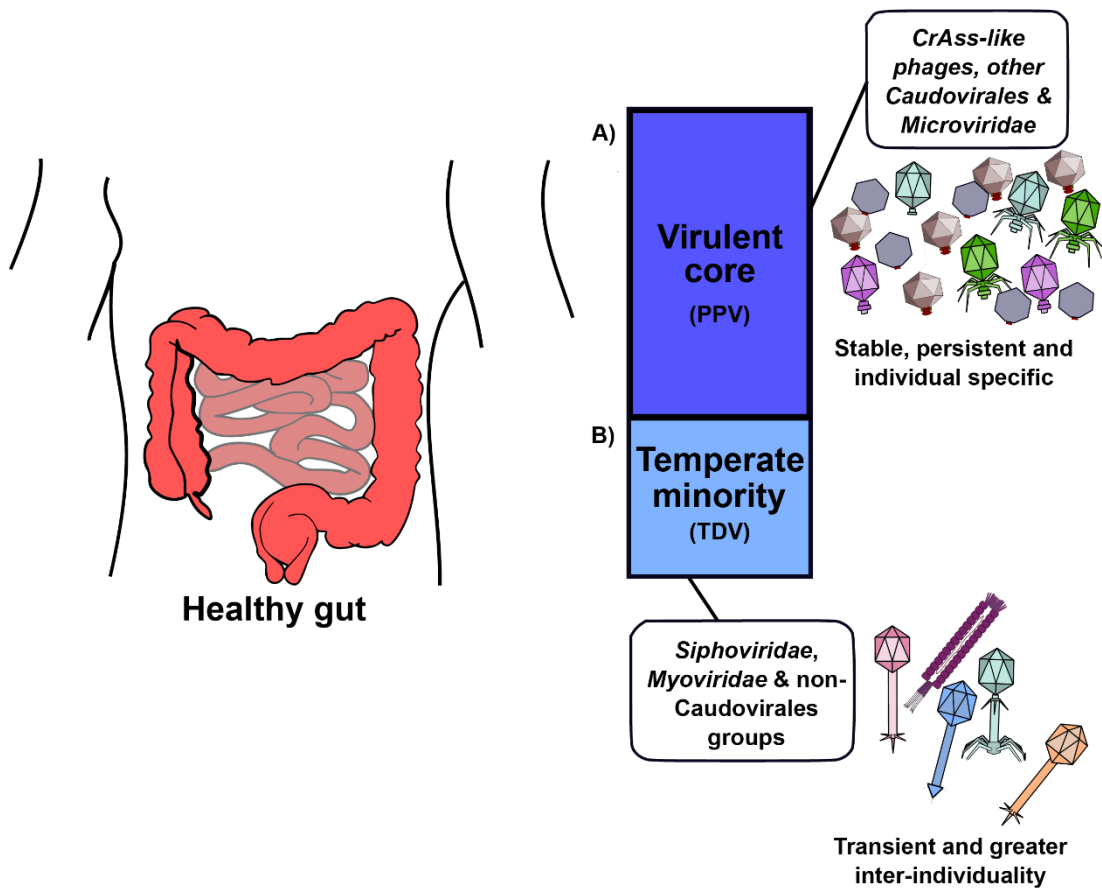
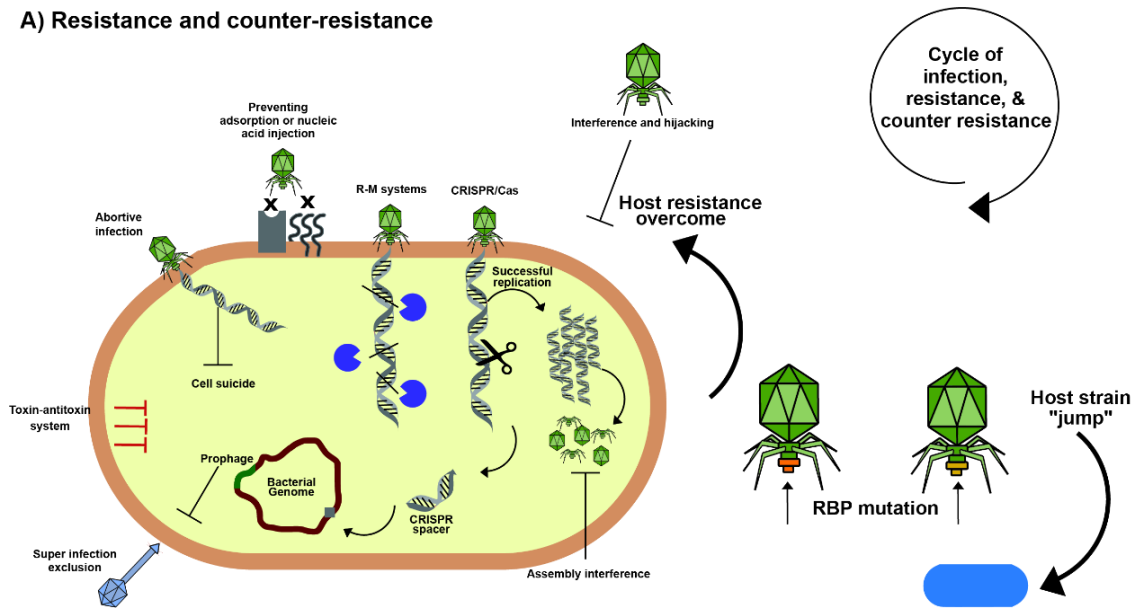


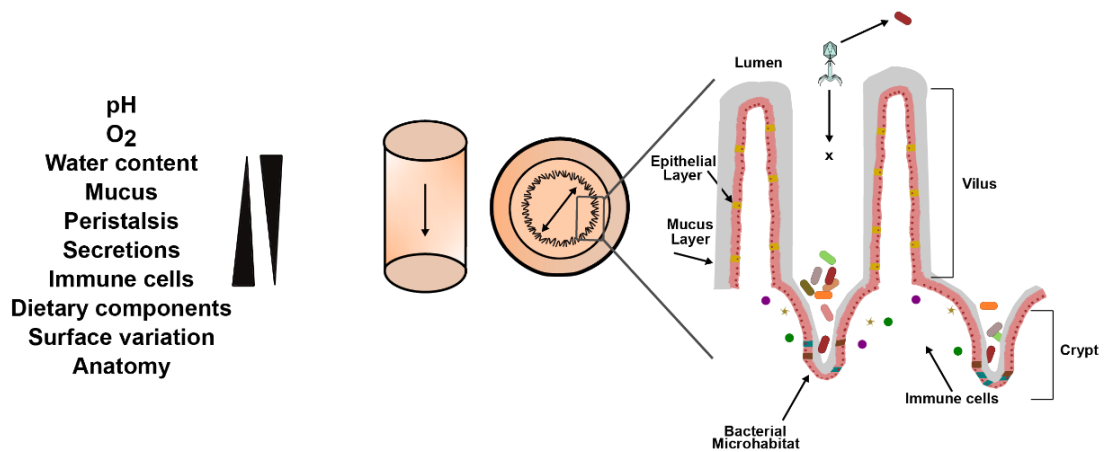
Figure 2. The phageome composition of a healthy human gut. (A). Virulent phages of the crAss-like phage family, other members of the Caudovirales order and the *Microviridae* family are among the most dominant and stable phages in the human gut forming the personal persistent virome (PPV). (B) Temperate phages and certain virulent phages are less abundant, less stable and more shared across individual viromes forming the transiently detected virome (TDV). Induction of specific members of the TDV is thought to be disease associated.

3.

A) Resistance and counter-resistance



B) GIT radial and longitudinal variation



C) Host metabolic state and phenotype

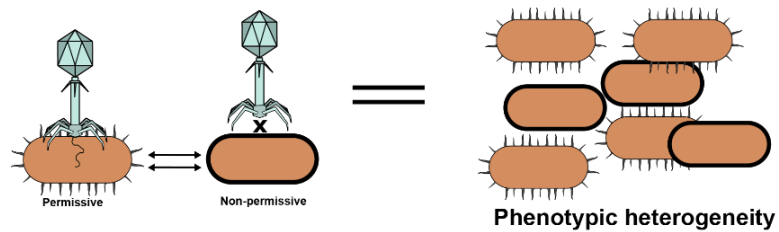


Figure 3. An overview of the factors that influence phage-host interactions in the human gut. (A). To overcome phage predation, bacteria possess an arsenal of defence mechanisms that target one or more stages of phage infection cycles. In retaliation phages have evolved an array of counter-defence mechanisms. This results in a cycle of infection, resistance and counter-resistance that leads to evolution and diversity in the human gut. (B) In the human gastrointestinal tract there is significant variation in biotic and abiotic factors both longitudinally and radially. This results in sectional variation in bacterial composition and spatial heterogeneity between phages and their hosts. Anatomical features of the gut such as mucus, crypts and villi create bacterial microhabitats which are inaccessible to phages allowing them to escape predation, gradually seed the lumen and maintain homeostasis. (C) Biotic and abiotic factors can also influence the metabolic state of the bacterial host. This can result in transient phenotypic changes mediated by mechanisms such as phase variation allowing adaptation to stress which includes phage infection. These phenotypic changes can result in an isogenic population of phage permissive and non-permissive hosts permitting the co-existence of phage and bacteria over time.

4.

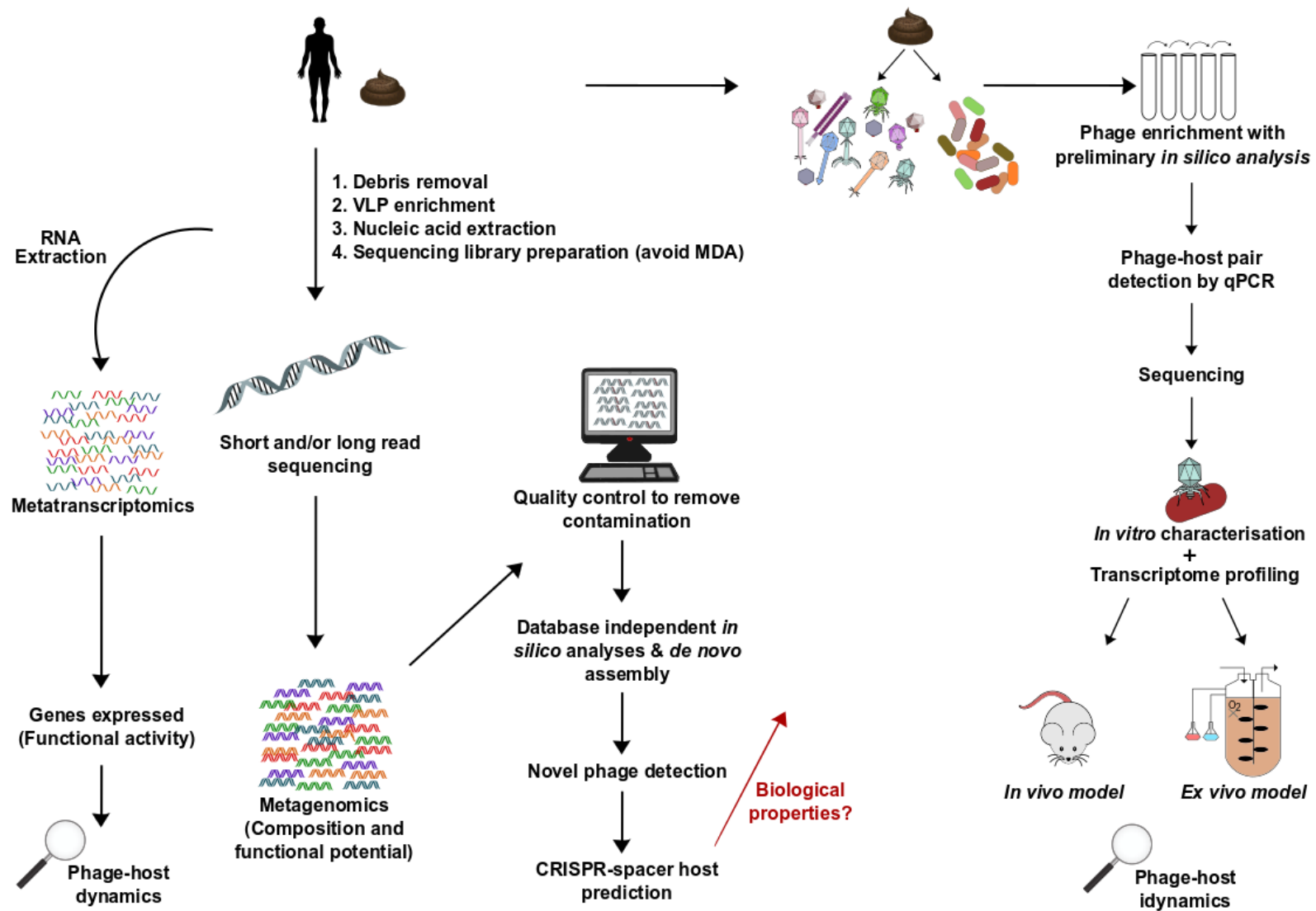


Figure 4. A generic overview of key experimental steps required in studying the human gut phageome from metagenomics, database-independent whole virome analyses, *in silico* identification of novel phages to *in vitro* isolation and *in vivo* characterisation. Linking bioinformatics and laboratory research provides important insights into phageome composition, novel phage detection and isolation with characterisation of biological properties and phage interactions with both the bacterial and mammalian host

5.

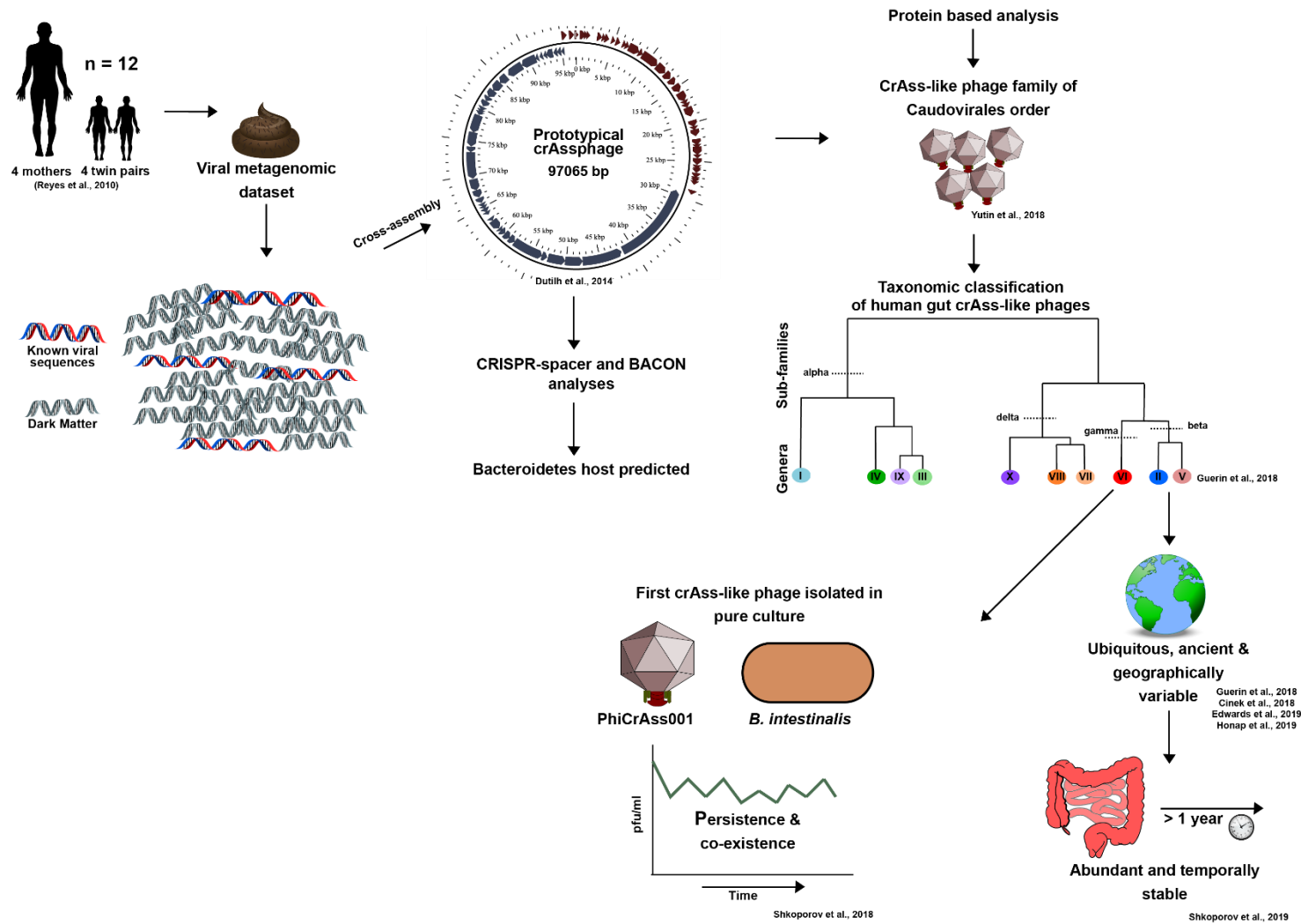


Figure 5. The crAssphage family timeline from *in silico* discovery to an *in vitro* reality.

7. References

- Abecasis, A.B., Serrano, M., Alves, R., Quintais, L., Pereira-Leal, J.B., and Henriques, A.O. (2013). A Genomic Signature and the Identification of New Sporulation Genes. *J. Bacteriol.* *195*, 2101–2115.
- Ackermann, H.-W. (1998). Tailed Bacteriophages: The Order Caudovirales. In *Advances in Virus Research*, K. Maramorosch, F.A. Murphy, and A.J. Shatkin, eds. (Academic Press), pp. 135–201.
- Ackermann, H.-W. (2012). Chapter 1 - Bacteriophage Electron Microscopy. In *Advances in Virus Research*, M. Łobocka, and W.T. Szybalski, eds. (Academic Press), pp. 1–32.
- Aggarwala, V., Liang, G., and Bushman, F.D. (2017). Viral communities of the human gut: metagenomic analysis of composition and dynamics. *Mob. DNA* *8*, 12.
- Ahmed, W., Lobos, A., Senkbeil, J., Peraud, J., Gallard, J., and Harwood, V.J. (2018). Evaluation of the novel crAssphage marker for sewage pollution tracking in storm drain outfalls in Tampa, Florida. *Water Res.* *131*, 142–150.
- Ahmed, W., Payyappat, S., Cassidy, M., and Besley, C. (2019). Enhanced insights from human and animal host-associated molecular marker genes in a freshwater lake receiving wet weather overflows. *Sci. Rep.* *9*, 1–13.
- Aiewsakun, P., Adriaenssens, E.M., Lavigne, R., Kropinski, A.M., and Simmonds, P. (2018). Evaluation of the genomic diversity of viruses infecting bacteria, archaea and eukaryotes using a common bioinformatic platform: steps towards a unified taxonomy. *J. Gen. Virol.* *99*, 1331–1343.

Allen, L.Z., Ishoey, T., Novotny, M.A., McLean, J.S., Lasken, R.S., and Williamson, S.J. (2011). Single Virus Genomics: A New Tool for Virus Discovery. *PLOS ONE* 6, e17722.

Argov, T., Sapir, S.R., Pasechnek, A., Azulay, G., Stadnyuk, O., Rabinovich, L., Sigal, N., Borovok, I., and Herskovits, A.A. (2019). Coordination of cohabiting phage elements supports bacteria–phage cooperation. *Nat. Commun.* 10, 1–14.

Baquero, F., and Nombela, C. (2012). The microbiome as a human organ. *Clin. Microbiol. Infect.* 18, 2–4.

Barr, J.J., Auro, R., Furlan, M., Whiteson, K.L., Erb, M.L., Pogliano, J., Stotland, A., Wolkowicz, R., Cutting, A.S., Doran, K.S., et al. (2013). Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc. Natl. Acad. Sci. U. S. A.* 110, 10771–10776.

Bauer, M.A., Kainz, K., Carmona-Gutierrez, D., and Madeo, F. (2018). Microbial wars: Competition in ecological niches and within the microbiome. *Microb. Cell Graz Austria* 5, 215–219.

Bayliss, C.D. (2009). Determinants of phase variation rate and the fitness implications of differing rates for bacterial pathogens and commensals. *FEMS Microbiol. Rev.* 33, 504–520.

Belkaid, Y., and Hand, T.W. (2014). Role of the microbiota in immunity and inflammation. *Cell* 157, 121–141.

Belleghem, J.D.V., Clement, F., Merabishvili, M., Lavigne, R., and Vaneechoutte, M. (2017). Pro- and anti-inflammatory responses of peripheral blood mononuclear cells

induced by *Staphylococcus aureus* and *Pseudomonas aeruginosa* phages. *Sci. Rep.* 7, 1–13.

Bergh, O., Børsheim, K.Y., Bratbak, G., and Heldal, M. (1989). High abundance of viruses found in aquatic environments. *Nature* 340, 467–468.

Bernheim, A., and Sorek, R. (2019). The pan-immune system of bacteria: antiviral defence as a community resource. *Nat. Rev. Microbiol.* 1–7.

Bertozzi Silva, J., Storms, Z., and Sauvageau, D. (2016). Host receptors for bacteriophage adsorption. *FEMS Microbiol. Lett.* 363.

Bilen, M., Dufour, J.-C., Lagier, J.-C., Cadoret, F., Daoud, Z., Dubourg, G., and Raoult, D. (2018). The contribution of culturomics to the repertoire of isolated human bacterial and archaeal species. *Microbiome* 6, 94.

Blower, T.R., Evans, T.J., Przybilski, R., Fineran, P.C., and Salmond, G.P.C. (2012). Viral Evasion of a Bacterial Suicide System by RNA-Based Molecular Mimicry Enables Infectious Altruism. *PLOS Genet.* 8, e1003023.

Bolduc, B., Jang, H.B., Doulcier, G., You, Z.-Q., Roux, S., and Sullivan, M.B. (2017). vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ* 5.

Brathwaite, K.J., Siringan, P., Connerton, P.L., and Connerton, I.F. (2015). Host adaption to the bacteriophage carrier state of *Campylobacter jejuni*. *Res. Microbiol.* 166, 504–515.

Breitbart, M., Bonnain, C., Malki, K., and Sawaya, N.A. (2018). Phage puppet masters of the marine microbial realm. *Nat. Microbiol.* 3, 754–766.

- Browne, H.P., Forster, S.C., Anonye, B.O., Kumar, N., Neville, B.A., Stares, M.D., Goulding, D., and Lawley, T.D. (2016). Culturing of ‘unculturable’ human microbiota reveals novel taxa and extensive sporulation. *Nature* 533, 543–546.
- Brüssow, H. (2019). Hurdles for Phage Therapy to Become a Reality—An Editorial Comment. *Viruses* 11, 557.
- Bruttin, A., Desiere, F., d’Amico, N., Guérin, J.P., Sidoti, J., Huni, B., Lucchini, S., and Brüssow, H. (1997). Molecular ecology of *Streptococcus thermophilus* bacteriophage infections in a cheese factory. *Appl. Environ. Microbiol.* 63, 3144–3150.
- Bull, J.J., Vegge, C.S., Schmerer, M., Chaudhry, W.N., and Levin, B.R. (2014). Phenotypic Resistance and the Dynamics of Bacterial Escape from Phage Control. *PLOS ONE* 9, e94690.
- Callanan, J., Stockdale, S.R., Shkoporov, A., Draper, L.A., Ross, R.P., and Hill, C. (2018). RNA Phage Biology in a Metagenomic Era. *Viruses* 10.
- Callanan, J., Stockdale, S.R., Shkoporov, A., Draper, L.A., Ross, R.P., and Hill, C. (2020). Expansion of known ssRNA phage genomes: From tens to over a thousand. *Sci. Adv.* 6, eaay5981.
- Cani, P.D. (2018). Human gut microbiome: hopes, threats and promises. *Gut* 67, 1716–1725.
- Casén, C., Vebø, H.C., Sekelja, M., Hegge, F.T., Karlsson, M.K., Cierniejewska, E., Dzankovic, S., Frøyland, C., Nestestog, R., Engstrand, L., et al. (2015). Deviations in human gut microbiota: a novel diagnostic test for determining dysbiosis in patients with IBS or IBD. *Aliment. Pharmacol. Ther.* 42, 71–83.

- Castellazzi, M., George, J., and Buttin, G. (1972). Prophage induction and cell division in *E. coli*. *Mol. Gen. Genet. MGG* 119, 153–174.
- Castro-Mejía, J.L., Muhammed, M.K., Kot, W., Neve, H., Franz, C.M.A.P., Hansen, L.H., Vogensen, F.K., and Nielsen, D.S. (2015). Optimizing protocols for extraction of bacteriophages prior to metagenomic analyses of phage communities in the human gut. *Microbiome* 3, 64.
- Chatterjee, S., and Rothenberg, E. (2012). Interaction of Bacteriophage λ with Its *E. coli* Receptor, LamB. *Viruses* 4, 3162–3178.
- Chen, Y.-C., Liu, T., Yu, C.-H., Chiang, T.-Y., and Hwang, C.-C. (2013). Effects of GC Bias in Next-Generation-Sequencing Data on De Novo Genome Assembly. *PLOS ONE* 8, e62856.
- Chibani-Chennoufi, S., Bruttin, A., Dillmann, M.-L., and Brüssow, H. (2004). Phage-Host Interaction: an Ecological Perspective. *J. Bacteriol.* 186, 3677–3686.
- Cieplak, T., Soffer, N., Sulakvelidze, A., and Nielsen, D.S. (2018). A bacteriophage cocktail targeting *Escherichia coli* reduces *E. coli* in simulated gut conditions, while preserving a non-targeted representative commensal normal microbiota. *Gut Microbes* 9, 391–399.
- Cinek, O., Mazankova, K., Kramna, L., Odeh, R., Alassaf, A., Ibekwe, M.U., Ahmadov, G., Mekki, H., Abdullah, M.A., Elmahi, B.M.E., et al. (2018). Quantitative CrAssphage real-time PCR assay derived from data of multiple geographically distant populations. *J. Med. Virol.* 90, 767–771.

Cisek, A.A., Dąbrowska, I., Gregorczyk, K.P., and Wyżewski, Z. (2017). Phage Therapy in Bacterial Infections Treatment: One Hundred Years After the Discovery of Bacteriophages. *Curr. Microbiol.* 74, 277–283.

Clooney, A.G., Sutton, T.D.S., Shkoporov, A.N., Holohan, R.K., Daly, K.M., O'Regan, O., Ryan, F.J., Draper, L.A., Plevy, S.E., Ross, R.P., et al. (2019). Whole-Virome Analysis Sheds Light on Viral Dark Matter in Inflammatory Bowel Disease. *Cell Host Microbe* 26, 764-778.e5.

Cohen, D., Melamed, S., Millman, A., Shulman, G., Oppenheimer-Shaanan, Y., Kacen, A., Doron, S., Amitai, G., and Sorek, R. (2019). Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature* 574, 691–695.

Conceição-Neto, N., Zeller, M., Lefrère, H., De Bruyn, P., Beller, L., Deboutte, W., Yinda, C.K., Lavigne, R., Maes, P., Ranst, M.V., et al. (2015). Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Sci. Rep.* 5, 16532.

Corbellino, M., Kieffer, N., Kutateladze, M., Balarjishvili, N., Leshkasheli, L., Askilashvili, L., Tsertsvadze, G., Rimoldi, S.G., Nizharadze, D., Hoyle, N., et al. (2019). Eradication of a multi-drug resistant, carbapenemase-producing *Klebsiella pneumoniae* isolate following oral and intra-rectal therapy with a custom-made, lytic bacteriophage preparation. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.*

Cornelissen, A., Ceyssens, P.-J., Krylov, V.N., Noben, J.-P., Volckaert, G., and Lavigne, R. (2012). Identification of EPS-degrading activity within the tail spikes of the novel *Pseudomonas putida* phage AF. *Virology* 434, 251–256.

Coward, C., Grant, A.J., Swift, C., Philp, J., Towler, R., Heydarian, M., Frost, J.A., and Maskell, D.J. (2006). Phase-Variable Surface Structures Are Required for

Infection of *Campylobacter jejuni* by Bacteriophages. *Appl. Environ. Microbiol.* 72, 4638–4647.

Cumby, N., Edwards, A.M., Davidson, A.R., and Maxwell, K.L. (2012). The Bacteriophage HK97 gp15 Moron Element Encodes a Novel Superinfection Exclusion Protein. *J. Bacteriol.* 194, 5012–5019.

Cumby, N., Reimer, K., Mengin-Lecreulx, D., Davidson, A.R., and Maxwell, K.L. (2015). The phage tail tape measure protein, an inner membrane protein and a periplasmic chaperone play connected roles in the genome injection process of *E. coli* phage HK97. *Mol. Microbiol.* 96, 437–447.

Czaplewski, L., Bax, R., Clokie, M., Dawson, M., Fairhead, H., Fischetti, V.A., Foster, S., Gilmore, B.F., Hancock, R.E.W., Harper, D., et al. (2016). Alternatives to antibiotics—a pipeline portfolio review. *Lancet Infect. Dis.* 16, 239–251.

Das, P., Ji, B., Kovatcheva-Datchary, P., Bäckhed, F., and Nielsen, J. (2018). In vitro co-cultures of human gut bacterial species as predicted from co-occurrence network analysis. *PLoS ONE* 13.

De Sordi, L., Khanna, V., and Debarbieux, L. (2017). The Gut Microbiota Facilitates Drifts in the Genetic Diversity and Infectivity of Bacterial Viruses. *Cell Host Microbe* 22, 801-808.e3.

De Sordi, L., Lourenço, M., and Debarbieux, L. (2019). The Battle Within: Interactions of Bacteriophages and Bacteria in the Gastrointestinal Tract. *Cell Host Microbe* 25, 210–218.

- Deng, L., Gregory, A., Yilmaz, S., Poulos, B.T., Hugenholtz, P., and Sullivan, M.B. (2012). Contrasting Life Strategies of Viruses that Infect Photo- and Heterotrophic Bacteria, as Revealed by Viral Tagging. *MBio* 3.
- Deng, L., Ignacio-Espinoza, J.C., Gregory, A.C., Poulos, B.T., Weitz, J.S., Hugenholtz, P., and Sullivan, M.B. (2014). Viral tagging reveals discrete populations in *Synechococcus* viral genome sequence space. *Nature* 513, 242–245.
- Denou, E., Berger, B., Barretto, C., Panoff, J.-M., Arigoni, F., and Brüssow, H. (2007). Gene Expression of Commensal *Lactobacillus johnsonii* Strain NCC533 during In Vitro Growth and in the Murine Gut. *J. Bacteriol.* 189, 8109–8119.
- Devoto, A.E., Santini, J.M., Olm, M.R., Anantharaman, K., Munk, P., Tung, J., Archie, E.A., Turnbaugh, P.J., Seed, K.D., Blekhman, R., et al. (2019). Megaphages infect *Prevotella* and variants are widespread in gut microbiomes. *Nat. Microbiol.* 4, 693.
- Donaldson, G.P., Lee, S.M., and Mazmanian, S.K. (2016). Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* 14, 20–32.
- Draper, L.A., Ryan, F.J., Smith, M.K., Jalanka, J., Mattila, E., Arkkila, P.A., Ross, R.P., Satokari, R., and Hill, C. (2018). Long-term colonisation with donor bacteriophages following successful faecal microbial transplantation. *Microbiome* 6, 220.
- Duan, Y., Llorente, C., Lang, S., Brandl, K., Chu, H., Jiang, L., White, R.C., Clarke, T.H., Nguyen, K., Torralba, M., et al. (2019). Bacteriophage targeting of gut bacterium attenuates alcoholic liver disease. *Nature* 1–7.

- Duerkop, B.A., Kleiner, M., Paez-Espino, D., Zhu, W., Bushnell, B., Hassell, B., Winter, S.E., Kyrpides, N.C., and Hooper, L.V. (2018). Murine colitis reveals a disease-associated bacteriophage community. *Nat. Microbiol.* 1.
- Duerr, D.M., White, S.J., and Schluesener, H.J. (2004). Identification of peptide sequences that induce the transport of phage across the gastrointestinal mucosal barrier. *J. Virol. Methods* 116, 177–180.
- Duncan, S.H., Hold, G.L., Harmsen, H.J., Stewart, C.S., and Flint, H.J. (2002). 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.
- Dutilh, B.E., Schmieder, R., Nulton, J., Felts, B., Salamon, P., Edwards, R.A., and Mokili, J.L. (2012). Reference-independent comparative metagenomics using cross-assembly: crAss. *Bioinformatics* 28, 3225–3231.
- Dutilh, B.E., Cassman, N., McNair, K., Sanchez, S.E., Silva, G.G.Z., Boling, L., Barr, J.J., Speth, D.R., Seguritan, V., Aziz, R.K., et al. (2014). A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* 5, 4498.
- Dy, R.L., Przybilski, R., Semeijn, K., Salmond, G.P.C., and Fineran, P.C. (2014). A widespread bacteriophage abortive infection system functions through a Type IV toxin–antitoxin mechanism. *Nucleic Acids Res.* 42, 4590–4605.
- Džunková, M., Low, S.J., Daly, J.N., Deng, L., Rinke, C., and Hugenholtz, P. (2019). Defining the human gut host–phage network through single-cell viral tagging. *Nat. Microbiol.* 1–12.

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., and Relman, D.A. (2005). Diversity of the Human Intestinal Microbial Flora. *Science* 308, 1635–1638.

Edwards, R.A., Vega, A.A., Norman, H.M., Ohaeri, M., Levi, K., Dinsdale, E.A., Cinek, O., Aziz, R.K., McNair, K., Barr, J.J., et al. (2019). Global phylogeography and ancient evolution of the widespread human gut virus crAssphage. *Nat. Microbiol.* 4, 1727–1736.

Eloe-Fadrosh, E.A. (2019). Towards a genome-based virus taxonomy. *Nat. Microbiol.* 4, 1249–1250.

Elzinga, J., Oost, J. van der, Vos, W.M. de, and Smidt, H. (2019). The Use of Defined Microbial Communities To Model Host-Microbe Interactions in the Human Gut. *Microbiol. Mol. Biol. Rev.* 83.

Enault, F., Briet, A., Bouteille, L., Roux, S., Sullivan, M.B., and Petit, M.-A. (2017). Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *ISME J.* 11, 237–247.

Eriksen, R.S., Svenningsen, S.L., Sneppen, K., and Mitarai, N. (2018). A growing microcolony can survive and support persistent propagation of virulent phages. *Proc. Natl. Acad. Sci. U. S. A.* 115, 337–342.

Faith, J.J., Guruge, J.L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A.L., Clemente, J.C., Knight, R., Heath, A.C., Leibel, R.L., et al. (2013). The Long-Term Stability of the Human Gut Microbiota. *Science* 341.

Forster, S.C., Kumar, N., Anonye, B.O., Almeida, A., Viciani, E., Stares, M.D., Dunn, M., Mkandawire, T.T., Zhu, A., Shao, Y., et al. (2019). A human gut bacterial genome

and culture collection for improved metagenomic analyses. *Nat. Biotechnol.* *37*, 186–192.

Furuse, K., Sakurai, T., Hirashima, A., Katsuki, M., Ando, A., and Watanabe, I. (1978). Distribution of ribonucleic acid coliphages in south and east Asia. *Appl. Environ. Microbiol.* *35*, 995–1002.

Galtier, M., Sordi, L.D., Maura, D., Arachchi, H., Volant, S., Dillies, M.-A., and Debarbieux, L. (2016). Bacteriophages to reduce gut carriage of antibiotic resistant uropathogens with low impact on microbiota composition. *Environ. Microbiol.* *18*, 2237–2245.

Galtier, M., De Sordi, L., Sivignon, A., de Vallée, A., Maura, D., Neut, C., Rahmouni, O., Wannerberger, K., Darfeuille-Michaud, A., Desreumaux, P., et al. (2017). Bacteriophages Targeting Adherent Invasive *Escherichia coli* Strains as a Promising New Treatment for Crohn's Disease. *J. Crohns Colitis* *11*, 840–847.

Gandon, S., Buckling, A., Decaestecker, E., and Day, T. (2008). Host-parasite coevolution and patterns of adaptation across time and space. *J. Evol. Biol.* *21*, 1861–1866.

García-Aljaro, C., Ballesté, E., Muniesa, M., and Jofre, J. (2017). Determination of crAssphage in water samples and applicability for tracking human faecal pollution. *Microb. Biotechnol.* *10*, 1775–1780.

Garmaeva, S., Sinha, T., Kurilshikov, A., Fu, J., Wijmenga, C., and Zhernakova, A. (2019). Studying the gut virome in the metagenomic era: challenges and perspectives. *BMC Biol.* *17*, 84.

Gencay, Y.E., Sørensen, M.C.H., Wenzel, C.Q., Szymanski, C.M., and Brøndsted, L. (2018). Phase Variable Expression of a Single Phage Receptor in *Campylobacter jejuni* NCTC12662 Influences Sensitivity Toward Several Diverse CPS-Dependent Phages. *Front. Microbiol.* 9.

Gevers, D., Kugathasan, S., Denson, L.A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S.J., Yassour, M., et al. (2014). The Treatment-Naive Microbiome in New-Onset Crohn's Disease. *Cell Host Microbe* 15, 382–392.

Gill, S.R., Pop, M., DeBoy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., Gordon, J.I., Relman, D.A., Fraser-Liggett, C.M., and Nelson, K.E. (2006). Metagenomic Analysis of the Human Distal Gut Microbiome. *Science* 312, 1355–1359.

Giongo, A., Gano, K.A., Crabb, D.B., Mukherjee, N., Novelo, L.L., Casella, G., Drew, J.C., Ilonen, J., Knip, M., Hyöty, H., et al. (2011). Toward defining the autoimmune microbiome for type 1 diabetes. *ISME J.* 5, 82–91.

Gkouskou, K.K., Deligianni, C., Tsatsanis, C., and Eliopoulos, A.G. (2014). The gut microbiota in mouse models of inflammatory bowel disease. *Front. Cell. Infect. Microbiol.* 4, 28.

Goerke, C., Köller, J., and Wolz, C. (2006). Ciprofloxacin and trimethoprim cause phage induction and virulence modulation in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 50, 171–177.

Gogokhia, L., Buhrke, K., Bell, R., Hoffman, B., Brown, D.G., Hanke-Gogokhia, C., Ajami, N.J., Wong, M.C., Ghazaryan, A., Valentine, J.F., et al. (2019). Expansion of Bacteriophages Is Linked to Aggravated Intestinal Inflammation and Colitis. *Cell Host Microbe* 25, 285-299.e8.

Goldfarb, T., Sberro, H., Weinstock, E., Cohen, O., Doron, S., Charpak-Amikam, Y., Afik, S., Ofir, G., and Sorek, R. (2015). BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J.* *34*, 169–183.

Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J.T., et al. (2014). Human Genetics Shape the Gut Microbiome. *Cell* *159*, 789–799.

Gorvitovskaia, A., Holmes, S.P., and Huse, S.M. (2016). Interpreting Prevotella and Bacteroides as biomarkers of diet and lifestyle. *Microbiome* *4*, 15.

Gregory, A.C., Zablocki, O., Howell, A., Bolduc, B., and Sullivan, M.B. (2019). The human gut virome database. *BioRxiv* 655910.

Guerin, E., Shkoporov, A., Stockdale, S.R., Clooney, A.G., Ryan, F.J., Sutton, T.D.S., Draper, L.A., Gonzalez-Tortuero, E., Ross, R.P., and Hill, C. (2018). Biology and Taxonomy of crAss-like Bacteriophages, the Most Abundant Virus in the Human Gut. *Cell Host Microbe* *24*, 653-664.e6.

Halfvarson, J., Brislawn, C.J., Lamendella, R., Vázquez-Baeza, Y., Walters, W.A., Bramer, L.M., D’Amato, M., Bonfiglio, F., McDonald, D., Gonzalez, A., et al. (2017). Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat. Microbiol.* *2*, 17004.

Hall, A.R., Scanlan, P.D., Morgan, A.D., and Buckling, A. (2011). Host-parasite coevolutionary arms races give way to fluctuating selection. *Ecol. Lett.* *14*, 635–642.

Hargreaves, K.R., Kropinski, A.M., and Clokie, M.R.J. (2014). What Does the Talking?: Quorum Sensing Signalling Genes Discovered in a Bacteriophage Genome. *PLOS ONE* *9*, e85131.

- Hartstra, A.V., Bouter, K.E.C., Bäckhed, F., and Nieuwdorp, M. (2015). Insights into the role of the microbiome in obesity and type 2 diabetes. *Diabetes Care* 38, 159–165.
- Helmkink, B.A., Khan, M.A.W., Hermann, A., Gopalakrishnan, V., and Wargo, J.A. (2019). The microbiome, cancer, and cancer therapy. *Nat. Med.* 25, 377–388.
- Hendrix, H., Kogadeeva, M., Zimmermann, M., Sauer, U., Smet, J.D., Muchez, L., Lissens, M., Staes, I., Voet, M., Wagemans, J., et al. (2019). Host metabolic reprogramming of *Pseudomonas aeruginosa* by phage-based quorum sensing modulation. *BioRxiv* 577908.
- d’Hérelle, F. (1917). Sur un microbe invisible antagoniste des bacilles dysentériques. *CR Acad Sci Paris* 165, 373–375.
- Hobbs, Z., and Abedon, S.T. (2016). Diversity of phage infection types and associated terminology: the problem with ‘Lytic or lysogenic.’ *FEMS Microbiol. Lett.* 363.
- Holt, R.D. (1985). Population dynamics in two-patch environments: Some anomalous consequences of an optimal habitat distribution. *Theor. Popul. Biol.* 28, 181–208.
- Honap, T.P., Sankaranarayanan, K., Schnorr, S.L., Ozga, A.T., Warinner, C., and Jr, C.M.L. (2020). Biogeographic study of human gut-associated crAssphage suggests impacts from industrialization and recent expansion. *PLOS ONE* 15, e0226930.
- Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the Immune System of Bacteria and Archaea. *Science* 327, 167–170.
- Howard-Varona, C., Hargreaves, K.R., Abedon, S.T., and Sullivan, M.B. (2017). Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J.* 11, 1511–1520.

Høyland-Kroghsbo, N.M., Mærkedahl, R.B., and Svenningsen, S.L. (2013). A Quorum-Sensing-Induced Bacteriophage Defense Mechanism. *MBio* 4.

Høyland-Kroghsbo, N.M., Paczkowski, J., Mukherjee, S., Broniewski, J., Westra, E., Bondy-Denomy, J., and Bassler, B.L. (2017). Quorum sensing controls the *Pseudomonas aeruginosa* CRISPR-Cas adaptive immune system. *Proc. Natl. Acad. Sci. U. S. A.* 114, 131–135.

Hoyles, L., McCartney, A.L., Neve, H., Gibson, G.R., Sanderson, J.D., Heller, K.J., and van Sinderen, D. (2014). Characterization of virus-like particles associated with the human faecal and caecal microbiota. *Res. Microbiol.* 165, 803–812.

Hryckowian, A.J., Merrill, B.D., Porter, N.T., Treuren, W.V., Nelson, E.J., Garlena, R.A., Russell, D.A., Martens, E.C., and Sonnenburg, J.L. (2020). *Bacteroides* thetaiotaomicron-infecting bacteriophage isolates inform sequence-based host range predictions. *BioRxiv* 2020.03.04.977157.

Hsu, B.B., Gibson, T.E., Yeliseyev, V., Liu, Q., Lyon, L., Bry, L., Silver, P.A., and Gerber, G.K. (2019). Dynamic Modulation of the Gut Microbiota and Metabolome by Bacteriophages in a Mouse Model. *Cell Host Microbe* 25, 803-814.e5.

d’Humières, C., Touchon, M., Dion, S., Cury, J., Ghoulane, A., Garcia-Garcera, M., Bouchier, C., Ma, L., Denamur, E., and P.C.Rocha, E. (2019). A simple, reproducible and cost-effective procedure to analyse gut phageome: from phage isolation to bioinformatic approach. *Sci. Rep.* 9, 1–13.

Hurwitz, B.L., Westveld, A.H., Brum, J.R., and Sullivan, M.B. (2014). Modeling ecological drivers in marine viral communities using comparative metagenomics and network analyses. *Proc. Natl. Acad. Sci.* 111, 10714–10719.

- Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J.H., Chinwalla, A.T., Creasy, H.H., Earl, A.M., FitzGerald, M.G., Fulton, R.S., et al. (2012). Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214.
- Hyman, P. (2019). Phages for Phage Therapy: Isolation, Characterization, and Host Range Breadth. *Pharmaceuticals* 12, 35.
- Jang, H.B., Bolduc, B., Zablocki, O., Kuhn, J.H., Roux, S., Adriaenssens, E.M., Brister, J.R., Kropinski, A.M., Krupovic, M., Lavigne, R., et al. (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* 37, 632–639.
- Jiang, X., Hall, A.B., Arthur, T.D., Plichta, D.R., Covington, C.T., Poyet, M., Crothers, J., Moses, P.L., Tolonen, A.C., Vlamakis, H., et al. (2019). Invertible promoters mediate bacterial phase variation, antibiotic resistance, and host adaptation in the gut. *Science* 363, 181–187.
- Joiner, K.L., Baljon, A., Barr, J., Rohwer, F., and Luque, A. (2019). Impact of bacteria motility in the encounter rates with bacteriophage in mucus. *Sci. Rep.* 9, 1–12.
- de Jonge, P.A., Nobrega, F.L., Brouns, S.J.J., and Dutilh, B.E. (2019a). Molecular and Evolutionary Determinants of Bacteriophage Host Range. *Trends Microbiol.* 27, 51–63.
- de Jonge, P.A., Meijenfeldt, F.A.B. von, Rooijen, L.E. van, Brouns, S.J.J., and Dutilh, B.E. (2019b). Evolution of BACON Domain Tandem Repeats in crAssphage and Novel Gut Bacteriophage Lineages. *Viruses* 11.
- Kai, M., Watanabe, S., Furuse, K., and Ozawa, A. (1985). *Bacteroides* bacteriophages isolated from human feces. *Microbiol. Immunol.* 29, 895–899.

- Karkman, A., Pärnänen, K., and Larsson, D.G.J. (2019). Fecal pollution can explain antibiotic resistance gene abundances in anthropogenically impacted environments. *Nat. Commun.* *10*, 80.
- Karlsson, F.H., Ussery, D.W., Nielsen, J., and Nookaew, I. (2011). A Closer Look at Bacteroides: Phylogenetic Relationship and Genomic Implications of a Life in the Human Gut. *Microb. Ecol.* *61*, 473–485.
- Kieser, S., Sarker, S.A., Sakwinska, O., Foata, F., Sultana, S., Khan, Z., Islam, S., Porta, N., Combremont, S., Betrisey, B., et al. (2018). Bangladeshi children with acute diarrhoea show faecal microbiomes with increased Streptococcus abundance, irrespective of diarrhoea aetiology. *Environ. Microbiol.* *20*, 2256–2269.
- Kim, M.-S., Park, E.-J., Roh, S.W., and Bae, J.-W. (2011). Diversity and Abundance of Single-Stranded DNA Viruses in Human Feces. *Appl. Environ. Microbiol.* *77*, 8062–8070.
- Kleiner, M., Hooper, L.V., and Duerkop, B.A. (2015). Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. *BMC Genomics* *16*.
- Knowles, B., Silveira, C.B., Bailey, B.A., Barott, K., Cantu, V.A., Cobián-Güemes, A.G., Coutinho, F.H., Dinsdale, E.A., Felts, B., Furby, K.A., et al. (2016). Lytic to temperate switching of viral communities. *Nature* *531*, 466–470.
- Korf, I.H.E., Meier-Kolthoff, J.P., Adriaenssens, E.M., Kropinski, A.M., Nimtz, M., Rohde, M., van Raaij, M.J., and Wittmann, J. (2019). Still Something to Discover: Novel Insights into Escherichia coli Phage Diversity and Taxonomy. *Viruses* *11*, 454.

Koskella, B., and Meaden, S. (2013). Understanding Bacteriophage Specificity in Natural Microbial Communities. *Viruses* 5, 806–823.

Koziolek, M., Grimm, M., Becker, D., Iordanov, V., Zou, H., Shimizu, J., Wanke, C., Garbacz, G., and Weitschies, W. (2015). Investigation of pH and Temperature Profiles in the GI Tract of Fasted Human Subjects Using the Intellicap(®) System. *J. Pharm. Sci.* 104, 2855–2863.

Labonté, J.M., Swan, B.K., Poulos, B., Luo, H., Koren, S., Hallam, S.J., Sullivan, M.B., Woyke, T., Wommack, K.E., and Stepanauskas, R. (2015). Single-cell genomics-based analysis of virus–host interactions in marine surface bacterioplankton. *ISME J.* 9, 2386–2399.

Labrie, S.J., Samson, J.E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* 8, 317–327.

Laganenka, L., Sander, T., Lagonenko, A., Chen, Y., Link, H., and Sourjik, V. (2019). Quorum Sensing and Metabolic State of the Host Control Lysogeny-Lysis Switch of Bacteriophage T1. *MBio* 10.

Lagier, J.-C., Armougom, F., Million, M., Hugon, P., Pagnier, I., Robert, C., Bittar, F., Fournous, G., Gimenez, G., Maraninchi, M., et al. (2012). Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin. Microbiol. Infect.* 18, 1185–1193.

Lagier, J.-C., Khelaifia, S., Alou, M.T., Ndongo, S., Dione, N., Hugon, P., Caputo, A., Cadoret, F., Traore, S.I., Seck, E.H., et al. (2016). Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat. Microbiol.* 1, 1–8.

- Lagier, J.-C., Dubourg, G., Million, M., Cadoret, F., Bilen, M., Fenollar, F., Levasseur, A., Rolain, J.-M., Fournier, P.-E., and Raoult, D. (2018). Culturing the human microbiota and culturomics. *Nat. Rev. Microbiol.* *16*, 540–550.
- Lawrence, D., Baldrige, M.T., and Handley, S.A. (2019). Phages and Human Health: More Than Idle Hitchhikers. *Viruses* *11*, 587.
- Lefkowitz, E.J., Dempsey, D.M., Hendrickson, R.C., Orton, R.J., Siddell, S.G., and Smith, D.B. (2018). Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.* *46*, D708–D717.
- Leiman, P.G., Battisti, A.J., Bowman, V.D., Stummeyer, K., Mühlenhoff, M., Gerardy-Schahn, R., Scholl, D., and Molineux, I.J. (2007). The structures of bacteriophages K1E and K1-5 explain processive degradation of polysaccharide capsules and evolution of new host specificities. *J. Mol. Biol.* *371*, 836–849.
- Lewis, R., and Hill, C. (2020). Overcoming barriers to phage application in food and feed. *Curr. Opin. Biotechnol.* *61*, 38–44.
- Liang, Y.Y., Zhang, W., Tong, Y.G., and Chen, S.P. (2016). CrAssphage is not associated with diarrhoea and has high genetic diversity. *Epidemiol. Infect.* *144*, 3549–3553.
- Lim, E.S., Zhou, Y., Zhao, G., Bauer, I.K., Droit, L., Ndao, I.M., Warner, B.B., Tarr, P.I., Wang, D., and Holtz, L.R. (2015). Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat. Med.* *21*, 1228–1234.
- Lima-Mendez, G., Toussaint, A., and Leplae, R. (2011). A modular view of the bacteriophage genomic space: identification of host and lifestyle marker modules. *Res. Microbiol.* *162*, 737–746.

- Lin, D.M., Koskella, B., and Lin, H.C. (2017). Phage therapy: An alternative to antibiotics in the age of multi-drug resistance. *World J. Gastrointest. Pharmacol. Ther.* 8, 162–173.
- Lloyd-Price, J., Abu-Ali, G., and Huttenhower, C. (2016). The healthy human microbiome. *Genome Med.* 8, 51.
- Łoś, M., and Węgrzyn, G. (2012). Pseudolysogeny. *Adv. Virus Res.* 82, 339–349.
- Lourenço, M., De Sordi, L., and Debarbieux, L. (2018). The Diversity of Bacterial Lifestyles Hampers Bacteriophage Tenacity. *Viruses* 10.
- Lourenço, M., Chaffringeon, L., Lamy-Besnier, Q., Campagne, P., Eberl, C., Bérard, M., Stecher, B., Debarbieux, L., and Sordi, L.D. (2019). The spatial heterogeneity of the gut limits bacteriophage predation leading to the coexistence of antagonist populations of bacteria and their viruses. *BioRxiv* 810705.
- Love, M.J., Bhandari, D., Dobson, R.C.J., and Billington, C. (2018). Potential for Bacteriophage Endolysins to Supplement or Replace Antibiotics in Food Production and Clinical Care. *Antibiotics* 7, 17.
- Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., and Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature* 489, 220–230.
- Lu, M.J., Stierhof, Y.D., and Henning, U. (1993). Location and unusual membrane topology of the immunity protein of the Escherichia coli phage T4. *J. Virol.* 67, 4905–4913.
- Ma, Y., You, X., Mai, G., Tokuyasu, T., and Liu, C. (2018). A human gut phage catalog correlates the gut phageome with type 2 diabetes. *Microbiome* 6, 24.

- Manning, A.J., and Kuehn, M.J. (2011). Contribution of bacterial outer membrane vesicles to innate bacterial defense. *BMC Microbiol.* *11*, 258.
- Manrique, P., Bolduc, B., Walk, S.T., van der Oost, J., de Vos, W.M., and Young, M.J. (2016). Healthy human gut phageome. *Proc. Natl. Acad. Sci.* *113*, 10400–10405.
- Marine, R., McCarren, C., Vorrasane, V., Nasko, D., Crowgey, E., Polson, S.W., and Wommack, K.E. (2014). Caught in the middle with multiple displacement amplification: the myth of pooling for avoiding multiple displacement amplification bias in a metagenome. *Microbiome* *2*, 3.
- Martínez, J.M., Swan, B.K., and Wilson, W.H. (2014). Marine viruses, a genetic reservoir revealed by targeted viromics. *ISME J.* *8*, 1079–1088.
- Martinez-Guryn, K., Leone, V., and Chang, E.B. (2019). Regional Diversity of the Gastrointestinal Microbiome. *Cell Host Microbe* *26*, 314–324.
- Martinez-Hernandez, F., Fornas, O., Gomez, M.L., Bolduc, B., Peña, M.J. de la C., Martínez, J.M., Anton, J., Gasol, J.M., Rosselli, R., Rodriguez-Valera, F., et al. (2017). Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat. Commun.* *8*, 1–13.
- Maslov, S., and Sneppen, K. (2017). Population cycles and species diversity in dynamic Kill-the-Winner model of microbial ecosystems. *Sci. Rep.* *7*, 39642.
- Maura, D., Morello, E., du Merle, L., Bomme, P., Le Bouguénec, C., and Debarbieux, L. (2012a). Intestinal colonization by enteroaggregative *Escherichia coli* supports long-term bacteriophage replication in mice. *Environ. Microbiol.* *14*, 1844–1854.

Maura, D., Galtier, M., Bouguénec, C.L., and Debarbieux, L. (2012b). Virulent Bacteriophages Can Target O104:H4 Enterohaggregative *Escherichia coli* in the Mouse Intestine. *Antimicrob. Agents Chemother.* *56*, 6235–6242.

McCann, A., Ryan, F.J., Stockdale, S.R., Dalmasso, M., Blake, T., Ryan, C.A., Stanton, C., Mills, S., Ross, P.R., and Hill, C. (2018). Viromes of one year old infants reveal the impact of birth mode on microbiome diversity. *PeerJ* *6*, e4694.

Meier-Kolthoff, J.P., and Göker, M. (2017). VICTOR: genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics* *33*, 3396–3404.

Meyer, J.R., Dobias, D.T., Weitz, J.S., Barrick, J.E., Quick, R.T., and Lenski, R.E. (2012). Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* *335*, 428–432.

Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2011). The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res.* *21*, 1616–1625.

Minot, S., Grunberg, S., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2012). Hypervariable loci in the human gut virome. *Proc. Natl. Acad. Sci.* *109*, 3962–3966.

Minot, S., Bryson, A., Chehoud, C., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2013). Rapid evolution of the human gut virome. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 12450–12455.

Mirzaei, M.K., and Maurice, C.F. (2017). Ménage à trois in the human gut: interactions between host, bacteria and phages. *Nat. Rev. Microbiol.* *15*, 397–408.

Monaco, C.L., Gootenberg, D.B., Zhao, G., Handley, S.A., Ghebremichael, M.S., Lim, E.S., Lankowski, A., Baldridge, M.T., Wilen, C.B., Flagg, M., et al. (2016).

Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. *Cell Host Microbe* 19, 311–322.

Moreno-Gallego, J.L., Chou, S.-P., Rienzi, S.C.D., Goodrich, J.K., Spector, T.D., Bell, J.T., Youngblut, N.D., Hewson, I., Reyes, A., and Ley, R.E. (2019). Virome Diversity Correlates with Intestinal Microbiome Diversity in Adult Monozygotic Twins. *Cell Host Microbe* 25, 261-272.e5.

Moustafa, A., Xie, C., Kirkness, E., Biggs, W., Wong, E., Turpaz, Y., Bloom, K., Delwart, E., Nelson, K.E., Venter, J.C., et al. (2017). The blood DNA virome in 8,000 humans. *PLOS Pathog.* 13, e1006292.

Moxon, E.R., Rainey, P.B., Nowak, M.A., and Lenski, R.E. (1994). Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* CB 4, 24–33.

Nakayama-Imaohji, H., Hirakawa, H., Ichimura, M., Wakimoto, S., Kuhara, S., Hayashi, T., and Kuwahara, T. (2009). Identification of the Site-Specific DNA Invertase Responsible for the Phase Variation of SusC/SusD Family Outer Membrane Proteins in *Bacteroides fragilis*. *J. Bacteriol.* 191, 6003–6011.

Nakayama-Imaohji, H., Hirota, K., Yamasaki, H., Yoneda, S., Nariya, H., Suzuki, M., Secher, T., Miyake, Y., Oswald, E., Hayashi, T., et al. (2016). DNA Inversion Regulates Outer Membrane Vesicle Production in *Bacteroides fragilis*. *PloS One* 11, e0148887.

Nanda, A.M., Thormann, K., and Frunzke, J. (2015). Impact of Spontaneous Prophage Induction on the Fitness of Bacterial Populations and Host-Microbe Interactions. *J. Bacteriol.* 197, 410–419.

- Ng, W.-L., and Bassler, B.L. (2009). Bacterial Quorum-Sensing Network Architectures. *Annu. Rev. Genet.* *43*, 197–222.
- Nguyen, S., Baker, K., Padman, B.S., Patwa, R., Dunstan, R.A., Weston, T.A., Schlosser, K., Bailey, B., Lithgow, T., Lazarou, M., et al. (2017). Bacteriophage Transcytosis Provides a Mechanism To Cross Epithelial Cell Layers. *MBio* *8*, e01874-17.
- Norman, J.M., Handley, S.A., Baldridge, M.T., Droit, L., Liu, C.Y., Keller, B.C., Kambal, A., Monaco, C.L., Zhao, G., and Fleshner, P. (2015). Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* *160*, 447–460.
- Obeng, N., Pratama, A.A., and Elsas, J.D. van (2016). The Significance of Mutualistic Phages for Bacterial Ecology and Evolution. *Trends Microbiol.* *24*, 440–449.
- O'Donnell, M.M., Rea, M.C., O'Sullivan, Ó., Flynn, C., Jones, B., McQuaid, A., Shanahan, F., and Ross, R.P. (2016). Preparation of a standardised faecal slurry for ex-vivo microbiota studies which reduces inter-individual donor bias. *J. Microbiol. Methods* *129*, 109–116.
- Oechslin, F. (2018). Resistance Development to Bacteriophages Occurring during Bacteriophage Therapy. *Viruses* *10*, 351.
- Ofir, G., Melamed, S., Sberro, H., Mukamel, Z., Silverman, S., Yaakov, G., Doron, S., and Sorek, R. (2018). DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nat. Microbiol.* *3*, 90–98.
- O'Hara, A.M., and Shanahan, F. (2006). The gut flora as a forgotten organ. *EMBO Rep.* *7*, 688–693.

- O’Sullivan, L., Bolton, D., McAuliffe, O., and Coffey, A. (2019). Bacteriophages in Food Applications: From Foe to Friend. *Annu. Rev. Food Sci. Technol.* *10*, 151–172.
- Ott, S.J., Waetzig, G.H., Rehman, A., Moltzau-Anderson, J., Bharti, R., Grasis, J.A., Cassidy, L., Tholey, A., Fickenscher, H., Seegert, D., et al. (2017). Efficacy of Sterile Fecal Filtrate Transfer for Treating Patients With *Clostridium difficile* Infection. *Gastroenterology* *152*, 799-811.e7.
- Oude Munnink, B.B., Canuti, M., Deijs, M., de Vries, M., Jebbink, M.F., Rebers, S., Molenkamp, R., van Hemert, F.J., Chung, K., Cotten, M., et al. (2014). Unexplained diarrhoea in HIV-1 infected individuals. *BMC Infect. Dis.* *14*, 22.
- Paez-Espino, D., Roux, S., Chen, I.-M.A., Palaniappan, K., Ratner, A., Chu, K., Huntemann, M., Reddy, T.B.K., Pons, J.C., Llabrés, M., et al. (2019). IMG/VR v.2.0: an integrated data management and analysis system for cultivated and environmental viral genomes. *Nucleic Acids Res.* *47*, D678–D686.
- Papenfort, K., and Bassler, B.L. (2016). Quorum sensing signal–response systems in Gram-negative bacteria. *Nat. Rev. Microbiol.* *14*, 576–588.
- Pereira, F.C., and Berry, D. (2017). Microbial nutrient niches in the gut. *Environ. Microbiol.* *19*, 1366–1378.
- Pitsalidis, C., Pappa, A.-M., Porel, M., Artim, C.M., Faria, G.C., Duong, D.D., Alabi, C.A., Daniel, S., Salleo, A., and Owens, R.M. (2018). Biomimetic Electronic Devices for Measuring Bacterial Membrane Disruption. *Adv. Mater.* *30*, 1803130.
- Porter, N.T., Hryckowian, A.J., Merrill, B.D., Fuentes, J.J., Gardner, J.O., Glowacki, R.W.P., Singh, S., Crawford, R.D., Snitkin, E.S., Sonnenburg, J.L., et al. (2020).

Multiple phase-variable mechanisms, including capsular polysaccharides, modify bacteriophage susceptibility in *Bacteroides thetaiotaomicron*. *BioRxiv* 521070.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65.

Rajilić-Stojanović, M., and de Vos, W.M. (2014). The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiol. Rev.* 38, 996–1047.

Ram, G., Chen, J., Kumar, K., Ross, H.F., Ubeda, C., Damle, P.K., Lane, K.D., Penadés, J.R., Christie, G.E., and Novick, R.P. (2012). Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. *Proc. Natl. Acad. Sci. U. S. A.* 109, 16300–16305.

Reyes, A., Haynes, M., Hanson, N., Angly, F.E., Heath, A.C., Rohwer, F., and Gordon, J.I. (2010). Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 466, 334–338.

Reyes, A., Wu, M., McNulty, N.P., Rohwer, F.L., and Gordon, J.I. (2013). Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. *Proc. Natl. Acad. Sci. U. S. A.* 110, 20236–20241.

Reyes, A., Blanton, L.V., Cao, S., Zhao, G., Manary, M., Trehan, I., Smith, M.I., Wang, D., Virgin, H.W., Rohwer, F., et al. (2015). Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc. Natl. Acad. Sci.* 112, 11941–11946.

- Riley, T.V., Brazier, J.S., Hassan, H., Williams, K., and Phillips, K.D. (1987). Comparison of alcohol shock enrichment and selective enrichment for the isolation of *Clostridium difficile*. *Epidemiol. Infect.* 99, 355–359.
- Roach, D.R., Leung, C.Y., Henry, M., Morello, E., Singh, D., Santo, J.P.D., Weitz, J.S., and Debarbieux, L. (2017). Synergy between the Host Immune System and Bacteriophage Is Essential for Successful Phage Therapy against an Acute Respiratory Pathogen. *Cell Host Microbe* 22, 38-47.e4.
- Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtyarev, S.K., Dryden, D.T.F., Dybvig, K., et al. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* 31, 1805–1812.
- Rodríguez, J.M., Murphy, K., Stanton, C., Ross, R.P., Kober, O.I., Juge, N., Avershina, E., Rudi, K., Narbad, A., Jenmalm, M.C., et al. (2015). The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb. Ecol. Health Dis.* 26, 26050.
- Rodriguez-Brito, B., Li, L., Wegley, L., Furlan, M., Angly, F., Breitbart, M., Buchanan, J., Desnues, C., Dinsdale, E., Edwards, R., et al. (2010). Viral and microbial community dynamics in four aquatic environments. *ISME J.* 4, 739–751.
- Rostøl, J.T., and Marraffini, L. (2019). (Ph)ighting Phages: How Bacteria Resist Their Parasites. *Cell Host Microbe* 25, 184–194.
- Roux, S., Krupovic, M., Debroas, D., Forterre, P., and Enault, F. (2013). Assessment of viral community functional potential from viral metagenomes may be hampered by contamination with cellular sequences. *Open Biol.* 3, 130160.

- Roux, S., Solonenko, N.E., Dang, V.T., Poulos, B.T., Schwenck, S.M., Goldsmith, D.B., Coleman, M.L., Breitbart, M., and Sullivan, M.B. (2016). Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ* 4, e2777.
- Roux, S., Emerson, J.B., Elie-Fadrosh, E.A., and Sullivan, M.B. (2017). Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* 5, e3817.
- Salem, M., Virtanen, S., Korkeala, H., and Skurnik, M. (2015). Isolation and characterization of Yersinia-specific bacteriophages from pig stools in Finland. *J. Appl. Microbiol.* 118, 599–608.
- Samson, J.E., Magadán, A.H., Sabri, M., and Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. *Nat. Rev. Microbiol.* 11, 675–687.
- Santiago-Rodriguez, T.M., Ly, M., Daigneault, M.C., Brown, I.H.L., McDonald, J.A.K., Bonilla, N., Vercoe, E.A., and Pride, D.T. (2015a). Chemostat culture systems support diverse bacteriophage communities from human feces. *Microbiome* 3, 58.
- Santiago-Rodriguez, T.M., Naidu, M., Abeles, S.R., Boehm, T.K., Ly, M., and Pride, D.T. (2015b). Transcriptome analysis of bacteriophage communities in periodontal health and disease. *BMC Genomics* 16, 549.
- Scanlan, P.D., Hall, A.R., Blackshields, G., Friman, V.-P., Davis, M.R., Goldberg, J.B., and Buckling, A. (2015). Coevolution with Bacteriophages Drives Genome-Wide Host Evolution and Constrains the Acquisition of Abiotic-Beneficial Mutations. *Mol. Biol. Evol.* 32, 1425–1435.
- Schrag, S.J., and Mittler, J.E. (1996). Host-Parasite Coexistence: The Role of Spatial Refuges in Stabilizing Bacteria-Phage Interactions. *Am. Nat.* 148, 348–377.

Schwechheimer, C., and Kuehn, M.J. (2015). Outer-membrane vesicles from Gram-negative bacteria: biogenesis and functions. *Nat. Rev. Microbiol.* *13*, 605–619.

Seed, K.D. (2015). Battling Phages: How Bacteria Defend against Viral Attack. *PLOS Pathog.* *11*, e1004847.

Seed, K.D., Lazinski, D.W., Calderwood, S.B., and Camilli, A. (2013). A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* *494*, 489–491.

Sender, R., Fuchs, S., and Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biol.* *14*, e1002533.

Shanahan, F., and Hill, C. (2019). Language, numeracy and logic in microbiome science. *Nat. Rev. Gastroenterol. Hepatol.* *16*, 387–388.

Shapiro, J.W., and Putonti, C. (2018). Gene Co-occurrence Networks Reflect Bacteriophage Ecology and Evolution. *MBio* *9*.

Shkoporov, A.N., and Hill, C. (2019). Bacteriophages of the Human Gut: The “Known Unknown” of the Microbiome. *Cell Host Microbe* *25*, 195–209.

Shkoporov, A.N., Khokhlova, E.V., Fitzgerald, C.B., Stockdale, S.R., Draper, L.A., Ross, R.P., and Hill, C. (2018a). ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. *Nat. Commun.* *9*.

Shkoporov, A.N., Ryan, F.J., Draper, L.A., Forde, A., Stockdale, S.R., Daly, K.M., McDonnell, S.A., Nolan, J.A., Sutton, T.D.S., Dalmaso, M., et al. (2018b). Reproducible protocols for metagenomic analysis of human faecal phageomes. *Microbiome* *6*, 68.

Shkoporov, A.N., Clooney, A.G., Sutton, T.D.S., Ryan, F.J., Daly, K.M., Nolan, J.A., McDonnell, S.A., Khokhlova, E.V., Draper, L.A., Forde, A., et al. (2019). The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific. *Cell Host Microbe* 26, 527-541.e5.

Silpe, J.E., and Bassler, B.L. (2019). A Host-Produced Quorum-Sensing Autoinducer Controls a Phage Lysis-Lysogeny Decision. *Cell* 176, 268-280.e13.

Silveira, C.B., and Rohwer, F.L. (2016). Piggyback-the-Winner in host-associated microbial communities. *Npj Biofilms Microbiomes* 2, 16010.

Simmonds, P., Adams, M.J., Benkő, M., Breitbart, M., Brister, J.R., Carstens, E.B., Davison, A.J., Delwart, E., Gorbalenya, A.E., Harrach, B., et al. (2017). Consensus statement: Virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol.* 15, 161–168.

Siranosian, B.A., Tamburini, F.B., Sherlock, G., and Bhatt, A.S. (2020). Acquisition, transmission and strain diversity of human gut-colonizing crAss-like phages. *Nat. Commun.* 11, 1–11.

Siringan, P., Connerton, P.L., Cummings, N.J., and Connerton, I.F. (2014). Alternative bacteriophage life cycles: the carrier state of *Campylobacter jejuni*. *Open Biol.* 4, 130200–130200.

Smeal, S.W., Schmitt, M.A., Pereira, R.R., Prasad, A., and Fisk, J.D. (2017a). Simulation of the M13 life cycle I: Assembly of a genetically-structured deterministic chemical kinetic simulation. *Virology* 500, 259–274.

- Smeal, S.W., Schmitt, M.A., Pereira, R.R., Prasad, A., and Fisk, J.D. (2017b). Simulation of the M13 life cycle II: Investigation of the control mechanisms of M13 infection and establishment of the carrier state. *Virology* 500, 275–284.
- Somerville, V., Lutz, S., Schmid, M., Frei, D., Moser, A., Irmeler, S., Frey, J.E., and Ahrens, C.H. (2019). Long-read based de novo assembly of low-complexity metagenome samples results in finished genomes and reveals insights into strain diversity and an active phage system. *BMC Microbiol.* 19, 143.
- Stachler, E., Kelty, C., Sivaganesan, M., Li, X., Bibby, K., and Shanks, O.C. (2017). Quantitative CrAssphage pcr assays for human fecal pollution measurement. *Environ. Sci. Technol.* 51, 9146–9154.
- Stämmler, F., Gläsner, J., Hiergeist, A., Holler, E., Weber, D., Oefner, P.J., Gessner, A., and Spang, R. (2016). Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. *Microbiome* 4, 28.
- Stockdale, S.R., Ryan, F.J., McCann, A., Dalmasso, M., Ross, P.R., and Hill, C. (2018). Viral Dark Matter in the Gut Virome of Elderly Humans.
- Strakosas, X., Bongo, M., and Owens, R.M. (2015). The organic electrochemical transistor for biological applications. *J. Appl. Polym. Sci.* 132.
- Su, H., Liu, H.-Y., Pappa, A.-M., Hidalgo, T.C., Cavassin, P., Inal, S., Owens, R.M., and Daniel, S. (2019). Facile Generation of Biomimetic-Supported Lipid Bilayers on Conducting Polymer Surfaces for Membrane Biosensing. *ACS Appl. Mater. Interfaces*.
- Sugden, R., Kelly, R., and Davies, S. (2016). Combatting antimicrobial resistance globally. *Nat. Microbiol.* 1, 1–2.

Sutton, T.D.S., and Hill, C. (2019). Gut bacteriophage: Current understanding and challenges. *Front. Endocrinol.* 10.

Sutton, T.D.S., Clooney, A.G., Ryan, F.J., Ross, R.P., and Hill, C. (2019). Choice of assembly software has a critical impact on virome characterisation. *Microbiome* 7, 12.

Swan, B.K., Tupper, B., Sczyrba, A., Lauro, F.M., Martinez-Garcia, M., González, J.M., Luo, H., Wright, J.J., Landry, Z.C., Hanson, N.W., et al. (2013). Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc. Natl. Acad. Sci.* 110, 11463–11468.

Tetz, G., Brown, S.M., Hao, Y., and Tetz, V. (2018). Parkinson's disease and bacteriophages as its overlooked contributors. *Sci. Rep.* 8, 1–11.

Thingstad, T.F. (2000). Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol. Oceanogr.* 45, 1320–1328.

Tock, M.R., and Dryden, D.T.F. (2005). The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.* 8, 466–472.

Tria, S.A., Ramuz, M., Huerta, M., Leleux, P., Rivnay, J., Jimison, L.H., Hama, A., Malliaras, G.G., and Owens, R.M. (2014). Dynamic Monitoring of *Salmonella typhimurium* Infection of Polarized Epithelia Using Organic Transistors. *Adv. Healthc. Mater.* 3, 1053–1060.

Turkington, C.J.R., Morozov, A., Clokie, M.R.J., and Bayliss, C.D. (2019). Phage-Resistant Phase-Variant Sub-populations Mediate Herd Immunity Against Bacteriophage Invasion of Bacterial Meta-Populations. *Front. Microbiol.* 10.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., and Gordon, J.I. (2007). The Human Microbiome Project. *Nature* 449, 804–810.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., et al. (2009). A core gut microbiome in obese and lean twins. *Nature* 457, 480–484.

Turovskiy, Y., Kashtanov, D., Paskhover, B., and Chikindas, M.L. (2007). Quorum Sensing: Fact, Fiction, and Everything in Between. *Adv. Appl. Microbiol.* 62, 191–234.

Twort, F.W. (1915). An investigation on the nature of ultra-microscopic viruses. *The Lancet* 186, 1241–1243.

Tzipilevich, E., Habusha, M., and Ben-Yehuda, S. (2017). Acquisition of Phage Sensitivity by Bacteria through Exchange of Phage Receptors. *Cell* 168, 186-199.e12.

Ursell, L.K., Metcalf, J.L., Parfrey, L.W., and Knight, R. (2012). Defining the Human Microbiome. *Nutr. Rev.* 70, S38–S44.

Van Belleghem, J.D., Clement, F., Merabishvili, M., Lavigne, R., and Vaneechoutte, M. (2017). Pro- and anti-inflammatory responses of peripheral blood mononuclear cells induced by *Staphylococcus aureus* and *Pseudomonas aeruginosa* phages. *Sci. Rep.* 7, 1–13.

Van Belleghem, J.D., Dąbrowska, K., Vaneechoutte, M., Barr, J.J., and Bollyky, P.L. (2019). Interactions between Bacteriophage, Bacteria, and the Mammalian Immune System. *Viruses* 11, 10.

Vandeputte, D., Falony, G., Vieira-Silva, S., Tito, R.Y., Joossens, M., and Raes, J. (2016). Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut* 65, 57–62.

Vandeputte, D., Kathagen, G., D’hoë, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., Wang, J., Tito, R.Y., De Commer, L., Darzi, Y., et al. (2017). Quantitative microbiome profiling links gut community variation to microbial load. *Nature* 551, 507–511.

Vogt, N.M., Kerby, R.L., Dill-McFarland, K.A., Harding, S.J., Merluzzi, A.P., Johnson, S.C., Carlsson, C.M., Asthana, S., Zetterberg, H., Blennow, K., et al. (2017). Gut microbiome alterations in Alzheimer’s disease. *Sci. Rep.* 7, 13537.

Warwick-Dugdale, J., Solonenko, N., Moore, K., Chittick, L., Gregory, A.C., Allen, M.J., Sullivan, M.B., and Temperton, B. (2019). Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands. *PeerJ* 7, e6800.

Weiss, M., Denou, E., Bruttin, A., Serra-Moreno, R., Dillmann, M.-L., and Brüssow, H. (2009). In vivo replication of T4 and T7 bacteriophages in germ-free mice colonized with *Escherichia coli*. *Virology* 393, 16–23.

Whitehead, N.A., Barnard, A.M.L., Slater, H., Simpson, N.J.L., and Salmond, G.P.C. (2001). Quorum-sensing in Gram-negative bacteria. *FEMS Microbiol. Rev.* 25, 365–404.

Wiedenheft, B. (2013). In defense of phage: viral suppressors of CRISPR-mediated adaptive immunity in bacteria. *RNA Biol.* 10, 886–890.

- Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482, 331–338.
- Wommack, K.E., and Colwell, R.R. (2000). Virioplankton: Viruses in Aquatic Ecosystems. *Microbiol. Mol. Biol. Rev.* 64, 69–114.
- Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al. (2012). Human gut microbiome viewed across age and geography. *Nature* 486, 222–227.
- Yutin, N., Makarova, K.S., Gussow, A.B., Krupovic, M., Segall, A., Edwards, R.A., and Koonin, E.V. (2018). Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. *Nat. Microbiol.* 3, 38–46.
- Zaleski, P., Wojciechowski, M., and Piekarowicz, A. (2005). The role of Dam methylation in phase variation of *Haemophilus influenzae* genes involved in defence against phage infection. *Microbiol. Read. Engl.* 151, 3361–3369.
- Zhang, F., Zhao, S., Ren, C., Zhu, Y., Zhou, H., Lai, Y., Zhou, F., Jia, Y., Zheng, K., and Huang, Z. (2018a). CRISPRminer is a knowledge base for exploring CRISPR-Cas systems in microbe and phage interactions. *Commun. Biol.* 1, 1–5.
- Zhang, L., Hou, X., Sun, L., He, T., Wei, R., Pang, M., and Wang, R. (2018b). *Staphylococcus aureus* Bacteriophage Suppresses LPS-Induced Inflammation in MAC-T Bovine Mammary Epithelial Cells. *Front. Microbiol.* 9.
- Zhao, G., Vatanen, T., Droit, L., Park, A., Kostic, A.D., Poon, T.W., Vlamakis, H., Siljander, H., Härkönen, T., Hämäläinen, A.-M., et al. (2017). Intestinal virome

changes precede autoimmunity in type I diabetes-susceptible children. *Proc. Natl. Acad. Sci. U. S. A.* *114*, E6166–E6175.

Zhao, G., Droit, L., Gilbert, M.H., Schiro, F.R., Didier, P.J., Si, X., Paredes, A., Handley, S.A., Virgin, H.W., Bohm, R.P., et al. (2019). Virome biogeography in the lower gastrointestinal tract of rhesus macaques with chronic diarrhea. *Virology* *527*, 77–88.

Zheng, D.-W., Dong, X., Pan, P., Chen, K.-W., Fan, J.-X., Cheng, S.-X., and Zhang, X.-Z. (2019). Phage-guided modulation of the gut microbiota of mouse models of colorectal cancer augments their responses to chemotherapy. *Nat. Biomed. Eng.* *3*, 717–728.

Zitomersky, N.L., Coyne, M.J., and Comstock, L.E. (2011). Longitudinal analysis of the prevalence, maintenance, and IgA response to species of the order Bacteroidales in the human gut. *Infect. Immun.* *79*, 2012–2020.

Zolfo, M., Pinto, F., Asnicar, F., Manghi, P., Tett, A., Bushman, F.D., and Segata, N. (2019). Detecting contamination in viromes using ViromeQC. *Nat. Biotechnol.* 1–5.

Zuo, T., Lu, X.-J., Zhang, Y., Cheung, C.P., Lam, S., Zhang, F., Tang, W., Ching, J.Y.L., Zhao, R., Chan, P.K.S., et al. (2019). Gut mucosal virome alterations in ulcerative colitis. *Gut* *68*, 1169–1179.

Chapter II

Biology and taxonomy of crAss-like bacteriophages, the most abundant virus in the human gut

This chapter was published as a research article in Cell Host & Microbe:

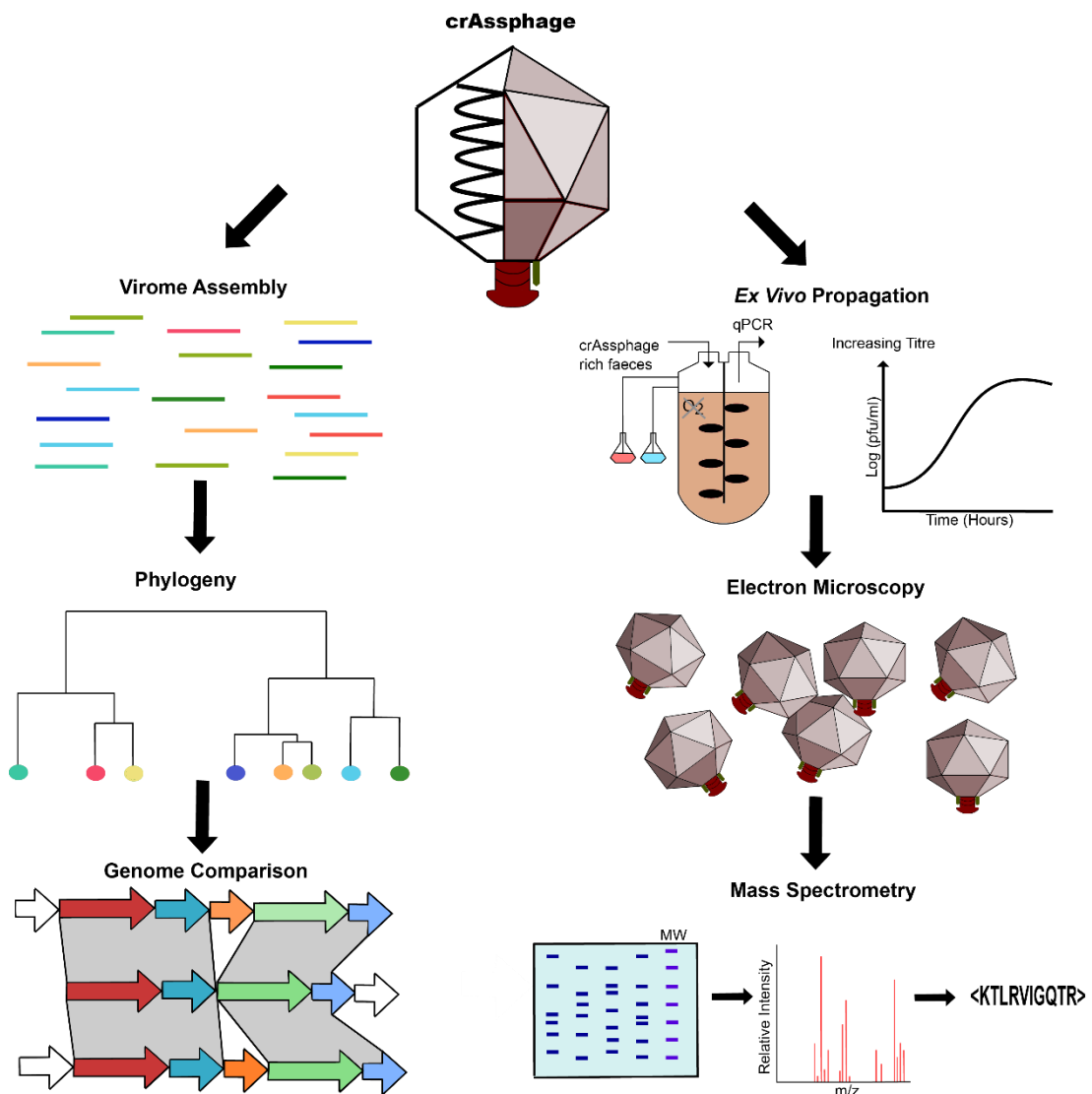
Guerin, E., Shkoporov, A., Stockdale, S.R., Clooney, A.G., Ryan, F.J., Sutton, T.D.S., Draper, L.A., Gonzalez-Tortuero, E., Ross, R.P., and Hill, C. (2018). Biology and Taxonomy of crAss-like Bacteriophages, the Most Abundant Virus in the Human Gut. *Cell Host Microbe* 24, 653-664.e6

<https://doi.org/10.1016/j.chom.2018.10.002>

Author contributions

Emma Guerin, Stephen Stockdale and Andrey Shkoporov equally contributed to this work. Stephen Stockdale performed the bioinformatic work. Emma Guerin performed the laboratory work which included crAss-rich faecal sample processing, faecal fermentations, virome and 16S rRNA gene sequencing preparation, qPCR primer and standard design, sample preparation for transmission electron microscopy and annotation of representatives from each of the defined candidate genera, with input from Andrey Shkoporov. Andrey Shkoporov performed elements of both the laboratory and bioinformatic work. Adam Clooney performed the 16S analysis. Feargal Ryan, Thomas Sutton, Lorraine Draper and Enrique Gonzalez-Tortuero assisted in the design, implementation of experiments. Electron microscopy imaging was performed by Dimitri Scholz and Tiina O'Neill at the UCD Conway Institute of Biomolecular and Biomedical Research. Mass spectrometry was also performed externally at the Metabolomics & Proteomics Technology Facility, University of York. Emma Guerin, Stephen Stockdale and Andrey Shkoporov wrote the paper and generated the figures. Adam Clooney, Feargal Ryan, Thomas Sutton, Lorraine Draper and Enrique Gonzalez-Tortuero reviewed drafts of the manuscript and provided constructive criticism for its improvement. Colin Hill and Paul Ross secured the funding and wrote the paper. All authors contributed to the analysis of the data.

Graphical abstract and highlights



- Screening of human faecal metagenomic samples reveals 249 crAss-like phage genomes
- The crAss-like phages were classified into 4 subfamilies composed of 10 candidate genera
- A crAss-like phage was propagated in ex vivo human faecal fermentations
- Short-tailed phage virions could be visualized by electron microscopy

2.1 Summary

CrAssphages represents the most abundant virus in the human gut microbiota, but the lack of available genome sequences for comparison has kept them enigmatic. Recently, sequence-based classification of distantly related crAss-like phages from multiple environments was reported, leading to a proposed familial level taxonomic group. Here, we assembled the metagenomic sequencing reads from 702 human faecal virome/phageome samples and analysed 99 complete circular crAss-like phage genomes and 150 contigs $\geq 70\text{kb}$. *In silico* comparative genomics and taxonomic analysis enabled a classification scheme of crAss-like phages from human faecal microbiomes into four candidate subfamilies composed of ten candidate genera. Laboratory analysis was performed on faecal samples from an individual harbouring seven distinct crAss-like phages. We achieved crAss-like phage propagation in *ex vivo* human faecal fermentations and visualised short-tailed podoviruses by electron microscopy. Mass spectrometry of a crAss-like phage capsid protein could be linked to metagenomic sequencing data, confirming crAss-like phage structural annotations.

2.2 Introduction

In recent years, increasing numbers of bacteria, archaea, fungi, protists and viruses residing on and within the human body have been associated with various states of human health and disease, including diet, age, weight, inflammatory bowel disease (IBD), diabetes, and cognition (Claesson et al., 2012; Cryan and Dinan, 2014; Everard and Cani, 2013; Frank et al., 2011; Norman et al., 2015; Reyes et al., 2010; Tremaroli and Bäckhed, 2012). A relatively small number of eukaryotic viruses present in the gastrointestinal tract can target human cells; however, much larger and more complex populations of viruses that target bacteria (bacteriophages or phages) are also present. The role of phages in the gut has been a subject of increased interest as initial investigations revealed substantial differences in phage populations between healthy and diseased cohorts (Manrique et al., 2016, 2017; Mills et al., 2013; Norman et al., 2015; Reyes et al., 2015). It is likely that phages have an important role in shaping our gut microbiome, but their precise role remains poorly understood.

In 2014, metagenomic studies of the viral fraction of the human gut microbiota identified a DNA phage, crAssphage, detectable in approximately 50% of individuals from specific human populations and reaching up to 90% of the total viral DNA load in faeces of certain individuals (Dutilh et al., 2014). Dutilh and colleagues noted that crAssphage had been overlooked in previous metagenomic studies as the vast majority of its genes do not match known sequences present in databases. Based on host co-occurrence and CRISPR-spacer profiling, it was predicted that prototypical crAssphage infects bacteria of the genus *Bacteroides* or other members of the Bacteroidetes phylum, an abundant human gut bacterial group which is important for the digestion of complex non-dietary carbohydrates.

Originally crAssphage was published as an individual genome following cross-assembly of several metagenomic samples (Dutilh et al., 2014). Analysis by Manrique *et al.*, of the healthy human gut phageome identified 4 circular crAssphage genomes and several related incomplete contigs (Manrique et al., 2016). PCR amplification and sequencing of the crAssphage polymerase gene by Liang and colleagues similarly demonstrated diversity amongst crAssphage-positive faecal samples (Liang et al., 2016). Recently, Cinek *et al.* described updated PCR primer sequences for the detection and evaluation of crAssphage diversity, while Stachler *et al.* developed primers targeting conserved genomic regions to evaluate the abundance of crAssphage as an indicator of human faecal pollution (Cinek et al., 2018; Stachler et al., 2017). Finally, an epidemiological survey of crAssphages conducted by Edwards, Dutilh and colleagues has suggested crAssphage is associated with humans and primates globally with significant diversity (Edwards et al., 2019).

A recent study provided the first detailed sequence-based taxonomic categorisation of crAss-like phages, proposing a novel familial level taxonomic group that would include prototypical crAssphage itself ('p-crAssphage'), as well as various related phages, from multiple environments (Yutin et al., 2018). Previous attempts to reconcile sequence-based and classical viral taxonomy have proposed that *Podoviridae* (viruses with a short-tail morphology) sharing >40% orthologous protein-coding genes should be grouped at the taxonomic rank of genus, while phages sharing only 20-40% orthologous protein-coding genes should be grouped at the higher taxonomic rank of subfamily (Lavigne et al., 2008). Other reports describe a phage genus as a cohesive group of viruses sharing >50% nucleotide sequence similarity (Adriaenssens and Brister, 2017). As crAssphages are not a

single entity, but rather a group of crAss-like phages that share similarity with the originally discovered p-crAssphage at various levels, a comparative analysis of crAss-like phage sequences is required to enable detailed taxonomic characterisation.

In this study, we combine several *in silico* and *in vitro* approaches to further explore the diversity of crAss-like phages in the human gut, and better understand their biological properties.

2.3 Experimental procedures

Written in the Cell Host and Microbe STAR★Methods format

2.3.1 Experiment model and subject details

CrAss-like phage rich faeces

Ethics for the collection of faecal samples from consenting donor subject ID 924, according to study protocol APC055, were approved by the Cork Research Ethics Committee. The 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. In general, all samples were processed into frozen standard inoculum immediately. The donor, denoted as subject ID 924, is a healthy female. Later metadata showed the subject suffered from gastritis and is vitamin B12 deficient. Recruitment of this individual was based on the consistent presence of crAss-like phages in faecal samples over a 12-month period.

2.3.2 Methods Details

Metagenomic datasets and contig assemblies

Sequencing reads from publicly available metagenomic datasets were downloaded from NCBI Sequence Read Archive (SRA) database. All published and unpublished metagenomic datasets that yielded crAss-like phage contigs, the DNA preparation protocol, the sequencing technology, the assembly program, and information related to contig nomenclature, are briefly described in (Table S1). All reads were processed using Trimmomatic v0.32 (Bolger et al., 2014) to remove adaptor sequences and to trim reads when the Phred quality score dropped below 30 for a 4bp sliding window. Trimmed reads were assembled using either SPAdes v3.6.2

(Bankevich et al., 2012) or metaSPAdes v3.10.0 (Nurk et al., 2017). Contigs from the assembly of 702 metagenomic samples were assigned a specific nomenclature, representing: [1] study/sample description, [2] SPAdes or metaSPAdes assembly, and [3] numerical rank of largest-to-smallest assembled contigs. The full list of contigs assembled in this study and available associated metadata are detailed in (Table S2).

Detection and curation of crAss-like phages

The detection of crAss-like phage contigs was performed as follows. The amino acid polymerase sequence of prototypical crAssphage (p-crAssphage; UGP_018, NC_024711.1) was queried using BLAST v2.2.28+ (Altschul et al., 1997) against a translated nucleotide database consisting of assembled metagenome contig sequences. The most conserved orthologous protein group detected in our initial putative crAss-like phage screening included p-crAssphage protein UGP_092, which was annotated through the HHPred homology and structural prediction web server (Söding et al., 2005; Zimmermann et al., 2017) as a phage terminase. This was then used as a second genetic signature of crAss-like phages and used in an additional BLAST search. All putative crAss-like phages selected for analysis met the following criteria: [1] a BLASTp hit against either p-crAssphage polymerase or terminase with an E-value less than 1E-05, [2] a BLAST query alignment length ≥ 350 bp, and [3] a minimum contig length of 70kb (representing near-complete crAss-like phage contigs).

Identification of crAss-like phage orthologous proteins and clusters

The encoded proteins of crAss-like phages were predicted using Prodigal v2.6.3 (Hyatt et al., 2010). Orthologous proteins shared between crAss-like phages were detected using OrthoMCL v2.0 using default parameters (Li et al., 2003). The

presence/absence of orthologous proteins between crAss-like phages was initially converted into a binary count matrix where the percentage of shared orthologous proteins was calculated. The optimum number of phage clusters was calculated using the percentage of shared homologous proteins using the NbClust v3.0 package for R. Hierarchical clustering was performed on the count matrix of percentage shared crAss-like phage orthologous proteins using Ward's minimum variance method ['Ward.D2' algorithm in R (Ward, 1963)]. The resulting dendrogram was cut at $k = 10$ based on the estimation of the number of crAss-like phage clusters (Figure 1).

As a verification of the 10 predicted crAss-like phage clusters, the original abundance matrix of crAss-like phage orthologous proteins was used to calculate Euclidean distances between sequences. These distance variations were calculated using the t-SNE machine learning algorithm ['tsne' v0.1-3 for R; (Maaten and Hinton, 2008)] and plotted using ggplot v2.2.1 (Figure 2).

Genomic comparisons of crAss-like phages

Complete circular genomes were annotated automatically using VIGA (<https://github.com/EGTortuero/viga>, genetic code table 11 or 15) and then manually using HHPred suit (Zimmerman et al., 2017) against the following databases: PDB_mm_CIF70_28_July, Pfam-A_v31.0, NCBI_CD_v3.16, TIGRFAMs_v15.0 (Table S3 for details). Genome comparison image was generated with Easyfig v2.2.2 (Sullivan et al., 2011), using tBLASTx algorithm with the following parameters: e-value cut-off 0.001, length filter 30 (Figure 3).

Conserved protein sequences were aligned using MUSCLE v3.8.31 and approximately-maximum-likelihood phylogenetic trees were generated using FastTree v2.1.7 with default parameters (Figure S2).

Alignment of virome metagenomic reads to crAss-like contigs

The quality filtered reads from 512 human faecal viromes (as subset of 702 viromes selected based on availability of sufficient metadata) were then aligned to the set of 131 nonredundant crAss-like phage genomes (with <90% identity and/or <90% overlap between them) using Bowtie2 v2.3.0 (Langmead and Salzberg, 2012) using the end-to-end alignment mode. A count table of reads aligned to contigs was generated with Samtools v0.1.19, which was then imported into R v3.3.1 for statistical analysis.

Recruitment of a crAssphage faecal donor and faecal fermentation

Human faecal viromes from a number of ongoing studies sequenced using Illumina HiSeq and MiSeq platforms were screened for crAss-like phages by aligning the obtained sequencing reads against prototypical crAssphage NC_024711.1 using Bowtie2 v2.3.0. One individual (subject ID 924) was found to carry crAssphage consistently at levels exceeding 30% of the total number of reads over a one-year period. A frozen standard inoculum (FSI) sample was processed as described by (O'Donnell et al., 2016) with the following modification: the sample was resuspended in 1X phosphate buffered saline (37 mM NaCl, 2.7 mM KCl, 8 mM Na₂HPO₄, and 2 mM KH₂PO₄), 0.05% (w/v) L-cysteine (Sigma Aldrich, Ireland) and (1 mg/L) resazurin (Sigma Aldrich, Ireland). The crAssphage-rich FSI was inoculated into 400 ml YCFA-GSCM broth in a 500 ml fermenter vessel at 5% (v/v). Fermentation media was prepared exactly as described by (Duncan et al., 2002) with the addition of glucose (2 g/L), soluble starch (2 g/L), cellobiose (2 g/L) and maltose (2 g/L). Fermentation was performed in batch format at approximately 37°C for 51 hours. Dissolved oxygen was sustained at <0.1% by constantly sparging the vessel with anaerobic gas mix (80%

(v/v) N₂, 10% (v/v) CO₂, 10% (v/v) H₂) and stirring at 200 rpm. Both 2M NaOH and HCl solutions were used to maintain pH at ~7. Samples were collected at the following time points; 0, 4, 21, 28, 45 and 51 hours. Collected samples were centrifuged at 4,700 rpm at +4°C for 10 minutes. The resulting supernatants were filtered once through a 0.45 µm pore syringe filter and stored at +4°C. Resultant pellets were stored at -80°C.

Extraction of viral nucleic acids and sequencing library preparation

Total virome extractions were performed on 0.45 µm pore filtered fermentation supernatants. Solid NaCl and polyethylene glycol 8000 were added to the filtrates to give a final concentration of 0.5M and 10% (w/v), respectively. After overnight incubation at +4°C samples were centrifuged at 4,700 rpm and +4°C for 20 minutes. The pellets were then resuspended in 400µl of SM buffer (1M Tris-HCl pH 7.5, 5M NaCl, 1M MgSO₄) and briefly vortexed with an equal volume of chloroform. This mixture was then centrifuged at 2,500g for 5 minutes using a standard desktop centrifuge. The resultant aqueous phase was then transferred into an Eppendorf to which 40µl DNase buffer (10mM CaCl₂ and 50mM MgCl₂) and 8U and 4U TURBO DNase (Ambion/ThermoFisher Scientific) and RNase I (ThermoFisher Scientific) were added, respectively. This was incubated at 37°C for 1 hour followed by an enzyme inactivation step at 70°C for 10 minutes. This was followed by the addition of 2µl proteinase K and 10% SDS and further incubation at 56°C for 20 minutes. Lastly, 100µl phage lysis buffer (4.5 M guanidinium isothiocyanate, 44 mM sodium citrate pH 7.0, 0.88% sarkosyl, 0.72% 2-mercaptoethanol) was added to lyse the viral particles. The final incubation was carried out at 65°C for 10 minutes. The resulting lysates were lightly vortexed with an equal volume of phenol/chloroform/isoamyl alcohol 25:24:1 (Fisher Scientific) and were centrifuged at room temperature for 5

minutes at 8,000g. This was again repeated with the resulting aqueous phase. Following the second extraction, the aqueous phase was passed through a DNeasy Blood and Tissue Kit (Qiagen) for final viral nucleic acid purification. The wash steps were each repeated twice, and the final elution was carried out in 50µl elution buffer. Viral DNA quantification was carried out with the Qubit HS DNA Assay Kit (Invitrogen/ThermoFisher Scientific) in a Qubit 3.0 Fluorometer (Life Technologies). The viral nucleic acids were then subjected to reverse transcription using SuperScript IV Reverse Transcriptase (RT) kit (Invitrogen/ThermoFisher Scientific). The protocol was carried out exactly as described in the manufacturer's protocol for random hexamer primers. Following this, 1µl of the reversed transcribed viral DNA was subjected to GenomiPhi V2 (GE Healthcare) Multiple Displacement Amplification (MDA). Finally, MDA and non-MDA viral DNA was prepared for sequencing using TruSeq DNA Library Preparation Kit (Illumina, Ireland). All steps were performed as per the manufacturer's instructions. Prepared libraries were sequenced on an Illumina HiSeq platform (Illumina, San Diego, California) with 2x300bp paired-end chemistry at GATC Biotech AG, Germany. Reads were filtered, trimmed and assembled into contigs as described above. A count matrix was created by aligning quality-filtered reads back to contigs using Bowtie2 and Samtools.

P-crAssphage PCR detection

Oligonucleotide primer pairs were designed based on the p-crAssphage DNA polymerase sequence UGP_018 (Dutilh et al., 2014) using PerlPrimer software (Marshall, 2004b). Primer sequences are as follows: pCrAss-DNAPol-Fwd5 5'-GCCTATTGTTGCTCAAGCTATTGAA-3' and pCrAss-DNAPol-Rev5 5'-ACAACAGAACCAGCTGCCAT-3'. PCR products were cloned into pCR2.1-TOPO

TA vector (ThermoFisher Scientific) and obtained plasmids at known concentrations were used to establish calibration curves through serial ten-fold dilutions. This plasmid was denoted as pCR2.1::pCrAssDNApol5. Subsequently, qPCR were run in 15 μ l reaction volumes using SensiFAST SYBR No-ROX Master Mix (Bioline) and LightCycler 480 thermocycler with the following conditions: initial denaturation at 95°C for 5 minutes, then 45 cycles of 94°C for 20 seconds, 55°C for 20 seconds and 72°C for 20 seconds, with a final extension at 72°C for 7 minutes. All samples were run in triplicate and the standard error was determined following calculation of DNA concentration based on the above standard curve.

Electron microscopy and detection of crAss-like phage proteins

A virus-enriched fraction of the crAssphage positive faecal sample, collected from subject ID 924, was prepared for electron microscopy imaging as follows. A 1:20 suspension (w/v) of faeces was prepared in SM buffer followed by vigorous vortexing until homogenised. The homogenised sample was chilled on ice for 5 minutes prior to centrifugation twice at 4,700 rpm for 10 minutes at +4°C. The resulting supernatant was then filtered twice through a 0.45 μ m pore syringe filters. The filtrate was ultracentrifuged at 120,000g for 3 hours using a F65L-6x13.5 rotor (ThermoScientific). The resulting pellets were resuspended in 5 ml SM buffer. The viral suspensions were ultracentrifuged again 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 Centrifugal Filter Units 10,000 MWCO (Merck). The purified fraction was then analysed by qPCR for the presence of crAssphage as

described above. 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 faecal viral fraction from subject ID 924 was further concentrated using Amicon Ultra-0.5 Centrifugal Filter Unit with 3 kDa MWCO membrane (Merck, Ireland). This concentrated fraction was loaded onto a premade 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. Six brightest bands with approximate molecular weights of 28, 35, 45, 55, 120 and 200 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).

16S rRNA gene library preparations

Total DNA was extracted from the pellets formed following centrifugation of fermentation samples. This was carried out using the QIAamp Fast DNA Stool Mini Kit (Qiagen, Hilden, Germany). All steps were carried out as per the manufacturer's protocol with the addition of a bead-beating step to aid total DNA extraction from the bacterial cells. Approximately 200mg of each pellet was placed in a 2ml screw-cap tube containing a mixture of one 3.5 mm glass bead, a 200µl scoop of 1mm zirconium beads and a 200µl scoop of 0.1mm zirconium beads (ThistleScientific) with 1ml of InhibitEX Buffer. Bead-beating was carried out three times for 30 seconds using the FastPrep-24 benchtop homogeniser (MP Biomedicals). Between each bead-beating the samples were cooled on ice for 30 seconds. The samples were then lysed at 95°C

for 5 minutes. All other steps were carried out as per the manufacturer's protocol. Following extraction of total bacterial DNA, the hypervariable regions of V3 and V4 16S ribosomal RNA genes were amplified from 15ng of the DNA using Phusion High-Fidelity PCR Master Mix (ThermoFisher Scientific) and 0.2 μ M of each of the following primers, containing Illumina-compatible overhang adapter sequences: MiSeq341F: 5'-TCGTCGGCAG CGTCAGATGT GTATAAGAGA CAGCCTACGG GNGGCWGCAG-3' and MiSeq805R 5'-GTCTCGTGGG CTCGGAGATG TGTATAAGAG ACAGGACTAC HVGGGTATCT AATCC-3' (56) The PCR program was run as follows: 98°C for 30 seconds, 25 cycles of 98°C for 10 seconds, 55°C for 15 seconds and 72°C for 20 seconds, with a final extension of 72°C for 5 minutes. The amplicons were then purified using Agencourt AMPure XP magnetic beads (Beckman-Coulter) followed by a second PCR to attach dual Illumina Nextera indices using the Nextera XT index kit v2 (Illumina). Purification was performed once again, and the libraries were quantified using a Qubit dsDNA HS Assay Kit. The libraries were then pooled in equimolar concentration and sent for sequencing on an Illumina MiSeq platform (Illumina, San Diego, California) at GATC Biotech AG, Germany. The quality of the raw reads was assessed with FastQC (v11.5) and initial quality filtering was performed using Trimmomatic v0.36. Filtered reads were imported into R (v3.4.3) for analysis with DADA2 v1.6.0. (Callahan et al., 2016) Further quality filtering and trimming (maxN of 0 and a maxEE of 2) was carried out on both the forward and reverse reads with only retention in cases of pairs being of sufficient high quality. Error correction was performed on forward and reverse reads separately and following this, reads were merged. The resulting unique Ribosomal Sequence Variants (RSVs) were subjected to further chimera filtering using USEARCH v8.1 (55) with the Chimera-Slayer gold database v20110519. The retained, high quality, chimera-free,

RSVs were classified with the RDP-classifier in mothur v1.34.4 (Schloss et al., 2009) against the RDP database v11.4 (phylum to genus) and SPINGO (Allard et al., 2015) for species assignment. Plots were generated using the R package ggplot2 v2.2.1.

2.3.3 Quantification and statistical analysis

Alignment of virome metagenomic reads to crAss-like contigs

The count table generated from Samtools v0.1.19 was then imported into R v3.3.1 for statistical analysis. The β -diversity of crAss-like viral populations in human cohorts was visualized using PCoA plot based on Spearman rank distances ($D = 1 - \rho$, where ρ is Spearman rank correlation coefficient of relative abundance of different crAss-like contigs between samples). Statistical analysis was performed using permutational multivariate analysis of variance (PERMANOVA) implemented in Vegan v2.4.3 package for R (Anderson, 2001) and non-parametric Kruskal-Wallis test.

Faecal fermentations

Statistical analysis was performed on the qPCR data acquired for the technical triplicates set-up for each fermentation time point. This was done to ensure minimal variation and error. The analysis was carried using GraphPad Prism v7.0 software using the standard error of the mean (SEM). Based on these bars we can have high confidence in the precision of the mean p-crAssphage copies/ μ l calculated for each time point.

2.3.4 Data and software availability

The data and R scripts required to reproduce the analyses of this study have provided as supplementary material. <https://doi.org/10.1016/j.chom.2018.10.002>

Data deposition:

The 249 crAss-like phage contigs analysed in this study are provided as supplementary material. <https://doi.org/10.1016/j.chom.2018.10.002>

2.4 Results

2.4.1 Detection of crAss-like phage contigs

Following the assembly of 702 human faecal virome/phageome metagenomic samples listed in (Table S1), contigs were screened for relatedness to the p-crAssphage. Initially, the polymerase of p-crAssphage (UGP_018, NC_024711.1) was used for crAss-like phage detection due to its use in several studies as a genetic signature to determine diversity of crAss-like phages (García-Aljaro et al., 2017; Liang et al., 2018, 2016). However, we extended our criteria in order to include partial genomes ($\geq 70\text{kb}$) that may not have included the polymerase gene in the assembly. Therefore, after an initial detection of crAss-like phages using the polymerase sequence, we identified the most frequently detected crAss-like phage protein in our dataset as the terminase protein, encoded by p-crAssphage UGP_092. This terminase was subsequently used as a second genetic signature for identifying crAss-like phages.

Initially, 239 contigs $\geq 70\text{kb}$ were detected with similarity to the p-crAssphage polymerase sequence. An additional 59 contigs $\geq 70\text{kb}$ were subsequently detected with relatedness to the p-crAssphage terminase sequence. Following an initial examination of these contigs, more stringent parameters were implemented. Only those contigs whose polymerase and/or terminase sequence(s) aligned across greater than 350bp were considered for further analysis as crAss-like phages. This reduced the total number of crAss-like phages to 256. In addition, as several assembled metagenomic samples were from the same person sequenced at multiple time points, redundant contigs were removed. When a contig aligned with 100% identity across the entire length or within a larger contig, the longer contig or that with the highest coverage was retained (see Methods section for a detailed description). This resulted

in a total of 244 crAss-like contigs, with 144 containing both a polymerase and terminase, 60 a polymerase only and 40 a terminase only (Figure S1).

Contigs $\geq 70\text{kb}$ from the family level taxonomic analysis of crAss-like phages conducted by Yutin *et al.* (Yutin et al., 2018) were analysed for inclusion into this study. Whole genome comparisons highlighted some sequences predicted as related at the familial taxonomic rank shared no significant homology at the nucleotide level (E-value $1\text{E-}05$), including one of this study's predicted 244 crAss-like phages. In addition, contig SRR073438_s_3, while sharing weak nucleotide homology to crAss-like phages, was more similar to larger putative *Caudovirales* phages ($\sim 170\text{kb}$) and was not predicted as a crAss-like phage. Therefore, a total of 249 contigs/genomes were analysed in this study as crAss-like phages. Metadata was available for the majority of the crAss-like phage originating faecal samples (see Table S2). CrAss-like phages were detected in healthy individuals across a wide age range (including infants 1 year of age and individuals ≥ 65 years of age) and individuals suffering from Crohn's disease, ulcerative colitis, HIV, cystic fibrosis, kwashiorkor and marasmus.

2.4.2 Taxonomy of crAss-like phages

Previously, studies have used the percentage of shared homologous proteins as a means of defining phage taxonomic ranks (Lavigne et al., 2008). Therefore, clusters of phages sharing between 20-40% of their protein-coding genes were categorised as related at the subfamily level, while phages sharing $>40\%$ protein-coding genes were grouped at the genus level. A heatmap based on the percentages of shared orthologous proteins suggests that crAss-like phages form 4 candidate subfamilies. The four subfamilies were assigned the nomenclature *Alphacrassvirinae* (which contains p-crAssphage), *Betacrassvirinae* (which contains IAS virus), *Gammacrassvirinae* and

Deltacrassvirinae (Figure 1). These subfamilies can be further subdivided into 10 candidate genera, with I containing p-crAssphage and Candidate Genus VI containing the IAS virus. Metadata of all crAss-like phages analysed in this study, including their categorisation into the various taxonomic divisions, is available in (Table S2).

An alternative approach for characterising the encoded proteome of crAss-like phages was performed by visualisation of genome clusters using the t-SNE machine learning algorithm with Euclidean distances of orthologous genes distribution between genomes as an input. Applying the previously determined 10 crAss-like phage candidate genera classifications to the t-SNE two-dimensional ordination demonstrated that some ellipses are tightly clustered (e.g. Candidate Genus II), suggesting uniformity. Other ellipses are larger relative to the number of sequences (e.g. Candidate Genus III), signifying heterogeneity (Figure 2A). In addition, no single cluster of crAss-like phages is exclusively associated with healthy or diseased individuals ($R^2 = 0.03$ in Adonis/PERMANOVA, $p = 0.006$).

Specific genera of crAss-like phages share similar G+C% nucleotide composition and may share related bacterial hosts, since phage G+C% content can be related to that of their hosts (Edwards et al., 2016; Lucks et al., 2008). Therefore, several groups of crAss-like phages with similar G+C% compositions, such as candidate genera I, IV, VII, IX and X, are likely to infect closely related bacterial taxa within the human microbiome (Figure 2B). Candidate Genus I is the most homogenous group of crAss-like phages containing p-crAssphage and 30 additional complete circular genomes and 32 linear contigs $\geq 70\text{kb}$ which all share a very similar G+C% nucleotide content ($29.12 \pm 0.14\%$). Candidate genera III and VI display the greatest heterogeneity, with G+C% contents of $29.07 \pm 3.07\%$ and $35.32 \pm 2.15\%$, respectively.

2.4.3 Genome structure of crAss-like phages

A set of representative complete circular genomes of the ten genera of crAss-like phages ranges in length from 91.3 to 104.5 kb with varying degrees of genomic synteny (Figure 3). A prominent feature of crAss-like phages is two clearly separated genome regions with opposite gene orientation, the smaller region encoding proteins involved in replication, the bigger region coding for proteins involved in transcription and virion assembly. CrAss-like phages encode large ORFs with sizes up to 18 kb (UGP_052 and UGP_053 in the genome of p-crAssphage), possibly coding for fused subunits of an RNA polymerase (Yutin et al., 2018). However, these large ORFs are only observed in Candidate Genus VII and VIII when a non-standard genetic code is used in ORF prediction (see Methods). All of the crAss-like phages of Candidate Genus I, II, and IV contain no tRNAs, while members of Candidate Genus VI had large sets of tRNA genes (up to 27; Table S2). Analysis of the crAss-like phage proteome suggests that four proteins are universally conserved across crAss-like phage subfamilies and have monophyletic evolutionary origin: major capsid protein (MCP, UGP_086), terminase (UGP_092), portal protein (UGP_091) and primase (UGP_025; Figures S2, Table S3). Phylogeny of crAss-like phages based on multiple alignment of these four proteins supports that based on percentage of shared proteins and clearly separates the four subfamilies. As clearly shown in Figure 3, blocks of genes responsible for tail morphogenesis are especially variable even between members of the same subfamily, both in the number of genes and sequence of the encoded protein products, highlighting potentially wide range of bacterial hosts infected by different crAss-like phage genera and strains.

2.4.4 Prevalence of crAss-like phages in human faecal virome samples

To obtain insights into relative abundance in various human populations we aligned quality filtered reads, representing 512 human faecal samples from the same datasets as used for assembly of crAss-like genomes, to a database of 131 nonredundant crAss-like phage genomic sequences (with <90% of homology and/or <90% overlap between them) representing all ten candidate genera.

CrAss-like phage colonization rates varied from 51-59% in Malawian infants to 98-100% of healthy individuals of various ages in the Western cohorts (Figure S3A). In total, 77% of all samples were positive for one or more crAss candidate genera. The relative phage abundance ranged from 0 to 95% of total reads per sample (Figure S3B) and depended significantly on the country of residence ($p = 4.2\text{E-}09$ in Kruskal-Wallis test) and age group of the donor ($p = 1.1\text{E-}10$). In ~8% of all virome samples, >50% of reads aligned to crAss-like phage genomes. The lowest overall crAss-like phage counts were seen in Irish and Malawian infants and in USA adults with IBD (Figure S3A). On a global scale, crAss-like candidate genera I, VIII, and IX seem to be the most prevalent with the highest mean percentage of reads aligned; 5%, 1.7% and 1.8%, respectively.

The specific composition of crAss-like phages in faeces partly separated a cohort of healthy and malnourished infants living in rural areas of Malawi from the healthy and diseased urban Western cohorts (Figure S3C). PERMANOVA analysis suggested that crAss-like phage composition was mostly driven by place of residence ($R^2 = 0.24$, $p = 0.001$) with condition and age group also having significant impact ($R^2 = 0.05$ and 0.01 respectively, $p = 0.001$). This observation is further supported by a clear difference in the distribution of specific crAss-like candidate genera across different populations (Figure 4). Specifically, Candidate Genus I, which includes p-crAssphage,

is by far the most prevalent type in Western population regardless of age. At the same time, the same genus was extremely scarce in Malawian cohort where Candidate Genus VIII and IX were the most common ($p = 4.7\text{E-}02$ and $4.4\text{E-}013$, respectively).

2.4.5 Faecal fermentations of a Candidate Genus I rich sample

During an ongoing longitudinal study of faecal viromes in healthy adults we identified one individual (subject ID 924), in which p-crAssphage consistently contributed >30% of virome metagenomic reads over a 12-month period. Thus, this donor was selected in order to investigate if p-crAssphage could be propagated in a batch faecal fermentation system. Quantitative PCR (qPCR) detection of a conserved fragment of the p-crAssphage DNA polymerase gene in the viral nucleic acid fractions throughout the fermentation revealed that p-crAssphage was effectively propagated. P-crAssphage was found to increase in titre 89-fold 21 hours into the fermentation (Figure 5A).

Interestingly, a shotgun metagenomic sequencing run, following multiple displacement amplification (MDA) of the viral enriched DNA from the fermentation supernatants, showed the presence of six other crAss-like phages in the study subject, in addition to a phage highly similar to p-crAssphage (Figure 5B; Table S2). Each of these crAss-like phage contigs were $\geq 70\text{kb}$ and grouped into five candidate genera (Figure 5B). Four of these contigs contributed to $\geq 1\%$ of the reads per sample. The most abundant phage of subject ID 924 detected after sequencing this viral DNA preparation was crAss-like phage Fferm_ms_6 (linear, 90.4kb), a member of Candidate Genus I and is closely related to p-crAssphage. Contig Fferm_ms_2 (linear, 88.8 kb) is the second most abundant in the sample and belongs to Candidate Genus

V. Five additional crAss-like phages were detected in the faeces of subject ID 924 (Table S2).

Analysis of the bacterial fraction of the microbiota in the fermentation vessel was performed using compositional 16S rRNA gene sequencing to investigate potential crAss-like phage hosts and to examine the hypothesis that their hosts are of the *Bacteroides* genus or Bacteroidetes phylum. The analysis showed a decrease in the relative abundance of *Bacteroides* up to time point 21 after which levels begin to recover. The converse was observed for p-crAssphage propagation for which titres increased up to time point 21 followed by a gradual decrease. Two *Bacteroides* species found to be abundant in the vessel included *B. dorei* and *B. uniformis*. However, attempts to isolate crAss-like phages on these strains did not yield plaques; therefore, we cannot definitively infer a host for p-crAssphage, but only confirm the presence of bacteria that other studies have suggested as putative hosts (Cinek et al., 2017; Reyes et al., 2013; Figure S4).

2.4.6 Biological characterisation of crAss-like phages

Transmission electron microscopy (TEM) of a faecal filtrate rich in a Candidate Genus I crAss-like phage showed a significant presence of short-tailed or non-tailed viral particles with icosahedral or isometric heads (53% with a *Podoviridae*-like short tail morphology and 29% of *Microviridae* or a smaller type of *Podoviridae*), with lower levels of tailed phages with a *Siphoviridae*-like morphology (15%; Figure 6A). The large *Podoviridae*-like icosahedral capsids with short-tails could be further classified into two types: type I, with head diameters of ~76.5 nm and short tails; and type II, with a similar head size but head-tail collar structures and slightly longer tails (Figure 6B). Sequencing, without MDA, of CsCl purified fraction of the same faecal

material as used for the TEM showed that approximately 40% of reads aligned to crAss-like genomic contigs (Figure 6C), with the Candidate Genus V crAss-like phage being the most abundant. Based on the size of the 7 crAss-like genomic contigs assembled (88.8-104.6 kb; termed 'Fferm' contigs within (Table S2) it is predicted that the predominant podovirus morphology observed corresponds to the crAss-like phages. For comparison, *Microviridae* phages have genomes 4.4-6.1 kb and icosahedral capsids of approx. 15-30 nm in diameter (Roux et al., 2012; Zhong et al., 2015).

The same CsCl fraction of faeces that was subjected to metagenomic sequencing without MDA and TEM visualisation was also analysed by SDS-PAGE followed by identification of major bands using MALDI-TOF mass spectrometry. A major structural protein of crAss-like phage Fferm_ms_2 was detected from a band excised from the ~55kDa area on a SDS-PAGE gel (Figure 6D). The obtained peptide profile corresponded to a protein of 490 amino acids and 55.4 kDa, with analysis showing the protein as having 37% identity with UGP_086, predicted as the major capsid protein of p-crAssphage (Yutin et al., 2018).

Finally, we attempted to independently establish the size of crAss-like phage virions by passing faecal filtrates through a series of filters with gradually decreasing pore sizes (Figure S5). Filtration through 0.1 µm pores (equivalent to 100 nm) resulted in partial retention of crAss-like phages while pores of 0.02 µm size completely removed crAssphage from the filtrate, as judged by qPCR.

2.5 Discussion

The overall objective of this study was to gain an insight into one of the most enigmatic phages discovered to date, crAssphage. This phage is highly abundant in the human microbiome on a global scale; however, it remains poorly understood. One reason why crAssphage has remained such a mystery is due to the lack of available genome sequences for comparison. When crAssphage was assigned a specific nomenclature and uploaded to a public repository by Dutilh and colleagues (Dutilh et al., 2014), it became a template for other studies.

P-crAssphage was the first identified representative of an expanding group of phages, associated with animal, soil and oceanic microbiomes (Dutilh et al., 2014; Yutin et al., 2018). While a previous study proposed a sequence-based classification of crAss-like phages at the familial level (Yutin et al., 2018), our *in silico* analysis focuses on human faecal-associated crAss-like phages. In this study, we present 244 new crAss-like phages from various metagenomic studies. Comparative genomics demonstrates an extensive degree of diversity among these phages, including the identification of four crAss-like phage subfamilies. While the *Alphacrassvirinae* subfamily currently has the greatest number of sequences of the 4 subfamilies, future studies looking for additional homologues of *Betacrassvirinae*, *Gammacrassvirinae* and *Deltacrassvirinae* members will expand and refine these taxonomic ranks. In particular, future faecal virome/phageome studies of humans from diverse geographical locations will likely expand the repertoire of known human-associated crAss-like phages significantly, as the large interpersonal differences in the human virome are likely multiplied by variations of diet and environmental exposure.

Assigning phage taxonomy, in the absence of a universal genetic marker such as 16S rRNA, is a difficult and potentially erroneous process. While crAss-like phages

will likely form their own taxonomic family (Yutin et al., 2018), they possess a dsDNA genome and a *Podoviridae*-like short tailed virion. The categorisation of crAss-like phages by the percentage of shared proteins identified ten candidate genera, with crAss-like phages in each genus originating from the faeces of putatively healthy individuals and people suffering from various metabolic, infectious, diet and gut-related disorders.

Several crAss-like phage genera proposed in this study have distinct nucleotide G+C% compositions. The nucleotide composition of obligate parasites, such as phages, can evolve in close association with their host bacterium (Lucks et al., 2008; Mavrich and Hatfull, 2017; Pride et al., 2006; Roux et al., 2015). Following this logic, Candidate Genera III and VI with diverse G+C% compositions are either heterogeneous groups of crAss-like phages that require further sequences to refine their taxonomic structure, or they are potentially capable of infecting across a broad host range. However, not all phages have a G+C% composition which mirrors their host; therefore, these results must be treated cautiously until further investigated (Henry et al., 2015).

Quantitative analysis of the crAss-like phage content in several cohorts revealed that in agreement with previous studies the vast majority of faecal viral metagenomic samples contained varied amounts of crAssphage DNA. P-crAssphage (Candidate Genus I) is by far most predominant type in Western populations, co-existing with other crAss-like phages in the majority of samples. By contrast, in the cohort of malnourished and healthy Malawian infants (Reyes et al., 2015; Smith et al., 2013), other candidate genera such as VI, VIII and IX are more abundant. It is well known that non-Western rural populations, which mostly consume high fibre, low fat and low animal protein diet are predominantly associated with high *Prevotella*/low *Bacteroides*

type of gut microbiota [known as enterotype II (Arumugam et al., 2011)], as opposed to *Bacteroides/Clostridia*-dominated microbiota (enterotype I) in urban populations consuming a western diet (Filippo et al., 2010; Gorvitovskaia et al., 2016). Indeed, our analysis of the Reyes et al. (2015) 16S rRNA gene sequencing data confirmed high prevalence of *Prevotella* in Malawian samples (Figure S6). Therefore, one can hypothesize that crAss-like phages of candidate genera VIII and IX might be associated with *Prevotella* or other members of the order Bacteroidales.

The *in vitro* analysis of samples obtained from subject ID 924 was particularly intriguing. By mapping metagenomic sequencing reads against p-crAssphage, it was initially thought that this donor only carried the prototypical crAssphage at levels exceeding 30% of total viral reads for a one-year period. A subsequent mining for phages related to p-crAssphage using metagenomic sequencing at later time points, with and without multiple displacement amplification, resulted in six additional crAss-like phages being simultaneously detected in this donor. It is possible many additional crAss-like phage genomes could be present within the metagenomic datasets that were examined in this study, but they were not included in our analysis because of the inclusion criteria chosen or even the choice of assembly program.

In total, subject ID 924 consistently carried seven crAss-like phages, which resolved in our taxonomic analysis into five candidate genera. Three belonged to Candidate Genus VI, supporting the notion this is a heterogeneous group and not simply composed of broad host range infecting phages. It is possible that there are more than seven crAss-like phages within subject ID 924. However, it is most probable that only a single representative of each candidate crAss-like phage genus (with the exception of the heterogeneous Candidate Genus VI) could assemble correctly, with two or more highly identical phages amalgamating their single nucleotide

polymorphisms into a single consensus representative sequence. This is a known feature of the chosen assembly program, where microdiversity is lost at the expense of assembling longer contigs/low coverage contigs within complex samples (Vollmers et al., 2017).

This study demonstrates the proliferation of crAss-like phages in a faecal fermenter, which represents the first evidence of a phage similar to p-crAssphage propagating under laboratory conditions. Furthermore, following our ability to propagate faecal crAss-like phages, we conducted the first transmission electron micrographs (TEMs) of these phages using faecal filtrate from sample ID 924 prepared prior to fermentation. The most abundant faecal viruses had short tails and were *Podoviridae*-like in structure. This agrees with the predictions made by Yutin *et al.*, following their detailed genome annotation of two crAss-like phages (Yutin et al., 2018). Interestingly, however, our TEMs suggest presence of two types of virions with short non-contractile tails (Figure 6B). Presumably, the more abundant type I virions with shorter tails may belong to members of Candidate Genus V, found as the most abundant crAss-like phage group sequenced in the same sample from subject ID 924 (Figure 6C). But without isolating these phages in pure culture, it is not possible to accurately assign which tail structure corresponds to which specific crAss-like phage subfamily or genera.

This work provides multiple levels of *in vitro* evidence confirming that crAss-like phages have a short-tailed podovirus structure. Experimentally, this is shown using the same CsCl fraction purified from a crAssphage rich faecal sample of a healthy human donor. A qPCR performed on this fraction using Candidate Genus I specific primers showed that approximately 1×10^9 copies per microlitre were present. To examine the size of the crAss-like phage virions *in vitro*, faecal filtrates from the

same sample were passed through a number of filters with decreasing pore sizes (ranging from 450-20nm). It was found that p-crAssphage could no longer be detected after filtration through a pore size of 20nm but was partially retained by filter sizes of 100nm. This supports the size prediction of crAss-like phage virions observed in TEMs. Sequencing, without MDA, of the same CsCl fraction that was visualised by TEM confirmed that almost 40% of the reads aligned to crAss-like phages, consistent with the percentage of short-tailed *Podoviridae*-like phage virions present in the sample. An examination of the protein content in the sample visualised by TEM detected the predominant major capsid protein of the Fferm_ms_2 crAss-like phage. The capsid protein was found to have similarity to other crAss-like phages of Candidate Genus V, as well as a moderate degree of similarity to p-crAssphage (Candidate Genus I).

Identifying a means of propagating crAss-like phages is of particular importance in expanding our knowledge on crAss-like phages. However, the primers applied in the qPCR analyses of viral nucleic acids were not suitable for targeting crAss-like phages associated with the various subfamilies and candidate genera other than p-crAssphage. With the availability of more crAss-like phage sequences, broad and narrow spectrum qPCR assays can be subsequently designed and applied to the analysis of these phages, which will be an important part of future work.

It is clear that human gut-associated crAss-like phages are not a single entity, but rather a group of diverse viruses sharing genomic traits, which target diverse bacterial taxa of the human microbiome. Previously, a member of the *Bacteroides* genus was hypothesised as being the host for crAssphage (Dutilh et al., 2014). In a study prior to the discovery of crAssphage (Reyes et al., 2013), a 95.9kb contig corresponding to a putative virus ϕ HSC05 was shown to be stably engrafted after

transplantation of human faecal virus fraction into germ-free mice colonized with an artificial defined community of 15 bacterial species. The retrospective analysis of contigs from that study conducted by ourselves showed that the ϕ HSC05 contig was 91.73% identical by its nucleotide sequence to p-crAssphage. The artificial bacterial community, among others, included: *Bacteroides thetaiotaomicron* (2 strains), *B. caccae*, *B. ovatus*, *B. vulgatus*, *B. cellulosilyticus* and *B. uniformis*. We suggest that one of these strains is more likely to have served as a host for this crAss-like phage propagation than the remaining eight strains of Gram-positive anaerobic bacteria. Since crAssphage had not been described at the time the article was published, this very interesting observation obviously could not have been made at that time. Recently, a crAss-like phage infecting *B. intestinalis* has been isolated in axenic culture (Shkoporov et al., 2018), allowing a preliminary investigation of its host range, replication strategy, virion morphology, and potential ecological impact on the human gastrointestinal microbiota.

With more divergent sequences, we could assume that different members of the *Bacteroides* genus, or even Bacteroidetes phylum for example, may serve as hosts for different crAss-like phages. One host that has been hypothesised for prototypical crAss-like phages is *B. dorei* (Cinek et al., 2017). This was inferred following the analysis of a dataset generated from infants and toddlers with islet autoimmunity. It was shown that crAssphage was only present when *B. dorei* also was detected within the samples. This was not true for other *Bacteroides* members tested, including *B. vulgatus* which is highly related to *B. dorei*. This correlation is compelling; however, it should be noted that there was no confirmation that crAssphage has any role in causing bacteriome alterations that lead to islet autoimmunity. Interestingly, one of the key *Bacteroides* species detected from our faecal fermentation 16S rRNA analysis was

B. dorei. Although our results cannot confirm this as a possible host of a crAss-like phage, this phage-host pair as well as the *Bacteroides* discussed above merit further investigation.

CrAss-like phages have also been defined as a part of the core human gut phageome (Manrique et al., 2016). This emphasises the importance of identifying hosts for diverse crAss-like phages belonging to different candidate genera proposed in this study. The ability to propagate crAss-like phages *in vitro* will prove a key step in gaining an insight into their biological significance including the possible role they play in shaping the bacterial composition of the human gut microbiome. This could be in a positive or negative manner, in the context of various disease states, such as inflammatory bowel disease, cancer, and obesity among others. Thus far, only a few studies have attempted to correlate crAss-like phages with a gastrointestinal disorder (Cinek et al., 2017; Liang et al., 2016; Norman et al., 2015).

In conclusion, our results expand the repertoire of known crAss-like phages significantly, providing a path towards the identification of further crAss-like phages and their hosts. This will lead to a better understanding of their role, if any, in human health and disease. Our work also provides an interesting insight into the diversity of these human gut-associated phages in various populations. In addition, we also demonstrate that these enigmatic phages can be efficiently propagated *in vitro* in a mixed culture as well as providing the first TEMs of crAss-like phages, giving an insight into their morphology. CrAss-like phages appear to be universally present in human populations, including those with various disease states. Due to the specificity of phage-host interactions, the diversity of crAss-like phages suggests they infect multiple diverse bacteria of the human gastrointestinal microbiota. However, more studies will be required to determine the biological significance and role of crAss-like

phages in the human gut and determine if its presence positively or negatively impacts human gastrointestinal health.

2.6 Figures.

1.

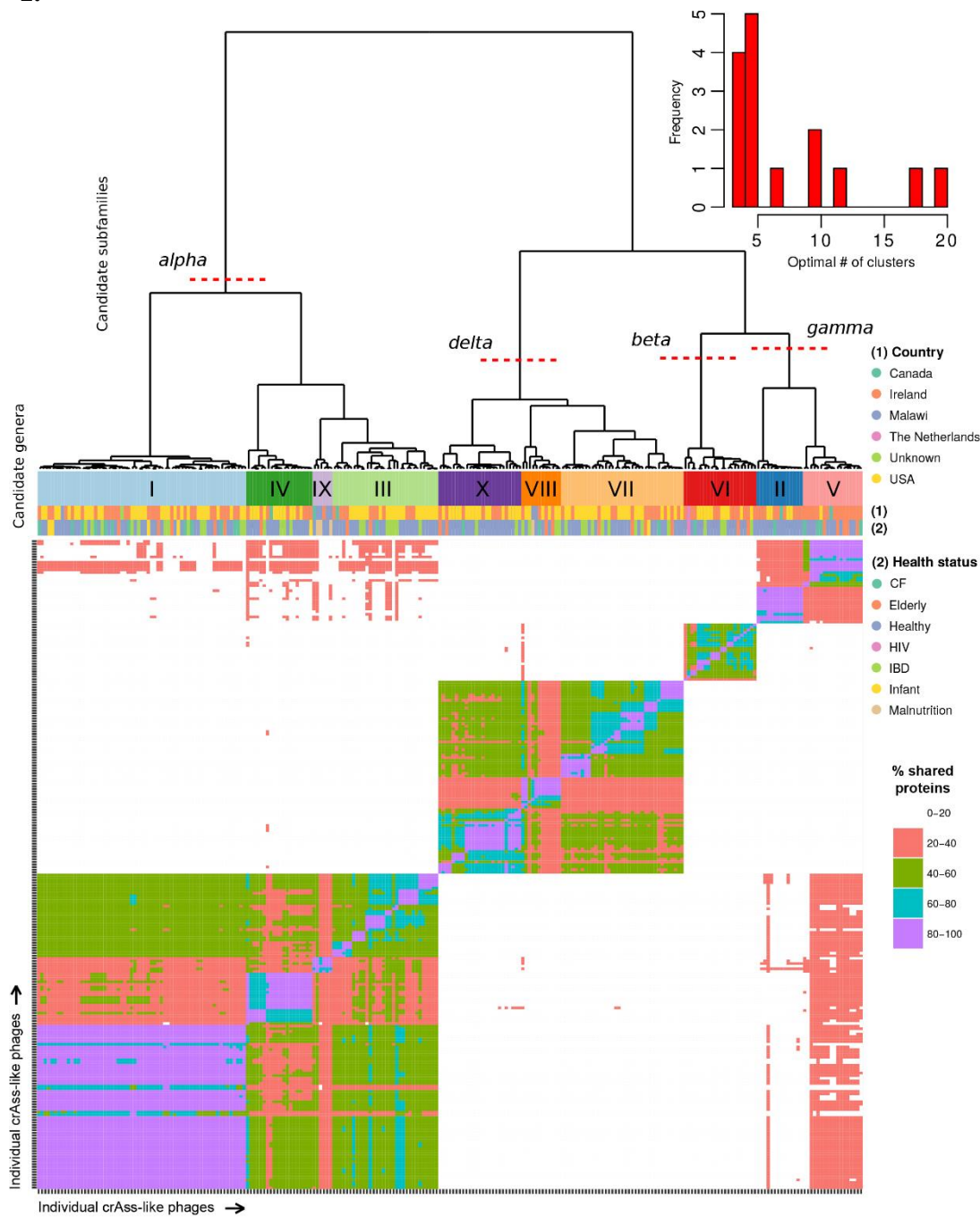
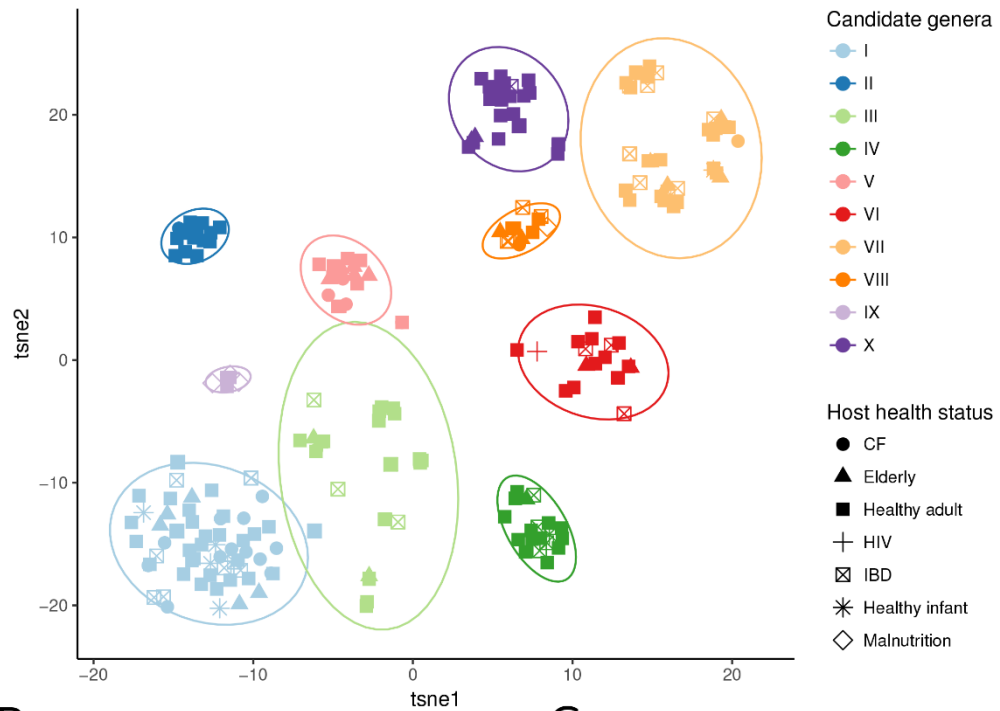


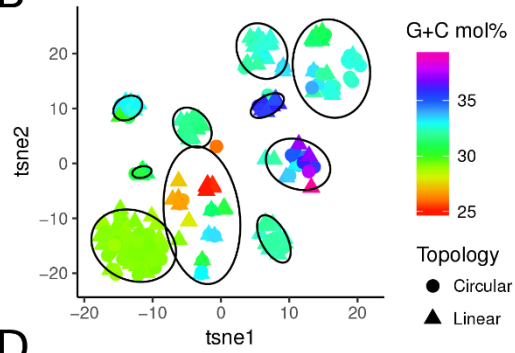
Figure 1. Determination of crAssphage candidate subfamilies and genera based on the percentage of shared protein-encoding genes. (Upper) The 4 red lines cut the hierarchical clustering dendrogram of crAss-like phage contigs into the 4 proposed candidate subfamilies of crAss-like phages. The histogram insert (top-right) represents the calculated optimal number of crAss-like phage clusters. The 10 optimal crAss-like phage clusters represent the putative candidate genera and are assigned specific colours. (Lower) Heatmap showing the percentage of shared protein-coding genes between crAss-like phage genomes. CrAss-like phages with 20-40% shared protein encoding genes are considered related at the subfamily level while phages with >40% similarity are believed to be related at the genus level, consistent with the calculated number of crAss-like phage clusters.

2.

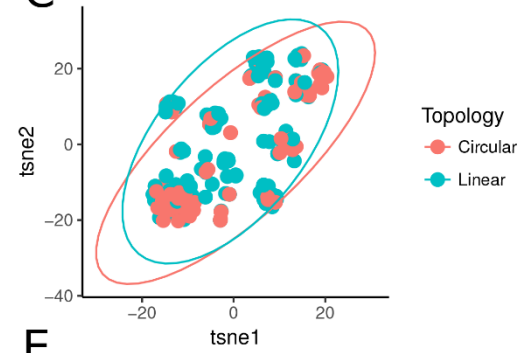
A



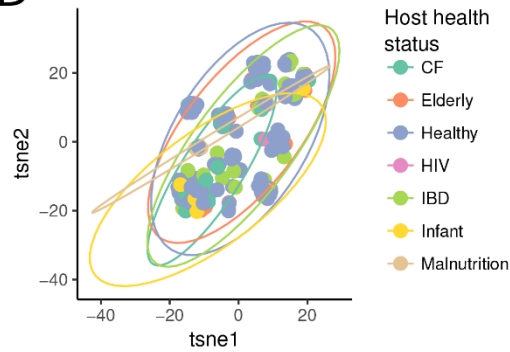
B



C



D



E

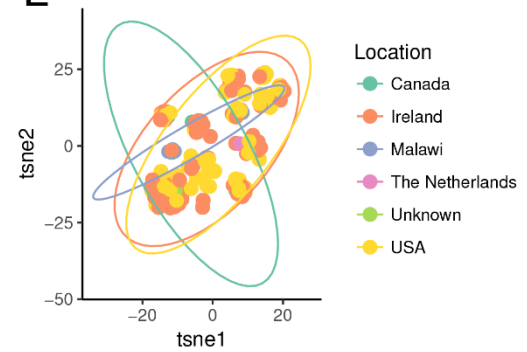


Figure 2. Two-dimensional ordination of crAss-like phages based on the abundance of their protein-encoded orthologous sequences was performed using t-SNE machine learning algorithm. (A) CrAss-like phages are coloured by candidate genus annotations and shape is determined by the health status of individuals carrying these crAss-like phages. (B) CrAss-like phages are coloured by the percentage G+C mol% nucleotide composition of their contig, while shape represents complete (circular) or partial (linear) genomes. (C) Ellipses highlighting the distribution of complete (circular) or near-complete (linear) crAss-like phage genomes. (D) Ellipses grouping the health status associated with the individual donating faecal samples which yielded crAss-like phage contigs. (E) Ellipses highlighting the geographical location of assembled crAss-like phage contigs.

3.

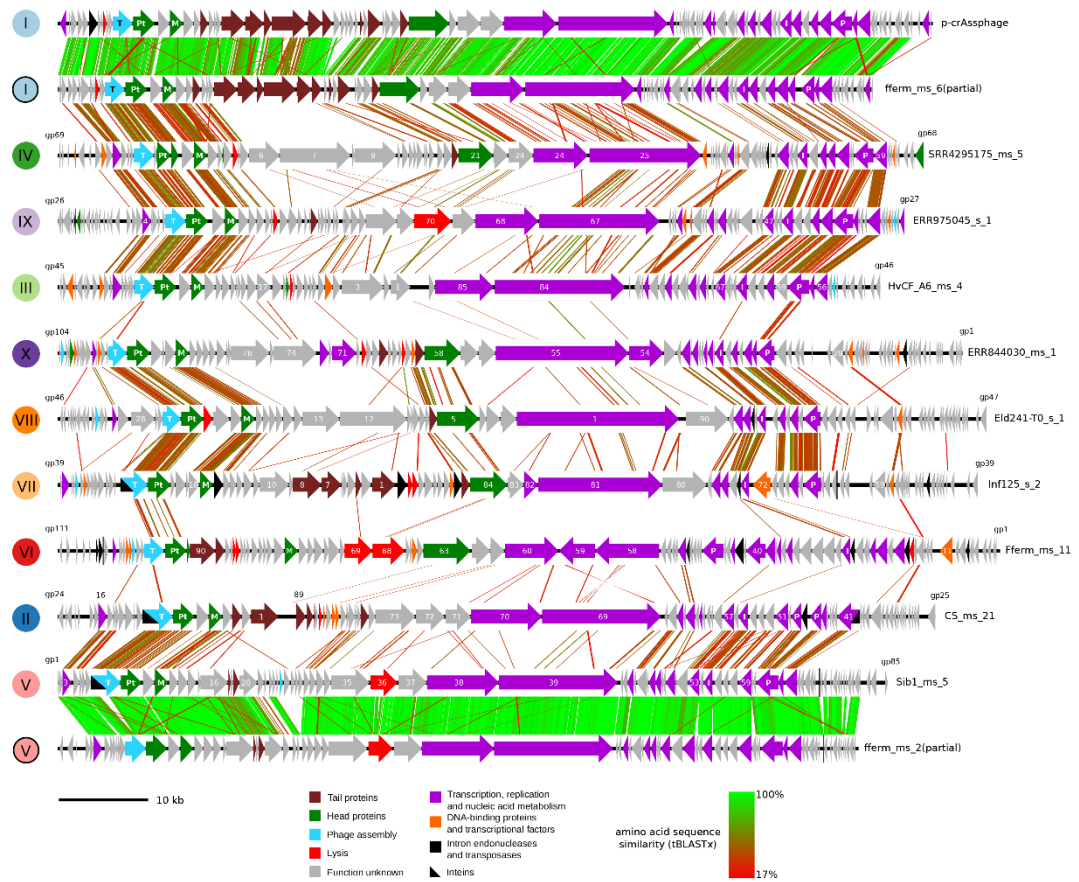


Figure 3. Whole genome comparisons of crAss-like phages from the proposed 10 candidate genera (including 10 complete circular genomes representative of each genus). Partial genomes of Fferm_ms_2 and Fferm_ms_6 of subject ID 924, which are prevalent during the *in vitro* characterisations of crAss-like phages, highlight the inter- and intra-relatedness of the different crAss-like genera. Circular genome maps were permuted in order to standardize for starting coordinate and gene order (gene product [gp] numbers indicate the first and the last gene on a map left to right). Protein coding sequences (CDS, arrows) are coloured by putative HHpred-predicted function. Numbers inside CDS arrows indicate gp numbers (see Table S3 for detailed information). Regions of tBLASTx homology between phage genomes are highlighted.

4.

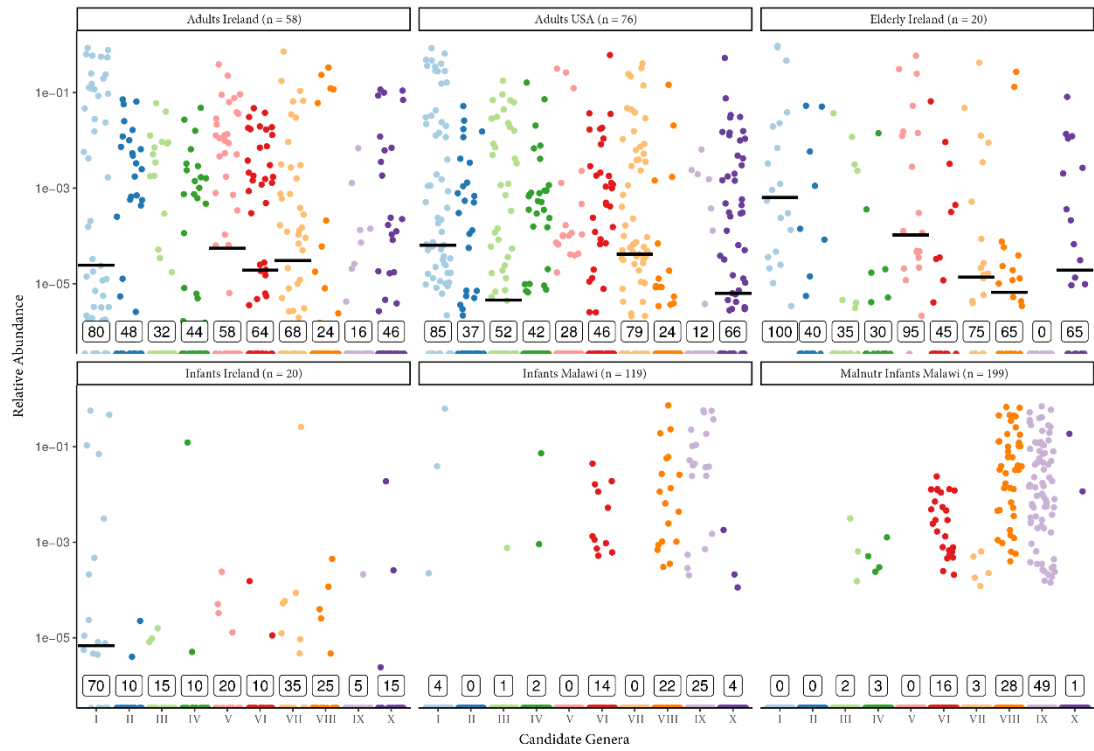


Figure 4. Relative abundance of the ten candidate genera of crAss-like phages in six different human cohorts based on the fraction of metagenomic reads aligned. Bars represent median relative abundances, while values within boxes represent percentage of positive samples.

5.

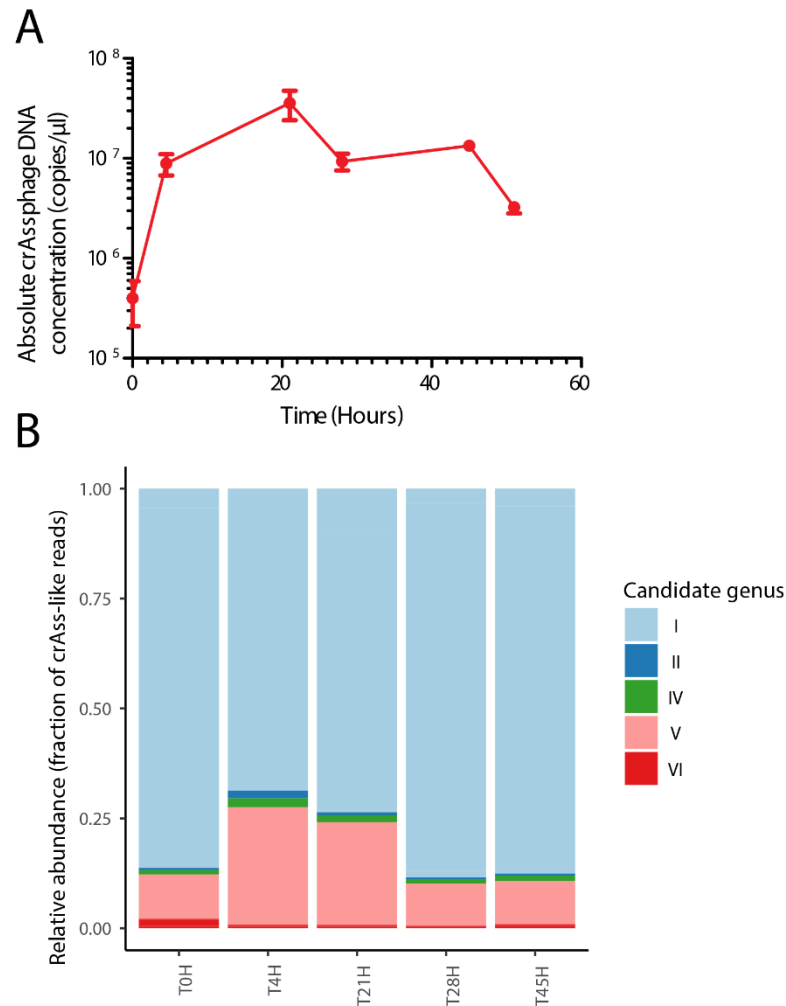


Figure 5. Analysis of crAss-like phage dynamics in a faecal fermenter. (A) Evidence of a Candidate Genus I crAss-like phage propagation following *in vitro* fermentations (standard error, n=3). The level of propagation was determined by qPCR analysis of viral-enriched DNA, respectively, using primers specific to a segment of the p-crAssphage DNA polymerase gene. (B) Six additional crAss-like phages, that group into five of the candidate genera, were identified following sequencing of the same viral-enriched DNA from the fermenter. The relative abundance of each of these crAss-like phages is skewed due to the biased amplification of other components of the viral-enriched DNA fraction that is associated with multiple displacement amplification.

6.

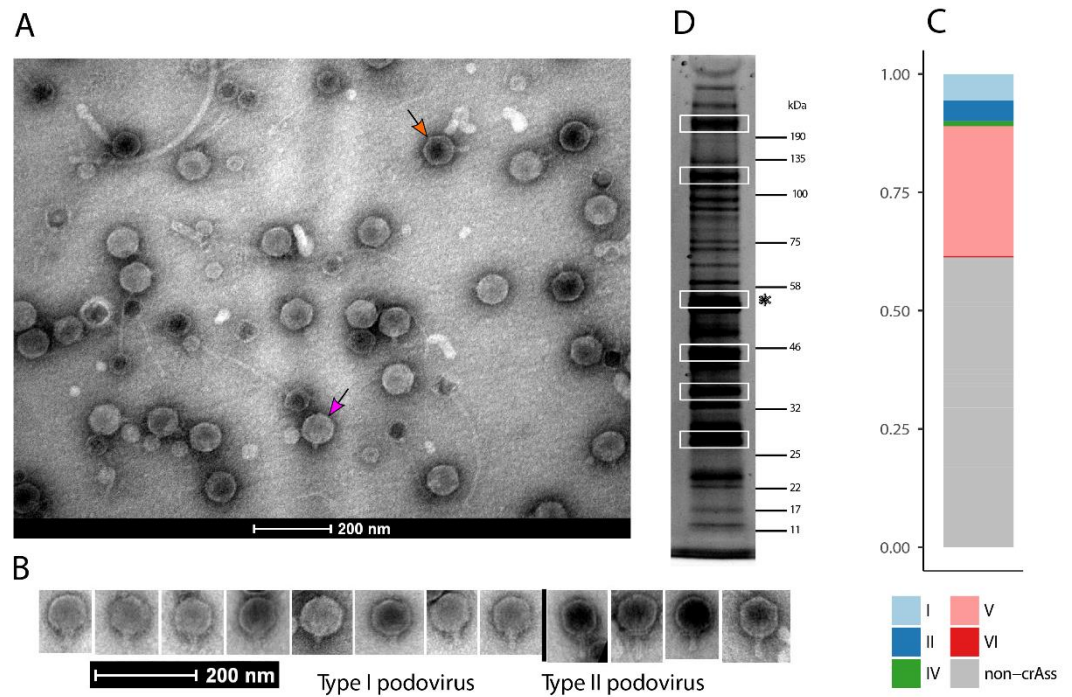


Figure 6. CrAss-like phage morphology was examined using a CsCl fraction purified from a crAss-like phage rich faecal filtrate of donor subject ID 924. (A) Analysis of the fraction through transmission electron microscopy (TEM) was performed, and is largely dominated by podoviruses (53%), microviruses (29%), siphoviruses (15%), and other phage morphologies (3%). (B) Further examination of the observed podovirus virions identifies two variants with differing tail morphologies, highlighted in panel (A) with a pink and orange arrow, respectively. Both variants having head diameters of ~76.5 nm. (C) Sequencing of the CsCl purified viral fraction, without multiple displacement amplification, showed that approximately 40% the reads aligned to crAss-like phages. (D) SDS-PAGE gel of the CsCl fraction, highlighting six bands which were excised and analysed by mass spectrometry. The major capsid protein of crAss-like phage Fferm_ms_2 was detected from the ~55 kDa band (the white box highlighted by *).

Supplementary figures

S1.

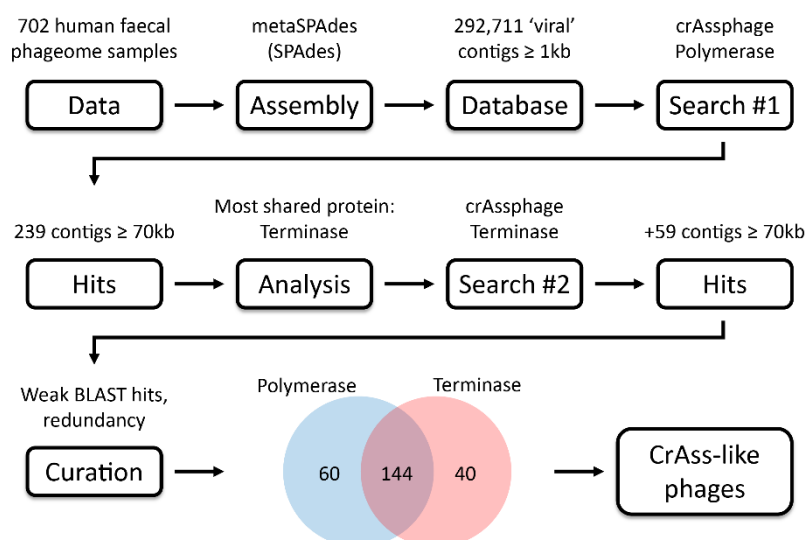


Figure S1. Work-flow overview for the detection of crAss-like phage contigs/genomes, related to Figure 1 and Star Methods sections ‘Metagenomic datasets and contig assemblies’ and ‘Detection and curation of crAss-like phages’.

All contigs were extracted from human faecal associated samples. Two of the 244 crAss-like phages found were removed from subsequent analyses.

S2.

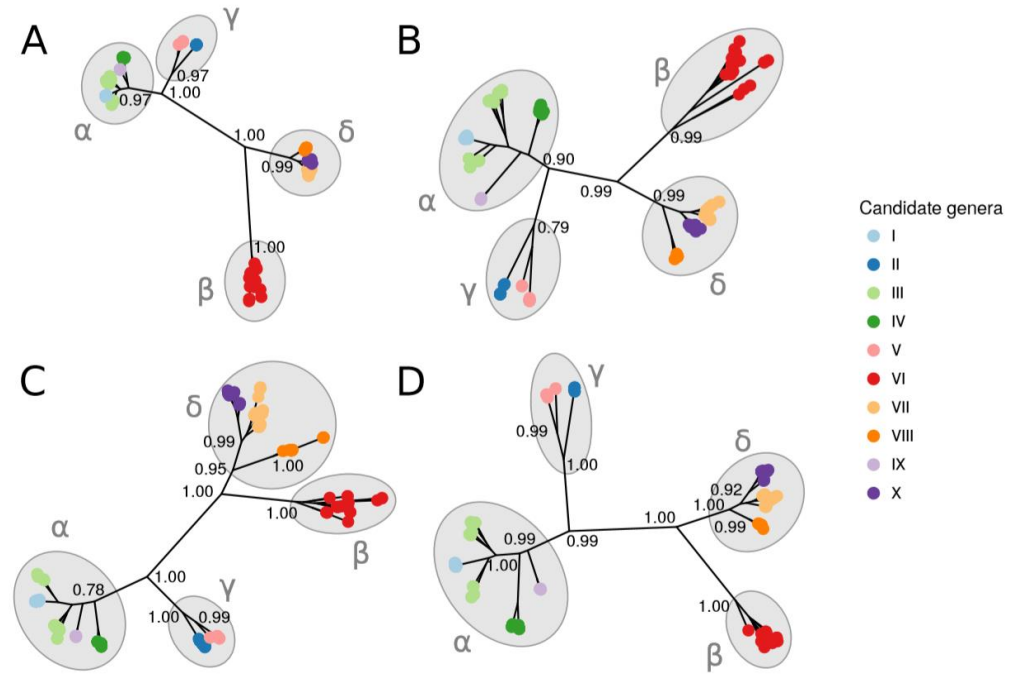


Figure S2. Phylogeny of the (A) MCP, (B) primase, (C) terminase and (D) portal proteins of crAss-like phages, related to Figure 3. Leaves of the phylogenetic NJ-tree represent the different candidate genera. The *Alpha*-, *Beta*-, *Gamma*- and *Deltacrassvirinae* subfamilies (α, β, γ, and δ) are grouped by ellipses (grey). Significant bootstrap values are given next to nodes.

S3.

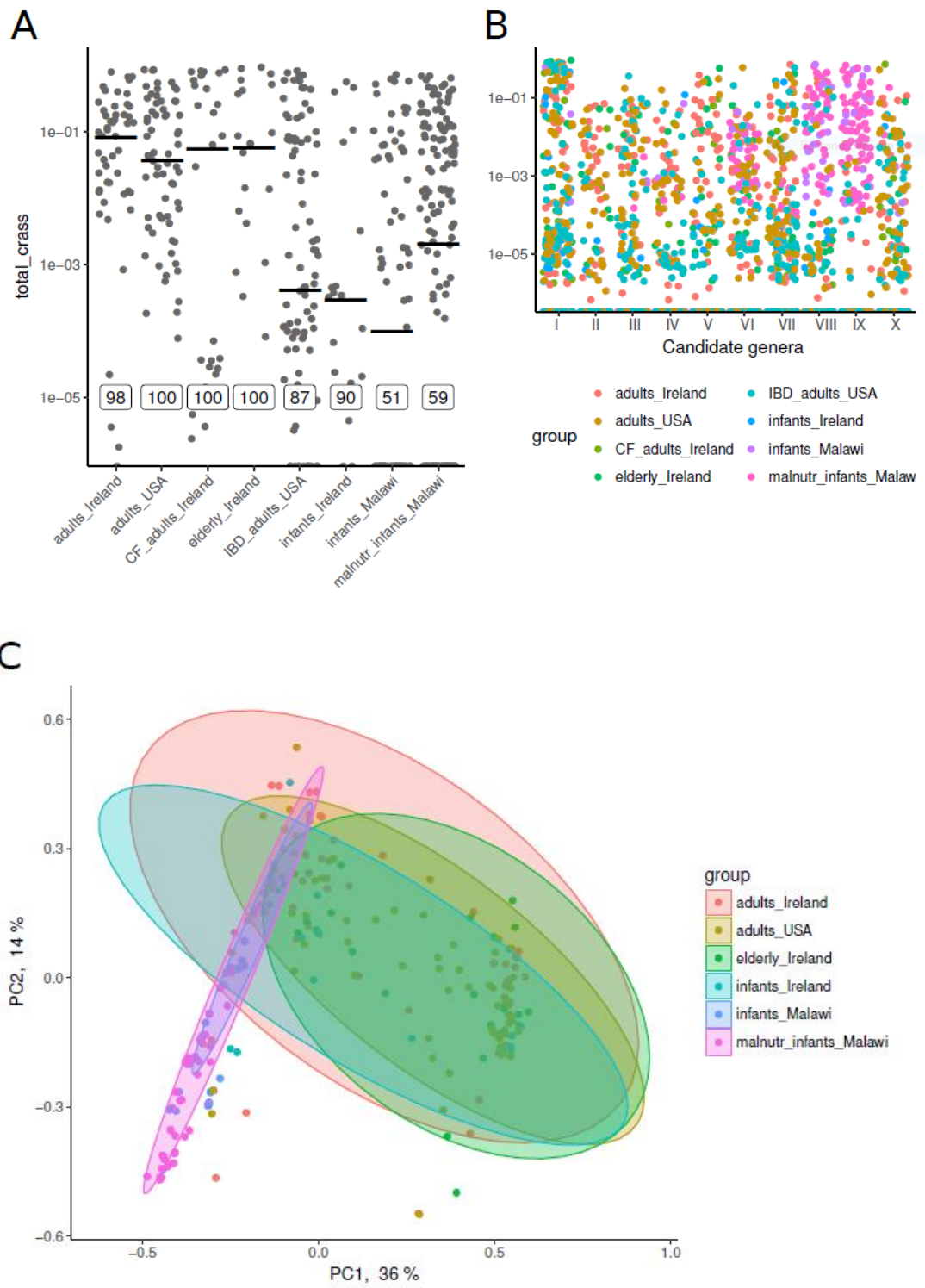


Figure S3. Prevalence of crAss-like phages in human faecal viromes, related to Figure 2 and Figure 4. (A) Relative abundance of total crAss-like phage in several cohorts differing in age, health status and country of origin, based on the fraction of metagenomic reads aligned. Bars represent median relative abundances, while the values within boxes represent percentage of positive samples. (B) Relative abundance of total crAss-like phage metagenomic reads across cohorts varying by age, health status and country of origin, with reads grouped by the proposed crAss-like candidate genera. (C) PCoA plot of crAss-like phages based on Spearman rank distances.

S4.

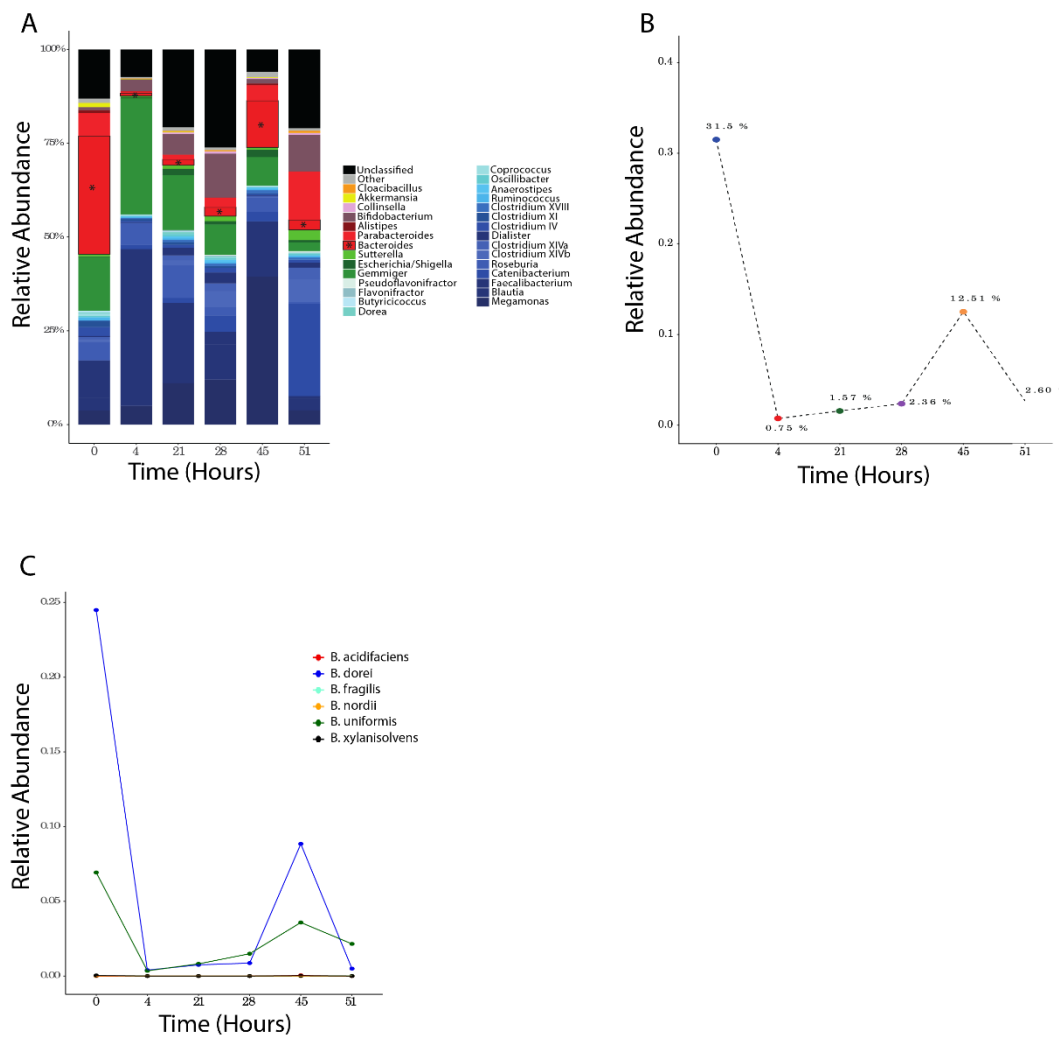


Figure S4. The relative abundance of 16S rRNA throughout the crAssphage-rich frozen standard inoculum initiated faecal fermentation, related to Figure 5. (A) The relative abundance of the major genera detected throughout the fermentation. *Bacteroides* (*), the genus hypothesised to be associated with p-crAssphage, can be seen to decrease between time points 0 and 4 of the fermentation after which levels gradually begin to increase again. (B) The relative abundance of total *Bacteroides* at each time point. (C) Abundances of individual *Bacteroides* species detected.

S5.

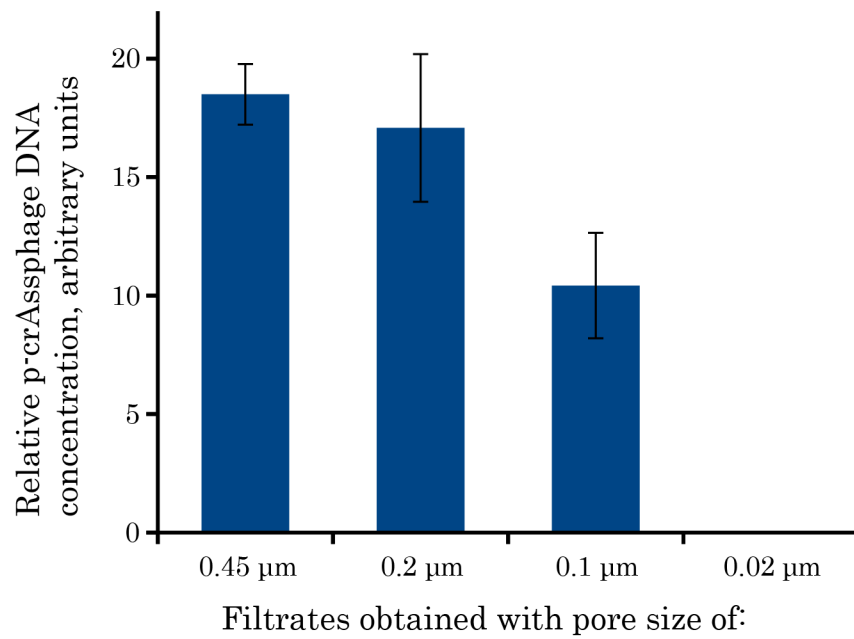


Figure S5. Quantitative PCR analysis of crAssphage rich filtrates, related to Figure 6.

Filtrates were obtained with different pore sizes from a crAssphage-rich faecal sample collected from subject ID 924.

S6.

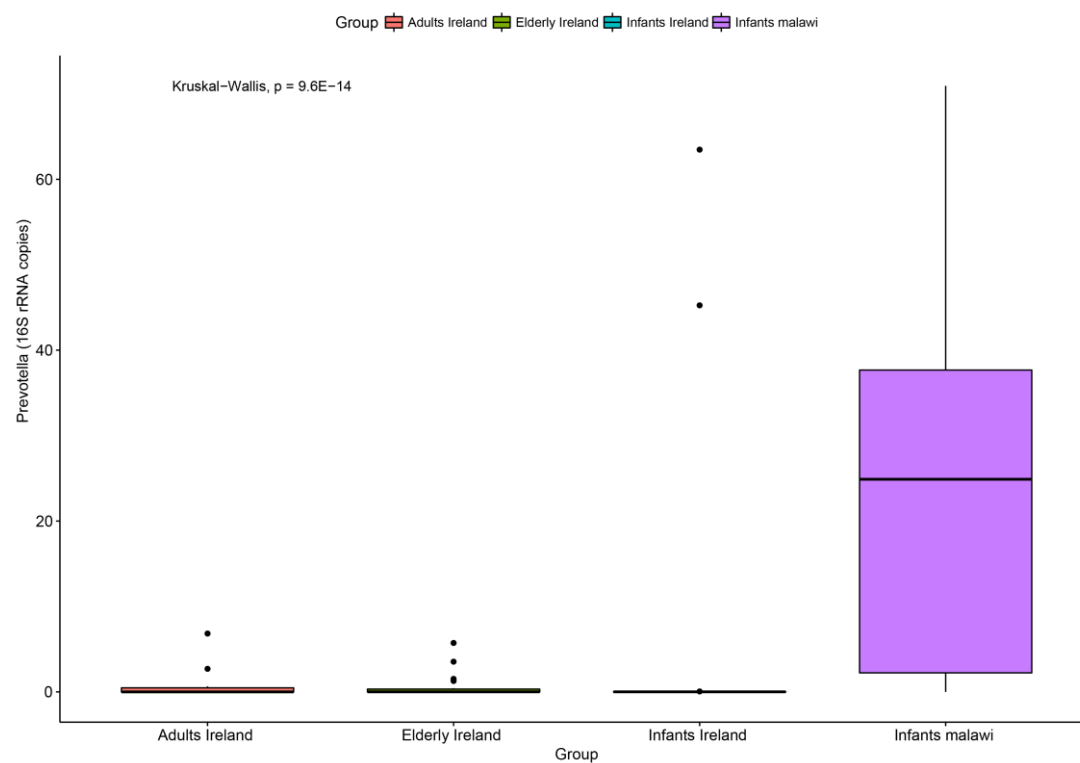


Figure S6. Comparison of 16S rRNA *Prevotella* abundances, related to Figure 4. *Prevotella* abundances among healthy Irish adults and infants were compared with Malawian infants.

2.7 Tables

Tables S1 -S4:

Table S1. Studies Examined for the Presence of crAss-like Phages, Related to STAR Methods Section “Metagenomic Datasets and Contig Assemblies”.

Table S2. Characteristics of crAss-like Contigs Assembled in This Study, Related to STAR Methods Section “Detection and Curation of crAss-like Phages”.

Table S3. Annotation of crAss-like Phage Candidate Genus Representative Sequences, Related to Figure 3.

Table S4. Core Proteome of crAss-like Phage Family, Related to STAR Methods Section “Genomic Comparisons of crAss-like Phage” and Figure S2.

Associated data can be found here <https://doi.org/10.1016/j.chom.2018.10.002> under “Supplemental information”.

2.8 References

Adriaenssens, E., and Brister, J.R. (2017). How to Name and Classify Your Phage: An Informal Guide. *Viruses* 9, 70.

Allard, G., Ryan, F.J., Jeffery, I.B., and Claesson, M.J. (2015). SPINGO: a rapid species-classifier for microbial amplicon sequences. *BMC Bioinformatics* 16, 324.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.

Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26, 32–46.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.-M., et al. (2011). Enterotypes of the human gut microbiome. *Nature* 473, 174–180.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* 19, 455–477.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583.

Casjens, S.R., and Gilcrease, E.B. (2009). Determining DNA Packaging Strategy by Analysis of the Termini of the Chromosomes in Tailed-Bacteriophage Virions. In *Bacteriophages*, (Humana Press), pp. 91–111.

Cinek, O., Kramna, L., Lin, J., Oikarinen, S., Kolarova, K., Ilonen, J., Simell, O., Veijola, R., Autio, R., and Hyöty, H. (2017). Imbalance of bacteriome profiles within the Finnish Diabetes Prediction and Prevention study: Parallel use of 16S profiling and virome sequencing in stool samples from children with islet autoimmunity and matched controls. *Pediatr. Diabetes* 18, 588–598.

Cinek, O., Mazankova, K., Kramna, L., Odeh, R., Alassaf, A., Ibekwe, M.U., Ahmadov, G., Mekki, H., Abdullah, M.A., Elmahi, B.M.E., et al. (2018). Quantitative CrAssphage real-time PCR assay derived from data of multiple geographically distant populations. *J. Med. Virol.* 90, 767–771.

Claesson, M.J., Jeffery, I.B., Conde, S., Power, S.E., O'Connor, E.M., Cusack, S., Harris, H.M.B., Coakley, M., Lakshminarayanan, B., O'Sullivan, O., et al. (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature* 488, 178–184.

Cryan, J.F., and Dinan, T.G. (2014). Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nat. Rev. Neurosci.* *13*, 701–712.

Duncan, S.H., Hold, G.L., Harmsen, H.J.M., Stewart, C.S., and Flint, H.J. (2002). 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.

Dutilh, B.E., Cassman, N., McNair, K., Sanchez, S.E., Silva, G.G.Z., Boling, L., Barr, J.J., Speth, D.R., Seguritan, V., Aziz, R.K., et al. (2014). A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* *5*, ncomms5498.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* *26*, 2460–2461.

Edwards, R.A., McNair, K., Faust, K., Raes, J., and Dutilh, B.E. (2016). Computational approaches to predict bacteriophage–host relationships. *FEMS Microbiol. Rev.* *40*, 258–272.

Edwards, R.A., Vega, A.A., Norman, H.M., Ohaeri, M., Levi, K., Dinsdale, E.A., Cinek, O., Aziz, R.K., McNair, K., Barr, J.J., et al. (2019). Global phylogeography and ancient evolution of the widespread human gut virus crAssphage. *Nat. Microbiol.* *4*, 1727–1736.

Everard, A., and Cani, P.D. (2013). Diabetes, obesity and gut microbiota. *Best Pract. Res. Clin. Gastroenterol.* *27*, 73–83.

Filippo, C.D., Cavalieri, D., Paola, M.D., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., and Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl. Acad. Sci.* *107*, 14691–14696.

Frank, D.N., Robertson, C.E., Hamm, C.M., Kpadeh, Z., Zhang, T., Chen, H., Zhu, W., Sartor, R.B., Boedeker, E.C., Harpaz, N., et al. (2011). Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflamm. Bowel Dis.* *17*, 179–184.

García-Aljaro, C., Ballesté, E., Muniesa, M., and Jofre, J. (2017). Determination of crAssphage in water samples and applicability for tracking human faecal pollution. *Microb. Biotechnol.* *10*, 1775–1780.

Garneau, J.R., Depardieu, F., Fortier, L.-C., Bikard, D., and Monot, M. (2017). PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. *Sci. Rep.* *7*, 8292.

Gorvitovskaia, A., Holmes, S.P., and Huse, S.M. (2016). Interpreting Prevotella and Bacteroides as biomarkers of diet and lifestyle. *Microbiome* *4*, 15.

- Henry, M., Bobay, L.-M., Chevallereau, A., Saussereau, E., Ceyssens, P.-J., and Debarbieux, L. (2015). The Search for Therapeutic Bacteriophages Uncovers One New Subfamily and Two New Genera of Pseudomonas-Infecting Myoviridae. *PLOS ONE* *10*, e0117163.
- Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* *11*, 119.
- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., and Glöckner, F.O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* *41*, e1–e1.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* *9*, 357–359.
- Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* *32*, 11–16.
- Lavigne, R., Seto, D., Mahadevan, P., Ackermann, H.-W., and Kropinski, A.M. (2008). Unifying classical and molecular taxonomic classification: analysis of the *Podoviridae* using BLASTP-based tools. *Res. Microbiol.* *159*, 406–414.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* 13, 2178–2189.

Liang, Y., Jin, X., Huang, Y., and Chen, S. (2018). Development and application of a real-time polymerase chain reaction assay for detection of a novel gut bacteriophage (crAssphage). *J. Med. Virol.* 90, 464–468.

Liang, Y.Y., Zhang, W., Tong, Y.G., and Chen, S.P. (2016). crAssphage is not associated with diarrhoea and has high genetic diversity. *Epidemiol. Amp Infect.* 144, 3549–3553.

Lucks, J.B., Nelson, D.R., Kudla, G.R., and Plotkin, J.B. (2008). Genome Landscapes and Bacteriophage Codon Usage. *PLOS Comput. Biol.* 4, e1000001.

Maaten, L. van der, and Hinton, G. (2008). Visualizing Data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.

Manrique, P., Bolduc, B., Walk, S.T., van der Oost, J., de Vos, W.M., and Young, M.J. (2016). Healthy human gut phageome. *Proc. Natl. Acad. Sci. U. S. A.* 113, 10400–10405.

Manrique, P., Dills, M., and Young, M.J. (2017). The Human Gut Phage Community and Its Implications for Health and Disease. *Viruses* 9, 141.

Marshall, O.J. (2004). PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics* 20, 2471–2472.

Mavrich, T.N., and Hatfull, G.F. (2017). Bacteriophage evolution differs by host, lifestyle and genome. *Nat. Microbiol.* 2, 17112.

Mills, S., Shanahan, F., Stanton, C., Hill, C., Coffey, A., and Ross, R.P. (2013). Movers and shakers: influence of bacteriophages in shaping the mammalian gut microbiota. *Gut Microbes* 4, 4–16.

Milne, I., Bayer, M., Cardle, L., Shaw, P., Stephen, G., Wright, F., and Marshall, D. (2010). Tablet—next generation sequence assembly visualization. *Bioinformatics* 26, 401–402.

Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2011). The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res.* 21, 1616–1625.

Norman, J.M., Handley, S.A., Baldrige, M.T., Droit, L., Liu, C.Y., Keller, B.C., Kambal, A., Monaco, C.L., Zhao, G., Fleshner, P., et al. (2015). Disease-specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. *Cell* 160, 447–460.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27, 824–834.

O'Donnell, M.M., Rea, M.C., O'Sullivan, Ó., Flynn, C., Jones, B., McQuaid, A., Shanahan, F., and Ross, R.P. (2016). Preparation of a standardised faecal slurry for ex-vivo microbiota studies which reduces inter-individual donor bias. *J. Microbiol. Methods* 129, 109–116.

Pride, D.T., Wassenaar, T.M., Ghose, C., and Blaser, M.J. (2006). Evidence of host-virus co-evolution in tetranucleotide usage patterns of bacteriophages and eukaryotic viruses. *BMC Genomics* 7, 8.

Reyes, A., Haynes, M., Hanson, N., Angly, F.E., Heath, A.C., Rohwer, F., and Gordon, J.I. (2010). Viruses in the fecal microbiota of monozygotic twins and their mothers. *Nature* 466, 334–338.

Reyes, A., Wu, M., McNulty, N.P., Rohwer, F.L., and Gordon, J.I. (2013). Gnotobiotic mouse model of phage–bacterial host dynamics in the human gut. *Proc. Natl. Acad. Sci. U. S. A.* 110, 20236–20241.

Reyes, A., Blanton, L.V., Cao, S., Zhao, G., Manary, M., Trehan, I., Smith, M.I., Wang, D., Virgin, H.W., Rohwer, F., et al. (2015). Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc. Natl. Acad. Sci. U. S. A.* 112, 11941–11946.

Roux, S., Krupovic, M., Poulet, A., Debroas, D., and Enault, F. (2012). Evolution and Diversity of the *Microviridae* Viral Family through a Collection of 81 New Complete Genomes Assembled from Virome Reads. *PLOS ONE* 7, e40418.

Roux, S., Hallam, S.J., Woyke, T., and Sullivan, M.B. (2015). Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. *ELife* 4, e08490.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., and Robinson, C.J. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541.

Shkoporov, A.N., Khokhlova, E.V., Fitzgerald, C.B., Stockdale, S.R., Draper, L.A., Ross, R.P., and Hill, C. (2018). Φ CrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. *Nat. Commun.* 9.

Smith, M.I., Yatsunenko, T., Manary, M.J., Trehan, I., Mkakosya, R., Cheng, J., Kau, A.L., Rich, S.S., Concannon, P., Mychaleckyj, J.C., et al. (2013). Gut Microbiomes of Malawian Twin Pairs Discordant for Kwashiorkor. *Science* 339, 548–554.

Söding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33, W244–W248.

Stachler, E., Kelty, C., Sivaganesan, M., Li, X., Bibby, K., and Shanks, O.C. (2017). Quantitative CrAssphage PCR Assays for Human Fecal Pollution Measurement. *Environ. Sci. Technol.* *51*, 9146–9154.

Sullivan, M.J., Petty, N.K., and Beatson, S.A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics* *27*, 1009–1010.

Tremaroli, V., and Bäckhed, F. (2012). Functional interactions between the gut microbiota and host metabolism. *Nature* *489*, 242.

Vollmers, J., Wiegand, S., and Kaster, A.-K. (2017). Comparing and Evaluating Metagenome Assembly Tools from a Microbiologist's Perspective - Not Only Size Matters! *PLOS ONE* *12*, e0169662.

Ward, J.H.J. (1963). Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* *58*, 236–244.

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis* (New York: Springer-Verlag).

Yutin, N., Makarova, K.S., Gussow, A.B., Krupovic, M., Segall, A., Edwards, R.A., and Koonin, E.V. (2018). Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. *Nat. Microbiol.* *3*, 38–46.

Zhong, X., Guidoni, B., Jacas, L., and Jacquet, S. (2015). Structure and diversity of ssDNA *Microviridae* viruses in two peri-alpine lakes (Annecy and Bourget, France). *Res. Microbiol.* *166*, 644–654.

Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A.N., and Alva, V. (2017). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.*

Chapter III

Isolation and characterization of a *Bacteroides xylanisolvens* infecting crAss-like phage from the human gut, Φ crAss002

Emma Guerin, Andrey Shkoporov, Stephen Stockdale, R. Paul Ross, Colin Hill

3.1 Summary

Bacteriophages are a major component of the human gut microbiome. However, we know little about the role and composition of these diverse biological entities. The gut phageome comprises a complex phage community of thousands of individual strains, with a few highly abundant bacteriophages. CrAss-like phages, which infect bacteria of the order Bacteroidales, are the most abundant bacteriophage family in the human gut and make an important contribution to an individual's core virome. Based on metagenomic data, crAss-like phages form a family, with four sub-families and ten candidate genera. To date, only three representatives have been reported to have been isolated in culture; Φ crAss001 and two closely related phages DAC15 and DAC17; all are members of the less abundant genus VI. The persistence at high levels of both crAss-like phage and their Bacteroidales hosts in the human gut has not been explained mechanistically and this phage-host relationship can only be properly studied with isolated phage-host pairs from as many genera as possible.

Faeces from a healthy donor with high levels of crAss-like phage was used to initiate a faecal fermentation in a chemostat, with selected antibiotics chosen to inhibit rapidly growing bacteria and selectively enrich for Gram-negative Bacteroidales. This had the objective of promoting the simultaneous expansion of crAss-like phages on their native hosts. The levels of seven different crAss-like phages expanded during the fermentation, indicating that their hosts were also present in the fermenter. The enriched supernatant was then tested against individual Bacteroidales strains isolated from the same faecal sample. This resulted in the isolation of a previously uncharacterised crAss-like phage of genus IV of the Alphacrassvirinae sub-family, Φ crAss002, that infects the gut commensal *Bacteroides xylanisolvens*. Φ crAss002 does not form plaques or spots on lawns of sensitive cells, nor does it lyse liquid

cultures, even at high titres. In keeping with the co-abundance of phage and host in the human gut, Φ crAss002 and *Bacteroides xylanisolvens* can also co-exist at high levels when co-cultured in laboratory media.

We report the isolation and characterisation of Φ crAss002, the first representative of the Alphacrassvirinae of the highly abundant crAss-like phage family. Φ crAss002 cannot form plaques or spots on bacterial lawns but can co-exist with its' host, *Bacteroides xylanisolvens*, at very high levels in liquid culture without impacting on bacterial numbers.

3.2 Introduction

It is over a century since the discovery of bacteriophages (phages), viruses capable of infecting bacterial cells (Twort, 1915; d'Hérelle, 1917). Phages are almost certainly the most abundant biological entities on Earth with an estimated total count of 10^{31} virions in the biosphere, potentially outnumbering their bacterial hosts by a factor of 10 (Breitbart and Rohwer, 2005; Dion et al., 2020). In recent years, interest in phage communities residing in the human gut (the gut phageome) has greatly increased due to the growing awareness of the role of gut microbiome in human health and a resurgence in interest in phage therapy (Shkoporov and Hill, 2019; Shkoporov et al., 2019). Recent reports also suggest that the human gut phageome may play a role in human health and disease (Reyes et al., 2015; Norman et al., 2015; Monaco et al., 2016; Manrique et al., 2017; Zhao et al., 2017; Carding et al., 2017; Kieser et al., 2018; Ma et al., 2018; Clooney et al., 2019).

The viral (which is overwhelmingly phage) fraction of the human gut microbiome remains its most elusive component (Mirzaei and Maurice, 2017; Reyes et al., 2010). In human faeces, the viral load has been estimated at only 10^9 - 10^{10} virus-like particles (VLPs) g^{-1} , meaning that unlike in other environments where phage dominate, phage in the human gut are outnumbered by their bacterial hosts (Shkoporov and Hill, 2019). Advances in sequencing technology have allowed us to generate vast numbers of viral sequences, but the majority are poorly annotated and unclassified taxonomically due to the lack of homology with known viruses in current databases. These sequences have been termed viral “dark matter” and can constitute up to 90% of total virome reads (Aggarwala et al., 2017). Perhaps the best example are the crAss-like phages, the most abundant phages in the human gut, but no representative had

ever been cultured and it was only identified by a database-independent approach in 2014 (Dutilh et al., 2014).

The prototypical crAssphage (p-crAssphage) is a 97 kb dsDNA phage of the Caudovirales order that was detected through cross-assembly (using crAss software) of human gut viral reads from metagenomic data. Intriguingly, this phage shared no homology with any other virus in databases at the time of its discovery, despite its extraordinary abundance (up to 90% of all phages in some individuals) (Dutilh et al., 2014). Protein sequence-based analysis of the prototypical crAssphage (p-crAssphage) demonstrated that this phage is the founder member of a family-level group of ‘crAss-like phages’ and was predicted to have a podovirus-like morphology (Yutin et al., 2018). We have proposed a taxonomic system for crAss-like phages from the human gut with four subfamilies (*Alpha*-, *Beta*-, *Gamma*- and *Deltacrassvirinae*) and ten candidate genera (I-X), with p-crAssphage belonging to genus I of the Alphacrassvirinae. We then confirmed the predicted podovirus-like morphology by transmission electron microscopy of a crAss-like phage rich faecal filtrate and followed this with the isolation in culture of the first representative member of the family, ΦcrAss001 (genus VI, subfamily *Betacrassvirinae*) (Guerin et al., 2018; Shkoporov et al., 2018a).

Although crAss-like phages are largely gut associated, they have also been detected in other diverse samples such as animal litter, surface/ground water and termite gut (Stachler et al., 2017; Yutin et al., 2018). In humans, the relative abundance of this phage family in the gut can be as high as 90% of the total viral load in some individuals. The geographical spread of this phage family has also been confirmed with p-crAssphage for example, being largely absent from hunter-gatherer gut populations compared to industrialized populations (Cinek et al., 2018; Edwards et al.,

2019; Guerin et al., 2018; Honap et al., 2020). This may be due to differences in dietary habits and bacteriome compositions. Our recent results showed that 77% of healthy Western adults carry one or more representatives of this phage family, although at widely variable abundances (Guerin et al., 2018). Despite the fact that hundreds of crAss-like phages genomes have been identified *in silico*, their bacterial hosts remain to be determined (Dutilh et al., 2014; Oude Munnink et al., 2014). The host phylum was hypothesised to be member of the phylum Bacteroidetes through co-abundance analysis, the presence of BACON-domain-containing proteins which are very characteristic of this phylum and partial matches in CRISPR-spacer sequences (Dutilh et al., 2014). Φ crAss001 that infects *Bacteroides intestinalis*, became the first crAss-like phage to be isolated in pure culture in 2018 (Shkoporov et al., 2018a).

Here we present the isolation and characterisation of Φ crAss002, that infects the gut commensal *Bacteroides xylanisolvens*. This is the first member from genus IV as well as the sub-family *Alphacrassvirinae* to be isolated. Initial attempts to isolate crAss-like phages using traditional methods, such as screening of crAss-rich faecal samples using plaque or spot assays, proved unsuccessful. Φ crAss002 was isolated following *ex vivo* enrichment in a faecal fermentation using antibiotics to selectively promote the growth of Bacteroidales, followed by liquid culturing, metagenomic sequencing, *in silico* analyses and quantitative real-time PCR. Biological characterisation of Φ crAss002 confirmed that this phage shares multiple traits with Φ crAss001, while also possessing several unique characteristics.

3.3 Experimental Procedure

3.3.1 Donor recruitment and sample collection

A healthy female donor in her forties, denoted as subject ID: 924, was recruited for faecal sample donation in October 2017. The individual was previously identified as being a persistent carrier of crAss-like phages over a period of two years (Guerin et al., 2018; Shkoporov et al., 2019). Therefore, this subject was deemed as a donor of interest for the isolation of potential novel crAss-like phages *in vitro*. Sample collection was in accordance with the study protocol APC055 and ethics approved by Cork Research Ethics Committee.

3.3.2 Faecal fermentation

On receipt, the sample was processed into frozen standard inoculum (FSI). This was done as described by (O'Donnell et al., 2016) with modifications as previously outlined (Guerin et al., 2018). Triplicate fermentations were run in batch format over 24 hours with conditions applied as by Guerin et al., (Guerin et al., 2018). Two fermenter vessels were set up in parallel, one with and one without the addition of antibiotics to the YCFA-GSCM broth post-autoclaving. Added antibiotics included 7.5 µg/ml vancomycin and 100 µg/ml kanamycin. The former was chosen based on its ability to suppress a broad range of Gram-positive bacteria (Wilhelm, 1991) and the latter to limit faster growing facultative anaerobes. Samples were collected at 0, 4.5, 17.5, 21 and 24 hours were directly processed after collection through centrifugation at 4,700 rpm for 10 minutes at +4°C. Following this, supernatants were passed through a 0.45µm pore polyethersulfone (PES) membrane filter and the resulting filtrates were stored at +4°C. The remaining bacterial rich pellets were stored at -80°C.

3.3.3 Extraction of viral nucleic acids and virome library preparation

VLPs were enriched and viral nucleic acids extracted from 10 ml of filtered fermentation supernatants using the protocol as previously described (Shkoporov et al., 2018b). One microlitre of the DNA was subjected to multiple displacement amplification (MDA) using Illustra GenomiPhi V2 kit (GE Healthcare) according to manufacturer's instructions. This was performed in triplicate for each sample. Further steps were performed as described previously (Shkoporov et al., 2019). The prepared libraries were then sequenced on an Illumina HiSeq 4000 platform with 2 x 150 nt paired-end chemistry at GATC Biotech AG, Germany.

The quality of the raw paired-end reads was analysed using FastQC v0.11.5. Trimming and filtered of the reads was performed with Trimmomatic v0.36. (Bolger et al., 2014). Parameters implemented were as follows: minimum length of 60, a sliding window size of 4 and a minimum Phread score of 33. The trimmed and filtered reads were then assembled into contigs using metaSPAdes v3.13.1 (Nurk et al., 2017). Contigs that correspond to crAss-like phages were then identified using BLASTn v2.2.28+ against a database of 249 crAss-like phages (Altschul et al., 1997; Guerin et al., 2018). A read count table was generated using Bowtie2 v2.3.0 (Langmead and Salzberg, 2012) and Samtools v0.1.19 to determine the relative abundance reads that resolve into crAss-like phage family genera. Virome diversity metrics for the antibiotic versus non-antibiotic containing vessels were calculated using R package Vegan v2.4.3

3.3.4 Total DNA extraction and 16S rRNA gene sequencing library preparation

Total DNA was extracted from faecal pellets collected from centrifugation of fermentation samples. Extractions were performed using the QIAamp Fast Stool Mini Kit (Qiagen, Hilden, Germany). Approximately 200mg of each pellet was weighed into a 2ml screw-cap tube containing a combination of glass beads varying in size – one 3.5mm glass bead, ~200 µl pf 1m zirconium beads and ~200 µl of 0.1 mm (Thistle Scientific). Proceeding extraction and library preparation steps were performed as described by Shkoporov and colleagues (Shkoporov et al., 2018b)

The quality of the reads was analysed using FastQC v0.11.5. Initial quality filtering was performed with Trimmomatic v0.36. The filtered reads were imported into R v3.4.3 and were analysed for errors using DADA2 package (v1.6.0) (Callahan et al., 2016). Identified errors were corrected via further quality filtering and trimming resulting in unique Ribosomal Sequence Variants (RSVs). The RSVs were subjected to chimera filtering using USEARCH v8.1 with the ChimeraSlayer gold database v20110519. The remaining RSVs were classified using RDP database v11.4 via the RDP-classifier in mothur v1.34.4 (Schloss et al., 2009). Species assignment was also performed using SPINGO (Allard et al., 2015). The resultant RSVs were further analysed. This data was then used to determine the relative abundance of bacterial orders within the antibiotic and non-antibiotic containing vessels, which was visualised using the R package ggplot2 v2.2.1. The 16S diversity metrics for the antibiotic versus non-antibiotic containing vessels were also calculated using R Package Vegan v2.4.3.

3.3.5 Absolute composition and quantitative real-time PCR

The absolute composition of detected crAss-like phages strains throughout the fermentations was determined using quantitative real-time PCR (qPCR) with the standard curve method. Primers were designed to target a consensus region of the terminase gene for each genus detected in the fermenter. Where possible, genus specific primers were designed based on terminase gene alignments as annotated by Guerin and colleagues (Guerin et al., 2018). Due to the heterogeneity of candidate genus VI and the genetic code variations observed for candidate genus VII crAss-like phages, primers were designed with phage specificity. Primer sequences are listed in (Table 1). PCR products generated from these primers were cloned into pCR2.1-TOPO TA vector (Thermo Fisher Scientific) to develop standards. Extracted plasmids were quantified using Qubit dsDNA BR Assay kit and diluted to 10^9 copies/ μ l based on molar mass of DNA. Ten-fold serial dilutions of the plasmids were used to build a standard calibration curve. Absolute quantification qPCR was performed with a 15 μ l reaction volume using SensiFAST SYBR No-ROX mastermix (Bioline) in a LightCycler 480 thermocycler with the following conditions: initial denaturation at 95°C for 5 minutes, then 45 cycles of 95°C for 20 seconds, 60°C for 20 seconds and 72°C for 20 seconds. Resulting Ct-values were converted to copies/ml based on the generated calibration curves. Results were visualised using GraphPad Prism v8.0 software.

3.3.6 Screening for novel crAss-like phages from faecal fermentates

Bacteroidales were enriched for from the FSI preparation. Ten-fold serial dilutions of the FSI were prepared in fresh Fastidious Anaerobe Broth (FAB, Neogen) and 100 μ l of each was spread plated on Fastidious Anaerobe Agar (FAA, Neogen),

YCFA-Agar, and Columbia Blood Agar (Oxoid) with 5% sheep blood supplemented with 25 µg/ml haemin and 100 µg/ml vitamin K. To each of these media 7.5 µg/ml vancomycin and 100 µg/ml kanamycin was added post autoclaving. The dilution plates were anaerobically incubated at 37°C for 48 hours and formed colonies were restreaked. Approximate species identification was performed by Sanger sequencing of the 16S rRNA region using the universal bacterial primers 1492R 5'-GGTTACCTTGTTACGACTT-3" (Reysenbach et al., 2000) and Bact 8 F 5'-AGAGTTTGATCCTGGCTCAG-3' (Edwards et al., 1989). Samples were prepared as per the instruction for LightRun Tube service (GATC) analysed via BLASTn against the NCBI 16S ribosomal RNA sequences (Bacteria and Archaea) database.

Phage-bacterium host pair screening was performed in biological triplicate by co-culturing. Overnight cultures were prepared from the purified Bacteroidales strains. Ten microlitres was sub-cultured into 400 µl of fresh FAB, with cofactors MgSO₄ and CaCl₂ at a final concentration of 1 mM, contained within deep well plates (Sigma-Aldrich). The cultures were incubated anaerobically at 37°C until early logarithmic phase of growth. An OD₆₀₀ = ~0.2 was measured approximately five hours post sub-culturing. To the early log phase cultures, 100 µl of crAss-like phage rich fermentate filtered of bacterial cells was added and incubated anaerobically at 37°C for 24 hours. Without centrifugation or filtering, 10 µl of the phage-bacteria mix was directly sub-cultured (1:50) into fresh FAB. Sub-culturing was repeated over three consecutive days. Total DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen) and analysed for phage propagation on a specific host by qPCR analysis (Table 1).

The detected phage-host pair, ΦcrAss002 and *B. xylanisolvens* APCS1/XY, was enriched using the above co-culturing method in 10 ml volumes. Following five

rounds of enrichment, viral nucleic acids were extracted from phage lysates using the protocol as described by (Shkoporov et al., 2018a). Library preparation for shotgun sequences was performed using the Accel-NGS 1S Plus DNA Library Kit (Swift Biosciences) according to the manufacturer's protocol. After index PCR an additional bead clean-up was performed using a ratio of 1:1 DNA/AMPure beads. Sequencing was performed using a 2 x 150 nt paired end run on an Illumina HiSeq 4000 platform at GATC Biotech AG, Germany.

3.3.7 Transmission electron microscopy

Ultra-centrifugation was performed using a 60 ml volume of Φ crAss002 filtrate. The supernatant was concentrated for a total of 4 hours at 120,000g using a F65L-6x13.5 rotor (Thermo Scientific). The resulting pellets were resuspended in a final volume of 5ml SM buffer. The suspensions were then applied onto to a step gradient of 5M and 3M CsCl solutions, followed by centrifugation at 105,000g for 2.5 hours at +4°C. The CsCl clean-up steps following this were previously described (Guerin et al., 2018).

Five microlitre aliquots of the concentrated 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, Ireland) by Tecnai G2 12 BioTWIN transmission electron microscope.

3.3.8 Shotgun sequencing of *B. xylanisolvens* APCS1/XY using Illumina and Oxford Nanopore platforms

Genomic DNA was extracted from 10 ml of *B. xylanisolvens* APCS1/XY overnight culture using phenol/chloroform extraction with precipitation in 3M sodium acetate and cold absolute ethanol. Cultures were centrifuged at 5,000 g for 10 minutes and pellets were resuspended in 1 ml deionised water. The protocol was then performed as described by Sambrook et al. with the modifications implemented by Bardina et al (Bardina et al., 2016; Sambrook et al., 1989). Following precipitation, the DNA was resuspended in 50 µl Tris-EDTA buffer and incubated at 37°C to aid resuspension.

A long read Oxford Nanopore library preparation was performed as described per the user manual for Rapid Barcoding Sequencing Kit (SQK-RPK004; Oxford Nanopore Technologies) with the following modifications: 800ng of each sample was used and final pellet resuspension was carried out using 20µl of nuclease-free water pre-warmed to 65°C followed by a 10 minute incubation at room temperature. Pooled samples were loaded into SpotON Flow Cell (Oxford Nanopore Technologies) and MinION sequenced for 48 hours (Oxford Nanopore Technologies). Short-read shotgun sequencing of the extracted DNA was performed as described above using Accel-NGS 1S Plus DNA Library Kit and Illumina HiSeq 4000 technology. Hybrid assembly of quality-filtered and trimmed Illumina and raw Nanopore reads was performed using SPAdes (v1.13.1) (Nurk et al., 2013; Antipov et al., 2016) to generate 3 circular scaffolds with a sequencing depth of 43.0x, corresponding to *B. xylanisolvens* APCS1/XY chromosome (6.4 Mbp) and two plasmids (5.6 and 4.1 kbp). The assembled and circularised scaffolds were annotated using NCBI Prokaryotic

Genome Annotation Pipeline (PGAP). The GenBank file for the genome was visualised using GView (v1.7) (Petkau et al., 2010).

Dynamic local recombination hotspots in the *B. xylanisolvens* APCS1/XY chromosome were detected by aligning Oxford Nanopore reads to the assembled scaffolds. Alignment was performed using BLASTn and only reads with length >1000 nt were considered (n = 155,425). Individual local alignments of >90% identity and >200nt length were kept. All internal inversions or shifts in alignment coordinates versus reference genomic scaffold >200 nt were deemed as recombinations. Recombination hotspots were identified when >8 reads with inconsistent alignment were present per 1000bp window of genome length.

3.3.9 *In silico* characterisation of Φ crAss002

Annotation of Φ crAss002 genome was performed using the *de novo* viral genome annotator VIGA (González-Tortuero et al., 2018). The predicted protein coding sequences were further analysed with HHPred using the following databases: PDB_mm_CIF70_28_Dec, Pfam-A_v31.0, NCBI_CD_v3.16, TIGRFAMs_v15.0 (Zimmermann et al., 2017). A genomic map of the Φ crAss002 genome was then generated using GView (v1.7). To examine the DNA termini and packing mechanism employed by Φ crAss002, the command line version of Phage Term v1.0.12 was used (Garneau et al., 2017).

The Φ crAss002 genome was examined against other crAss-like phages of candidate genus IV using BLASTn to examine relatedness (Guerin et al., 2018). Average nucleotide identity (ANI) was determined using the default settings in PYANI (Leighton Pritchard et al., 2017). The output was exported into R environment to

generate a heatmap with Bioconductor Complex Heatmap (v1.20.0) and Circilize (v0.4.5) packages (Gentleman et al., 2004; Gu et al., 2014). A pairwise comparison of the genomes was performed to examine synteny using tBLASTx in Easyfig (v2.2.2) with a minimum alignment length of 50bp and 30 percent identity. Genome comparison were also performed investigate the synteny of Φ crAss002 in comparison to Φ crAss001 (GenBank MH675552.1), and p-crAssphage (GenBank NC_024711.1).

3.3.10 Biological characterisation of Φ crAss002

Plaque assays were performed using 10-fold serial dilutions of Φ crAss002 lysate prepared in SM buffer with an overlay of 0.3% FAA agar (0.3% agar w/v), containing MgSO₄ and CaCl₂ at a final concentration of 1 mM. To 3 ml of molten overlay, 300 μ l *B. xylanisolvens* APCS1/XY overnight culture was added and 50 μ l of phage dilution. This mixture was vortexed and poured onto pre-prepared FAA base agar (1.5% agar w/v). Plates were incubated anaerobically at 37°C. Plaque formation was checked at 24 and 48 hours. Spot assays were performed as described for plaque assays but without addition of phage to the molten overlay agar. A 10 μ l drop of phage was directly applied to the solidified lawn of *B. xylanisolvens* APCS1/XY and dried prior to incubation.

Attempts to generate a one-step growth curve were performed by infecting an early logarithmic phase culture of *B. xylanisolvens* APCS1/XY with Φ crAss002 at a multiplicity of infection (MOI) of 1. Following incubation at room temperature for 5 mins, centrifugation was performed at 5000 rpm in a swing bucket rotor for 15 minutes at 20°C. The supernatant was removed, and the resultant pellet was resuspended with FAB. Anaerobic conditions were maintained at 37°C for 3 hours with 1 ml sample collection every 15 mins. Samples were centrifuged and filtered through 0.45 μ M pore

syringe filters. Analysis was performed using qPCR as described above using CGIV_Fwd and CGIV_Rev primers (Table 1).

The ability of Φ crAss002 to infect another commercially available *B. xylanisolvens* strain, DSM 18836 (DSMZ), and the Φ crAss001 host, *B. intestinalis* 919/174, was examined by co-culture and the standard propagation method of centrifuging and filtering phage lysate between propagations and re-exposure to the host. Efficiency of lysogeny was performed using 100 μ l of Φ crAss002 at $\sim 10^9$ pfu/ml spread plated on FAA agar plates. One hundred microlitres of 10-fold serially diluted *B. xylanisolvens* APCS1/XY overnight culture was added to 3ml of molten 0.3% FAA soft agar with co-factors and was poured onto the phage seeded plates. Negative controls were prepared in the same manner but without the addition of phage to the plate. The plates were incubated anaerobically at 37°C for 48 hours. Efficiency of lysogeny was calculated as the percentage of colonies on the phage seeded plate versus counts for the equivalent negative control. Thirty resistant colonies were restreaked three times. Standard PCR was performed using Φ crAss002 specific primers to test for potential lysogens.

3.3.11 Co-cultivation of Φ crAss002 and *Bacteroides xylanisolvens* APCS1/XY

Co-cultivation of Φ crAss002 and *B. xylanisolvens* APCS1/XY was performed to examine propagation dynamics over time *in vitro* via serial sub-culturing of phage and host. This was initiated using 10 ml of culture ($OD_{600} = \sim 0.2$) prepared from naïve *B. xylanisolvens* cells i.e. ~ 10 generations without exposure to Φ crAss002 and 1ml of phage lysate at $\sim 10^9$ pfu/ml. Subsequent rounds of sub-culturing were performed by introducing the prior co-culture into fresh FAB at a ratio of 1:50. This was repeated

over 11 days and Φ crAss002 titre quantification was performed via qPCR. The phage lysate generated from the final time point was then used to initiate another round of serial co-cultivation over 30 days. Having observed a propagation pattern, a third round of experimentation was initiated over 6 days. On day six, the co-culture was centrifuged for 15 mins at 5000 rpm in a swing bucket rotor at 4°C. The supernatant was filtered through 0.45 μ M pore syringe filters. The resultant pellet was retained and t-streaked for three generations to purify the bacteria of phage. The absence of adsorbed or integrated Φ crAss002 was confirmed via qPCR. Ten *B. xylanisolvens* colonies, that were recently exposed to Φ crAss002 but were free of any remaining phage, were used to initiate a final cycle of co-cultivation. Absolute quantification of Φ crAss002 titre was performed via qPCR (Table 1). Significant differences between the Φ crAss002 titres sustained by the cultures was evaluated using one-way ANOVA followed by Tukey's post hoc comparisons. Spot assays of Φ crAss002 were performed with each of the clones to test for resistance or changes in spot opacity.

3.3.12 Examination of Φ crAss001 and Φ crAss002 dynamics in a fermenter system with a defined bacterial consortium

An unusual phage-host equilibrium has been observed both of the crAss-like phages isolated to date. The was further examined in a fermenter system with a defined community of commensal bacteria. This defined community was compiled of six bacteria in addition to *B. xylanisolvens* APCS1/XY and *B. intestinalis* 919/174: *E. coli* LF82, *E. faecalis* OG1RF, *Ruminococcus gnavus* ATCC 29149, *Faecalibacterium prausnitzii* A2-165, *Lactobacillus plantarum* WCFS1, and *Bifidobacterium longum* subsp. *longum* ATCC 15707. Collectively these human gut derived bacteria are

referred to as a simplified human consortium (SIHUMI) (Eun et al., 2014). Our community excluded *Bacteroides vulgatus* ATCC 8482 to avoid issues with discrimination between the *Bacteroides* phage hosts.

Fermentations were performed in batch format for the first 24 hours to allow establishment of the bacterial community. Overnight cultures were grown to $\sim 10^9$ cfu/ml and combined at a ratio of 1:1. The fermenter vessels, containing 200 ml of YCFA-GSCM broth, were then inoculated with 1ml of this mixture. After 24 hours 10 ml of Φ crAss001 ($\sim 10^9$ pfu/ml) and 10 ml of Φ crAss002 ($\sim 10^8$ pfu/ml) were inoculated into one vessel. The other remained without phage in parallel to act as a control. For 4 days, a continuous fermentation was performed with 400 ml of media exchanged in every 24 hours with equivalent waste removal. Samples were collected at the following time points: 0, 6, 24, 48, and 72 hours post-phage inoculation and were frozen directly at -80°C . Fermentation runs were performed in triplicate using aliquots from the same phage lysate. Total DNA was extracted using the QIAamp Fast Stool Mini Kit (Qiagen, Hilden, Germany). All DNA samples were normalised prior to analyses by diluting to 5 ng/ μl . Bacterial primers were designed targeting unique genes (Table 2) and were used to develop qPCR standards with the method described previously. Primers specific to the portal protein were used in Φ crAss001 qPCR analyses (Shkoporov et al., 2018a). Conditions for qPCR were as follows: initial denaturation at 95°C for 5 minutes, then 45 cycles of 95°C for 20 seconds, 62°C for 20 seconds and 72°C for 20 seconds. For bacterial analyses an annealing temperature of 62°C was used and when analysing both phages, 60°C . Results were visualised using GraphPad Prism v8.0 software.

3.4 Results

3.4.1 Schematic overview of workflow

A faecal fermentation was initiated using faeces from an individual (subject ID 924) identified as a persistent carrier of multiple crAss-like phages. The fermentation was performed under anaerobic conditions with vancomycin and kanamycin added in an attempt to suppress the growth of Gram-positive and facultative anaerobic bacteria and favour the growth of strictly anaerobic Gram-negative Bacteroidales order bacteria (Figure 1 depicts the workflow).

3.4.2 Bacterial composition following antibiotic enrichment in faecal fermenters

Faecal fermentations were conducted in the presence or absence of antibiotics designed to enrich for members of the Order Bacteroidales. 16S rRNA gene analysis of the bacterial composition confirmed a significant increase in the relative abundance of Bacteroidales in the presence of antibiotics (Figure 2A). After 4.5 hours the fermenter was dominated by members of this order and remained so for the remainder of the run. *Bacteroides* and *Parabacteroides* dominated at the genus level with the former representing as much as 75% of 16S rRNA gene reads, and the latter up to 15-20% of reads. The vessels without added antibiotics were largely dominated by Gram-positive bacterial orders. There was a rapid expansion of Erysipelotrichales after 4.5 hours but by 17.5 hours the relative abundance of this order decreased and Clostridiales once again became the dominant order. Diversity indices further highlight the shift in the overall bacterial communities under selective conditions (Figure S1A).

3.4.3 Enrichment of crAss-like phages

Shotgun metagenomic sequencing data, generated from viral enriched DNA extracted from three separate fermentation runs for each condition, was examined for the presence of crAss-like phages. Assembled viral contigs were analysed against a database of 249 crAss-like phages via BLASTn v2.2.28+ (Guerin et al., 2018). This database includes seven crAss-like phages previously detected in donor subject ID: 924 six months prior to donating a sample for this study (Guerin et al., 2018). Six of the original seven crAss-like phages could be detected but one phage (originally denoted as Fferm_ms_1 of candidate genus II) was no longer detectable. However, a previously undetected crAss-like phage of candidate genus VII was now present, meaning that five different crAss-like phage genera were represented in the sample. The relative abundance of all the crAss-like phages represented 38% of the total viral reads generated from VLPs. The relative abundance of each genus was calculated as a percentage of the reads identified as crAss-like (Figure 2B). In the presence and absence of antibiotics the vessels were dominated by a phage of candidate genus I (p-crAssphage). The relative abundance of a phage of candidate genus IV was found to increase under the selective conditions. In line with similar alterations in the bacteriome, we observed a reduction in virome alpha-diversity, species richness and evenness when antibiotics were included in the fermenter (Figure S1B).

qPCR directed at the conserved terminase or primase genes was performed to determine an approximate titre in copies/ml of the seven crAss-like phages (Figure 2C). Primers were either genus specific (genera I, IV, and V) or phage strain specific (genera IV and VII) depending on the level of homo- or heterogeneity of the terminase gene sequence in each genus (or primase in the case of candidate genus VII). For each

candidate genus, higher phage titres were detected in the presence of antibiotics, although to varying extents. Candidate genus I was 6-fold higher in titre in comparison to the control vessel by the end of the fermentation, candidate genus IV 17-fold, candidate genus V 8-fold, candidate genus VI phage A 10-fold, candidate genus VI phage B 24-fold, candidate genus VI phage C 14-fold and candidate genus VII 5-fold higher. The statistical significance of the differences between the highest titres attained for each phage under both conditions tested was examined by a two-tailed paired t-test. This showed that a statistically significant titre difference occurred for three of the phages: candidate genus IV and candidate genus VI phage B (P-value ≤ 0.001) and candidate genus VI phage C (P-value ≤ 0.01).

3.4.4 Screening the 924 faecal sample for potential bacterial hosts

The same crAss-rich faecal sample was plated on antibiotic agar selective for Bacteroidetes and 48 colonies were chosen based on variations in colony morphology. Sequencing of the 16S rRNA gene fragment assigned the 48 isolates to six species: *Bacteroides uniformis*, *Bacteroides ovatus*, *Bacteroides dorei*, *Bacteroides fragilis*, *Bacteroides xylanisolvens* and *Parabacteroides. distasonis* (Table 3). The crAss-like phage enriched fermenter filtrate was added to a pure culture of each of the 48 strains and following incubation anaerobically at 37°C genus- or phage-specific qPCR was used to detect increases in individual phage strains. We detected propagation of a crAss-like phage of candidate genus IV on *B. xylanisolvens* APCS1/XY following three consecutive days of sub-culturing. Despite the efficient propagation of the phage during this and subsequent enrichment, the liquid culture failed to clear. When faecal filtrate prepared from subject ID:924 faeces prior to fermentation was spotted onto a

lawn of *B. xylanisolvens* APCS1/XY no zone of clearing or individual plaques were observed.

3.4.5 Genome analysis of Φ crAss002

Shotgun sequencing of the phage propagated in pure culture confirmed the isolation of a novel crAss-like phage, designated Φ crAss002, with a circular genome of 93,030 bp (NCBI GenBank MN917146) (Figure 3, Table 4). Φ crAss002 genome has 81 protein-coding genes in two oppositely orientated gene modules (possibly two transcriptional units) but less than half could be assigned a function (Additional file 4: Table S2). No genes for lysogeny functions were identified, and we did not detect integrated copies of this or related phages in bacterial genomes (NCBI RefSeq). Module 1 spanning 45-93kb is largely dominated by functions associated with replication and nucleic acid metabolism, whereas module 2 spanning 0-45kb includes phage structural genes, as well as those encoding lysis and packaging functions. Two large genes (gp32 and gp33) located at the beginning of module 1 are predicted to encode RNA polymerase subunits. The Φ crAss001 genome also has large genes in a similar location that were assigned the same function (Shkoporov et al., 2018a). The G+C content of the Φ crAss002 genome is 31.92 mol%, approximately 10 mol% lower than the host G+C content of 42.24 mol%. The tail components and the receptor binding proteins remain to be identified, but are likely to be associated with gp11, 13-15, and/or 17-31. The DNA packaging mechanism of Φ crAss002 was predicted to be headful packaging with terminase initiation occurring at a *pac* site and so the packaged genomes are circularly permuted with redundant termini. The average nucleotide identity (ANI) was determined between Φ crAss002 and twenty other phage genomes

of candidate genus IV (Figure 4A). The metadata associated with these phages is listed in (Table 5). Φ crAss002 is the phage designated fferm_ms_10 in the original sample collection. There are three obvious clades with >95% nucleotide similarity, recommended by The International Committee on Taxonomy of Viruses (ICTV) as the cut-off level for species (Adriaenssens and Brister, 2017). Therefore, amongst the crAss-like candidate genus IV phage sequences, there are potentially four species, with err844056 potentially forming its own species. Genome comparison of a subset of these phages, with different degrees of relatedness, revealed a high degree of genome synteny (Figure 4B). Phages are in descending order of decreasing ANI in relation to Φ crAss002. Even the most distant representative shown here, ϕ eld18-t3_s_1, shares a genetic organisation highly syntenic to that of Φ crAss002.

The genomic synteny of Φ crAss002, ϕ crAss001 and p-crAssphage (prototypical crAssphage generated only from *in silico* data) were also compared (Figure 4C). Φ crAss002 shares no tBLASTx homology with ϕ crAss001 but there is much greater homology and synteny between Φ crAss002 and p-crAssphage (candidate genus I), which could be expected as these phages both belong the *Alphacrassvirinae* sub-family.

3.4.6 Biological characterisation of Φ crAss002

Transmission electron microscopy confirmed that Φ crAss002 has a Podovirus morphology with a capsid diameter of 77.0 ± 2.0 nm and tails of 18.1 ± 2.3 nm. Unlike crAss001, the tail structure is short and has no obvious appendages (Figure 5A). Plaque assays and spot assays were performed using enriched cultures of Φ crAss002.

Plaques failed to form despite testing several media modifications and spots with concentrated phage suspensions were opaque and only barely visible. The clarity of spots was variable between independently generated overnight cultures of *B. xylanisolvens* APCS1/XY. Furthermore, the phage only propagated if co-cultured for a minimum of 3-5 days.

To examine the growth dynamics of Φ crAss002, the phage and host were serially propagated via daily sub-culturing over 11 days. The phage load was quantified daily in copies/ml using qPCR. The phage failed to propagate between day 1 and day 3 during which a reduction in titre of about 4 to 6 logs occurred. This is largely consistent with a dilution effect on the inoculum. After day 3, the titre increased until it stabilised at approximately 10^8 copies/ml from day 5 onwards (Figure 5B). The observed stability is consistent with the persistence of crAss-like phages observed in human gut viromes over time (Shkoporov et al., 2019). The commercially available strain, *B. xylanisolvens* DSM18836 failed to support replication of Φ crAss002, as did the Φ crAss001 host, *B. intestinalis* 919/174 (Figure 5C). When *B. xylanisolvens* APCS1/XY was plated on an agar plate coated with a high-density preparation of Φ crAss002, 76% of bacterial cells gave rise to a colony. This suggests that a dominant fraction of the population does not support phage replication (perhaps due to phase variation). Ten of these colonies were triple streaked and tested for the presence of episomally replicating or integrated copies of Φ crAss002. PCR confirmed all clones were phage negative. When isolates were tested with the phage by spot assay, the spots presented with varying turbidity.

3.4.7 Characterisation of the Φ crAss002 host *Bacteroides xylanisolvens* APCS1/XY

The genome of *B. xylanisolvens* APCS1/XY consists of a single circular chromosome (6,461,058 bp, GenBank CP042282) and two circular plasmids (pBXS1-1, 5,595 bp, GenBank CP042281 and pBXS1-2, 4,148 bp, GenBank CP042283). A considerable number of elements were identified in the genome that could potentially drive phase variation of surface structures and contribute to phage resistance/sensitivity. These included multiple genes coding for site-specific tyrosine recombinases (9), tyrosine-type DNA invertases (4) and site-specific integrases (28) (Figure S2A). In certain cases, these genes were in proximity to genes encoding bacterial surface molecules that have been previously identified as being subjected to phase variation and affecting surface composition and phage sensitivity in *Bacteroides* (Porter et al., 2020). These bacterial features included nutrient uptake genes such as the products of the *sus* gene family, genes coding for TonB-dependent nutrient transporters or capsule polysaccharide biosynthesis genes (Sonnenburg et al., 2005; Nakayama-Imaohji et al., 2009; Horino et al., 2009; Porter et al., 2020). Examples of cell-surface associated genes co-localised with site-specific recombinases were observed at the following loci: FNQN58_01735-01770 (TonB gene family + tyrosine-type recombinase/integrase), FNQN58_07900-07925 (TonB/Sus/Rag gene family + tyrosine-type recombinase/integrase), FNQN58_12195-12225 (lipopolysaccharide and capsule biosynthesis + tyrosine-type recombinase). In order to obtain preliminary evidence of phase variation in *B. xylanisolvens* APCS1/XY associated with dynamic recombinations in the genome, we performed analysis of individual Oxford Nanopore reads. Reads were aligned to the assembled chromosome scaffold and recombination hotspots, indicating potential phase-variable loci were identified (Figure S2B-D).

Interestingly, many of the recombination hotspots overlapped with or occur in proximity to genes that function as Sus or TonB family transporters or receptors, site-specific integrases, restriction endonucleases or transposases. For example, a recombination hotspot that extends from genome position 6151000-6152000 bp overlaps a TonB-dependent receptor and occurs in proximity to RagB/SusD family nutrient uptake outer membrane protein encoding genes and site-specific integrase. Another example extends from 909000-910000 bp overlapping a site-specific integrase which occurs upstream of a capsule biosynthesis gene. The hotspot identified across genome position 916000-917000 bp overlaps a glycosyltransferase gene which is in proximity to lipopolysaccharide synthesis genes.

Other features of interest noted in the *B. xylanisolvens* APCS1/XY genome were over 139 transposases associated with thirteen or more insertion sequence (IS) families, the presence of three xylanases characteristic of this bacterial species, over one hundred *sus/tonB* associated genes and three capsule polysaccharide biosynthesis operons. Annotation of plasmids pBXS1-1 and pBXS1-2 showed that both carry genes coding for toxin-antitoxin systems which may have a role in phage defences by mechanisms such as abortive infection. Other roles of these systems include post-segregational killing or persistent formation which allows transient physiological changes that increase tolerance to antibacterial substances such as antibiotics (Harms et al., 2018). Plasmid pBXS1-2 also carries a gene that encodes for vesicle formation. This may be linked to the outer membrane vesicle-like structures (OMVs) observed in micrographs from cross-sections of soft agar lawns prepared with cultures of *B. xylanisolvens* APCS1/XY, with and without Φ crAss002 exposure (Figure S3). The formation of such vesicles from the outer membrane occurs naturally among Gram-negative bacteria and they are thought to have multiple roles including secretion and

transport of soluble and insoluble molecules, DNA transfer, stress adaptation, virulence and phage defence (Kulp and Kuehn, 2010; Toyofuku et al., 2019). The precise role of the vesicles observed on the surface of *B. xylanisolvens* APCS1/XY cells remains to be elucidated.

The genetic accessibility of *B. xylanisolvens* APCS1/XY was confirmed via conjugation of pSAM-Bt with donor S17-1 λ pir. This is significant as it allows the potential development of random mutant libraries. This may lead to the isolation of *B. xylanisolvens* APCS1/XY variants that are more sensitive to Φ crAss002. Ultimately, this genetic accessibility could prove a useful tool in the further characterisation of Φ crAss002 and its hosts.

3.4.8 Co-cultivation of Φ crAss002 and *Bacteroides xylanisolvens* APCS1/XY

We have already described the phenomenon in which Φ crAss002 fails to propagate for several days in the presence of its host, before accumulating to and maintaining high levels (Figure 5B). This suggests that either the phage or the host has undergone some adaptation within the first days of propagation. This was confirmed via serial sub-culturing over 30 days in which the same phenomenon was observed. Following recovery and stabilisation, the phage propagated at approximately 10^8 copies/ml for 21 days (Figure 6A). Phage lysate collected at the end of the 30 days was then propagated on a naïve host (a bacterial culture which had not been in recent contact with the phage) and once again there was an initial drop in titre, consistent with a dilution effect, before recovery and maintenance at high titres. (Figure 6B). This suggests that phage variants have not been selected, and that bacterial host

adaptation may be responsible for the stable co-propagation. To examine this, ten individual colonies were selected from the bacterial pellet formed after centrifugation of the 6-day co-culture on naïve cells (grown in the presence of phage) and used as the starting material for another co-culture cycle. Three behaviours were observed. Two clones immediately supported Φ crAss002 propagation at high titres, whereas the phage titre dropped significantly in the presence of one clone (Figure 6C).

The remaining seven clones gave an intermediate response. By day two, all cultures supported a high titre of Φ crAss002 of $\sim 10^9$ copies/ml. On naïve cells, 4-5 days of co-culturing is required to achieve similar titres. This suggests that the bacterial population is heterogenous in terms of its phage permissiveness and that counterintuitively the presence of the phage selects for phage sensitive host variants. When Φ crAss002 was spotted on lawns the phage exposed clones, spot turbidity was reduced in varying amounts compared to spots on lawns of naïve cells.

3.4.9 Impact of crAssphage on hosts in a defined community

A continuous fermentation model was initiated in triplicate with a defined bacterial community constructed from eight different species representing a simplified human microbiota consortium (SIHUMI) (Figure 7). Following inoculation at similar levels, a community structure formed within hours that remained stable for 72 hours (Figure 7A). When Φ crAss002 and Φ crAss001 were added from time point 0, there was no impact on either the levels of their individual hosts, or on the community structure (Figure 7B). Φ crAss002 levels decreased as observed previously consistent with a washout due to media replacement, before recovering within 72 hours (Figure

7C). Φ crAss001 achieved high levels within a few hours and were maintained at high titres thereafter (Figure 7D).

3.5 Discussion

Further advances in understanding the biology of the human gut phageome will depend on the isolation, propagation and characterisation of individual phage-host pairs. This is particularly true for the most abundant representatives of this viral community, the crAss-like phages. With the exception of Φ crAss001 and two close relatives of this phage, DAC15 and DAC17, all other crAss-like genomes described to date are the result of composite assemblies and have never been propagated in pure culture in a laboratory (Dutilh et al., 2014; Norman et al., 2015; Manrique et al., 2016; Yutin et al., 2018; Guerin et al., 2018; Shkoporov et al., 2018a; McCann et al., 2018; Cervantes-Echeverría et al., 2018; Shkoporov et al., 2019; Edwards et al., 2019; Siranosian et al., 2020; Hryckowian et al., 2020). Biological characterisation of Φ crAss001 revealed several intriguing traits. Although shown to be virulent and capable of forming plaques on agar plates, it fails to clear liquid cultures of its host where both phage and host can stably co-exist and propagate to high levels (Shkoporov et al., 2018a). CrAss-like phages also form part of the personal persistent virome (PPV), a consistently present, individual-specific core of mostly virulent phages in the viromes of healthy individuals (Clooney et al., 2019; Shkoporov et al., 2019). In addition, crAss-like phages have been demonstrated to engraft and persist in the microbiome of faecal microbiota transplantation recipients and can undergo vertical transmission between mother and infant (Draper et al., 2018; McCann et al., 2018; Siranosian et al., 2020). This suggests that ecological models yet to be characterised are at play in the human gut that allow this persistence. It may be that virulent phages employ a “piggyback-the-winner” strategy to take advantage of the success of well-adapted bacterial hosts in the gut. Originally this model was described by Rohwer and colleagues proposing that in thriving bacterial populations with a low virus-to-microbe

ratio (VMR), phages can make a lytic to lysogenic lifestyle switch to “piggyback on” the success of their host and thus ensure their maintenance in the ecosystem (Knowles et al., 2016; Silveira and Rohwer, 2016). In the context of the human gut, similar strategies may be employed that do not necessarily involve lysogeny but allow virulent phages persist in the human gut. The human gut is thought to have a VMR of ~0.1:1, which is lower than the VMR predicted in other ecosystems (Shkoporov and Hill, 2019). Virulent gut phages may have evolved less aggressive infection strategies to ensure efficient and sustained replication without host elimination thus ensuring their persistence in the gut. Elucidation of the mechanisms behind this persistence requires further investigation.

Isolation of these dominant gut bacteriophages is a challenge given that they are likely to be specialists in terms of their host range, and the high levels of variation observed in their predicted receptor-binding proteins (Shkoporov et al., 2018a). Furthermore, many anaerobic gut bacteria are difficult to cultivate and variations in host sensitivity/resistance may make it difficult to identify phage host pairs by standard agar-based methods such as plaquing and spot assays. The isolation of further crAss-like phages will be important in helping us understand their interaction with their host(s) and how certain phages can persist in the human gut over such extensive time periods.

The overall objective of this study was the isolation of novel phages from the gut of an individual previously identified as being rich in several different crAss-like phages. The antibiotic-driven crAss-like phage enrichment implemented in this study confirmed that suitable hosts for crAss-like phage propagation were present and viable in the faecal samples. This led to the isolation of Φ crAss002, a novel member of the crAss-like phage family isolated in pure culture. Φ crAss002 infects *B. xylanisolvens*,

a Gram-negative, strictly anaerobic, non-pathogenic, xylan-degrading bacterium, which was first isolated from human faeces in 2008 (Chassard et al., 2008; Ulsemer et al., 2012). Recently, *B. xylanisolvens* has been demonstrated to be capable of boosting production of natural TFA sugar antigen-specific IgM antibodies in healthy humans, an antibody response believed to be involved in cancer immune surveillance (Ulsemer et al., 2016). Due to the associated beneficial properties of this bacterial species, *B. xylanisolvens* DSM23964 is one of the few bacterial strains approved by the European Food Safety Authority for applications in a novel food; heat-treated milk products fermented with the strain in a non-viable form, under the Novel Food Regulation No 258/97 (European Commission, 2015).

ΦcrAss002 is a member of candidate genus IV of the *Alphacrassvirinae* sub-family, according to our recently proposed taxonomic scheme for gut-associated crAss-like phages (Guerin et al., 2018). The first crAss-like phage isolated, ΦcrAss001, infects *B. intestinalis* and is a member of the more heterogeneous candidate genus VI of the sub-family *Betacrassvirinae* (Shkoporov et al., 2018a). The isolation of two crAss-like phages which are from the same genus VI has been reported, DAC15 and DAC17, that both infect *Bacteroides thetaiotaomicron* (Hryckowian et al., 2020). Detailed biological characterisation of these two phages, isolated from sewer-adjacent pond water collected in Bangladesh will provide further insights into this phage family. ΦcrAss002 is more closely related to p-crAssphage, the founder member of the family and most abundant bacteriophage in the Western populations, which also belongs to the *Alphacrassvirinae* sub-family (Figure 4C). The cultured crAss-like phages share a podovirus-like morphology but the tail structure of ΦcrAss001 appears to be more elaborate than that of ΦcrAss002 (Figure 5A) (Shkoporov et al., 2018a). ΦcrAss001 produces large plaques whereas ΦcrAss002 does not and even at high titre

only forms opaque zones of clearing when spotted in high concentration on a lawn of its host bacterium. Both phages also employ different mechanisms for DNA packaging (short direct terminal repeats in Φ crAss001 versus pac-type headful packaging in Φ crAss002). Despite their apparent differences, both phages share notable similarities (Table 6). Both phages infect *Bacteroides* hosts, appear to be specialists in host range and neither possess lysogeny associated genes nor are they able to form stable lysogens or pseudolysogens. Their genomes also differ in G + C content by about 10% mol in comparison to their host genomes. While Φ crAss001 and Φ crAss002 are virulent in nature, both fail to clear liquid cultures yet still reach high titres. Intriguingly, both can co-exist in high levels with their host over prolonged periods.

It has become apparent that many virulent phages persist at high levels in concert with their bacterial hosts. This has been indicated by observations in multiple studies (Weiss et al., 2009; Maura et al., 2012a; Siringan et al., 2014; Shkoporov et al., 2018a; Lourenço et al., 2019). Overall, the “kill-the-winner” dynamics has not been observed in the human gut virome, at least at the level of resolution of genus or species taxonomic levels (Reyes et al., 2010; Minot et al., 2011; Reyes et al., 2013; Moreno-Gallego et al., 2019; Shkoporov et al., 2019). The observed persistence of the crAss-like phages is consistent with an in-depth longitudinal study of the virome which followed ten healthy individuals over a one-year period. The stable abundant component of a healthy virome largely consists of virulent phages (including crAss-like phages) with a minority of temperate phages (Shkoporov et al., 2019). This study strongly supports the notion that phage communities employ strategies that allow stable co-existence with their hosts.

Continuous co-culture of Φ crAss002 revealed that propagation does not occur efficiently on initial host contact. Following a period in which no replication is

observed, phage titre recovers and stabilizes (Figure 6A, B). Counterintuitively, this suggests that growth in the presence of phage selects for a phage permissive population composition, whereas growth in the absence of phage selects for a phage resistant population structure. When added to a simplified bacteria community grown in a chemostat host levels are unaffected by the presence and propagation of the phages (Figure 7B). Similar findings were observed for phages infecting enteroaggregative O104:H4 *Escherichia coli* in a conventional mouse model. These phages, isolated from sewage, were able to propagate stably and continuously over a number of weeks *in vivo*. Interestingly, faecal bacterial counts for phage treated and non-phage treated mice remained the same despite phage propagation (Maura et al., 2012a). In the fermenter, no knock-on effects were observed for off-target community members. This is both supported and contradicted by other *in vitro* and *in vivo* studies (Lourenço et al., 2019; Cieplak et al., 2018; Hsu et al., 2019). The observed stability and persistence may be due “piggyback-the-winner” style dynamics which describes the way in which phages can adopt less disruptive infection strategies and “piggyback” on the success of their bacterial host in an ecosystem (Knowles et al., 2016; Silveira and Rohwer, 2016). This allows more efficient replication and permits co-existence with that cognate host that would not be achieved with aggressive lytic replication. In the context of the gut, this would be favourable as it guarantees persistence. It may also be that “Royal Family” ecological dynamics are at play which is the occurrence of “kill-the-winner” dynamics at a strain or sub-strain level thus resulting phage-host fluctuations going undetected at genus or species level (Breitbart et al., 2018). This is supported by the observed “host-jumping” of phages at the strain level which requires as little as a single point mutation in the tail fiber gene (Sordi et al., 2017). This phenomenon most likely occurs due to the inability of the phage to access the original

host strain or as a result of a reduction in the cognate host strain counts due to variations in the gut environment. Interestingly, enrichment of non-synonymous mutations in the tail fibre gene of p-crAssphage have been observed with a greater incidence compared to other genes (Siranosian et al., 2020). Strain level variation can be difficult to identify as 16S rRNA gene sequencing does not detect below species level (Sutton and Hill, 2019). Long-read sequencing, with platforms such as Oxford Nanopore, which can generate reads that are representative of near complete genomes may aid strain-level analyses (Somerville et al., 2019; Warwick-Dugdale et al., 2019). Bioinformatic pipelines that allow strain-resolved metagenomics have also been described (Edwards et al., 2019; Segata, 2018). It is still unclear why *B. xylanisolvens* APCS1/XY seems to select for the presence of high titres of the phage while the bacterial count remains unaffected by its presence. Perhaps it is favourable to the host to prevent extinction of the phage. It may be that some ecological advantage is conveyed or that sub-strain level variation imposed by the presence of the phage is important in more complex situations than those examined in this study.

Phase variation (PV) is one possible mechanism that could allow hosts and phages to co-exist stably. This allows the host to transiently switch between phage permissive and non-permissive phenotypes through the reversible inversion of promotor-containing DNA regions called invertons at loci such as those encoding cell surface features which can act as phage receptors (Jiang et al., 2019; Porter et al., 2020). Bacterial hosts have been shown to use PV to their advantage by developing herd immunity on phage exposure via phenotypic switch control of predating phage viral load (Turkington et al., 2019). This is possible due to the constant presence of a transiently phage permissive sub-population with the other portion of the population in a non-permissive state. This permits phage propagation but limits the viral load so

that the host is never completely eliminated. A sub-population of each culture may revert to non-phage permissive at a certain threshold and thus control viral load. It has also been demonstrated in murine models that host permissiveness to phage infection is not uniform throughout the gut and is influenced by ecophysiology (Maura et al., 2012b; Galtier et al., 2017). Recently, the significance of transient resistance conveyed by PV and how it can dictate phage-host interactions has become of interest (Jiang et al., 2019; Turkington et al., 2019; Porter et al., 2020). Examination of the *B. xylanisovlens* APCS1/XY genome revealed a large number of genes coding for outer membrane proteins and capsular polysaccharide biosynthesis enzymes, associated in many cases with potential invertons (GenBank CP042282, Figure S2) (Zaleski et al., 2005; Nakayama-Imaohji et al., 2009; Zitomersky et al., 2011; Nakayama-Imaohji et al., 2016; Porter et al., 2020). One study demonstrated that PV invertons are particularly dominant among human gut Bacteroidetes with ~19 invertons per genome (Jiang et al., 2019). Another noteworthy characteristic of PV detected *in vitro* is the inability of the phage to clear liquid culture despite reaching a high titre (Porter et al., 2020). This is consistent with behaviour observed for both Φ crAss001 and Φ crAss002 (Shkoporov et al., 2018a). Furthermore, this may explain the observed variability of opacity in zones of clearing and co-culture titres using cultures produced on different days and originating from different single colonies of *B. xylanisovlens* APCS1/XY.

In vivo studies with relevant conditions and transcriptomics could potentially expand our understanding of how *B. xylanisovlens* and Φ crAss002 interact and co-exist. The persistence of virulent gut phages and their bacterial hosts was examined in a murine model using the defined Oligo-Mouse-Microbiota (OMM) bacterial consortium with the addition of two *E. coli* strains (murine commensal strain Mt1B1 and enteroaggregative strain 55989) with three virulent phages infecting the former

and one the latter (Lourenço et al., 2019; Brugiroux et al., 2016). Interestingly, it was found that the radial variation in the murine gut due to anatomical features and condition gradients allows for variable virulent phage-host accessibility. Phage replication appears to largely occur in the lumen whereas the mucosa crypts provide a site of refuge for part of the target bacterial population which gradually migrate into the lumen. Further supporting this was the identification of an increasing mucosa to lumen gradient of lytic phages. Therefore, hosts are exposed to variable phage concentrations in the gut highlighting the significance of spatial heterogeneity (Lourenço et al., 2019). This is also supported by observations of non-uniform phage propagation and composition throughout the GIT (Galtier et al., 2017; Maura et al., 2012b; Zhao et al., 2019). This work also suggests that arm-race dynamics and extension of host range do not have a role in persistence of virulent phages (Lourenço et al., 2019). With strain variation among gut bacterial species often linked with genetic changes associated with phage resistance, the rate of and occurrence of genetic versus transient phage resistance requires further analysis (Scanlan, 2017). Overall, the above study specifically provides valuable insights into how persistent propagation of virulent phages can occur in the human gut without host elimination. Spatial separation in parallel with strain level interactions and transient host resistance may have important roles in this persistence.

Conclusions

We report the isolation of Φ crAss002 from the human gut following antibiotic driven enrichment of its host, *B. xylanisolvens* APCS1/XY, in a faecal fermenter. Biological and *in silico* characterization of Φ crAss002 revealed a number of interesting traits including the inability to form plaques or clear liquid cultures of its host despite the

phage being lytic in nature and attaining high titres. Like Φ crAss001, Φ crAss002 can co-exist with its host over time without impacting host levels at the species level. In the context of the gut, we hypothesize that multiple phenomena are occurring in parallel to allow such persistence and co-existence including, “piggyback-the-winner” and “Royal Family” ecological dynamics, transient host phenotypic variations and spatial heterogeneity. The isolation of more crAss-like phages will be key to expanding our understanding of the most abundant phage family in the human gut. They infect one of the most abundant and important bacterial groups in the gut, Bacteroidales, and so these phages and their hosts provide an opportunity to study the dynamics involved in phage-bacterium interactions in microbial functionality within the gut. Understanding such interactions will be necessary if we are to comprehend the role of the phageome in bacterial homeostasis in the gastrointestinal tract.

3.6 Figures

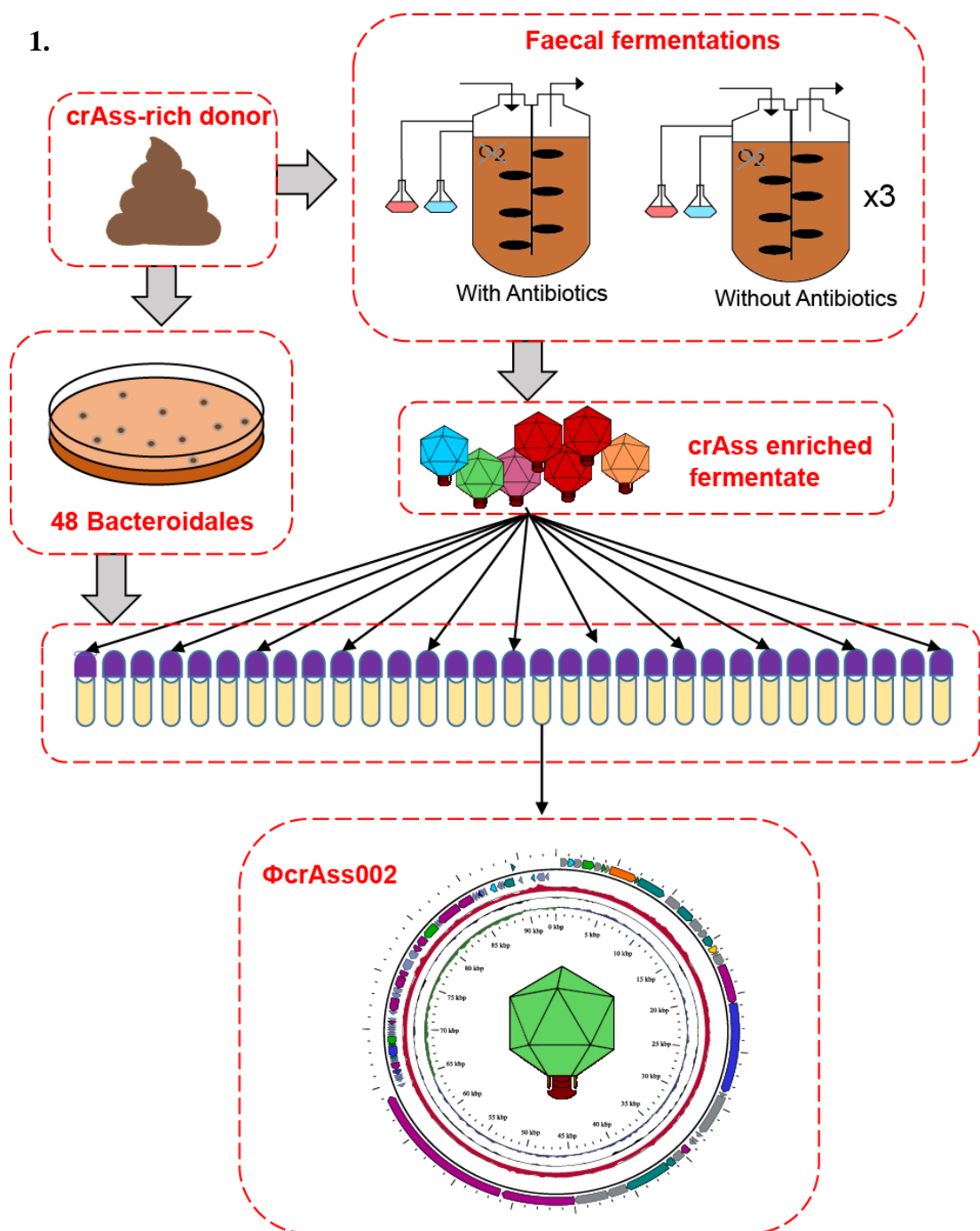
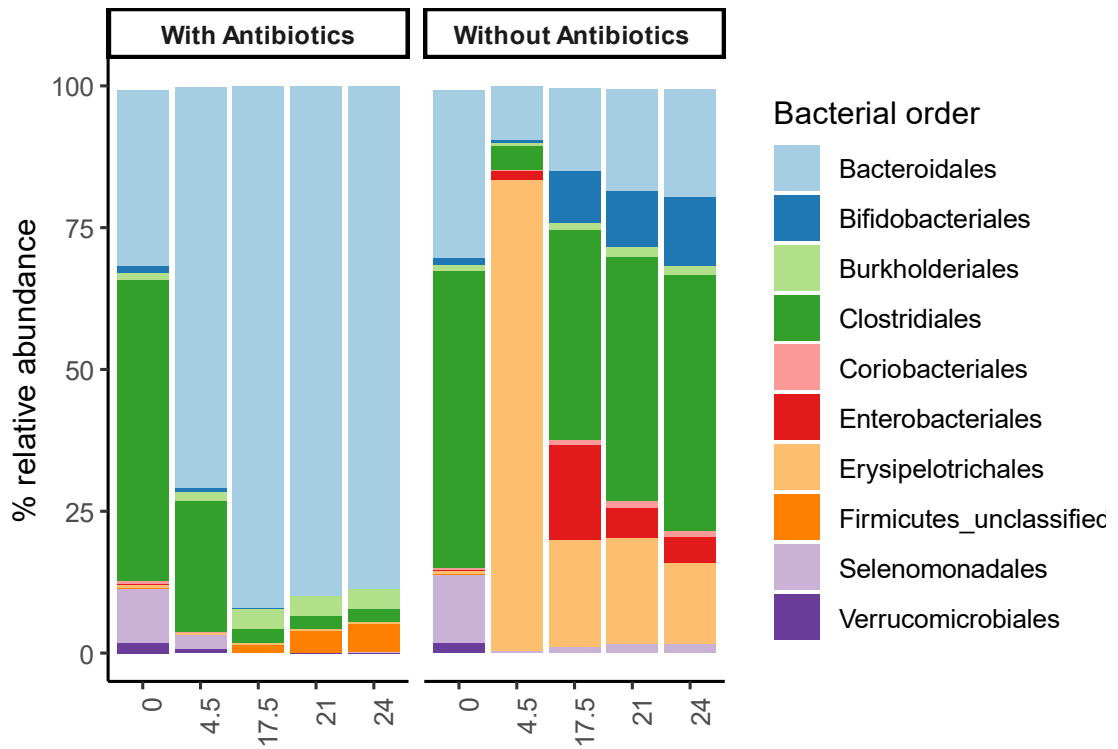


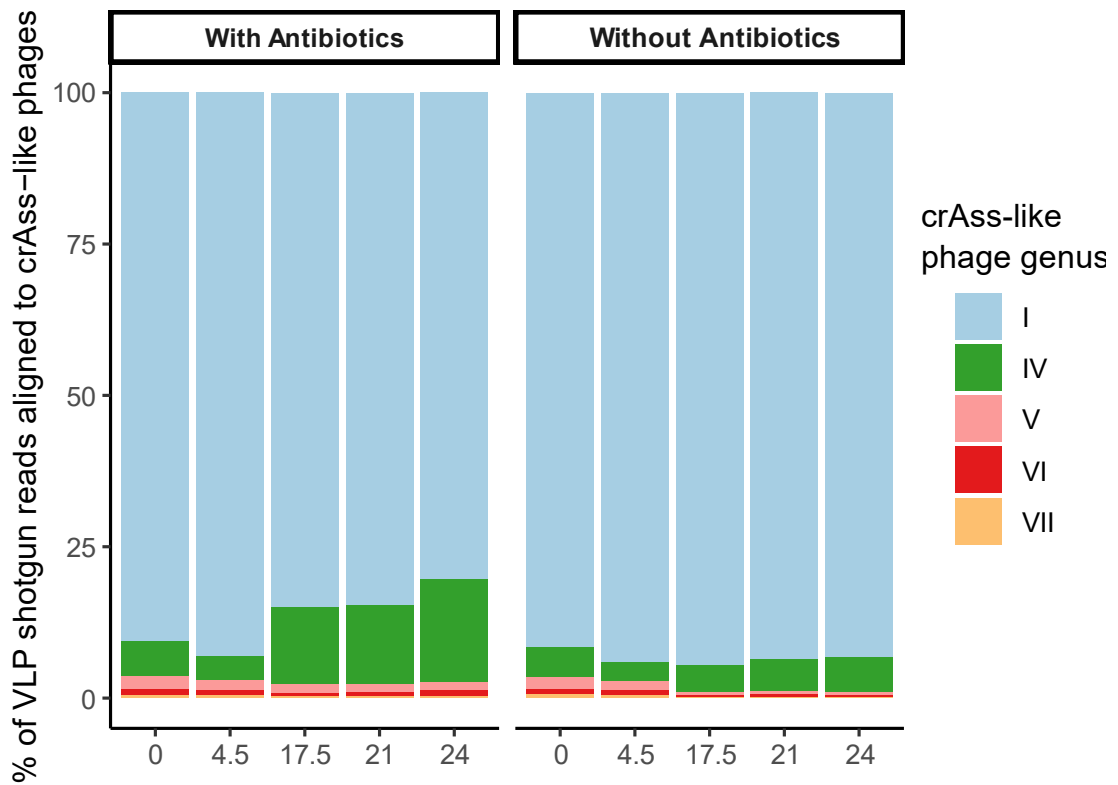
Figure 1. Graphical representation of the key experimental steps taken in the isolation of ΦcrAss002.

2.

A



B



C

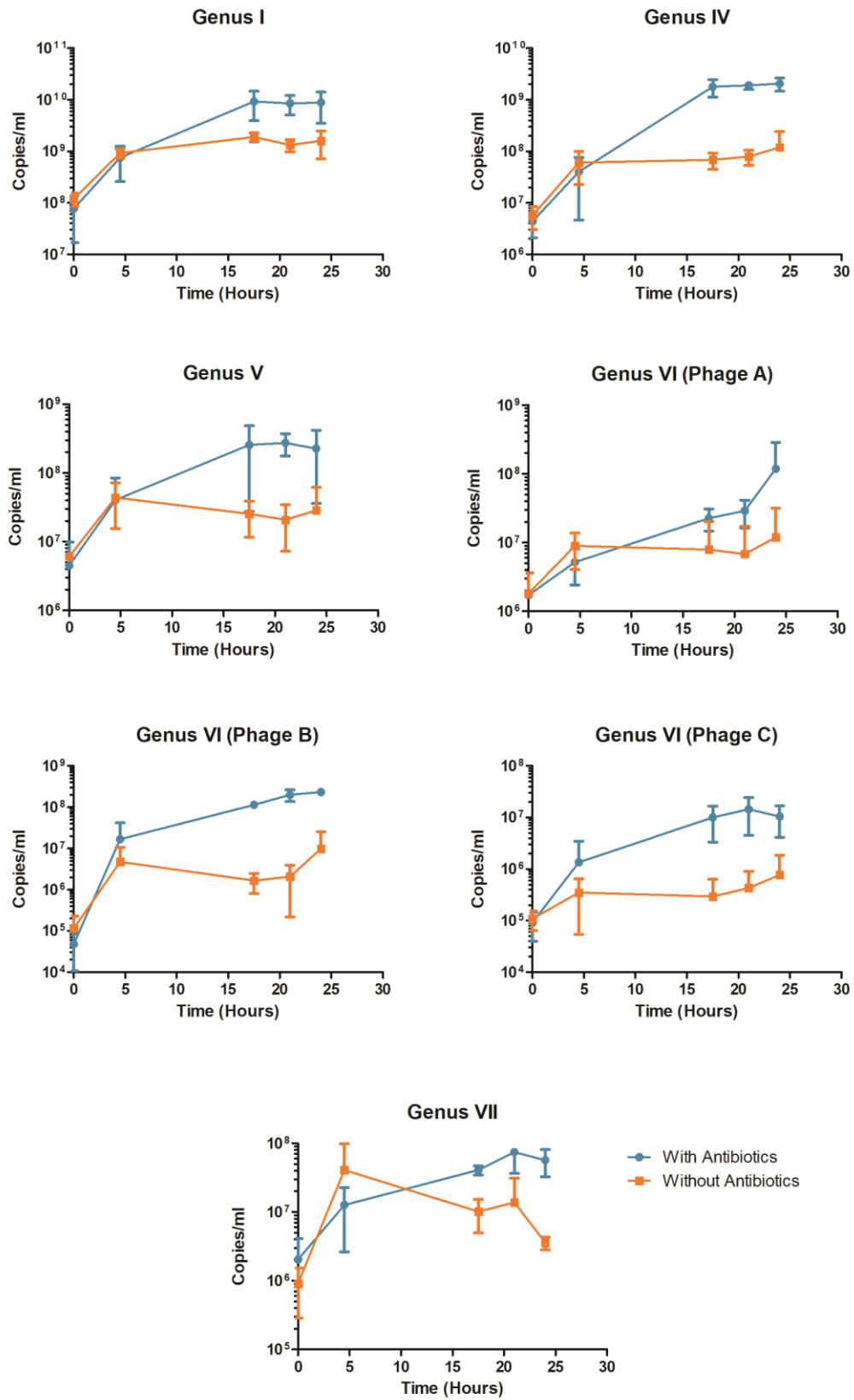


Figure 2. The effect of antibiotic selective enrichment in faecal fermenter on the abundance of bacterial orders and the parallel effect on different crAss-like phage abundances. (A) The mean (across three experimental runs) relative abundances of the key bacterial orders under both conditions tested. (only orders with relative abundance of 1% in any of the samples are shown). (B) The mean relative abundance of crAss-like phage contigs per genus as a percentage of total crAss-like reads. The crAss-like phage contigs are coloured based on candidate genus. CrAss-like phages that resolves into five of the ten crAss-like family candidate genera were detected. (C) Absolute quantification of each crAss-like phage detected using qPCR with phage or genus specific primers targeting a segment of the terminase or primase gene. Error bars indicate standard deviation ($n = 3$).

3.

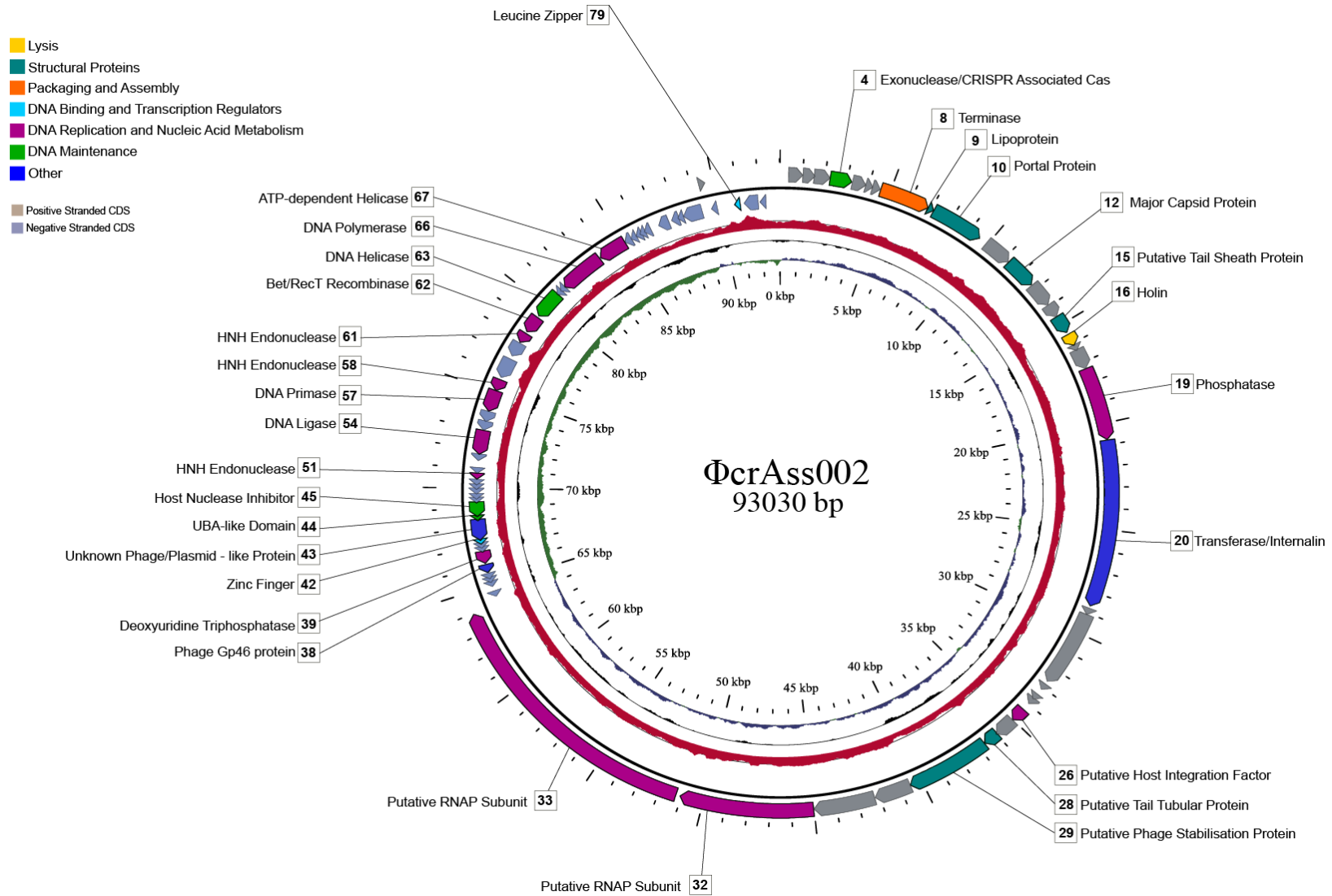
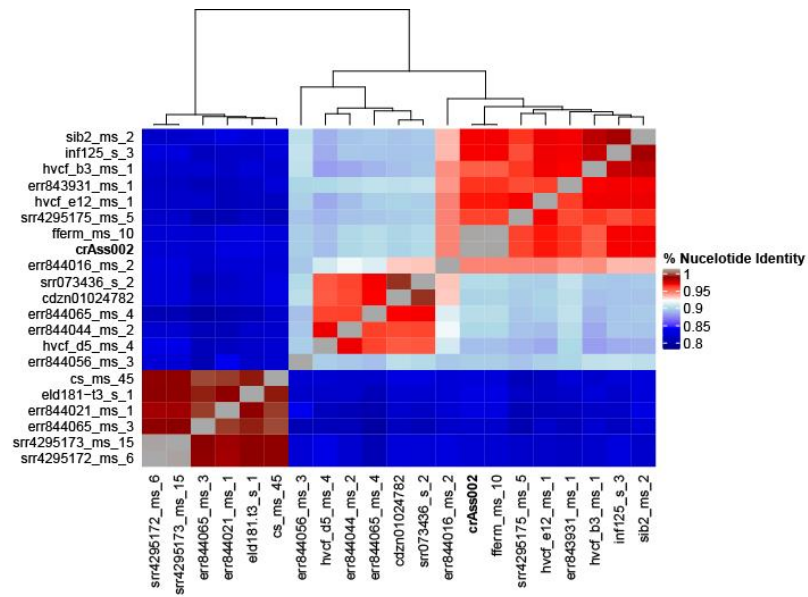


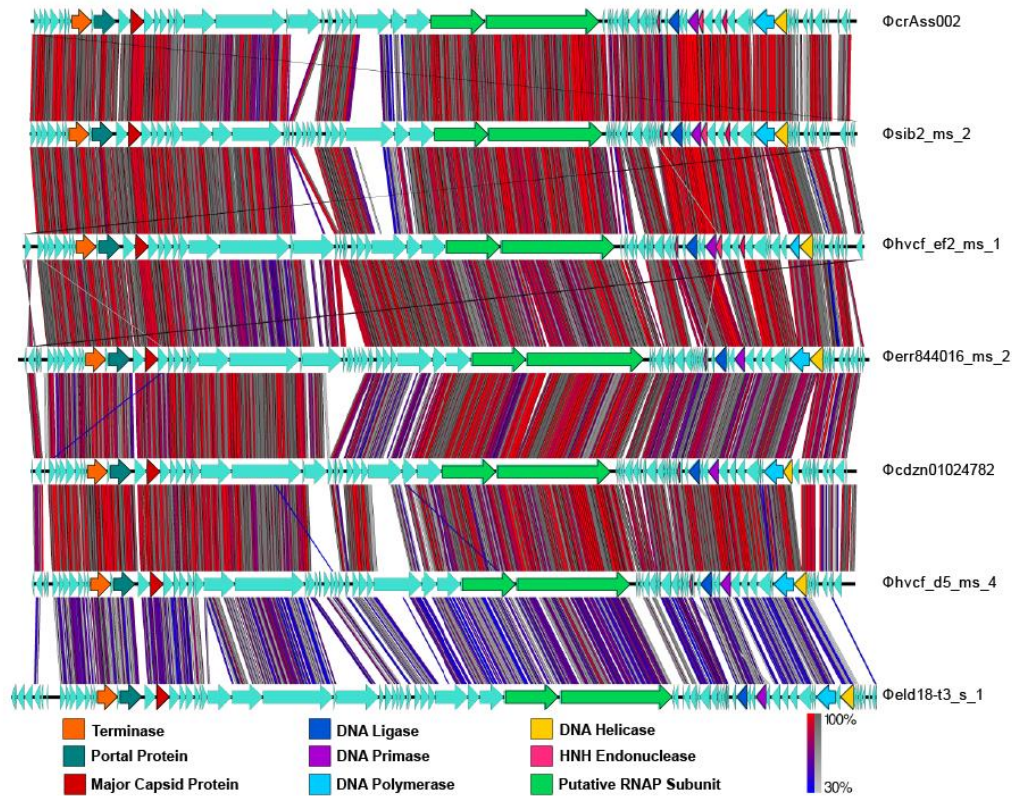
Figure 3. Circular genome map of the Φ crAss002 genome. The innermost ring (blue; positive strand, green; negative strand) depicts G + C skew, the central ring (black) shows G + C content and the outer ring (red) highlights Illumina read coverage along the genome. The outermost circle shows coding genes (CDS) which are labelled on HHpred function predictions. CDS are coloured based on general function which corresponds to the legend. No function predictions were possible for genes which are unlabelled.

4.

A



B



C

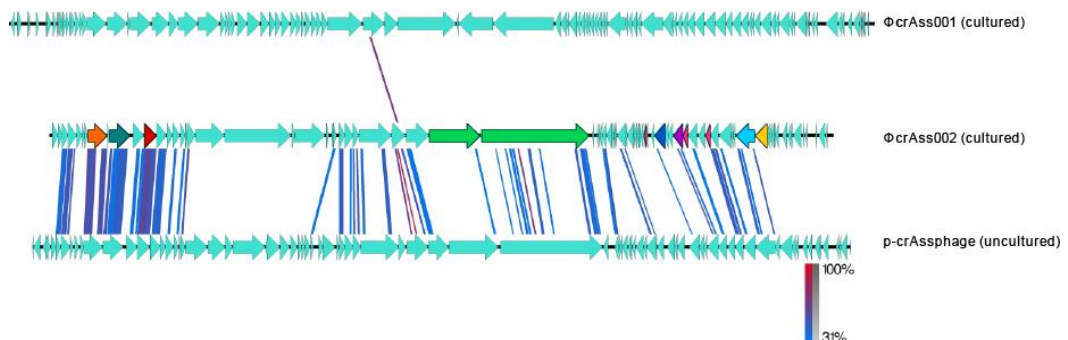


Figure 4. *In silico* characterisation of Φ crAss002. (A) To examine relatedness of Φ crAss002 with twenty other phages of candidate genus IV identified *in silico*, a heatmap was generated based on average nucleotide identity (ANI). (B) Whole genome comparisons of Φ crAss002 and a subset of related phages to highlight synteny and genome organisation. In decreasing order from the top are phages with higher to lower ANI/relatedness. Genes with predicted functions are colour coded based on generalised function. Areas of tBLASTx homology between the genomes is highlighted. (C) Whole genome comparison of Φ crAss002 with Φ crAss001 (sequence from pure isolate) and the prototypical crAssphage (sequence solely *in silico*) to examine synteny and homology. Regions of homology (tBLASTx) are highlighted.

5.

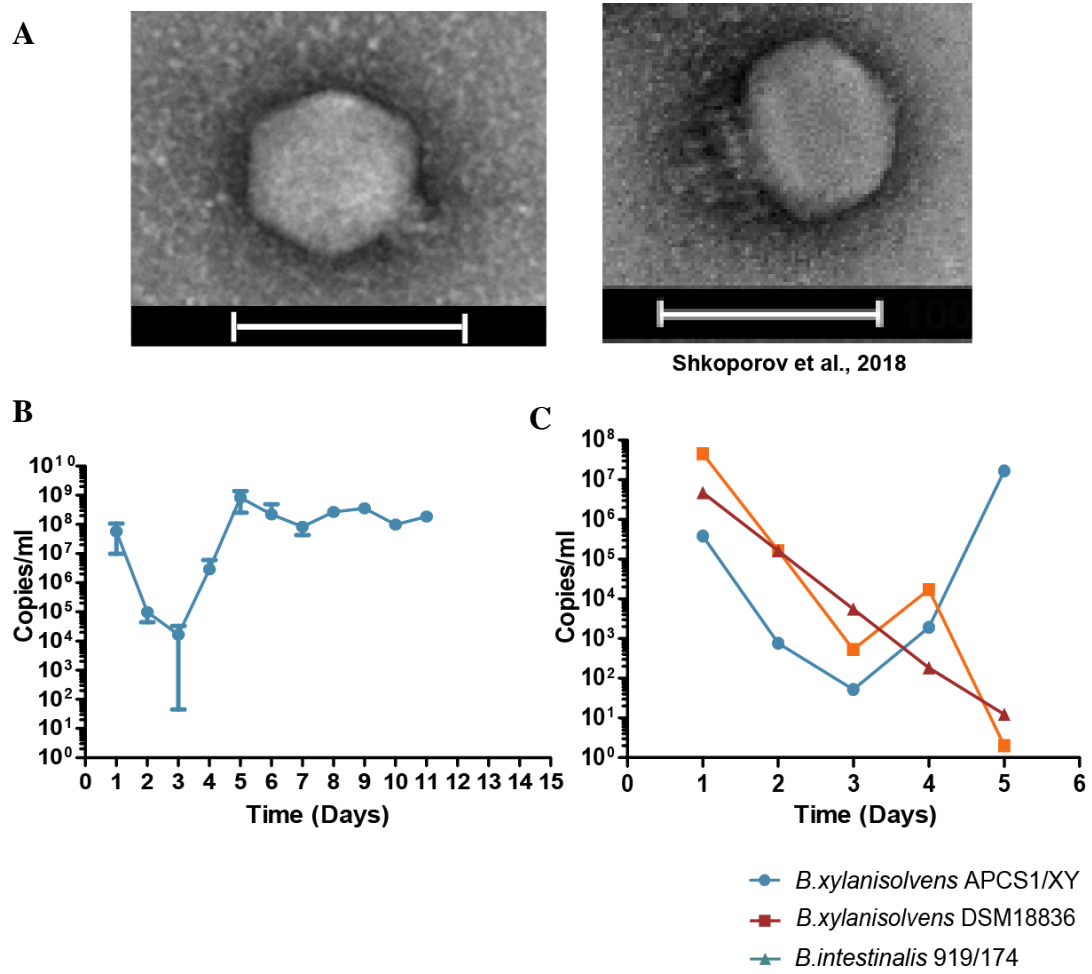


Figure 5. Biological characterisation of Φ CrAss002. (A) Transmission electron micrographs generated from Φ CrAss002 enriched lysate, stained with uranyl acetate. Micrographs show podovirus virions with a diameter of ~ 77 nm and a simple tail structure. Scale bars represent 100 nm. A micrograph of ϕ crAss001 included for comparative purposes. (B) *In vitro* propagation of Φ CrAss002 over 11 days to interaction with its host. Titre quantification was performed using qPCR. (C) Investigation of the ability of Φ CrAss002 to propagate on commercial *B. xylanisolvens* DSM18836 and *B. intestinalis* 919/174 via liquid propagation over five days. Error bars indicate standard deviation (n = 3)

6.

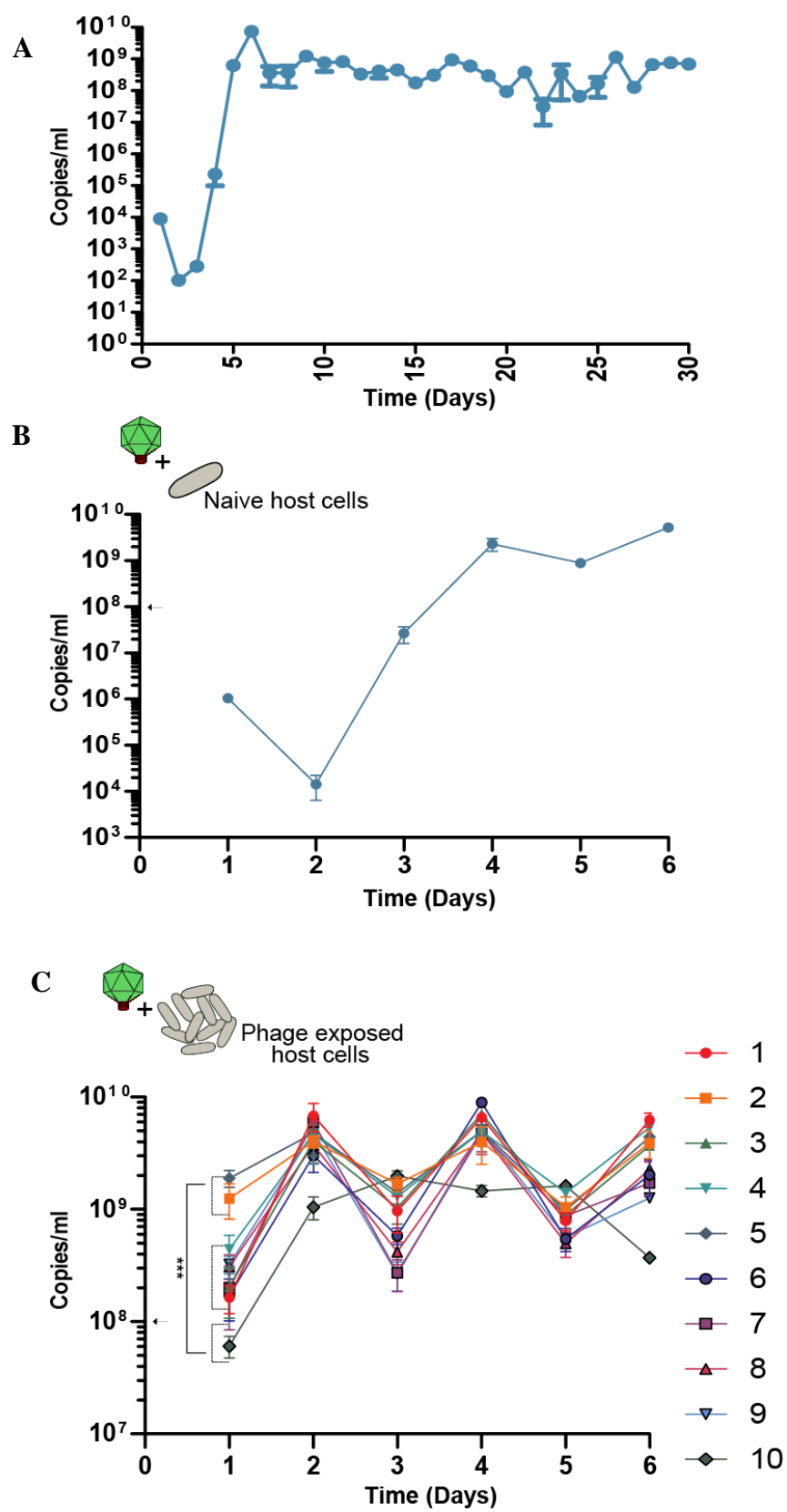


Figure 6. Continuous co-culture of Φ crAss002 and *Bacteroides xylanisolvens* APCS1/XY. (A) Serial co-culturing of Φ crAss002 over 30 days. (B) Serial propagation of the phage on naïve host cells (absent of phage exposure for ~10 generations). (C) Equivalent experiment using ten *B. xylanisolvens* cultures with recent phage exposure. Φ crAss002 titre is shown in copies/ml, determined by absolute qPCR. Statistical analysis was performed by the one-way ANOVA ($P < 0.001$) with Tukeys as post-test comparing titres. Statistical significance of the difference between the lowest titre and the highest titre sustained shown ($P < 0.001$, ***). Arrows indicate the approximate titre Φ crAss002 at the initiation of each propagation cycle. Error bars represent standard deviation ($n = 3$).

7.

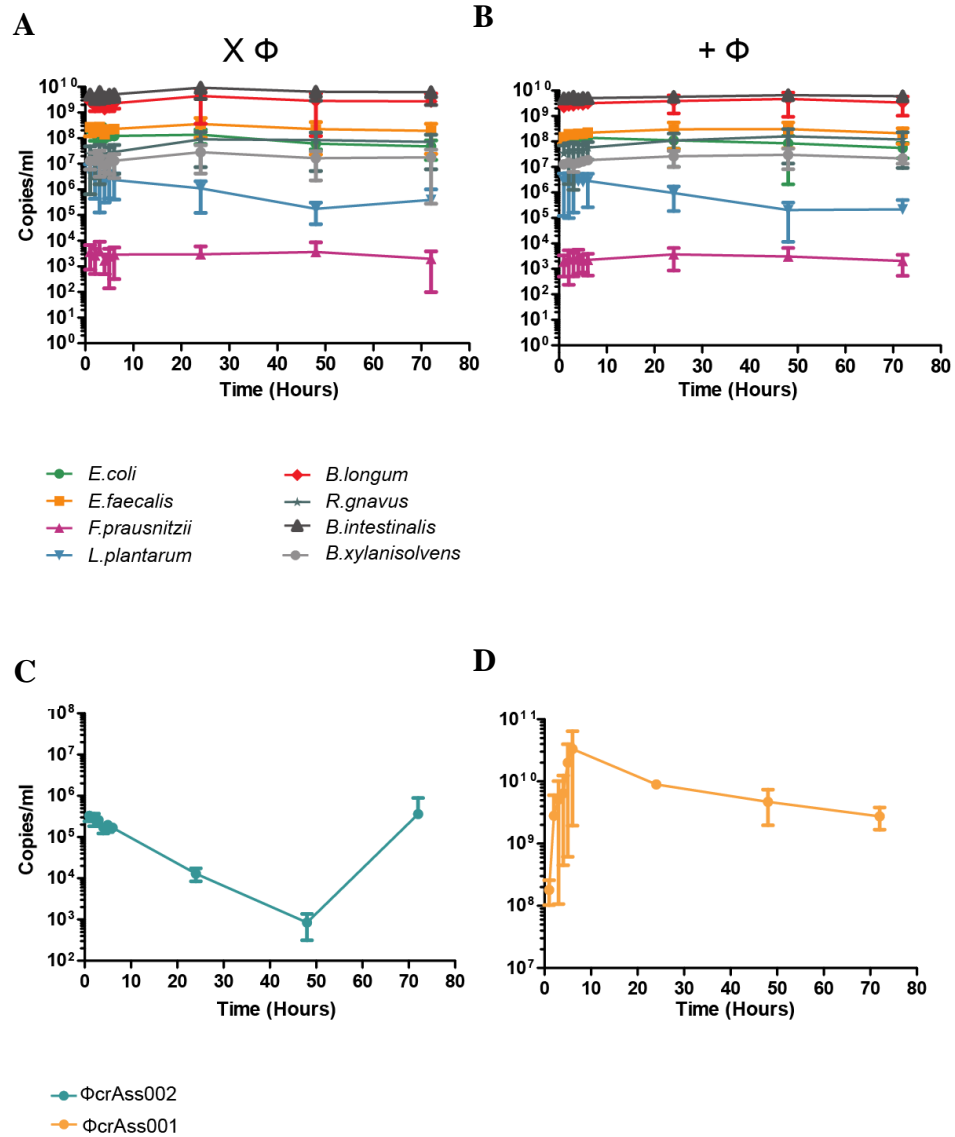
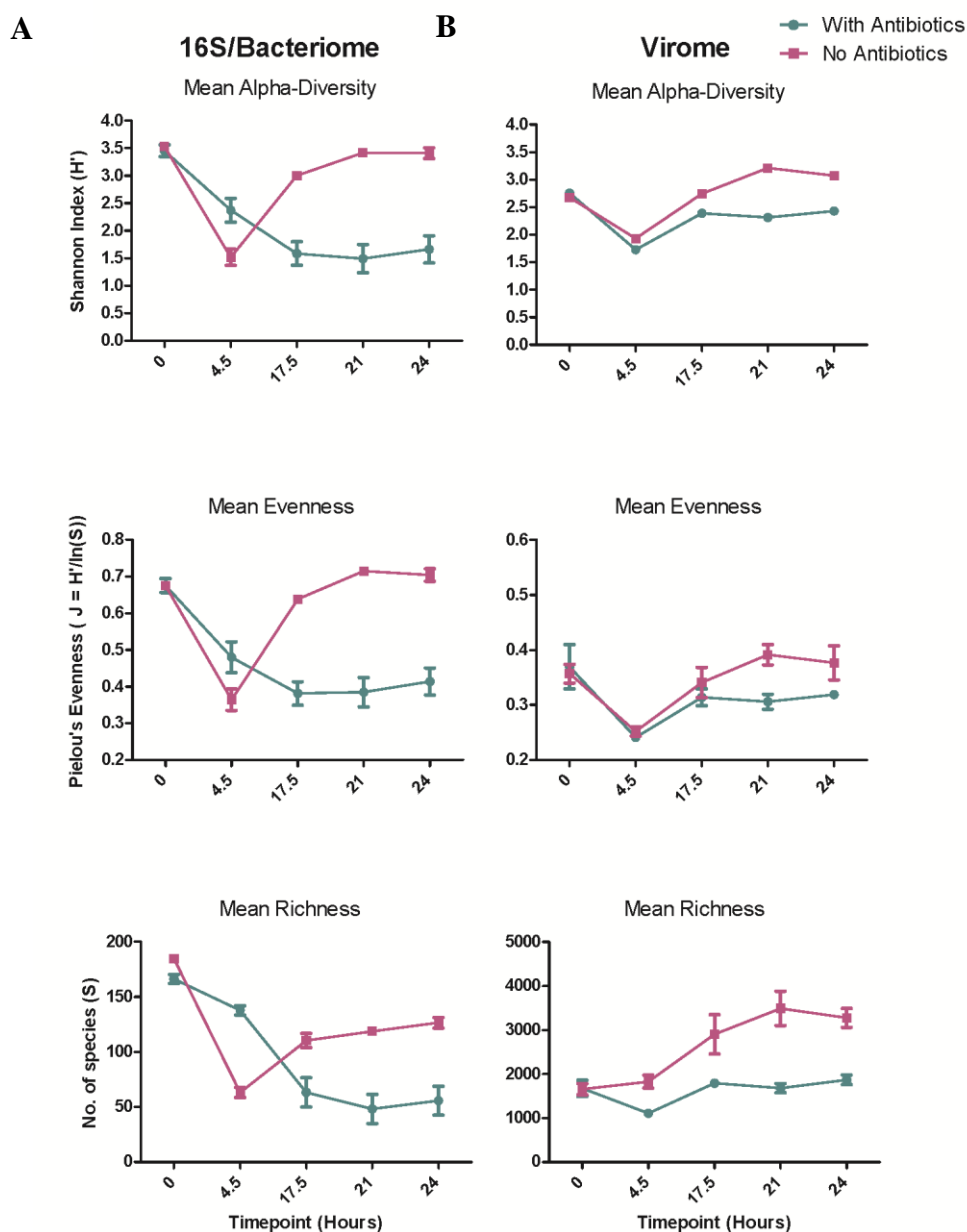


Figure 7. The impact of Φ crAss001 and Φ crAss002 on host counts in a defined bacterial community. Continuous fermentations were performed in parallel with and without phage addition. Respective hosts were included; *B. xylanisolvans* (Φ crAss002) and *B. intestinalis* (Φ crAss001). Absolute quantification, by qPCR, was performed on total DNA. (A) Absolute quantification of the bacterial community structure without phage addition. (B) Equivalent graph showing the community structure in the presence of the phages. (C) The titre and propagation dynamic of Φ crAss002. (D) Equivalent graph for Φ crAss001. Error bars represent standard deviation ($n = 3$).

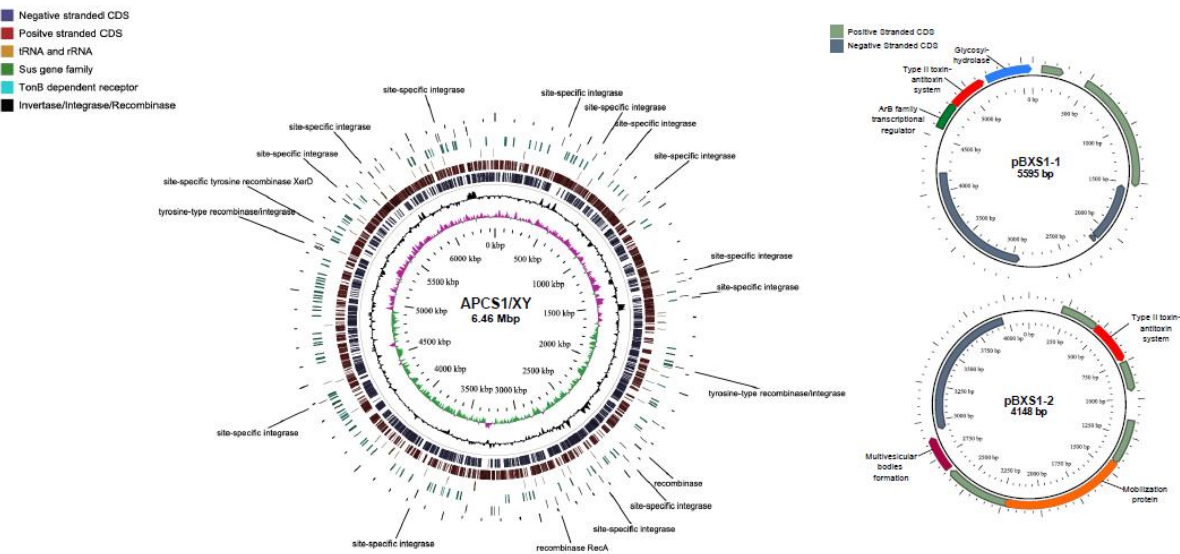
Supplementary Figures.

S1.

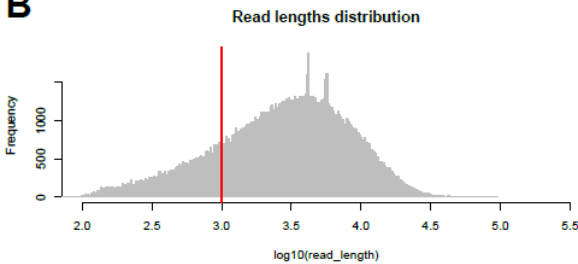


Supplementary Figure 1. Diversity index of fermentates generated with and without selective conditions. (A) 16S diversity index. In the presence of antibiotics alpha-diversity, evenness and richness are decreased. (B) Virome diversity index. Under selective enrichment there is a reduction for each index in parallel with bacteriome reduction. Error bars indicate standard deviation between triplicate fermentations ($n = 3$).

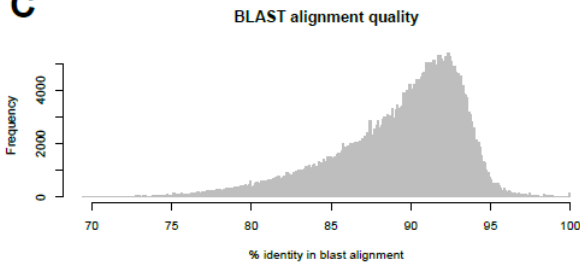
A



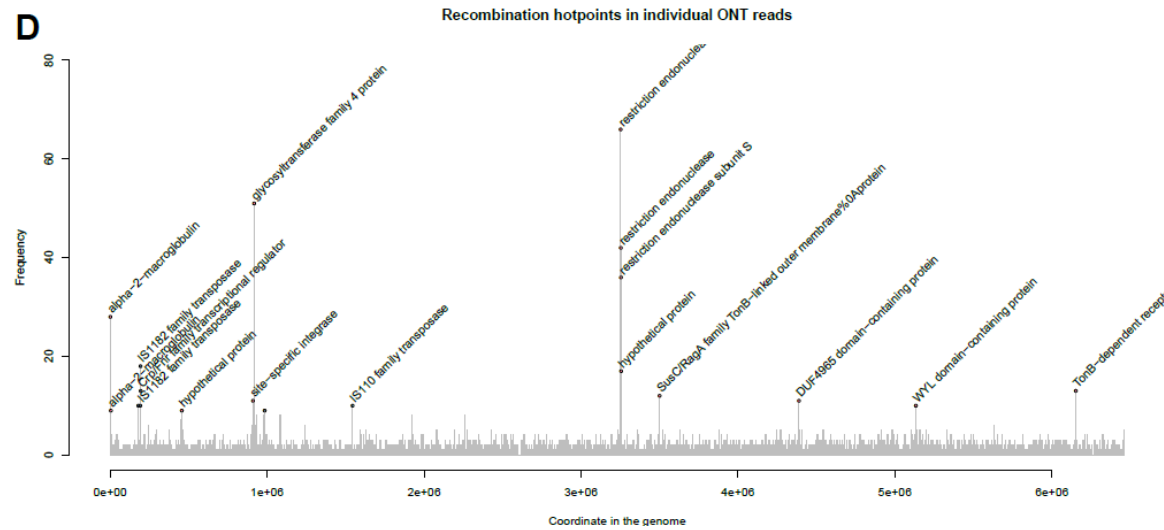
B



C



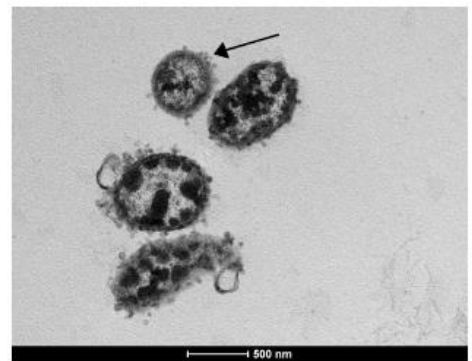
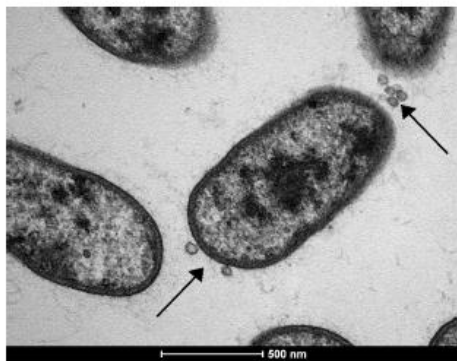
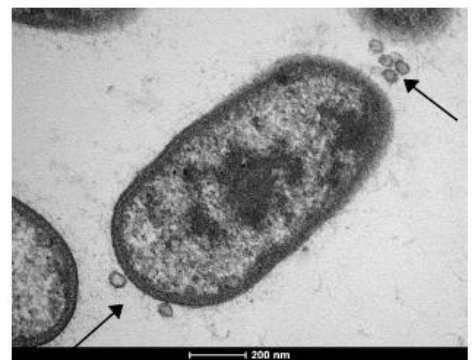
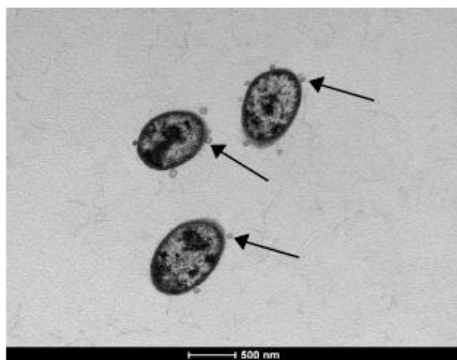
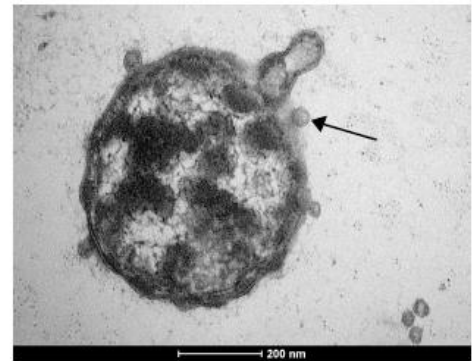
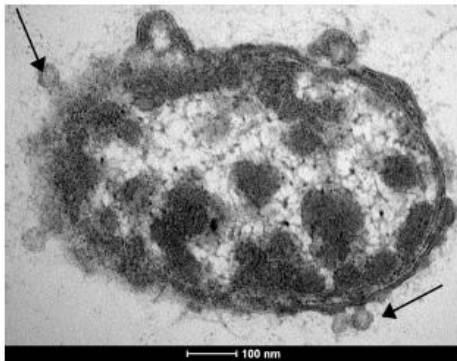
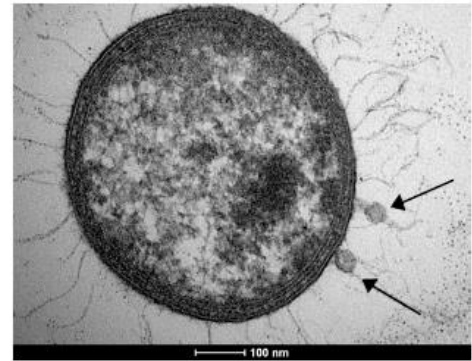
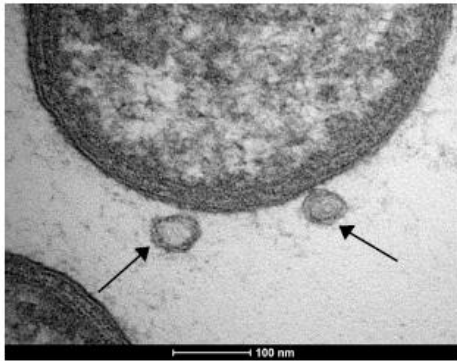
D



Supplementary Figure 2. Multiple site-specific recombinase-encoding genes and evidence of dynamic recombinations in the genome of *B. xylanisolvens* APCS1/XY.

(A) Circular map of genome (the innermost circle [green and purple], GC skew; circle two [black], relative G+C content; circles three and four [red and dark blue], open reading frames identified on the positive and negative DNA strands respectively; circle five [orange], tRNA and rRNA genes. circle six, genes annotated as Sus-like surface-associated glycan utilisation proteins [green] and TonB-dependent nutrient receptor [light blue]; circle seven [black], genes annotated as invertases, integrases and recombinases). To the right of the genome map, circular maps of the two associated circular plasmids are shown; pBXS1-1 and pBXS1-2. Annotated features are coloured and labelled; (B) Distribution of length of Oxford Nanopore sequencing reads used for dynamic genome recombination analysis; (C) Distribution of percentage identity in Oxford Nanopore reads aligned using BLASTn to the chromosome scaffold; (D) Frequency of detected recombinations at a single read level (reads of at least 1000nt, with individual alignments of >90% identity and >200nt length; all inversions or shifts in coordinates >200 nt were deemed as recombinations) versus coordinates in the chromosome scaffold (histogram bin size = 1000bp); recombination hotspots were identified when >8 reads with inconsistent alignment were present per 1000bp bin; gene products overlapping with hotspots are marked on the plot.

S3.



Supplementary Figure 3. Transmission electron micrographs showing vesicle-like structures on the surface of *B. xylanisolvens* APCS1/XY cells. Micrographs were prepared from cross-sections of soft agar collected lawns of *B. xylanisolvens* APCS1/XY with and without spotting of Φ crAss002 lysates.

3.7 Tables

Table 1. Primer sequences specific to each crAss-like phage detected following sequencing of subject ID: 924 faeces post fermentation.

Primer	Sequence 5' to 3'	Target	Size (bp)
^aCGI_Fwd	GCTAGAACATATCAAGC CAC	Terminase	244
CGI_Rev	GCAGTAGCTCCATGATT G		
CGIV_Fwd	GTTATATGGAAGCTATTG GTTCTGC	Terminase	186
CGIV_Rev	CTAGCATCAATCTTAGCT ATACCTC		
CGV_Fwd	GTACGGGTGGTACAAAA GGTG	Terminase	285
CGV_Rev	GACTGTTAGCACGTTGAC CTAC		
CGVI_Fferm8_F wd	CCTCAAGAAGTCCAGGA TCAAC	Terminase	180
CGVI_Fferm8_R ev	GGTAATGTAAGGCAGTA GGTCTG		
CG VI _Fferm9_Fwd	GGAGTTCGGTGCTTTCAA TAAATTC	Terminase	200
CGVI_Fferm9_R ev	GTAAATAGTGCGCTTGGC AATAG		
CGVI_Fferm11_ Fwd	GACTTATGGACTGACAA GATTGTTG	Terminase	250

CGVI_Fferm11_Rev	GTTAGTACCCTTAGCCTT ATTGC		
CGVII_Fwd	GACTCAGTTACTAAGTGG TATGAAGAG	Primase	189
CGVII_Rev	CATAAGGATTAACCTTGG CATCTC		

CG = candidate genus; Fwd = forward primer; Rev = reverse primer

Table 2. Primers specific to the SIHUMI bacterial community, Φ crAss001 host *B. intestinalis* 919/174, and Φ crAss002 host *B. xylanisolvens* APCS1/XY.

Primer	Sequence 5' to 3'	Product Size (bp)
E.coli_LF82_Fwd	CGGGTGTGTGTCCTAACTGCT	107
E.coli_LF82_Rev	CGAGTGGTCATTGGCCTCAT	
E.faecalis_OG1RF_Fwd	ACGGAGATTGTACGCTTAGT	122
E.faecalis_OG1RF_Rev	TCGGCATTATCTGGGTGGTC	
R.gnavus_ATCC29149_Fwd	GCCTGAACAGTTGCTTTCGG	115
R.gnavus_ATCC29149_Rev	GCGTGCTTGTATTCCGGATG	
F.prau_A2-165_Fwd	TGGATAAGAAACCGGGTCGC	94
F.prau_A2-165_Rev	ACGGACACAGCGATTTCCTT	
L.plant_WCFS1_Fwd	AATGTGGCAAGCATGGAAGC	94
L.plant_WCFS1_Rev	TTCATCCTCTCCGTCGGTCT	
B.longum_ATCC15707_Fwd	ACGCGAAGAACCTTACCTGG	120
B.longum_ATCC15707_Rev	CCCAACATCTCACGACACGA	
B.xylan_APCS1/XY_Fwd	ACAACTCCTCACTGAATCCCGCAT TTATCC	165
B.xylan_APCS1/XY_Rev	GCCACATCGGCAGATTATGACAAG ACAAC	
B.intest_919/174_Fwd	GAAGCAGGAGAAGCCACACCCA	263
B.intest_919/174_Rev	TTGACGAACGGGTCGGCTGT	

Table 3.

Bacterial isolates identified following Sanger sequencing. The isolates were enriched from subject ID: 924 faeces with the aid of antibiotic selective enrichment to promote Bacteroidales growth. FAA; Fastidious anaerobic agar, YCFA; yeast extract, casitone, fatty acids agar, CBA; Columbia blood agar.

Strain Code	Size (bp)	Identity (%)	Coverage (%)	Top Hit Species	Media (+ vancomycin and kanamycin)
FAA					
FFA S1	968	99	100	<i>B.uniformis</i>	FAA
FFA S2	1186	99	99	<i>B.uniformis</i>	FAA
FAA S3	1120	99	100	<i>B.uniformis</i>	FAA
FFA S4	1105	99	100	<i>B.uniformis</i>	FAA
FFA S5	1122	99	100	<i>B.ovatus</i>	FAA
FAA S6	>1000	99	100	<i>B.dorei</i>	FAA
FFA S7	1279	99	99	<i>B.uniformis</i>	FAA
FAA S8	1115	98	99	<i>P.distasonis</i>	FAA
FAA B1	673	95	100	<i>P.distasonis</i>	FAA
FAA B2	977	97	99	<i>B.uniformis</i>	FAA
FAA B3	1071	95	96	<i>P.distasonis</i>	FAA
FAA B4	947	97	99	<i>P.distasonis</i>	FAA
FAA B5	1000	98	99	<i>P.distasonis</i>	FAA
FAA B6	1138	97	100	<i>P.distasonis</i>	FAA
FAA B7	1057	99	97	<i>P.distasonis</i>	FAA
FAA B8	1111	98	100	<i>P.distasonis</i>	FAA
YCFA					
YCFA S1	776	99	100	<i>B.uniformis</i>	YCFA
YCFA S2	1101	99	99	<i>B.uniformis</i>	YCFA
YCFA S3	1171	91	100	<i>B.uniformis</i>	YCFA
YCFA S4	1109	99	99	<i>B.uniformis</i>	YCFA
YCFA S5	965	99	100	<i>B.uniformis</i>	YCFA
YCFA S6	1087	99	98	<i>B.uniformis</i>	YCFA
YCFA S7	1195	99	100	<i>B.uniformis</i>	YCFA
YCFA S8	1107	99	100	<i>B.uniformis</i>	YCFA
YCFA B1	901	98	100	<i>B.uniformis</i>	YCFA
YCFA B2	658	99	99	<i>B.dorei</i>	YCFA
YCFA B3	1081	99	100	<i>B.uniformis</i>	YCFA
YCFA B4	906	99	99	<i>P.distasonis</i>	YCFA
YCFA B5	895	99	100	<i>B.uniformis</i>	YCFA
YCFA B6	698	99	100	<i>B.uniformis</i>	YCFA
YCFA B7	742	99	100	<i>B.uniformis</i>	YCFA
YCFA B8	884	99	100	<i>B.uniformis</i>	YCFA

CBA					
CBA S1	800	99	100	<i>B.xylanisolvens</i>	CBA (horse blood & vitK)
CBA S2	936	96	100	<i>P.distasonis</i>	CBA (horse blood & vitK)
CBA S3	1000	99	100	<i>B.uniformis</i>	CBA (horse blood & vitK)
CBA S4	701	99	100	<i>P.distasonis</i>	CBA (horse blood & vitK)
CBA S5	917	99	100	<i>B.uniformis</i>	CBA (horse blood & vitK)
CBA S6	938	98	99	<i>P.distasonis</i>	CBA (horse blood & vitK)
CBA S7	574	99	100	<i>B.uniformis</i>	CBA (horse blood & vitK)
CBA S8	790	99	100	<i>B.uniformis</i>	CBA (horse blood & vitK)
CBA B1	893	99	100	<i>B.uniformis</i>	CBA (horse blood & vitK)
CBA B2	>1000	99	99	<i>B.fragilis</i>	CBA (horse blood & vitK)
CBA B3	647	93	99	<i>B.dorei</i>	CBA (horse blood & vitK)
CBA B4	351	94	100	<i>P.distasonis</i>	CBA (horse blood & vitK)
CBA B5	173	90	98	<i>B.dorei</i>	CBA (horse blood & vitK)
CBA B6	394	81	100	<i>B.dorei</i>	CBA (horse blood & vitK)
CBA B7	922	99	100	<i>B.dorei</i>	CBA (horse blood & vitK)
CBA B8	481	82	100	<i>B.dorei</i>	CBA (horse blood & vitK)

Table 4. (A) HHPred and (B) BLASTp functional annotation of Φ crAss002 protein-coding genes.

A.

LOC/CDS Name	Protein Length (aa)	Hit	Probability	E-value	Function
LOC_1_1	213	Unknown	Unknown	Unknown	Unknown
LOC_1_2	181	cd13833	89.55	1.3	HU_IHF_like/DNA-binding protein HU-beta/Topoisomerase II subunit
LOC_1_3	239	Unknown	Unknown	Unknown	Unknown
LOC_1_4	332	5ZYU_A	99.78	5.40E-20	Mitochondrial genome maintenance exonuclease X2, CRISPR-associated exonuclease Csa1
LOC_1_5	204	Unknown	Unknown	Unknown	Unknown
LOC_1_6	111	PF00430.18	78.76	26	ATP synthase
LOC_1_7	121	N/+C8:K8A	Unknown	Unknown	Unknown
LOC_1_8	765	3CPE_A	100	1.80E-33	Terminase (large subunit)
LOC_1_9	88	PF10043.9	98.82	4.30E-10	Predicted periplasmic lipoprotein (DUF2279)/VanZ like family/Unknown function (DUF2585)
LOC_1_10	810	PF16510.5	100	7.00E-41	Portal protein
LOC_1_11	436	Unknown	Unknown	Unknown	Unknown
LOC_1_12	491	PF17236.2	99.84	3.80E-23	Major Capsid Protein
LOC_1_13	384	Unknown	Unknown	Unknown	Unknown
LOC_1_14	217	2BSQ_G	56.01	74	Trafficking protein
LOC_1_15	293	Unknown	Unknown	Unknown	Unknown
LOC_1_16	196	PF10960.8	65.59	90	Holin (BhIA)

LOC_1_17	72	Unknown	Unknown	Unknown	Unknown
LOC_1_18	322	Unknown	Unknown	Unknown	Unknown
LOC_1_19	1153	3ZK4_B	100	2.60E-37	Phosphatase (purple acid)
LOC_1_20	2622	5JDA_A	98.36	1.20E-06	Transferase/Internalin-J; leucine rich repeat
LOC_1_21	106	Unknown	Unknown	Unknown	Unknown
LOC_1_22	1197	3VTM_A	67.45	34	Iron-regulated surface determinant protein
LOC_1_23	103	Unknown	Unknown	Unknown	Unknown
LOC_1_24	57	Unknown	Unknown	Unknown	Unknown
LOC_1_25	118	Unknown	Unknown	Unknown	Unknown
LOC_1_26	230	Unknown	Unknown	Unknown	Unknown
LOC_1_27	323	Unknown	Unknown	Unknown	Unknown
LOC_1_28	243	Unknown	Unknown	Unknown	Unknown
LOC_1_29	1287	Unknown	Unknown	Unknown	Unknown
LOC_1_30	557	Unknown	Unknown	Unknown	Unknown
LOC_1_31	922	Unknown	Unknown	Unknown	Unknown
LOC_1_32	2020	4QHJ_A	90.48	2.1	Peptidase
LOC_1_33	4245	5IJO_Y	44.14	300	Nuclear pore complex
LOC_1_34	71	Unknown	Unknown	Unknown	Unknown
LOC_1_35	61	Unknown	Unknown	Unknown	Unknown
LOC_1_36	38	Unknown	Unknown	Unknown	Unknown
LOC_1_37	98	Unknown	Unknown	Unknown	Unknown
LOC_1_38	150	PF13876.6	100	1.90E-32	Phage gp49_66
LOC_1_39	218	3C2T_A	100	1.30E-32	Deoxyuridine triphosphatase (E.C.3.6.1.23)
LOC_1_40	64	Unknown	Unknown	Unknown	Unknown
LOC_1_41	81	Unknown	Unknown	Unknown	Unknown
LOC_1_42	50	PF13719.6	98.39	1.70E-09	Zinc finger

LOC_1_43	362	TIGR03299	100	2.10E-57	Phage/plasmid-like protein/Unknown
LOC_1_44	35	cd14416	94.73	0.04	UBA-like domain found in nascent polypeptide-associated complex
LOC_1_45	257	PF05565.11	99.9	1.00E-25	Siphovirus Gp157/Mu Gam like protein/Host-nuclease inhibitor protein
LOC_1_46	60	Unknown	Unknown	Unknown	Unknown
LOC_1_47	89	Unknown	Unknown	Unknown	Unknown
LOC_1_48	101	Unknown	Unknown	Unknown	Unknown
LOC_1_49	58	Unknown	Unknown	Unknown	Unknown
LOC_1_50	99	Unknown	Unknown	Unknown	Unknown
LOC_1_51	111	1U3E_M	99.89	1.10E-25	HNH endonuclease
LOC_1_52	92	Unknown	Unknown	Unknown	Unknown
LOC_1_53	117	Unknown	Unknown	Unknown	Unknown
LOC_1_54	421	6GDR_A	100	1.10E-38	DNA ligase
LOC_1_55	160	Unknown	Unknown	Unknown	Unknown
LOC_1_56	179	Unknown	Unknown	Unknown	Unknown
LOC_1_57	377	TIGR01391	100	7.40E-33	DNA primase
LOC_1_58	203	1U3E_M	100	4.30E-32	Zinc-binding loop region of homing endonuclease/ HNH endonuclease
LOC_1_59	347	Unknown	Unknown	Unknown	Unknown
LOC_1_60	251	Unknown	Unknown	Unknown	Unknown
LOC_1_61	185	1U3E_M	100	4.90E-32	HNH catalytic motif/HNH endonuclease
LOC_1_62	306	TIGR01913	99.43	8.60E-15	Recombination protein Bet/RecT recombinase family
LOC_1_63	488	2FWR_C	100	6.50E-34	DNA repair protein/UnknownTP-dependent DNA helicase/DNA phosphorothioation system restriction enzyme

LOC_1_64	77	Unknown	Unknown	Unknown	Unknown
LOC_1_65	79	3DOA_A	77.42	7.7	Fibrinogen binding protein
LOC_1_66	753	2GV9_B	100	2.00E-61	DNA polymerase
LOC_1_67	485	3UPU_B	100	2.60E-36	ATP-dependent DNA helicase
LOC_1_68	121	Unknown	Unknown	Unknown	Unknown
LOC_1_69	83	Unknown	Unknown	Unknown	Unknown
LOC_1_70	103	5JBH_Y	96.41	3.30E-04	30S ribosomal protein/Zinc ribbon domain
LOC_1_71	48	Unknown	Unknown	Unknown	Unknown
LOC_1_72	113	Unknown	Unknown	Unknown	Unknown
LOC_1_73	176	4H9Q_C	75.51	21	DNA binding protein
LOC_1_74	127	Unknown	Unknown	Unknown	Unknown
LOC_1_75	50	Unknown	Unknown	Unknown	Unknown
LOC_1_76	314	6GAJ_C	95.48	0.027	Outer capsid protein sigma-1 (cell attachment protein)
LOC_1_77	60	3DUZ_A	35.09	45	Major envelope glycoprotein;
LOC_1_78	67	Unknown	Unknown	Unknown	Unknown
LOC_1_79	59	PF06698.11	87.36	2.3	Leucine zipper
LOC_1_80	230	Unknown	Unknown	Unknown	Unknown
LOC_1_81	60	Unknown	Unknown	Unknown	Unknown

B.

LOC/CDS Name	Protein Length (aa)	Best Hit	Max Score	Query Cover	E-value	% Identity	Accession
LOC_1_15	293	Putative tail sheath protein [CrAssphage sp.]	134	96%	1.00E-33	34.83%	AXF52196.1
LOC_1_19	1153	Metallophosphoesterase	361	32%	1.00E-110	48.31%	CCZ13600.1
LOC_1_22	1197	Hypothetical protein [Bacteroides xylanisolvens]	318	66%	3.00E-89	33.33%	WP_087318698.1
LOC_1_26	230	Putative Integration host factor IHF	234	97%	2.00E-74	49.57%	YP_009052530.1
LOC_1_28	243	Putative tail tubular protein [CrAssphage sp.]	173	96%	5.00E-50	40.39%	AXF52199.1
LOC_1_29	1287	Putative Phage stabilization protein, P22_gp10 homolog	224	84%	2.00E-55	28.02%	YP_009052527.1
LOC_1_30	557	Putative transmembrane protein [Bacteroides phage crAss001]	106	13%	3.00E-20	68.42%	AXQ62688.1
LOC_1_32	2020	Putative RNAP catalytic subunit	219	34%	5.00E-53	30.13%	AXF52200.1
LOC_1_33	4245	Putative RNAP catalytic subunit	441	71%	1.00E-119	29.18%	YP_009052522.2
LOC_1_59	347	Putative ssb single stranded DNA-binding protein	198	100%	3.00E-57	38.78%	YP_009052501.1

Table 5. Metadata associated with Φ crAss002 and related phages of candidate genus IV identified by *in silico* analyses.

	Length, bp	G+C, mol%	Topology	Health status	Location
srr4295172_ms_6	96564	31.92	Linear	Healthy	USA
srr4295173_ms_15	96117	31.94	Linear	Healthy	USA
err844065_ms_3	95257	32.14	Linear	IBD	USA
err844021_ms_1	94809	32.02	Linear	Healthy	USA
eld181-t3_s_1	97503	32.02	Linear	Elderly	Ireland
cs_ms_45	71412	32.49	Linear	Healthy	Ireland
err844056_ms_3	97150	32.14	Circular	IBD	USA
hvcf_d5_ms_4	92723	32.09	Circular	Healthy	Ireland
err844044_ms_2	96498	31.96	Linear	Healthy	USA
err844065_ms_4	94590	31.92	Linear	IBD	USA
cdzn01024782	93052	32	Linear	Healthy	Canada
srr073436_s_2	95598	32.14	Circular	Healthy	USA
err844016_ms_2	95883	32.29	Linear	Healthy	USA
Φ crAss002	93030	31.92	Circular	Healthy	Ireland
fferm_ms_10	101844	32.09	Linear	Healthy	Ireland
srr4295175_ms_5	96082	32.07	Circular	Healthy	USA
hvcf_e12_ms_1	94697	32.3	Linear	Healthy	Ireland
err843931_ms_1	94475	32.19	Circular	IBD	USA
hvcf_b3_ms_1	72230	32.3	Linear	Healthy	Ireland

Table 6. Summary of Φ crAss001 and Φ crAss002 characteristics

	Φ crAss001	Φ crAss002
Isolation source	Faeces	Faeces
Genome size	102kb	93kb
Host	<i>B.intestinalis</i> 919/174	<i>B.xylanisolvens</i> APCS1/XY
Preferred propagation method	Standard or continuous co- culture	Continuous co- culture
Clearing of culture	No	No
Spot formation	Yes	Yes
Plaque formation	Yes	No
Lifestyle	Persistence/Co- existence with host	Persistence/Co- existence with host
Shared genes of interest	RNAP associated genes,	RNAP associated genes
Unique genes of interest	Extensive tail- associated proteins	Two large genes: LOC_19 and LOC_20, specific function unknown
Lysogenic associated genes	No	No
tRNA genes	25	None
DNA packaging	Short direct terminal repeats	Pac-type headful

3.8 References

- Adriaenssens, E., and Brister, J.R. (2017). How to name and classify your phage: an informal guide. *Viruses* 9, 70.
- Aggarwala, V., Liang, G., and Bushman, F.D. (2017). Viral communities of the human gut: metagenomic analysis of composition and dynamics. *Mobile DNA* 8, 12.
- Allard, G., Ryan, F.J., Jeffery, I.B., and Claesson, M.J. (2015). SPINGO: a rapid species-classifier for microbial amplicon sequences. *BMC Bioinformatics* 16, 324.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- Antipov, D., Korobeynikov, A., McLean, J.S., and Pevzner, P.A. (2016). hybridSPAdes: an algorithm for hybrid assembly of short and long reads. *Bioinformatics* 32, 1009–1015.
- Bardina, C., Colom, J., Spricigo, D.A., Otero, J., Sánchez-Osuna, M., Cortés, P., and Llagostera, M. (2016). Genomics of Three New Bacteriophages Useful in the Biocontrol of Salmonella. *Front. Microbiol.* 7.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Breitbart, M., and Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* 13, 278–284.

Breitbart, M., Bonnain, C., Malki, K., and Sawaya, N.A. (2018). Phage puppet masters of the marine microbial realm. *Nat Microbiol* 3, 754–766.

Brugiroux, S., Beutler, M., Pfann, C., Garzetti, D., Ruscheweyh, H.-J., Ring, D., Diehl, M., Herp, S., Lötscher, Y., Hussain, S., et al. (2016). Genome-guided design of a defined mouse microbiota that confers colonization resistance against *Salmonella enterica* serovar Typhimurium. *Nat Microbiol* 2, 16215.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583.

Carding, S.R., Davis, N., and Hoyles, L. (2017). Review article: the human intestinal virome in health and disease. *Alimentary Pharmacology & Therapeutics* 46, 800–815.

Cervantes-Echeverría, M., Equihua-Medina, E., Cornejo-Granados, F., Hernández-Reyna, A., Sánchez, F., López-Contreras, B.E., Canizales-Quinteros, S., and Ochoa-Leyva, A. (2018). Whole-genome of Mexican-crAssphage isolated from the human gut microbiome. *BMC Research Notes* 11, 902.

Chassard, C., Delmas, E., Lawson, P.A., and Bernalier-Donadille, A. (2008). *Bacteroides xylanisolvens* sp. nov., a xylan-degrading bacterium isolated from human faeces. *International Journal of Systematic and Evolutionary Microbiology* 58, 1008–1013.

Cieplak, T., Soffer, N., Sulakvelidze, A., and Nielsen, D.S. (2018). A bacteriophage cocktail targeting *Escherichia coli* reduces *E. coli* in simulated gut conditions, while preserving a non-targeted representative commensal normal microbiota. *Gut Microbes* 9, 391–399.

Cinek, O., Mazankova, K., Kramna, L., Odeh, R., Alassaf, A., Ibekwe, M.U., Ahmadov, G., Mekki, H., Abdullah, M.A., Elmahi, B.M.E., et al. (2018). Quantitative CrAssphage real-time PCR assay derived from data of multiple geographically distant populations. *J. Med. Virol.* 90, 767–771.

Clooney, A.G., Sutton, T.D.S., Shkoporov, A.N., Holohan, R.K., Daly, K.M., O'Regan, O., Ryan, F.J., Draper, L.A., Plevy, S.E., Ross, R.P., et al. (2019). Whole-Virome Analysis Sheds Light on Viral Dark Matter in Inflammatory Bowel Disease. *Cell Host & Microbe* 26, 764-778.e5.

Dion, M.B., Oechslin, F., and Moineau, S. (2020). Phage diversity, genomics and phylogeny. *Nature Reviews Microbiology* 18, 125–138.

Draper, L.A., Ryan, F.J., Smith, M.K., Jalanka, J., Mattila, E., Arkkila, P.A., Ross, R.P., Satokari, R., and Hill, C. (2018). Long-term colonisation with donor bacteriophages following successful faecal microbial transplantation. *Microbiome* 6, 220.

Dutilh, B.E., Cassman, N., McNair, K., Sanchez, S.E., Silva, G.G.Z., Boling, L., Barr, J.J., Speth, D.R., Seguritan, V., Aziz, R.K., et al. (2014). A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat Commun* 5, 4498.

Edwards, R.A., Vega, A.A., Norman, H.M., Ohaeri, M., Levi, K., Dinsdale, E.A., Cinek, O., Aziz, R.K., McNair, K., Barr, J.J., et al. (2019). Global phylogeography and ancient evolution of the widespread human gut virus crAssphage. *Nat Microbiol* 1–10.

Edwards, U., Rogall, T., Blöcker, H., Emde, M., and Böttger, E.C. (1989). Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res.* *17*, 7843–7853.

Eun, C.S., Mishima, Y., Wohlgemuth, S., Liu, B., Bower, M., Carroll, I.M., and Sartor, R.B. (2014). Induction of Bacterial Antigen-Specific Colitis by a Simplified Human Microbiota Consortium in Gnotobiotic Interleukin-10^{-/-} Mice. *Infect Immun* *82*, 2239–2246.

European Commission (2015). Scientific Opinion on the safety of ‘heat-treated milk products fermented with *Bacteroides xylanisolvens* DSM 23964’ as a novel food. *EFSA Journal* *13*, 3956.

Galtier, M., De Sordi, L., Sivignon, A., de Vallée, A., Maura, D., Neut, C., Rahmouni, O., Wannerberger, K., Darfeuille-Michaud, A., Desreumaux, P., et al. (2017). Bacteriophages Targeting Adherent Invasive *Escherichia coli* Strains as a Promising New Treatment for Crohn’s Disease. *J Crohns Colitis* *11*, 840–847.

Garneau, J.R., Depardieu, F., Fortier, L.-C., Bikard, D., and Monot, M. (2017). PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. *Sci Rep* *7*, 8292.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* *5*, R80.

González-Tortuero, E., Sutton, T.D.S., Velayudhan, V., Shkoporov, A.N., Draper, L.A., Stockdale, S.R., Ross, R.P., and Hill, C. (2018). VIGA: a sensitive, precise and automatic de novo Viral Genome Annotator. *BioRxiv* 277509.

- Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize Implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812.
- Guerin, E., Shkoporov, A., Stockdale, S.R., Clooney, A.G., Ryan, F.J., Sutton, T.D.S., Draper, L.A., Gonzalez-Tortuero, E., Ross, R.P., and Hill, C. (2018). Biology and Taxonomy of crAss-like Bacteriophages, the Most Abundant Virus in the Human Gut. *Cell Host Microbe* 24, 653-664.e6.
- Harms, A., Brodersen, D.E., Mitarai, N., and Gerdes, K. (2018). Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. *Molecular Cell* 70, 768–784.
- d’Hérelle, F. (1917). Sur un microbe invisible antagoniste des bacilles dysentériques. *CR Acad. Sci. Paris* 165, 373–375.
- Honap, T.P., Sankaranarayanan, K., Schnorr, S.L., Ozga, A.T., Warinner, C., and Jr, C.M.L. (2020). Biogeographic study of human gut-associated crAssphage suggests impacts from industrialization and recent expansion. *PLOS ONE* 15, e0226930.
- Horino, A., Kenri, T., Sasaki, Y., Okamura, N., and Sasaki, T. (2009). Identification of a site-specific tyrosine recombinase that mediates promoter inversions of phase-variable *mpl* lipoprotein genes in *Mycoplasma penetrans*. *Microbiology*, 155, 1241–1249.
- Hryckowian, A.J., Merrill, B.D., Porter, N.T., Treuren, W.V., Nelson, E.J., Garlena, R.A., Russell, D.A., Martens, E.C., and Sonnenburg, J.L. (2020). Bacteroides thetaiotaomicron-infecting bacteriophage isolates inform sequence-based host range predictions. *BioRxiv* 2020.03.04.977157.

Hsu, B.B., Gibson, T.E., Yeliseyev, V., Liu, Q., Lyon, L., Bry, L., Silver, P.A., and Gerber, G.K. (2019). Dynamic Modulation of the Gut Microbiota and Metabolome by Bacteriophages in a Mouse Model. *Cell Host & Microbe* 25, 803-814.e5.

Jiang, X., Hall, A.B., Arthur, T.D., Plichta, D.R., Covington, C.T., Poyet, M., Crothers, J., Moses, P.L., Tolonen, A.C., Vlamakis, H., et al. (2019). Invertible promoters mediate bacterial phase variation, antibiotic resistance, and host adaptation in the gut. *Science* 363, 181–187.

Kieser, S., Sarker, S.A., Sakwinska, O., Foata, F., Sultana, S., Khan, Z., Islam, S., Porta, N., Combremont, S., Betrisey, B., et al. (2018). Bangladeshi children with acute diarrhoea show faecal microbiomes with increased *Streptococcus* abundance, irrespective of diarrhoea aetiology. *Environmental Microbiology* 20, 2256–2269.

Knowles, B., Silveira, C.B., Bailey, B.A., Barott, K., Cantu, V.A., Cobián-Güemes, A.G., Coutinho, F.H., Dinsdale, E.A., Felts, B., Furby, K.A., et al. (2016). Lytic to temperate switching of viral communities. *Nature* 531, 466–470.

Kulp, A., and Kuehn, M.J. (2010). Biological Functions and Biogenesis of Secreted Bacterial Outer Membrane Vesicles. *Annu Rev Microbiol* 64, 163–184.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.

Leighton Pritchard, Peter Cock, and YT (2017). widdowquinn/pyani v0.2.3 (Zenodo).

Lourenço, M., Chaffrignon, L., Lamy-Besnier, Q., Campagne, P., Eberl, C., Bérard, M., Stecher, B., Debarbieux, L., and Sordi, L.D. (2019). The spatial heterogeneity of

the gut limits bacteriophage predation leading to the coexistence of antagonist populations of bacteria and their viruses. *BioRxiv* 810705.

Ma, Y., You, X., Mai, G., Tokuyasu, T., and Liu, C. (2018). A human gut phage catalog correlates the gut phageome with type 2 diabetes. *Microbiome* 6, 24.

Manrique, P., Bolduc, B., Walk, S.T., van der Oost, J., de Vos, W.M., and Young, M.J. (2016). Healthy human gut phageome. *Proceedings of the National Academy of Sciences* 113, 10400–10405.

Manrique, P., Dills, M., and Young, M.J. (2017). The Human Gut Phage Community and Its Implications for Health and Disease. *Viruses* 9, 141.

Maura, D., Morello, E., Merle, L. du, Bomme, P., Bouguénec, C.L., and Debarbieux, L. (2012a). Intestinal colonization by enteroaggregative *Escherichia coli* supports long-term bacteriophage replication in mice. *Environmental Microbiology* 14, 1844–1854.

Maura, D., Galtier, M., Bouguénec, C.L., and Debarbieux, L. (2012b). Virulent Bacteriophages Can Target O104:H4 Enteroaggregative *Escherichia coli* in the Mouse Intestine. *Antimicrobial Agents and Chemotherapy* 56, 6235–6242.

McCann, A., Ryan, F.J., Stockdale, S.R., Dalmasso, M., Blake, T., Ryan, C.A., Stanton, C., Mills, S., Ross, P.R., and Hill, C. (2018). Viromes of one year old infants reveal the impact of birth mode on microbiome diversity. *PeerJ* 6, e4694.

Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2011). The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res.* 21, 1616–1625.

- Mirzaei, M.K., and Maurice, C.F. (2017). Ménage à trois in the human gut: interactions between host, bacteria and phages. *Nat. Rev. Microbiol.* *15*, 397–408.
- Monaco, C.L., Gootenberg, D.B., Zhao, G., Handley, S.A., Ghebremichael, M.S., Lim, E.S., Lankowski, A., Baldrige, M.T., Wilen, C.B., Flagg, M., et al. (2016). Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. *Cell Host & Microbe* *19*, 311–322.
- Moreno-Gallego, J.L., Chou, S.-P., Rienzi, S.C.D., Goodrich, J.K., Spector, T.D., Bell, J.T., Youngblut, N.D., Hewson, I., Reyes, A., and Ley, R.E. (2019). Virome Diversity Correlates with Intestinal Microbiome Diversity in Adult Monozygotic Twins. *Cell Host & Microbe* *25*, 261-272.e5.
- Nakayama-Imaohji, H., Hirakawa, H., Ichimura, M., Wakimoto, S., Kuhara, S., Hayashi, T., and Kuwahara, T. (2009). Identification of the Site-Specific DNA Invertase Responsible for the Phase Variation of SusC/SusD Family Outer Membrane Proteins in *Bacteroides fragilis*. *Journal of Bacteriology* *191*, 6003–6011.
- Nakayama-Imaohji, H., Hirota, K., Yamasaki, H., Yoneda, S., Nariya, H., Suzuki, M., Secher, T., Miyake, Y., Oswald, E., Hayashi, T., et al. (2016). DNA Inversion Regulates Outer Membrane Vesicle Production in *Bacteroides fragilis*. *PloS One* *11*, e0148887.
- Norman, J.M., Handley, S.A., Baldrige, M.T., Droit, L., Liu, C.Y., Keller, B.C., Kambal, A., Monaco, C.L., Zhao, G., and Fleshner, P. (2015). Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* *160*, 447–460.

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A., Korobeynikov, A., Lapidus, A., Prjibelsky, A., Pyshkin, A., Sirotkin, A., Sirotkin, Y., et al. (2013). Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. In *Research in Computational Molecular Biology*, M. Deng, R. Jiang, F. Sun, and X. Zhang, eds. (Springer Berlin Heidelberg), pp. 158–170.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Research* 27, 824–834.

O'Donnell, M.M., Rea, M.C., O'Sullivan, Ó., Flynn, C., Jones, B., McQuaid, A., Shanahan, F., and Ross, R.P. (2016). Preparation of a standardised faecal slurry for ex-vivo microbiota studies which reduces inter-individual donor bias. *Journal of Microbiological Methods* 129, 109–116.

Oude Munnink, B.B., Canuti, M., Deijis, M., de Vries, M., Jebbink, M.F., Rebers, S., Molenkamp, R., van Hemert, F.J., Chung, K., Cotten, M., et al. (2014). Unexplained diarrhoea in HIV-1 infected individuals. *BMC Infectious Diseases* 14, 22.

Petkau, A., Stuart-Edwards, M., Stothard, P., and Van Domselaar, G. (2010). Interactive microbial genome visualization with GView. *Bioinformatics* 26, 3125–3126.

Porter, N.T., Hryckowian, A.J., Merrill, B.D., Fuentes, J.J., Gardner, J.O., Glowacki, R.W.P., Singh, S., Crawford, R.D., Snitkin, E.S., Sonnenburg, J.L., et al. (2020). Multiple phase-variable mechanisms, including capsular polysaccharides, modify bacteriophage susceptibility in *Bacteroides thetaiotaomicron*. *BioRxiv* 521070.

Reyes, A., Haynes, M., Hanson, N., Angly, F.E., Heath, A.C., Rohwer, F., and Gordon, J.I. (2010). Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 466, 334–338.

Reyes, A., Wu, M., McNulty, N.P., Rohwer, F.L., and Gordon, J.I. (2013). Gnotobiotic mouse model of phage–bacterial host dynamics in the human gut. *PNAS* 110, 20236–20241.

Reyes, A., Blanton, L.V., Cao, S., Zhao, G., Manary, M., Trehan, I., Smith, M.I., Wang, D., Virgin, H.W., Rohwer, F., et al. (2015). Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *PNAS* 112, 11941–11946.

Reysenbach, A.-L., Longnecker, K., and Kirshtein, J. (2000). Novel Bacterial and Archaeal Lineages from an In Situ Growth Chamber Deployed at a Mid-Atlantic Ridge Hydrothermal Vent. *Appl Environ Microbiol* 66, 3798–3806.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual. *Molecular Cloning: A Laboratory Manual*.

Scanlan, P.D. (2017). Bacteria–Bacteriophage Coevolution in the Human Gut: Implications for Microbial Diversity and Functionality. *Trends in Microbiology* 25, 614–623.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., and Robinson, C.J. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* 75, 7537–7541.

Segata, N. (2018). On the Road to Strain-Resolved Comparative Metagenomics. *MSystems* 3.

Shkoporov, A.N., and Hill, C. (2019). Bacteriophages of the Human Gut: The “Known Unknown” of the Microbiome. *Cell Host & Microbe* 25, 195–209.

Shkoporov, A.N., Khokhlova, E.V., Fitzgerald, C.B., Stockdale, S.R., Draper, L.A., Ross, R.P., and Hill, C. (2018a). Φ CrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. *Nat Commun* 9.

Shkoporov, A.N., Ryan, F.J., Draper, L.A., Forde, A., Stockdale, S.R., Daly, K.M., McDonnell, S.A., Nolan, J.A., Sutton, T.D.S., Dalmasso, M., et al. (2018b). Reproducible protocols for metagenomic analysis of human faecal phageomes. *Microbiome* 6, 68.

Shkoporov, A.N., Clooney, A.G., Sutton, T.D.S., Ryan, F.J., Daly, K.M., Nolan, J.A., McDonnell, S.A., Khokhlova, E.V., Draper, L.A., Forde, A., et al. (2019). The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific. *Cell Host & Microbe* 26, 527-541.e5.

Silveira, C.B., and Rohwer, F.L. (2016). Piggyback-the-Winner in host-associated microbial communities. *Npj Biofilms and Microbiomes* 2, 16010.

Siranosian, B.A., Tamburini, F.B., Sherlock, G., and Bhatt, A.S. (2020). Acquisition, transmission and strain diversity of human gut-colonizing crAss-like phages. *Nat Commun* 11, 1–11.

- Siringan, P., Connerton, P.L., Cummings, N.J., and Connerton, I.F. (2014). Alternative bacteriophage life cycles: the carrier state of *Campylobacter jejuni*. *Open Biol* 4, 130200–130200.
- Somerville, V., Lutz, S., Schmid, M., Frei, D., Moser, A., Irmeler, S., Frey, J.E., and Ahrens, C.H. (2019). Long-read based de novo assembly of low-complexity metagenome samples results in finished genomes and reveals insights into strain diversity and an active phage system. *BMC Microbiol.* 19, 143.
- Sonnenburg, J.L., Xu, J., Leip, D.D., Chen, C.-H., Westover, B.P., Weatherford, J., Buhler, J.D., and Gordon, J.I. (2005). Glycan Foraging in Vivo by an Intestine-Adapted Bacterial Symbiont. *Science* 307, 1955–1959.
- Sordi, L.D., Khanna, V., and Debarbieux, L. (2017). The Gut Microbiota Facilitates Drifts in the Genetic Diversity and Infectivity of Bacterial Viruses. *Cell Host & Microbe* 22, 801-808.e3.
- Stachler, E., Kelty, C., Sivaganesan, M., Li, X., Bibby, K., and Shanks, O.C. (2017). Quantitative CrAssphage pcr assays for human fecal pollution measurement. *Environmental Science & Technology* 51, 9146–9154.
- Sutton, T.D.S., and Hill, C. (2019). Gut bacteriophage: Current understanding and challenges. *Front. Endocrinol.* 10.
- Toyofuku, M., Nomura, N., and Eberl, L. (2019). Types and origins of bacterial membrane vesicles. *Nature Reviews Microbiology* 17, 13–24.

Turkington, C.J.R., Morozov, A., Clokie, M.R.J., and Bayliss, C.D. (2019). Phage-Resistant Phase-Variant Sub-populations Mediate Herd Immunity Against Bacteriophage Invasion of Bacterial Meta-Populations. *Front. Microbiol.* 10.

Twort, F.W. (1915). An investigation on the nature of ultra-microscopic viruses. *The Lancet* 186, 1241–1243.

Ulsemer, P., Toutounian, K., Schmidt, J., Karsten, U., and Goletz, S. (2012). Preliminary Safety Evaluation of a New *Bacteroides xylanisolvens* Isolate. *Appl. Environ. Microbiol.* 78, 528–535.

Ulsemer, P., Toutounian, K., Kressel, G., Goletz, C., Schmidt, J., Karsten, U., Hahn, A., and Goletz, S. (2016). Impact of oral consumption of heat-treated *Bacteroides xylanisolvens* DSM 23964 on the level of natural TFA-specific antibodies in human adults. *Beneficial Microbes* 7, 485–500.

Warwick-Dugdale, J., Solonenko, N., Moore, K., Chittick, L., Gregory, A.C., Allen, M.J., Sullivan, M.B., and Temperton, B. (2019). Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands. *PeerJ* 7, e6800.

Weiss, M., Denou, E., Bruttin, A., Serra-Moreno, R., Dillmann, M.-L., and Brüssow, H. (2009). In vivo replication of T4 and T7 bacteriophages in germ-free mice colonized with *Escherichia coli*. *Virology* 393, 16–23.

Wilhelm, M.P. (1991). Vancomycin. *Mayo Clinic Proceedings* 66, 1165–1170.

Yutin, N., Makarova, K.S., Gussow, A.B., Krupovic, M., Segall, A., Edwards, R.A., and Koonin, E.V. (2018). Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. *Nat Microbiol* 3, 38–46.

Zaleski, P., Wojciechowski, M., and Piekarowicz, A. (2005). The role of Dam methylation in phase variation of *Haemophilus influenzae* genes involved in defence against phage infection. *Microbiology (Reading, Engl.)* 151, 3361–3369.

Zhao, G., Vatanen, T., Droit, L., Park, A., Kostic, A.D., Poon, T.W., Vlamakis, H., Siljander, H., Härkönen, T., Hämäläinen, A.-M., et al. (2017). Intestinal virome changes precede autoimmunity in type I diabetes-susceptible children. *Proc. Natl. Acad. Sci. U.S.A.* 114, E6166–E6175.

Zhao, G., Droit, L., Gilbert, M.H., Schiro, F.R., Didier, P.J., Si, X., Paredes, A., Handley, S.A., Virgin, H.W., Bohm, R.P., et al. (2019). Virome biogeography in the lower gastrointestinal tract of rhesus macaques with chronic diarrhea. *Virology* 527, 77–88.

Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A.N., and Alva, V. (2017). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology*.

Zitomersky, N.L., Coyne, M.J., and Comstock, L.E. (2011). Longitudinal analysis of the prevalence, maintenance, and IgA response to species of the order Bacteroidales in the human gut. *Infect. Immun.* 79, 2012–2020.

Chapter IV

Isolation and characterization of a novel *Parabacteroides distasonis* bacteriophage isolated from the human gut

Emma Guerin, Andrey Shkoporov, Stephen Stockdale, Paul Ross, Colin Hill

4.1 Summary

The human gut microbiome is thought to play a significant role in health and disease. The virome, predominantly composed of bacteriophages, has received significantly less attention in comparison to the bacteriome. This is largely due to the challenges associated with the isolation and characterization of novel gut bacteriophages, and bioinformatic challenges that include the lack of a universal bacteriophage marker gene and the absence of homologs in viral databases. Here we describe the isolation of a novel virulent siphovirus infecting *Parabacteroides distasonis*; Φ PDS1. Key to the isolation of this phage was the antibiotic driven selective enrichment of the bacterial host in a faecal fermenter that permitted parallel phage expansion. We also present the *in silico* and biological characterization of this phage. To date there have been no detailed reports of *Parabacteroides distasonis* siphoviruses or the genomes of such phages deposited in the NCBI Taxonomy database. Multiple studies have shown that perturbations of this gut commensal can be linked to various disease states, making this novel phage-host pair and their interactions of particular interest.

4.2 Introduction

Bacteriophages (phages), were discovered a century ago by Twort and d'Herelle (Twort, 1915; d'Hérelle, 1917). In recent years, these bacteria infecting viruses have become the subject of renewed attention in light of their potential role in the microbiome. Phages form the largest component of the human gut virome, also referred to as the gut phageome. These viruses could have important implications for health by shaping the composition of the microbiome through interactions with the gut bacterial community or indirectly by interactions with the human host immune system (Norman et al., 2015; Manrique et al., 2016; Nguyen et al., 2017; Belleghem et al., 2017; Gogokhia et al., 2019). Alterations in phageome composition have also been associated with multiple disease states (Reyes et al., 2015; Tetz and Tetz, 2018; Kieser et al., 2018; Ma et al., 2018; Zhao et al., 2019; Zuo et al., 2019; Khan Mirzaei et al., 2020). For example, in IBD patients there is shift from predominantly virulent phages in health towards induction of temperate phages (Clooney et al., 2019). However, little is known about the mechanisms by which phages might support homeostasis or help to resist perturbations of the gut microbiome.

Despite this revived interest in phages, the gut phageome remains largely unknown. Modern sequencing technology and viral metagenomics have played an important part in improving our insight into the composition of the phageome. However, bottlenecks associated with these methods include the absence of a universal marker gene and the lack of sequence homology with phages on currently available public databases the use of which can result in an incomplete analysis (Chibani et al., 2019; Roux et al., 2015; Shkoporov and Hill, 2019; Sutton and Hill, 2019). Most phage contigs or genomes cannot be assigned with any certainty to a specific taxonomic position, nor can their bacterial hosts be predicted with any accuracy. These

uncharacterized viral sequences have been termed the “viral dark matter” (Aggarwala et al., 2017). The isolation of novel gut phages can be hindered due to difficulty in culturing associated hosts, traditional screening methods involving plaquing can result in novel phage-host pairs being overlooked, mimicking gut conditions is difficult and bacterial host factors, such as genetic resistance and transient resistance due to phase variation can further heighten this challenge (Hyman, 2019; Porter et al., 2020).

The isolation of further human gut phage-host pairs permits the elucidation of novel biological properties and dynamics. Bacteroidales form the most abundant bacterial order in a healthy human gut, with *Bacteroides* and *Parabacteroides* being two of the most important genera (García-Bayona and Comstock, 2019). CRISPR-spacer sequence analyses suggests that a significant proportion of persistent virulent phages in the human virome infect hosts of the Bacteroidales order (Shkoporov et al., 2019). Bacteroidales infecting crAss-like phages, the most abundant phage family in the human gut, provide one of the most notable examples of how the isolation and characterisation of a phage originally detected *in silico* was key in gaining insight into biological properties (Dutilh et al., 2014; Shkoporov et al., 2018a).

Here we conducted an antibiotic enrichment of a faecal bacterial community in a fermenter system to promote the growth of Bacteroidales and parallel expansion of associated phages. We report the isolation a novel *Parabacteroides distasonis* phage; ΦPDS1. This phage is the first lytic *P. distasonis* infecting siphovirus to be biologically characterised and to have its genome deposited on NCBI Taxonomy. Isolation of novel gut phages in pure culture that interact with abundant members of the microbiome will play an important role in improving our understanding of phage-host interactions. Given that *P. distasonis* has been linked to human disease and health, characterisation of the phages with which it interacts takes on an added significance.

4.3 Experimental procedures

4.3.1 Donor recruitment and faecal fermentation

A healthy individual, denoted as subject ID:924, was recruited for faecal sample donation. Collection of samples from this individual was in accordance with the study protocol APC055 and ethics approved by Cork Research Ethics Committee. On receipt, the sample was processed immediately into frozen standard inoculum (FSI). This was done exactly as described by (O'Donnell et al., 2016) with a few minor modifications as outlined by (Guerin et al., 2018). The FSI was aliquoted into three volumes so the fermentations could be performed in triplicate. The prepared FSI is the same that was used to initiate a faecal fermentation investigating crAss-like phages (refer to chapter 3, methods 3.3.1). The fermentation was run in batch format over 24 hours with conditions as applied by (Guerin et al., 2018). This was implemented as in (chapter 3, methods 3.3.2). Two fermenter vessels were set up in parallel, one with and one without the addition of antibiotics to the YCFA-GSCM broth post-autoclaving. The antibiotics used were 7.5 µg/ml vancomycin and 100 µg/ml kanamycin chosen to selectively promote the growth of Bacteroidales via the elimination of gram-positive bacteria as well as facultative anaerobes. Samples were collected at 0, 4.5, 17.5, 21 and 24 hours. The samples were directly processed after collection through centrifugation at 4,700 rpm at +4°C for 10 minutes. Following this, supernatants were passed through a 0.45µm pore polyethersulfone (PES) membrane filter and stored +4°C. The remaining bacterial rich pellets were sorted at -80°C.

4.3.2 Extraction of virus-like particles and sequencing of fermenter virome

The extraction of virus-like particles (VLPs) from 10 ml of collected filtered fermentates, virome library preparation from VLPs, processing of the resultant viral reads via trimming and filtering followed by assembly into contigs was done exactly as described previously (Chapter 3; methods 3.3.3.).

4.3.3 Extraction of total DNA and 16S rRNA gene library preparation

Total DNA was extracted from faecal pellets generated following centrifugation of samples collected from the fermenter. The extractions were performed using the QIAamp Fast Stool Mini Kit (Qiagen, Hilden, Germany). The extraction protocol as well as 16S rRNA gene library preparation of the extracted DNA was implemented as described previously (Chapter 3; methods 3.3.4.).

4.3.4 Novel phage-host pair screening

Selective enrichment of anaerobic Bacteroidales species using Fastidious Anaerobe Agar (FAA), Columbia Blood Agar (CBA) and YCFA-Agar containing vancomycin and kanamycin, colony purification and identification via Sanger sequencing of the 16S rRNA region was performed as described previously (Chapter 3; methods 3.3.6.). To ensure the absence of aerobes, incubation of isolated colonies was performed aerobically and anaerobically. Following identification of the bacterial species, 48 pure cultures were used in the phage-host pair screening. Unlike the screening method implemented in the search for novel crAss-like phage-host pairs (chapter 3), no preliminary *in silico* analyses were performed to aid the screening

process such as via primer development to target sequences detected in the fermenter. The screening implemented here was more traditional via spot assays.

Overnight cultures of the Bacteroidales were prepared in 10 ml of Fastidious Anaerobe Broth (FAB, Neogen). From each of these cultures, 300 µl was added to 0.3% FAA agar overlay (0.3% agar w/v), with MgSO₄ and CaCl₂ added to give final concentration of 1 mM. This was followed by pouring of the mixture onto pre-prepared FAA base agar (1.5% agar w/v) after gentle vortexing. After drying of the overlay, 5 µl of the filtered fermentates collected from the antibiotic containing and antibiotic absent vessels at time point 21 were spotted onto the lawns of each culture and dried. In addition to this, the original faecal filtrate prepared from subject ID:924 faeces without enrichment was spotted on the lawns. The plates were incubated anaerobically at 37°C for 48 hours. Formed spots were picked using an inoculation loop, placed into 100 µl of SM buffer, vortexed and incubated at room temperature for 5 hours. Following this the resuspended spots were spun in a desktop centrifuge at maximum speed for 10 minutes. The resultant supernatant was filtered through a 0.45µm pore polyethersulfone (PES) membrane filter. Spot assays were then repeated using each lysate on lawns of the culture it originally spotted on. Plaque assays were also performed using 3 ml of 0.3% FAA agar (0.3% agar w/v), with the addition of MgSO₄ and CaCl₂ at a final concentration of 1 mM. Ten-fold serial dilutions of the lysates were prepared in SM buffer. To the molten agar, 300 µl of each overnight culture was added and 50 µl of phage dilution. Following this, spot picking and spot assays were repeated a third time. It was observed that spots formed consistently on all the cultures identified as *Parabacteroides distasonis*, determined by Sanger sequencing of the 16S rRNA region. Four *Parabacteroides distasonis* cultures, which

produced the least turbid zones of clearing were chosen for phage enrichment. These cultures were designated as FAA-B5, FAA-B8, FAA-S8 and CBA-S2 (Table 1).

4.3.5 Novel phage enrichment and shotgun sequencing

Six rounds of phage enrichment were performed with the *P. distasonis* cultures to increase titre of the detected phage. Phage filtrates was added to *P. distasonis* cultures at $OD_{600} = \sim 0.2$ which was achieved approximately 5 to 6 hours post sub-culturing of 100 μ l of overnight culture in 10ml of fresh FAB (plus $MgSO_4$ and $CaCl_2$ added to give final concentration of 1 mM) with anaerobic incubation at 37°C. Once the culture was at the desired optical density, 1ml of phage lysate was added at an unknown titre to each culture and incubated anaerobically overnight at 37°C. The lysates were then centrifuged at 4,700rpm and passed once through a 0.45 μ m membrane syringe filter. Filtrates were then stored at +4°C. Each round of propagation was performed following the same procedure. To ensure maintenance of phage throughout the serial propagations, spot assays were performed as described above. On completion of the enrichment spot and plaque assays were repeated to check for improved clarity of zones clearing.

Nucleic acids were extracted from 10 ml of filtered phage lysate collected on the final round of enrichment. Solid NaCl and PEG-8000 were added to the lysates give a final concentration of 0.5M and 10% w/v. Following this, the powders were completely dissolved, and the samples were placed at +4°C for overnight incubation. The following VLP purification and DNA extraction steps proceeding this were performed as described by (Shkoporov et al., 2018b).. The DNA was then directly subjected to random shotgun library preparation using Nextera XT DNA Library

Preparation Kit (Illumina) without preliminary multiple displacement amplification (MDA). Normalisation was performed as per the manufacturer's protocol using the bead-based method. The prepared libraries were sequenced using 2 x 300 bp paired-end chemistry on an Illumina MiSeq platform (Illumina, San Diego, California) at GATC Biotech AG, Germany. The quality of the raw reads was analysed using FastQC v0.11.5. Removal of Nextera adaptors was performed Trimmomatic v0.36. (Bolger et al., 2014). Parameters implemented were as follows: minimum length of 60, a sliding window size of 4 and a minimum Phread score of 33. The trimmed and filtered reads were then assembled into contigs using SPAdes (v1.13.1) (Antipov et al., 2016; Nurk et al., 2017).

4.3.6 *In silico* analyses and characterisation of Φ PDS1

Sequencing revealed that the novel phage, denoted as Φ PDS1, resolves into the Siphovirus family of the Caudovirales order. Prediction of protein coding genes and preliminary annotation of the Φ PDS1 genome was performed using *de novo* viral genome annotator VIGA (González-Tortuero et al., 2018). The predicted protein coding sequences then underwent manual functional analyses using HHPred to generate more detailed annotations. HHPred annotations were performed with the following databases: PDB_mm_CIF70_3_Aug, Pfam-A_v32.0, NCBI_CD_v3.16, TIGRFAMs_v15.0 (Zimmermann et al., 2017). A genomic map of the Φ PDS1 genome was then generated using GView (v1.7) with annotations incorporated (Petkau et al., 2010).

To analyse the incidence of Φ PDS1 in the gut phageome of different human cohorts, the phage genome was queried against an in-house database generated from

multiple human virome studies using BLASTn v2.2.28+ (Altschul et al., 1997). Hits used for further analyses met the following inclusion criteria: [1] a contig length of $\geq 40\text{kb}$ and [2] an E-value of $< 1\text{E-}05$. Metadata of the sixteen detected phages was examined to investigate possible cohort associations.

To investigate the relatedness of the sixteen *in silico* detected phages and ΦPDS1 , average nucleotide identity (ANI) was determined using the default settings in PYANI (Leighton Pritchard et al., 2017). The ANI output was exported into R environment and used to generate a complex heatmap with Bioconductor Complex Heatmap (v1.20.0) and Circlize (v0.4.5) packages (Gentleman et al., 2004; Gu et al., 2014, 2016). The topology of the contigs was determined and the sixteen ΦPDS1 -like phage genomes were annotated using VIGA (González-Tortuero et al., 2018). A highly similar, an intermediately related and a more distantly related phage of ΦPDS1 were chosen as representatives for a whole genome pairwise comparison to examine synteny using Easyfig (v2.2.2) by inputting the resultant GenBank files from VIGA. The EasyFig image was generated using BLASTn with a minimum length of 50bp and percentage identity of 30bp. In addition to this, the occurrence of ΦPDS1 associated CRISPR-spacers in bacterial genomes was analysed using an in-house database. This was also carried out for the sixteen ΦPDS1 -like phages identified *in silico* to examine if they also infect *P. distasonis* or other bacterial hosts.

4.3.7 Transmission electron microscopy

Ultra-centrifugation was performed using a 50 ml pool of ΦPDS1 lysate. The lysate was concentrated for a total of 3 hours at 120,000g using a F65L-6x13.5 rotor (Thermo Scientific). This was done using 25ml of lysate in the initial 1.5 hour spin

followed by removal of the supernatant. The remaining 25 ml was added to the resultant pellets without resuspension and a second round of ultra-centrifugation was then performed. Finally, the pellets were resuspended in a final volume of 5ml SM buffer. The suspensions were then applied onto to a step gradient of 5M and 3M CsCl solutions and centrifuged at 105,000g for 2.5 hours at +4°C. The steps following this were performed as described by Guerin et al. (Guerin et al., 2018). Five microlitre aliquots of the concentrated 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 Tecnai G2 12 BioTWIN transmission electron microscope

4.3.8 Quantitative real-time PCR: primer and standard development

Following sequencing and annotation of Φ PDS1, the phage portal protein (gene gp_48) was chosen as a suitable target for qPCR primer and standard development. Primers were designed using PERL Primer software and CLC Sequence Viewer 8.0. The following primers; Fwd 5'GGAACAACGGGACGATTG-3' and Rev 5'-CAATCACGGACGCAATAGG-3' to produce a product of 193bp using standard PCR with the following conditions: initial denaturation at 95°C for 5 minutes, then 35 cycles of 95°C for 20 seconds, 60°C for 20 seconds 72°C for 20 seconds and 72°C for 1 minute. To develop standards for qPCR calibration curves, the products generated from the above primers were cloned into pCR2.1-TOPO TA vector (Thermo Fisher Scientific). Extracted plasmids were quantified using Qubit dsDNA BR Assay kit and diluted to 10⁹ copies/μl based on molar mass of DNA. Ten-fold serial dilutions of the

plasmids were used to build a standard calibration curve. Following this, the absolute composition of Φ PDS1 in the faecal fermenters was performed via qPCR of the viral nucleic extracted from the filtered fermentates. This was performed in a 15 μ l reaction volume using SensiFAST SYBR No-ROX mastermix (Bioline) in a LightCycler 480 thermocycler with the following conditions: initial denaturation at 95°C for 5 minutes, then 45 cycles of 95°C for 20 seconds, 60°C for 20 seconds and 72°C for 20 seconds. Resulting Ct-values were converted to copies/ml based on the generated calibration curves. The relative abundance of Φ PDS1 was also calculated. This was expressed as the percentage of total viral reads that align to the Φ PDS1 genome. Results were visualised using package ggplot2 v2.2.

4.3.9 Biological characterisation of Φ PDS1

A one-step growth curve was performed in triplicate to determine the latent period and burst size of Φ PDS1. An early logarithmic phase culture of *P. distasonis* APCS2/PD was infected with Φ PDS1 with a multiplicity of infection (MOI) of \sim 1. This was followed by incubation at room temperature for 5 mins and centrifugation at 5000 rpm in a swing bucket rotor for 15 minutes at 20°C. The supernatant was removed, and the resultant pellet was resuspended with FAB. The phage-host pair were maintained at 37°C under anaerobic conditions for 3 hours with 1 ml sample collection every 15 mins. Samples were centrifuged and filtered through 0.45 μ M pore syringe filters. Analysis was performed using absolute qPCR with the standard curve method as described above.

Efficiency of lysogeny was performed by spreading 100 μ l of Φ PDS1 lysate on FAA agar plates. After drying, 3ml of 0.3% FAA molten overlay containing 10-

fold serial dilutions of *P. distasonis* APCS2/PD was poured onto the Φ PDS1 spread plates. Negative controls were prepared in the same manner but without the addition of phage to the base plate. The plates were incubated anaerobically at 37°C for 48 hours. Efficiency of lysogeny was calculated as the percentage of colonies on the phage seeded plates versus counts for the equivalent negative control. Colony morphology was observed, and sixteen colonies were picked and restreaked four times. Standard PCR was performed using primers specific to Φ PDS1 as described above to check for the potential integration of the phage into the host genome. Spot assays were performed on clones after restreaking to check for host resistance after phage exposure.

Co-culturing of Φ PDS1 *P. distasonis* APCS2/PD was performed by serial sub-culturing over ten days in triplicate. This was initiated using 1ml of Φ PDS1 lysate at 10^{10} copies/ml that was introduced into 10 ml of *P. distasonis* APCS2/PD culture in FAB at $OD_{600} = 0.2$. Subsequent rounds of sub-culturing were performed by introducing the prior co-culture into 10 ml fresh FAB at a ratio of 1:50. Φ PDS1 titre was quantified using qPCR. The bacterial pellet collected following generation of the final phage lysate collected on day ten was serially streaked three times and tested for resistance against Φ PDS1 by spot assay. The resultant colonies were also examined for presence of episomally replicating or integrated copies of the phage by standard PCR.

The ability of the Φ PDS1 to infect other commercially available *P. distasonis* strains DSM 29491 and DSM 20701 (ATCC 8503) and *Odoribacter splanchnicus* DSM 20712, following CRISPR-spacer analysis results, was investigated via liquid propagation using the same method as described above for phage enrichment.

Propagation of Φ PDS1 on *P. distasonis* APCS2/PD was performed in parallel as a control. Propagation was determined by examining titre increase using absolute qPCR.

4.3.10 Shotgun sequencing of *P. distasonis* APCS2/PD using Illumina and Oxford Nanopore platforms

For the generation of long and short read sequences for hybrid assembly, genomic DNA was extracted from 10 ml of *P. distasonis* APCS2/PD overnight culture using phenol/chloroform with precipitation in 3M sodium acetate and cold absolute ethanol (Sambrook et al., 1989; Bardina et al., 2016). DNA was quantified using Qubit BR DNA Assay Kit in a Qubit 3.0 Fluorometer. All proceeding steps were performed as described previously (Chapter 3; methods 3.3.8.) including library preparation, sequencing, read processing and assembly. Following assembly using SPAdes (v1.13.1), nine scaffolds greater than 1kb were generated (Nurk et al., 2013; Antipov et al., 2016). Scaffolds were then manually curated, joined and circularised using CLC Sequence Viewer. The assembled and circularised genome and associated plasmids were submitted to NCBI Prokaryotic Genome Annotation Pipeline. The GenBank file for the genome was visualised using GView (v1.7) with a style sheet optimised to highlight a number of potentially phase variable associated genes (Petkau et al., 2010).

4.4. Results

4.4.1 Schematic overview

A graphical overview of the protocol implemented to screen of novel Bacteroidales phages is presented in Figure 1. A healthy donor was recruited for faecal sample donation and a frozen standard inoculum (FSI) was prepared immediately. Triplicate fermentations were performed using two vessels in parallel, one of which contained vancomycin and kanamycin. In conjunction with this, individual Bacteroidales were enriched and purified from the FSI on various media containing the same antibiotics. To screen for novel phage-host pairs, the phage enriched fermentate filtered of bacterial cells was screened against the isolated Bacteroidales species using spot on lawn assays.

4.4.2 Bacterial composition following antibiotic enrichment

16S rRNA gene sequencing of total DNA extracted from bacterial pellets, collected following centrifugation of faecal fermentates, revealed a significant difference in the composition and relative abundance of bacterial genera with and without selective enrichment (Figure 2). At T0 hours, both vessels are dominated by genera that resolve into the following orders: Clostridales (~50%), Bacteroidales (~30%) and Selenomonadales (12.5%). By T4.5 hours, the composition of the bacterial genera changes significantly under antibiotic enrichment with Bacteroidales associated genera beginning to dominate whereas the control vessel is dominated by gram-positive *Catenibacterium*. By T24 hours, the relative abundance of *Bacteroides* and *Parabacteroides* constitutes up to 90% of the total 16S rRNA gene reads under selective conditions whereas these genera contribute to less than 20% of the total reads

in the control vessel. These findings are consistent with the detected bacterial species isolated from the same faecal sample following plating on agar containing the same antibiotics. The majority of the isolates are of the *Bacteroides* genus as well as a significant enrichment of the species *Parabacteroides distasonis* (Table 1). In the absence of antibiotics mostly gram-positive bacteria were recovered. This confirms that vancomycin and kanamycin were able to promote selective enrichment of Bacteroidales species. The phage-rich faecal fermentate collected from the antibiotic containing vessel formed more readily visible spots when screening for phage-host pairs, confirming that phages were also enriched using this approach (Figure S1).

4.4.3 Genome analysis of Φ PDS1 and *Parabacteroides distasonis* APCS2/PD

When the phage-rich faecal fermentate was spotted on lawns of the isolated Bacteroidales species it was observed that clearing consistently occurred on lawns of *P. distasonis*. This phage was subsequently enriched on its host and shotgun sequencing confirmed the isolation of a novel siphovirus. The reads generated assembled into a circularised genome of 44,691 bp with a read coverage of 347 (GenBank MN929097) (Figure 3). Shotgun sequencing of the host, named *Parabacteroides distasonis* APCS2/PD (Figure S2), revealed that it has a genome size of 5,345,521 bp (deposited in the NCBI database under the following GenBank CP042285, BioProject PRJNA556872) with one associated plasmid, pPDS2-1 (GenBank CP042284). The G + C content of the phage is 45.28 G + C mol% and its host genome is 45.30 G + C mol%. It has been reported that phages generally tend to have a G + C content slightly lower than that of their host although this is not true in all cases (Almpanis et al., 2018).

Preliminary annotation of the Φ PDS1 genome identified 61 protein coding genes (CDS), half of which could be assigned functional annotations using HHpred (Figure 3, Table 2). These included three head structural proteins, five tail associated proteins, five DNA replication genes, two terminase genes and one gene for lysis. The right-hand side of the genome was dominated by structural genes and the left-hand side genes encoding functions associated with DNA replication and maintenance. A lysogenic gene module was not identified which is typically composed of a serine or tyrosine integrase, a repressor of the lytic-cycle and excisionase (Luo et al., 2017; Shkoporov and Hill, 2019). These modules are also often in the opposite orientation to proximal gene modules and again this was not observed. One capsular polysaccharide synthesis gene, over one hundred *sus/tonB* associated genes and more than ten tyrosine recombinase genes were identified, characteristic of phase variable gene expression.

4.4.4 Abundance of Φ PDS1 in faecal fermentates

The relative and absolute abundance of Φ PDS1 was determined for both condition variables examined during the fermentations. It is evident that the selective enrichment of the host greatly aided the propagation and expansion of this phage which in turn facilitated its isolation (Figure S2). Relative abundance of the phage in relation to the total viral reads sequenced showed that Φ PDS1 contributed to ~13-14% of reads sequenced from the later time point samples collected from the antibiotic vessel (Figure 4A). Absolute quantification using qPCR with the standard curve method show that at these later time points the phage achieved a titre of $\sim 4 \times 10^8$ copies/ml.

4.4.5 Genome analysis of Φ PDS1

Comparison of Φ PDS1 against an in-house database of human viral sequences generated from faeces identified sixteen phages with significant similarity (Figure 5A). The G + C content of these phages ranged from 43.19 - 45.76 mol%, size 42,181 - 46,687 bp and average nucleotide identity (ANI) 81 - 98%. The detected phages originate from four independent studies; Enteric Virome IBD (Norman et al., 2015), Longitudinal Virome Study (Shkoporov et al., 2019), Cystic Fibrosis Sibling (NCBI BioProject: PRJNA498332) and Cystic Fibrosis (unpublished). These phages do not correlate with any disease state or other metadata (Table 3). Despite slight variations in ANI, overall genome organisation and genetic synteny among these phages is largely conserved, even for phage Sib2_ms_18 which is the most distantly related to Φ PDS1 (Figure 5B).

The presence of CRISPR-spacers targeting Φ PDS1 were found in *P. distasonis* and in *Odoribacter splanchnicus* (Figure 5C). CRISPR-spacer analysis for the sixteen identified Φ PDS1-like phages found that the two phages most related to Φ PDS1, LS_917t11 and ERR944031, also had CRISPR-spacer hits for *P. distasonis* (Figure 5C). CRISPR-spacer hits associated with the fourteen other phages were detected in *O. splanchnicus* and *P. distasonis* with the phage least related to Φ PDS1, Sib2_ms_18, found to have a CRISPR-spacer hit in *Bacteroides fragilis*.

4.4.6 Biological characterisation of Φ PDS1

Transmission electron micrographs of Φ PDS1-rich lysate show that this phage has a characteristic siphovirus morphology with a long, non-contractile tail (Figure 6A). The arrow shaped tail tip of the phage is unusual and larger than typically

observed for other siphoviruses. The capsid diameter is approximately 53 ± 2.0 nm and tail length 150 ± 10.0 nm. Zones of clearing on host lawns produced by high titres of the phage are readily visible, but clearing is incomplete (Figure 6B). The phage also produces plaques on its sensitive host, but they are pin prick in size making visualisation and enumeration difficult. Φ PDS1 one-step growth curve was performed in triplicate with a MOI of ~ 1 (Figure 6C). It was found that the phage has a latent period of approximately 90 minutes with a relatively small burst size of ~ 14 to 20 virions per infected cell.

Continuous co-culture of Φ PDS1 showed that after 24 hours the phage titre recovered to $\sim 1 \times 10^{10}$ copies/ml following initial reductions consistent with dilution associated with sub-culturing (Figure 6D). By day two, the titre reduces to $\sim 4 \times 10^8$ copies/ml and remains at this approximate titre for subsequent rounds of sub-culturing. This suggests that the presence of the phage selects for an increase in non-phage permissive host variants but as the phage continues to propagate without becoming extinct, a permissive sub-population of the host culture must co-exist to sustain the high titres observed. The bacterial pellet collected at the end of the experiment was serially streaked three times. PCR confirmed that all colonies were negative for Φ PDS1 integration or maintenance episomally. Ten clones were tested for phage resistance by spot assay. It was found that zones of clearing had varying increases in turbidity compared to zones of clearing on naïve host lawns.

We examined the ability of the Φ PDS1 to propagate on two strain bank *P. distasonis* strains, DSM 29491 and DSM 20701, and an *O. splanchnicus* strain, DSM 20712, chosen following findings from CRISPR-spacer analyses (Figure 6E). The phage failed to propagate on the tested strains despite efficient propagation on *P. distasonis* APCS2/PD, determined by qPCR.

An efficiency of lysogeny style experiment (plating the host on agar seeded with high titres of phage) found that ~15% of *P. distasonis* APCS2/PD cells are resistant on initial exposure to Φ PDS1. This is consistent with lack of clearing of broth cultures and hazy zones of clearing on host lawns. In the absence of the phage, the host appeared to form two colony morphologies: large round creamy colonies and smaller granular colonies. A greater count of the former morphology type was present on the phage seeded plate, suggesting that possible phenotypic alterations have a role to play in morphology changes and the observed resistance. When Φ PDS1 was screen against lawns of the phage exposed clones, the zones of clearing had greater turbidity compared to the zones those formed on lawns of naïve *P. distasonis* APCS2/PD: 10% evident but incomplete clearing, 20% opaque zones and 70% no clearing

4.5. Discussion

In recent years many studies have revealed interesting insights into the human gut phageome. However, the majority of phages continue to exist only as sequence “dark matter” and phages which have been characterised *in silico* provide little insight into their biological properties. Isolating and characterising these phages is key in advancing our understanding of phage-host interactions in the human gut. The human gut virome is largely composed of dsDNA phages of the Caudovirales order and ssDNA phages of the *Microviridae* family (Creasy et al., 2018; Lim et al., 2015; Minot et al., 2013; Shkoporov et al., 2019). The bacterial hosts of the majority of these phages are unknown. With the goal of isolating novel human gut phages and potentially gain insight into phage-host interactions, this study focused on Bacteroidales due to the importance of this bacterial order in the human gut.

CRISPR-spacer analyses of virulent gut phages predicted that they are particularly associated with hosts of the Clostridiales and Bacteroidales orders, including *Bacteroides* and *Parabacteroides* (Shkoporov et al., 2019). The majority of Bacteroidales species belong to the *Bacteroides* and *Parabacteroides* genera (Zitomersky et al., 2011). These form two important and abundant genera in the human gut and are associated with both health and disease (García-Bayona and Comstock, 2019). There have been multiple reports of *Bacteroides* infecting phages (Booth et al., 1979; Kai et al., 1985; Tartera and Jofre, 1987; Klieve et al., 1991; Puig and Gironés, 1999; Hawkins et al., 2008; Ogilvie et al., 2012; Gilbert et al., 2017; Shkoporov et al., 2018a; Benler et al., 2018; Porter et al., 2020; Hryckowian et al., 2020). However, few studies have examined *Parabacteroides* infecting phages. Currently, there are only two *Parabacteroides* associated prophage genomes deposited on NCBI Taxonomy; Φ Parabacteroides YZ-2015a and Φ Parabacteroides YZ-2015b. The former infects *P.*

distasonis and the latter *P. merdae* and both are ssDNA phages of the *Microviridae* family. These phages were first detected in Sphagnum-dominated peat viromes (Quaiser et al., 2015). An additional seven *Parabacteroides* prophages of the *Microviridae* family were detected in the gut of *Ciona robusta*, a marine vertebrate (Creasy et al., 2018). This highlights how little is known about human gut *Parabacteroides* phages.

The selective enrichment implemented in this study successfully selected for the growth of Bacteroidales (Figure 2). This led to the isolation of a novel lytic *Parabacteroides distasonis* phage of the *Siphoviridae* family, Φ PDS1. There are two brief reports of apparent virulent phages associated with *Parabacteroides distasonis*, formerly known as *Bacteroides distasonis*, however neither of these phages were characterized (Eggerth and Gagnon, 1933; Sabiston and Cohl, 1969; Booth et al., 1979). Therefore, Φ PDS1 is the first *P. distasonis* targeting siphovirus to be isolated and characterised.

P. distasonis is a Gram-negative, obligate anaerobe with rod-shaped morphology that is largely human gut associated (Sakamoto and Benno, 2006). These bacteria have been reported as important commensals with anti-inflammatory roles in the gut, alleviation of metabolic disorders and obesity (Kverka et al., 2011; Koh et al., 2018; Wang et al., 2019). Reduced abundance of the *Parabacteroides* genus has also been observed in the faeces of patients suffering from inflammatory bowel disease (IBD) and multiple sclerosis patients (Lewis et al., 2015; Cekanaviciute et al., 2017; Olbjørn et al., 2019). However, multiple studies have also linked *P. distasonis* to disease states suggesting that it may be an opportunistic pathogen. It was found to be the most abundant bacterial species in faecal samples collected from individuals suffering from Crohn's Disease (CD) and DSS induced enhanced colitis mice

(Dziarski et al., 2016; Lopetuso et al., 2018). Furthermore, *P. distasonis* was isolated in abundance from a lesion removed from the ileum of a CD patient (Yang et al., 2019). This bacterial species was also shown to exacerbate symptoms in an amyotrophic lateral sclerosis mouse model (Blacher et al., 2019). Considering the health and disease implications associated with this bacterial species, *P. distasonis* and ΦPDS1 prove to be a phage-host pair of interest.

Annotation of the ΦPDS1 genome revealed that nearly half of the genes remain unidentified (Figure 3). However, of the genes successfully annotated, some were associated with functions of interest including a quorum sensing (QS) regulated transcription factor associated with virulence (gp_26). Several QS response regulator homologs have been detected in sequenced phage genomes deposited on NCBI (Hargreaves et al., 2014). Furthermore, gene gp_34 encodes radical SAM enzymes which are involved in synthesis and metabolism of many cell compounds. They are also involved in pathways that lead to synthesis of key players associated with quorum sensing; autoinducers (Parveen and Cornell, 2011). QS is a phenomenon used by bacteria that allows population density dependent controlled expression of specific genes through cell-to-cell communication for processes such as virulence and biofilm formation (Whitehead et al., 2001; Turovskiy et al., 2007; Ng and Bassler, 2009; Papenfort and Bassler, 2016). There are several examples of phages hijacking this process. One such example includes a *Vibrio cholerae* infecting phage, VP882, which encodes a QS receptor homologous to that of the host. As a result, the phage can “listen in” on its hosts and monitor population densities thus allowing an informed lytic/lysogenic lifestyle switch. In high host cell density, lysis occurs resulting in optimal propagation (Silpe and Bassler, 2019). It has also been demonstrated in a lytic *Pseudomonas aeruginosa* phage that QS can be used to interfere with host phage

defences to aid infection (Hendrix et al., 2019). In addition, Φ PDS1 encodes a single-stranded binding protein (SSB) which suggests that it does not depend on a host SSB for replication (Tang et al., 2013). Genes gp_25 and gp_31 both encode both encode Nin family recombinases, also known as *orf* and *rap*. These are associated with lytic growth (Tarkowski et al., 2002; Paepe et al., 2014). An auxillary metabolic gene, MazG, is encoded by gp_39. Such genes are of host origin and allow phages to alter host metabolic processes to their advantage. They have been mostly detected in marine phages and are associated with both lytic and temperate lifestyles (Breitbart et al., 2018; Warwick-Dugdale et al., 2019). MazG has been linked to maintenance of phage propagation in starved host cells by aiding metabolism (Bryan et al., 2008; Kang et al., 2013).

Analysis of Φ PDS1 against an in-house human phage database identified sixteen Φ PDS1 related phages (Figure 5A). CRISPR-spacer analyses suggest that these phages have evolved to infect one of two members of the Bacteroidales order: *Parabacteroides distasonis* or *Odoribacter splanchnicus*. As Φ PDS1 CRISPR-spacers were detected in *P. distasonis* ATCC 8503 (DSM 20701) and *O. splanchnicus* DSM 20712, these specific strains were ordered from DSMZ in addition to the only other *P. distasonis* strain available from culture collections, *P. distasonis* DSM 2949. None of these strains could support propagation of Φ PDS1 (Figure 6E). This suggests that Φ PDS1 is highly strain specific in its infection strategy and has specialised to propagate on the *P. distasonis* APCS2/PD on which it was isolated. Interestingly, when these findings are considered in parallel with whole genome comparisons, we can predict that the receptor binding protein is encoded by gene gp_60 (Figure 5B). Φ PDS1 and its closest relative, LS_917t11, both have top CRISPR-spacers hits in *P. distasonis* and at gene gp_60 both share significant nucleotide similarity. However, an

intermediate relative, LS_922til, had CRISPR-spacers detected in *Odoribacter splanchnicus* and a more distant relative, Sib2_ms_18, had spacer hits for *B. fragilis*. Both share no similarity with Φ PDS1 at gene gp_60 despite intermediate to high nucleotide similarity at the majority of other genes.

Φ PDS1 fails to clear liquid cultures of its host despite efficient propagation and zones of clearing remain slightly hazy. The phage is also unable to form stable lysogens and its genome has no detectable integrase, strongly indicating that the phage is lytic. Although, with infrequent plaque formation pseudolysogeny may be possible but unlikely (Siringan et al., 2014). The phage can also stably co-exist with its host during serial co-culturing despite initially selecting for an increase in resistant host variants as indicated by the titre decrease on day two of the co-culture cycle (Figure 6D). Following this initial titre decrease, the phage continues to stably propagate for days with titres maintained at $\sim 6 \times 10^8$ pfu/ml and without clearing of the host culture. This could be indicative of host population heterogeneity in terms of phage permissivity. This may also explain the variation in turbidity for zones of clearing and phage kinetics observed for different *P. distasonis* APCS2/PD cultures. The inability of an apparently lytic phage to clear liquid culture of its host at high titres was previously observed for a *B. intestinalis* phage, Φ crAss001, and *B. thetaiotaomicron* phages (Shkoporov et al., 2018a; Porter et al., 2020). In the case of the latter, phase-variable expression of host capsule polysaccharide loci and surface features created transient phenotypic heterogeneity within an isogenic population. This heterogeneity resulted in a mixture of phage permissive and non-permissive host variants thus allowing both phage and host to co-exist (Porter et al., 2020). This highlights the importance of bacterial host factors in influencing phage-host interactions. Phase-variable regions have previously been reported among gut *P. distasonis* (Coyne and

Comstock, 2008; Fletcher et al., 2007). Potential phase-variable mediators such as invertases, integrases and serine/tyrosine recombinases were observed in the *P. distasonis* APCS2/PD genome (Figure S2). Invertible promotor regions are characteristic of Bacteroidales species residing in the human gut and are not generally conserved among Bacteroidales occupying other niches (Coyne and Comstock, 2008; Jiang et al., 2019). Transcriptomics may provide important insights into the role of these features. The mechanisms that mediate the interaction between Φ PDS1 and *P. distasonis* APCS2/PD merit further investigation and this phage-host pair should be examined in *in vivo* models with naturally relevant conditions.

In conclusion, we report the isolation, biological and *in silico* characterization of Φ PDS1. This phage is the first lytic siphovirus isolated from the human gut that infects *P. distasonis* to be characterised in detail. Φ PDS1 is also the first phage genome of this type to be deposited on NCBI Taxonomy. Considering the health and disease associations of *P. distasonis*, further characterisation of this phage-host pair could provide interesting insights into phage-host interactions in the human gut, particularly the persistence of virulent phages.

4.6 Figures

1.

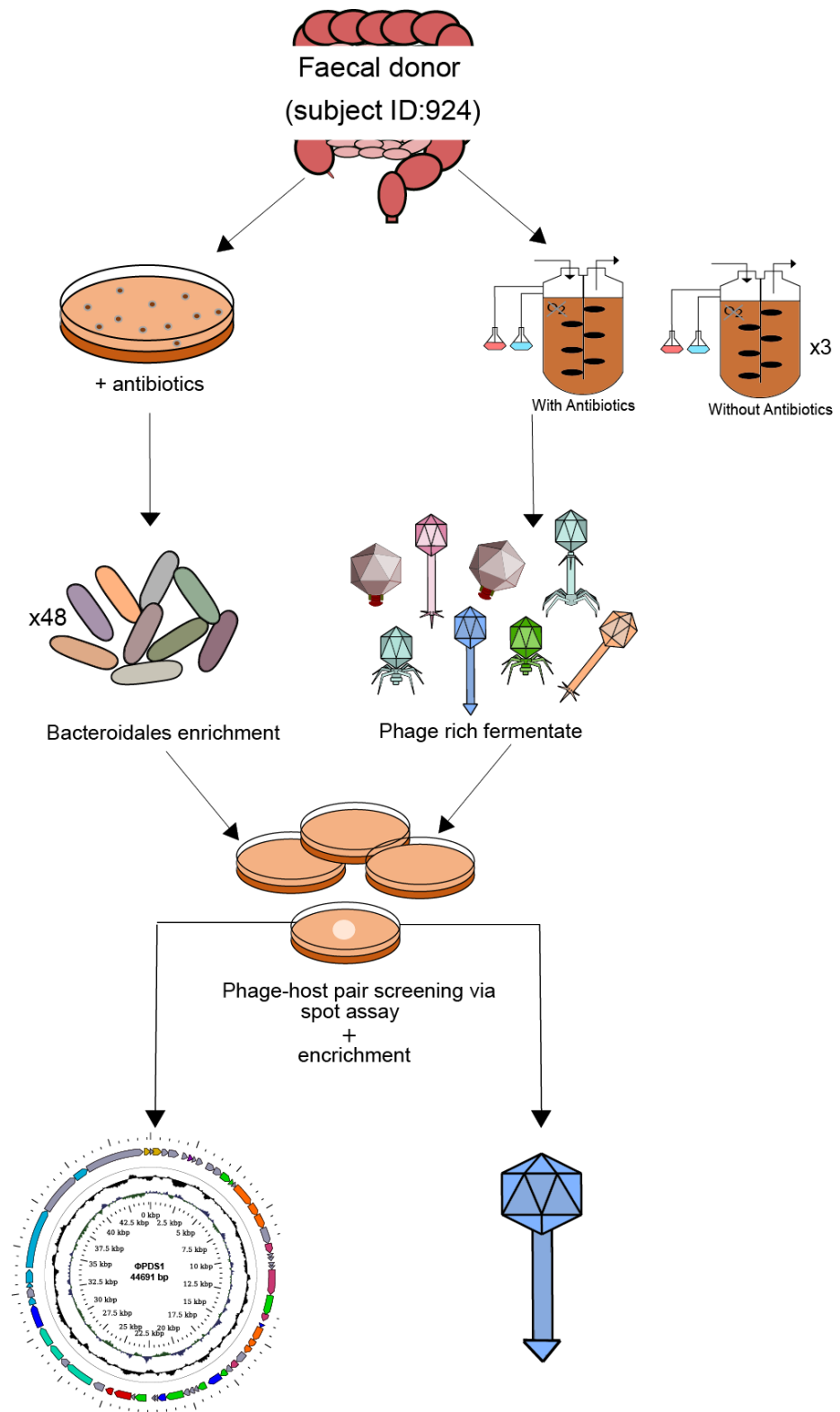


Figure 1. Schematic overview of the key experimental steps implemented to screen for novel Bacteroidales phage-host pairs from the human gut.

2.

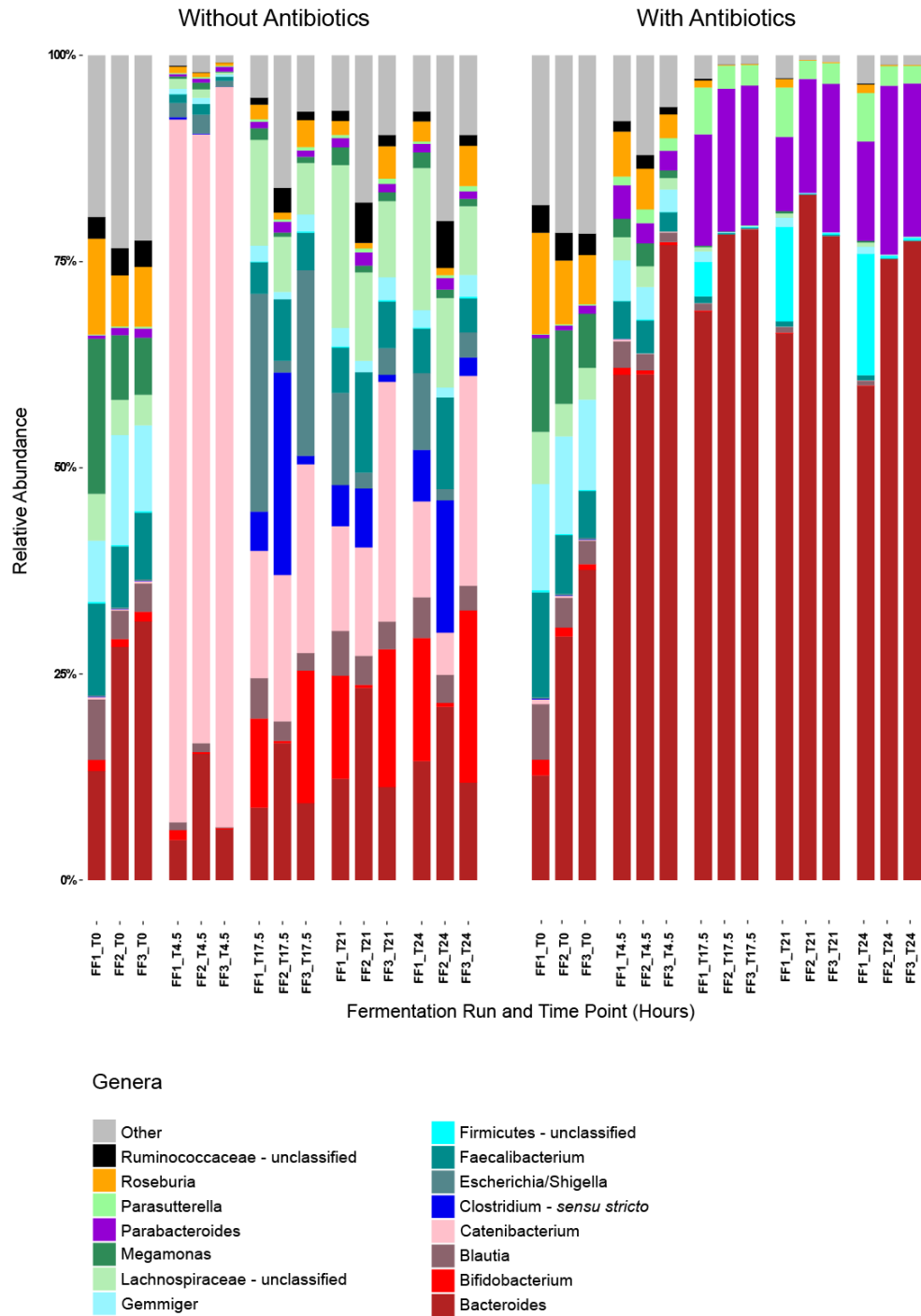


Figure 2. The relative abundance of the bacterial genera detected with and without antibiotic driven selective enrichment of Bacteroidales throughout the faecal fermentations. Biological triplicates are shown for each time point (FF1, FF2, FF3).

3.

I Proteins

- Packaging and Assembly
- Tail Morphogenesis
- DNA Replication
- DNA-Binding and Transcription Regulators
- Metabolic Genes
- Other
- DNA Maintenance
- Unknown Function

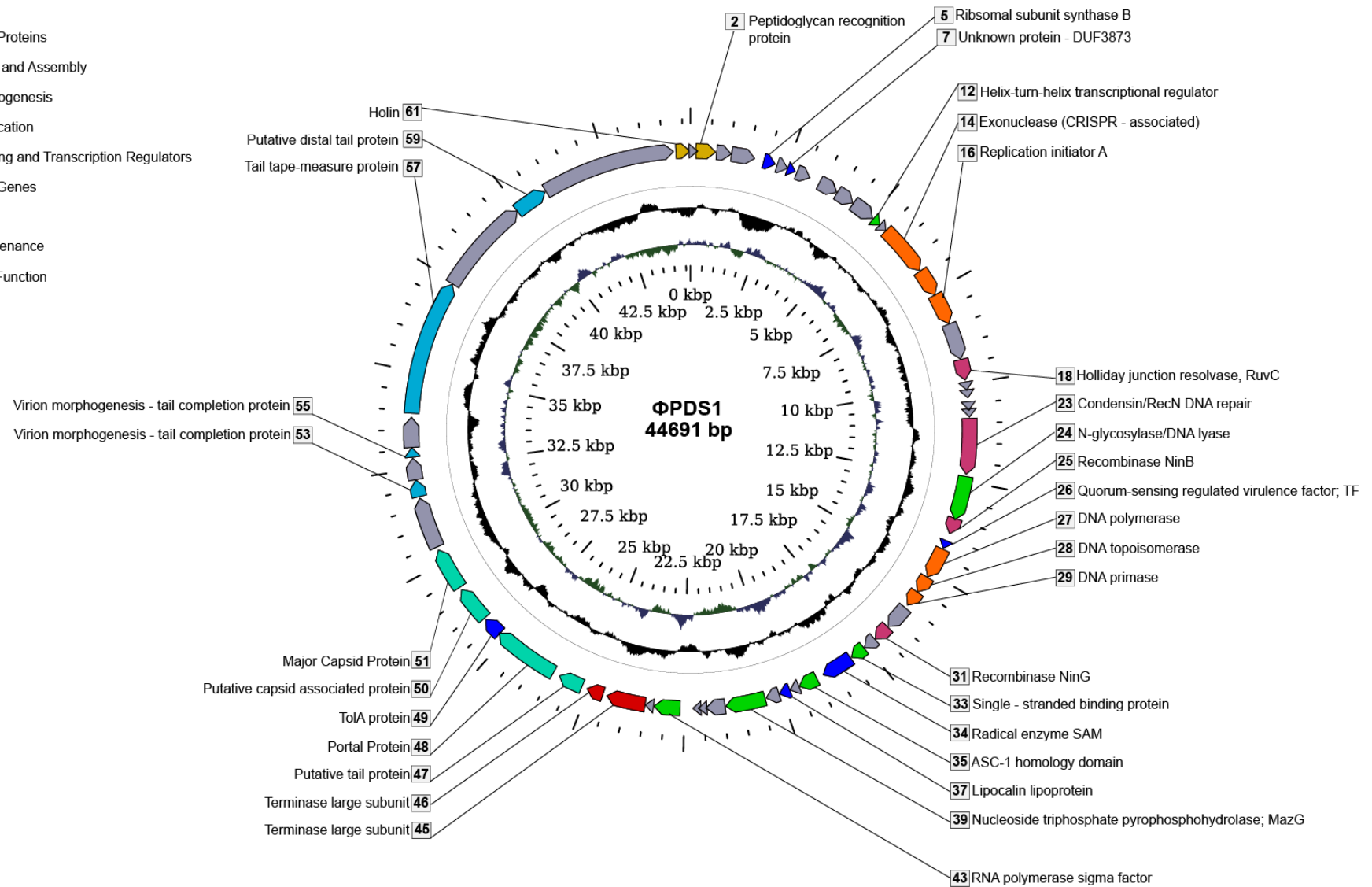
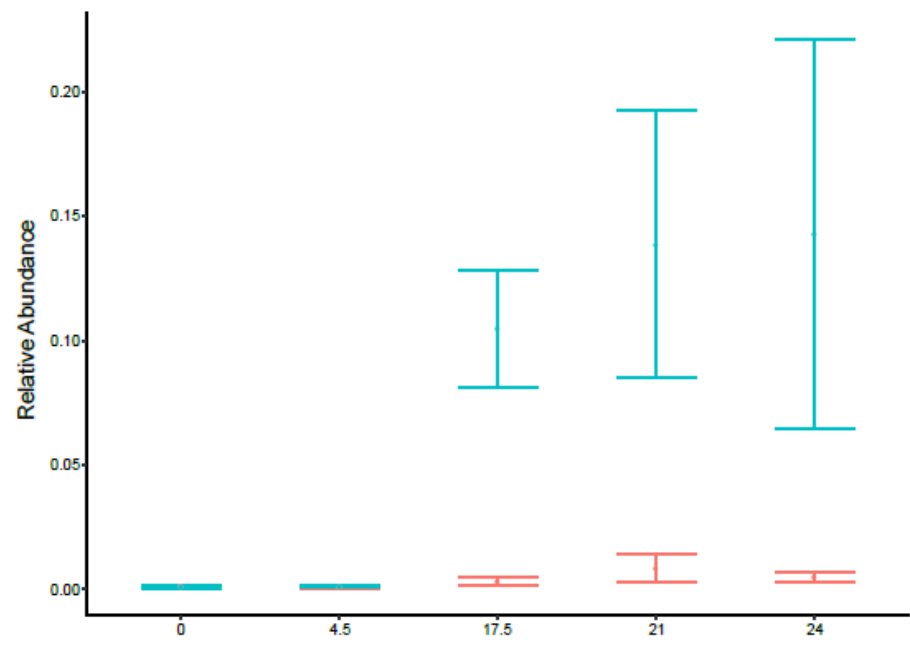


Figure 3. Circular genome map of novel Φ PDS1. Genome size is 44,691bp. The innermost ring (blue; positive strand, green; negative strand) depicts G + C skew, the central ring (black) shows G + C content. The outermost circle shows protein coding genes (CDS) which are labelled with HHpred predicted function. CDS are coloured based on general function which corresponds to the legend. Where function could not be determined, genes are coloured grey and are unlabelled.

4.

A



B

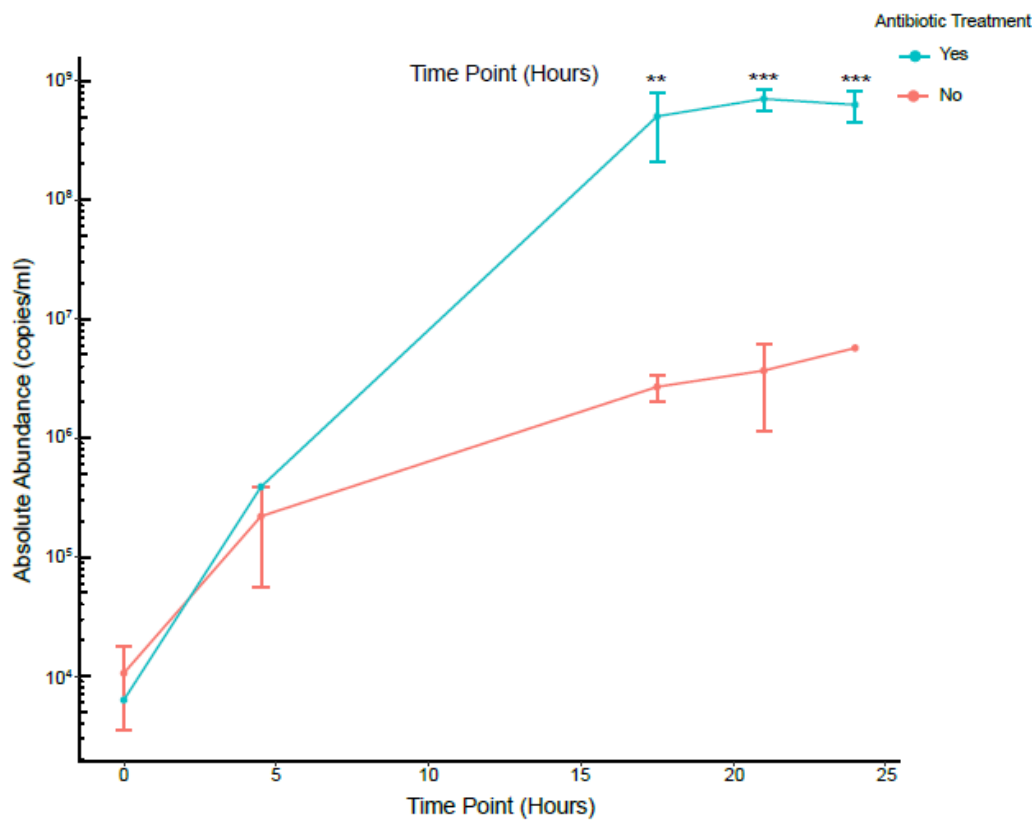
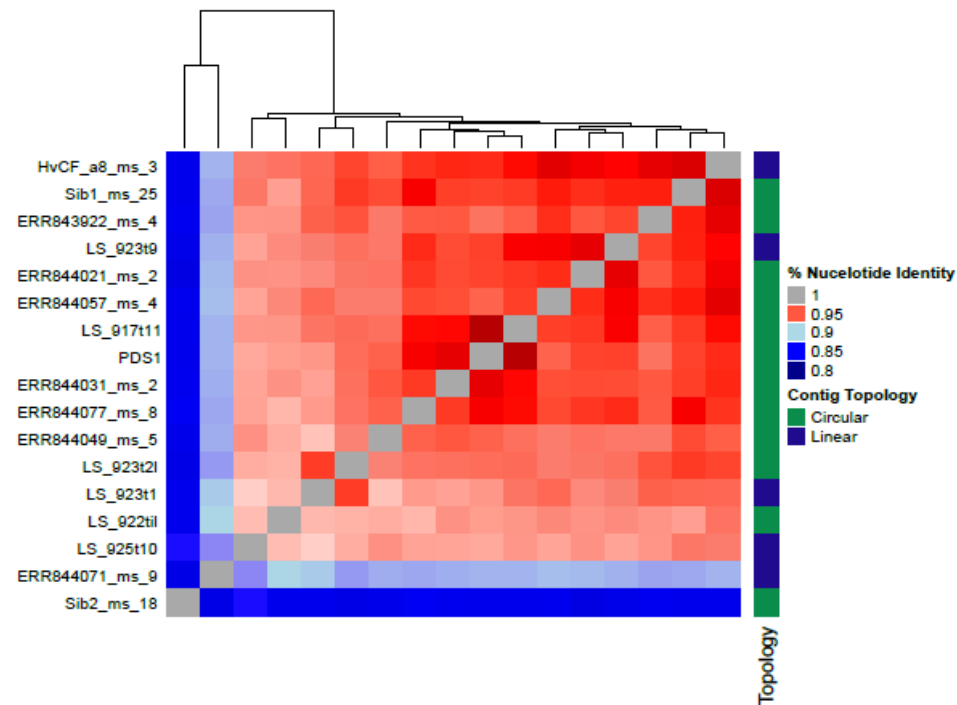


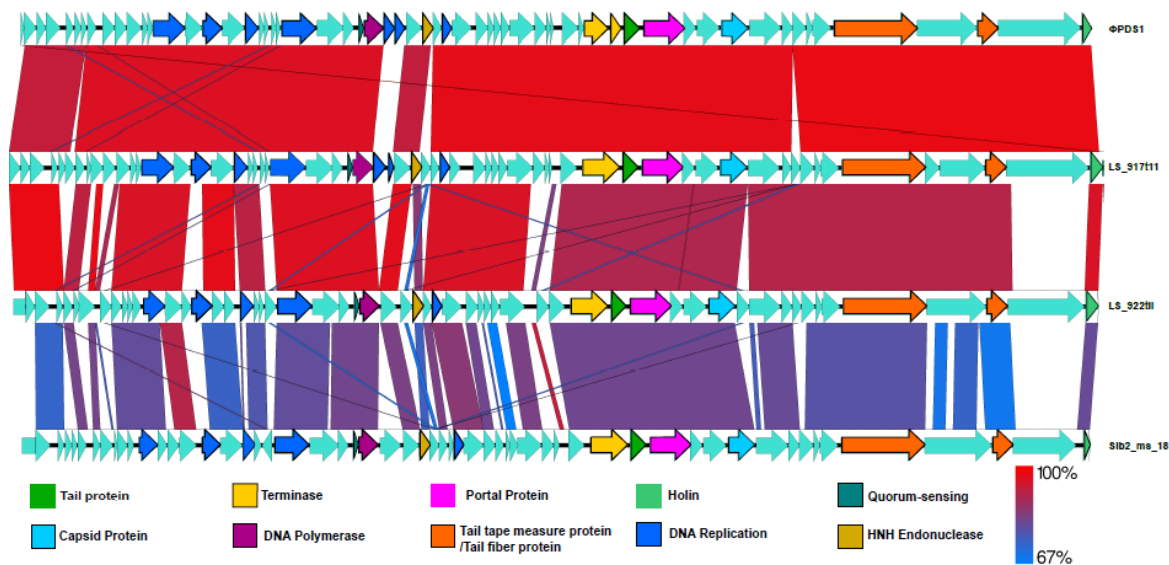
Figure 4. The influence of host selective enrichment on the abundance of Φ PDS1 throughout faecal fermentation (A) The relative abundance of Φ PDS1 with and without antibiotic enrichment of host, represented as the percentage of total viral reads aligning to the phage. (B) Absolute quantification of Φ PDS1 by qPCR targeting a segment of the phage portal protein. Titre determined in copies/ml using the standard curve method. Both results highlight that the selective conditions greatly aided Φ PDS1 propagation as a result of host expansion. Error bars represent standard deviation (n = 3). Statistical significance of titre difference was determined using the paired T-test (P-value = ≤ 0.01 (**), ≤ 0.001 (***)).

5.

A



B



C

Top 15 Genome Hits with Φ PD51 CRISPR Detected

Phage	Bacterial Genomes with Φ PD51 CRISPR-Spacer Hits	% Identity	Alignment Length	Number of Mismatches	Number of Gap Openings	Query Start	Query End	Subject Start	Subject End	E-value	Bit Score
Φ PD51	Parabacteroides distasonis ATCC 8503	100	37	0	0	21011	21047	1	37	4.00E-10	69.4
Φ PD51	Odoribacter splanchnicus DSM20712	100	37	0	0	2022	2058	37	1	4.00E-10	69.4
Φ PD51	Odoribacter splanchnicus DSM20712	100	37	0	0	16828	16864	37	1	4.00E-10	69.4
Φ PD51	Odoribacter splanchnicus DSM20712	100	37	0	0	19715	19751	1	37	4.00E-10	69.4
Φ PD51	Odoribacter splanchnicus DSM20712	100	37	0	0	2022	2058	1	37	4.00E-10	69.4
Φ PD51	Odoribacter splanchnicus DSM20712	100	37	0	0	19642	19678	37	1	4.00E-10	69.4
Φ PD51	Odoribacter splanchnicus DSM20712	100	37	0	0	19642	19678	1	37	4.00E-10	69.4
Φ PD51	Odoribacter splanchnicus DSM20712	100	37	0	0	18375	18411	37	1	4.00E-10	69.4
Φ PD51	Parabacteroides distasonis ATCC 8503	100	37	0	0	21011	21047	1	37	4.00E-10	69.4
Φ PD51	Parabacteroides distasonis ATCC 8503	100	37	0	0	21011	21047	1	37	4.00E-10	69.4
Φ PD51	Parabacteroides distasonis ATCC 8503	100	37	0	0	21011	21047	38	2	4.00E-10	69.4
Φ PD51	Odoribacter splanchnicus DSM20712	100	37	0	0	2022	2058	37	1	4.00E-10	69.4
Φ PD51	Parabacteroides distasonis ATCC 8503	100	37	0	0	21401	21437	37	1	4.00E-10	69.4
Φ PD51	Bacteroides fragilis YCH46	100	36	0	0	4535	4570	1	36	2.00E-09	67.6
Φ PD51	Bacteroides fragilis YCH46	100	36	0	0	25582	25617	36	1	2.00E-09	67.6
Φ PD51	Unknown	100	36	0	0	29837	29872	1	36	2.00E-09	67.6

Other Hits											
Φ PD51	Bacteroides vulgatus ATCC8482	100	35	0	0	9194	9228	1	35	5.00E-09	65.8
Φ PD51	Bacteroides dorei isolate HS1_L_1_B_010	100	30	0	0	39476	39505	3	32	3.00E-06	56.5
Φ PD51	Bacteroides helcogenes P36-108	100	30	0	0	18712	18741	1	30	3.00E-06	56.5

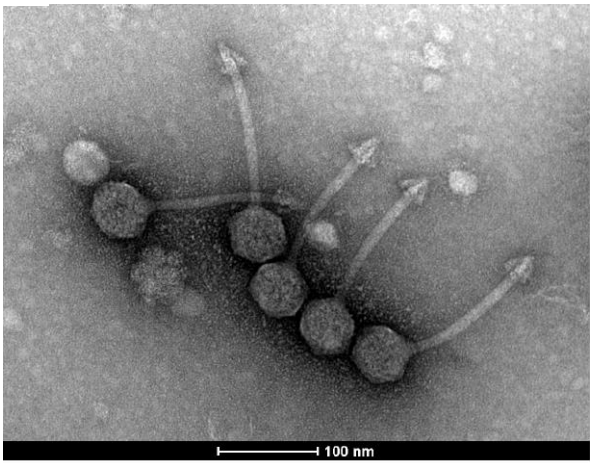
Top 15 Genome Hits with Φ PD51-like CRISPR Detected

Phage (Order Φ PD51)	Bacterial Genomes with CRISPR-Spacer Hits	% Identity	Alignment Length	Number of Mismatches	Number of Gap Openings	Query Start	Query End	Subject Start	Subject End	E-value	Bit Score
917111	Parabacteroides distasonis ATCC8503	100	37	0	0	2622	2658	1	37	4.00E-10	69.4
ERR844031_ms_2	Parabacteroides distasonis ATCC8503	100	37	0	0	26480	26516	37	1	4.00E-10	69.4
ERR844077_ms_8	Odoribacter splanchnicus DSM20712	100	37	0	0	8016	8052	37	1	4.00E-10	69.4
HvCF_a8_ms_3	Odoribacter splanchnicus DSM20712	100	37	0	0	7314	7350	37	1	4.00E-10	69.4
Sib1_ms_25	Odoribacter splanchnicus DSM20712	100	37	0	0	24498	24534	37	1	4.00E-10	69.4
ERR843922_ms_4	Parabacteroides distasonis ATCC8503	100	37	0	0	30598	30634	37	1	4.00E-10	69.4
92319	Odoribacter splanchnicus DSM20712	100	37	0	0	12526	12562	37	1	4.00E-10	69.4
ERR844021_ms_2	Odoribacter splanchnicus DSM20712	100	37	0	0	13900	13936	37	1	4.00E-10	69.4
ERR844021_ms_5	Odoribacter splanchnicus DSM20713	100	37	0	0	13900	13936	37	1	4.00E-10	69.4
ERR844049_ms_5	Parabacteroides distasonis ATCC8503	100	38	0	0	35113	35150	38	1	1.00E-10	71.3
92312l	Odoribacter splanchnicus DSM20712	100	37	0	0	37542	37578	37	1	4.00E-10	69.4
9231il	Odoribacter splanchnicus DSM20712	100	37	0	0	43828	43864	37	1	4.00E-10	69.4
922til	Odoribacter splanchnicus DSM20713	100	38	0	0	26660	26697	38	1	1.00E-10	71.3
925t10	Odoribacter splanchnicus DSM20712	100	37	0	0	33794	33830	1	37	4.00E-10	69.4
ERR844071_ms_9	Odoribacter splanchnicus DSM20712	100	37	0	0	9362	9398	37	1	4.00E-10	69.4
Sib2_ms_18	Bacteroides fragilis YCH46	100	36	0	0	19020	19055	1	36	2.00E-09	67.6

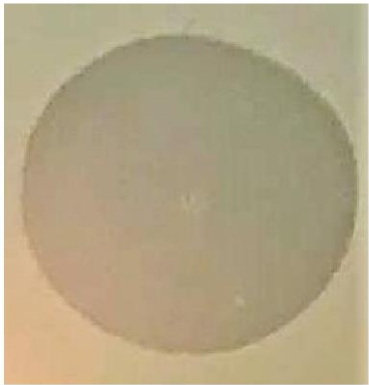
Figure 5. *In silico* analysis of Φ PDS1 and related phages. BLASTn of the Φ PDS1 genome against a database generated from multiple human gut virome studies identified sixteen closely related phage contigs. (A) Average nucleotide identity of the phages was determined to gain insight into relatedness. The closest relative of Φ PDS1 was identified as LS_917t11 and the most distant Sib2_ms_118. (B) Whole genome comparison of Φ PDS1 against a highly, intermediately and more distantly related phage to examine synteny and conservation of genome organisation. (C) Analysis of Φ PDS1 and the sixteen related phages for the presence of CRISPR-spacer regions in bacterial genomes. The top 10 bacterial genome hits with Φ PDS1 CRISPR-spacers are shown. The top bacterial genome hit for each Φ PDS1-related phage is shown.

6.

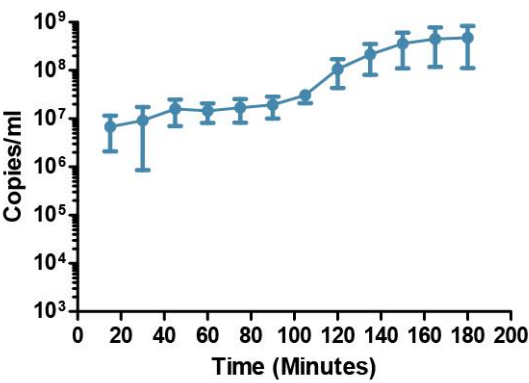
A



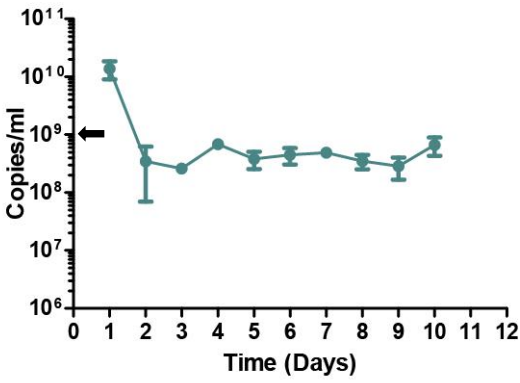
B



C



D



E

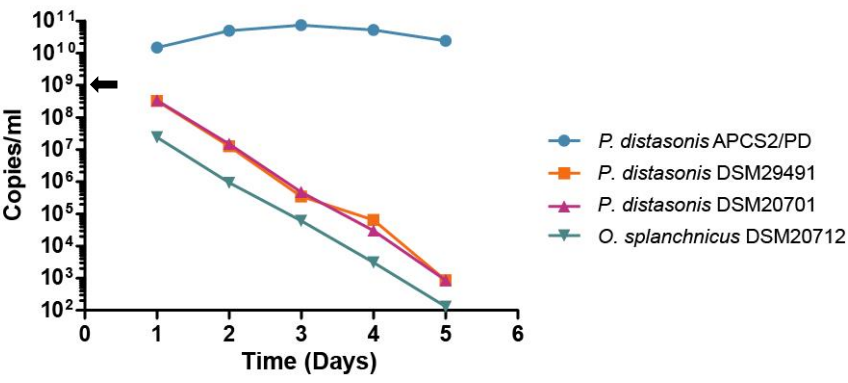
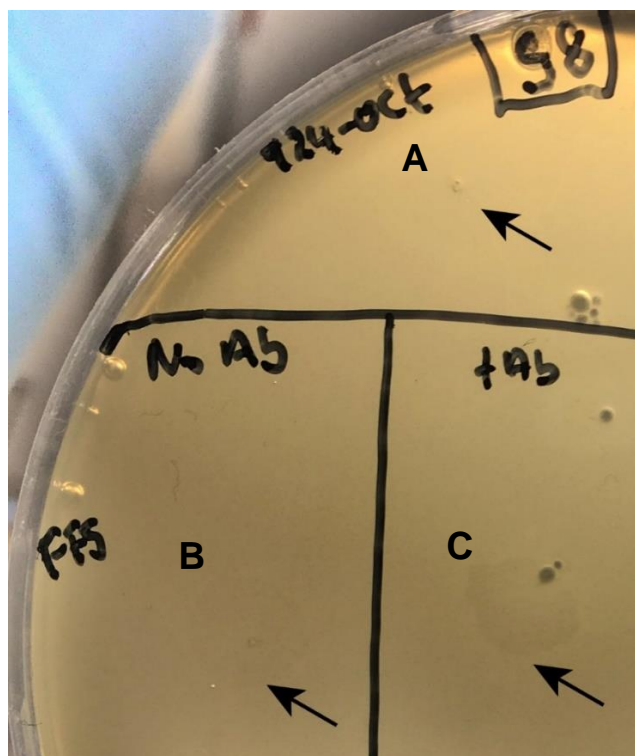


Figure 6. Biological characterisation of Φ PDS1 (A.) Transmission electron micrograph of Φ PDS1 generated from enriched lysate, stained with uranyl acetate showing that Φ PDS1 has a siphovirus morphology. The capsid diameter is approximately 53 ± 2.0 nm and tail length 150 ± 10.0 nm. (B) Spot morphology with incomplete clearing. (C) Φ PDS1 one-step growth curve. Sampling was performed every 15 mins over 3 hours. Latent period is 90 mins and burst size of ~14-20 virions per infected cell. (D) The titre of Φ PDS1 following continuous co-culture on host *P. distasonis* APCS2/PD over 10 days. (E) Liquid propagation of Φ PDS1 on two commercial *P. distasonis* strains (DSM 29491 and DSM 20701) and *O. splanchnicus* (DSM 20712) was examined over five days with host *P. distasonis* APCS2/PD used as a control. Titres determined in copies/ml by qPCR. Arrows on y-axis indicate phage titre on initiation of experiment. Error bars indicate standard deviation (n = 3).

Supplementary Figures

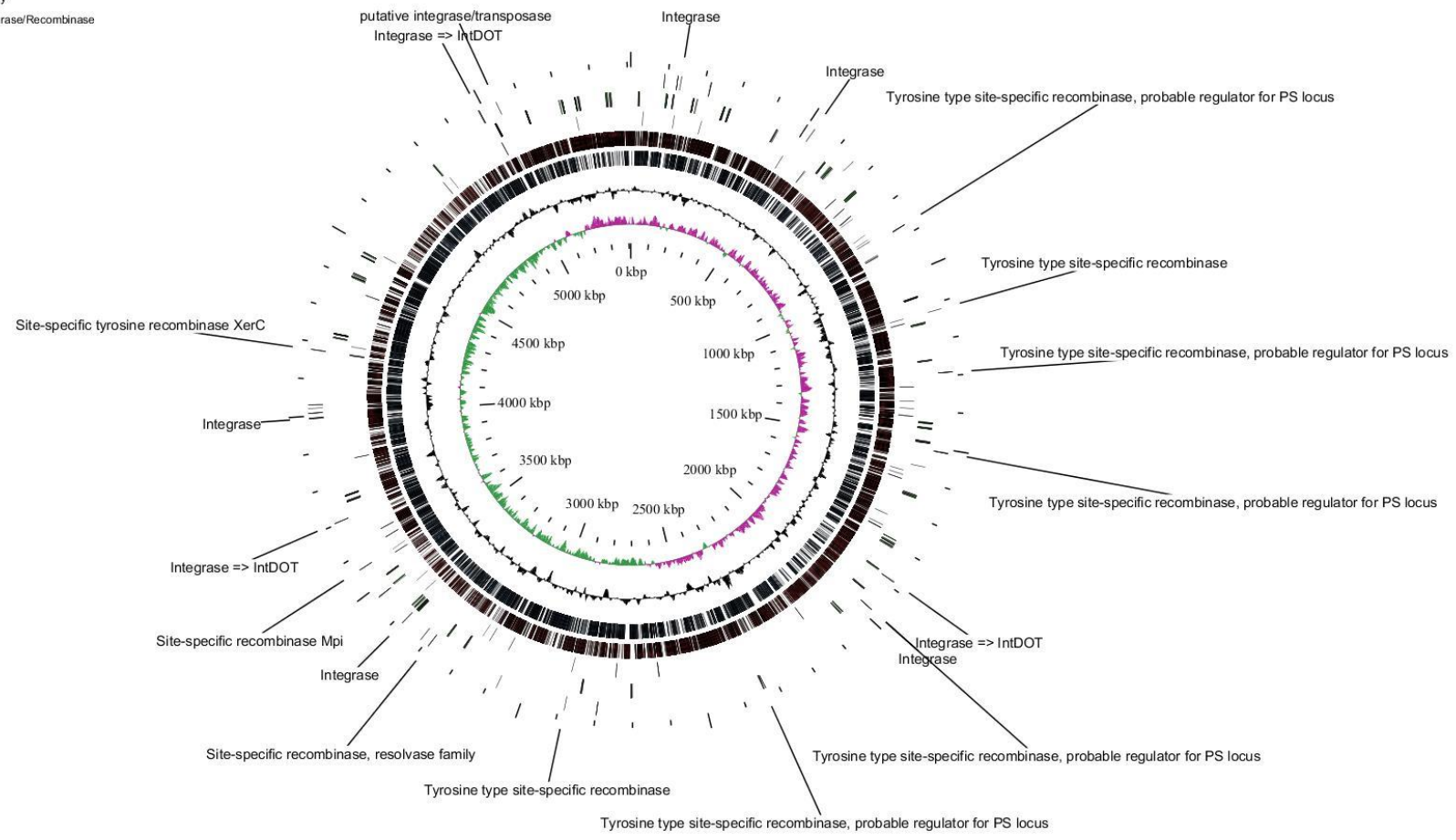
S1.



Supplementary figure 1. Spot assay of ϕ PDS1 on lawn of *P. distasonis* APCS2/PD. Samples examined originate from the same subject ID:924 faecal sample and were tested as follows: (A) Faecal filtrate prepared directly from subject ID:924 faeces. (B) Phage rich fermentate collected at time point 21 hours. Selective enrichment to promote the growth of Bacteroidales was not performed in the case of A + B (C) Phage rich fermentate collected at time point 21 hours following selective enrichment. These results show that the enrichment aided the phage-host pair screening process.

S2.

- Negative stranded CDS
- Positive stranded CDS
- tRNA and rRNA
- Sus gene family
- Invertase/Integrase/Recombinase



Supplementary figure 2. Circular map of Φ PDS1 host *P. distasonis* APCS2/PD genome 5.35 Mbp in size. The innermost circle (green and purple) depicts GC skew and circle two (black), G + C content. Circle 3 and 4 (red and dark blue), opening reading frames identified on the positive and negative DNA strands. Circle 5 (orange), tRNA and rRNA genes. Circle 6, genes annotated as *sus* gene family associated (green) and TonB dependent receptor (light blue). Circle 7 (black), genes annotated as invertase, integrase and recombinase

4.7 Tables.

Table 1. Species identified from Sanger sequencing of bacteria enriched from subject ID: 924 faecal sample following selective enrichment. FAA; Fastidious anaerobic agar, YCFA; yeast extract, casitone, fatty acids agar, CBA; Columbia blood agar.

Strain Code	Size (bp)	Identity (%)	Coverage (%)	Top Hit Species	Media (+ vancomycin and kanamycin)
FAA					
FFA S1	968	99	100	<i>B.uniformis</i>	FAA
FFA S2	1186	99	99	<i>B.uniformis</i>	FAA
FAA S3	1120	99	100	<i>B.uniformis</i>	FAA
FFA S4	1105	99	100	<i>B.uniformis</i>	FAA
FFA S5	1122	99	100	<i>B.ovatus</i>	FAA
FAA S6	>1000	99	100	<i>B.dorei</i>	FAA
FFA S7	1279	99	99	<i>B.uniformis</i>	FAA
FAA S8	1115	98	99	<i>P.distasonis</i>	FAA
FAA B1	673	95	100	<i>P.distasonis</i>	FAA
FAA B2	977	97	99	<i>B.uniformis</i>	FAA
FAA B3	1071	95	96	<i>P.distasonis</i>	FAA
FAA B4	947	97	99	<i>P.distasonis</i>	FAA
FAA B5	1000	98	99	<i>P.distasonis</i>	FAA
FAA B6	1138	97	100	<i>P.distasonis</i>	FAA
FAA B7	1057	99	97	<i>P.distasonis</i>	FAA
FAA B8	1111	98	100	<i>P.distasonis</i>	FAA
YCFA					
YCFA S1	776	99	100	<i>B.uniformis</i>	YCFA
YCFA S2	1101	99	99	<i>B.uniformis</i>	YCFA
YCFA S3	1171	91	100	<i>B.uniformis</i>	YCFA
YCFA S4	1109	99	99	<i>B.uniformis</i>	YCFA
YCFA S5	965	99	100	<i>B.uniformis</i>	YCFA
YCFA S6	1087	99	98	<i>B.uniformis</i>	YCFA
YCFA S7	1195	99	100	<i>B.uniformis</i>	YCFA
YCFA S8	1107	99	100	<i>B.uniformis</i>	YCFA
YCFA B1	901	98	100	<i>B.uniformis</i>	YCFA
YCFA B2	658	99	99	<i>B.dorei</i>	YCFA
YCFA B3	1081	99	100	<i>B.uniformis</i>	YCFA
YCFA B4	906	99	99	<i>P.distasonis</i>	YCFA
YCFA B5	895	99	100	<i>B.uniformis</i>	YCFA
YCFA B6	698	99	100	<i>B.uniformis</i>	YCFA
YCFA B7	742	99	100	<i>B.uniformis</i>	YCFA
YCFA B8	884	99	100	<i>B.uniformis</i>	YCFA

CBA					
CBA S1	800	99	100	<i>B.xylanisolvans</i>	CBA (horse blood & vitK)
CBA S2	936	96	100	<i>P.distasonis</i>	CBA (horse blood & vitK)
CBA S3	1000	99	100	<i>B.uniformis</i>	CBA (horse blood & vitK)
CBA S4	701	99	100	<i>P.distasonis</i>	CBA (horse blood & vitK)
CBA S5	917	99	100	<i>B.uniformis</i>	CBA (horse blood & vitK)
CBA S6	938	98	99	<i>P.distasonis</i>	CBA (horse blood & vitK)
CBA S7	574	99	100	<i>B.uniformis</i>	CBA (horse blood & vitK)
CBA S8	790	99	100	<i>B.uniformis</i>	CBA (horse blood & vitK)
CBA B1	893	99	100	<i>B.uniformis</i>	CBA (horse blood & vitK)
CBA B2	>1000	99	99	<i>B.fragilis</i>	CBA (horse blood & vitK)
CBA B3	647	93	99	<i>B.dorei</i>	CBA (horse blood & vitK)
CBA B4	351	94	100	<i>P.distasonis</i>	CBA (horse blood & vitK)
CBA B5	173	90	98	<i>B.dorei</i>	CBA (horse blood & vitK)
CBA B6	394	81	100	<i>B.dorei</i>	CBA (horse blood & vitK)
CBA B7	922	99	100	<i>B.dorei</i>	CBA (horse blood & vitK)
CBA B8	481	82	100	<i>B.dorei</i>	CBA (horse blood & vitK)

Table 2: HHpred determined functional annotation of Φ PDS1 genome

LOC/CDS Name	Protein Length (aa)	Hit	Probability	E-value	Function
gp_1	45	-	-	-	-
gp_2	171	2EAX_B	99.84	1.20E-22	Peptidoglycan recognition protein
gp_3	128	-	-	-	-
gp_4	204	-	-	-	-
gp_5	105	-	-	-	Ribosomal large subunit pseudouridine synthase B (Determined by VIGA and BLASTp)
gp_6	92	-	-	-	-
gp_7	68	PF12989.7	99.94	8.70E-31	Domain of - function (DUF3873)
gp_8	123	-	-	-	-
gp_9	172	-	-	-	-
gp_10	162	-	-	-	-
gp_11	206	-	-	-	-
gp_12	80	2EWT_A	99.29	5.20E-13	Helix-turn-helix transcriptional regulator
gp_13	70	-	-	-	-
gp_14	447	PF12684.7	98.96	2.20E-10	PDDEXK-like domain/CRISPR-associated exonuclease Csa1
gp_15	242	-	-	-	-
gp_16	274	4PT7_A	98.62	4.40E-09	Replication initiator A family protein
gp_17	318	-	-	-	-
gp_18	184	1HJR_B	99.82	2.10E-21	Holliday junction resolvase, RuvC
gp_19	69	-	-	-	-

gp_20	63	-	-	-	-
gp_21	67	-	-	-	-
gp_22	60	-	-	-	-
gp_23	481	5XEI_A	99.68	8.30E-16	Condensin - chromosome partition/DNA repair protein RecN
gp_24	382	TIGR02757	100	8.40E-75	Protein of - function/ N-glycosylase/DNA lysase
gp_25	143	1PC6_A	99.83	4.00E-22	Protein ninB/Nuclease
gp_26	68	PF12843.7	99.02	4.00E-12	Putative quorum-sensing regulated virulence factor TF (QSregVF)
gp_27	272	5M1S_D	99.71	4.80E-18	DNA Polymerase
gp_28	152	1D3Y_A	95.97	8.30E-04	DNA topoisomerase
gp_29	143	cd00223	97.57	1.30E-06	Topoisomerase-primase nucleotidyl transferase/hydrolase
gp_30	219	-	-	-	-
gp_31	150	PF05766.12	99.04	2.40E-12	NinG recombination protein/HNH Endonuclease
gp_32	117	-	-	-	-
gp_33	137	3LGJ_A	99.93	1.80E-26	Single-strand binding protein
gp_34	278	TIGR04471	99.92	3.60E-28	Radical SAM enzyme
gp_35	168	2DP9_A	99.58	1.80E-16	ASC-1 homology domain
gp_36	84	-	-	-	-
gp_37	98	5CYB_A	92.9	0.12	Lipocalin lipoprotein
gp_38	119	-	-	-	-
gp_39	345	3CRA_B	99.51	1.10E-16	Nucleoside triphosphate pyrophosphohydrolase, MazG
gp_40	162	-	-	-	-

gp_41	56	-	-	-	-
gp_42	45	-	-	-	-
gp_43	227	PF07638.11	95.23	0.031	RNA polymerase sigma factor
gp_44	72	-	-	-	-
gp_45	337	4BIJ_C	99.93	6.70E-27	Terminase large subunit
gp_46	148	TIGR01630	99.67	7.10E-18	Terminase large subunit
gp_47	216	-	-	-	Putative tail protein (Determined by VIGA)
gp_48	575	PF13264.6	99.42	2.00E-14	Portal protein
gp_49	172	TIGR02794	95.33	6.60E-02	TolA inner membrane protein (TonB paralog)
gp_50	316	-	-	-	Putative capsid associated protein (Determined by VIGA)
gp_51	370	3J7W_D	93.08	5	Major capsid protein
gp_52	440	-	-	-	-
gp_53	149	PF05069.13	96.49	1.90E-03	Phage virion morphogenesis protein
gp_54	189	-	-	-	-
gp_55	85	PF07098.11	98.11	2.10E-07	Protein of - function (DUF1360)
gp_56	257	-	-	-	-
gp_57	1164	-	-	-	Tail tape meausre protein (Determined by VIGA and BLASTp)
gp_58	836	-	-	-	-
gp_59	291	-	-	-	Putative tail fibre protein (Determined by VIGA and BLASTp)
gp_60	1151	-	-	-	-
gp_61	126	PF05105.12	98.62	2.10E-08	Holin

Table 3. Metadata associated with Φ PDS1 and sixteen related phages identified *in silico*.

	Study	Health Status	Gender	Age	Location/Nationality
LS_922t1	Longitudinal	Healthy	M	29	Romanian
Sib2_ms_18	Cystic Fibrosis Sibling	Cystic Fibrosis	Unknown	Unknown	Irish
ERR844031_ms_2	Norman	Healthy	F	51	Chicago, IL
LS_917t11	Longitudinal	Healthy	F	40	Irish
ERR844049_ms_5	Norman	Healthy	F	48	Boston, MA
ERR844071_ms_9	Norman	Healthy	M	57	Cambridge, MA
ERR844057_ms_4	Norman	UC	F	32	Chicago, IL
ERR844021_ms_2	Norman	Healthy	F	24	Chicago, IL
ERR843922_ms_4	Norman	Healthy	M	45	Cambridge, MA
ERR844077_ms_8	Norman	CD	M	<30	Cambridge, MA
HvCF_a8_ms_3	Cystic Fibrosis	Cystic Fibrosis	M	Unknown	Irish
LS_923t21	Longitudinal	Healthy	M	54	English
PDS1	Longitudinal	Healthy	F	43	Irish
LS_923t9	Longitudinal	Healthy	M	54	English
LS_925t10	Longitudinal	Healthy	F	45	Irish
LS_923t1	Longitudinal	Healthy	M	54	English
Sib1_ms_25	Cystic Fibrosis Sibling	Cystic Fibrosis	Unknown	Unknown	Irish

4.8 References

- Aggarwala, V., Liang, G., and Bushman, F.D. (2017). Viral communities of the human gut: metagenomic analysis of composition and dynamics. *Mobile DNA* 8, 12.
- Almpanis, A., Swain, M., Gatherer, D., and McEwan, N. (2018). Correlation between bacterial G+C content, genome size and the G+C content of associated plasmids and bacteriophages. *Microb Genom* 4.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- Antipov, D., Korobeynikov, A., McLean, J.S., and Pevzner, P.A. (2016). hybridSPAdes: an algorithm for hybrid assembly of short and long reads. *Bioinformatics* 32, 1009–1015.
- Bardina, C., Colom, J., Spricigo, D.A., Otero, J., Sánchez-Osuna, M., Cortés, P., and Llagostera, M. (2016). Genomics of Three New Bacteriophages Useful in the Biocontrol of Salmonella. *Front. Microbiol.* 7.
- Belleghem, J.D.V., Clement, F., Merabishvili, M., Lavigne, R., and Vaneechoutte, M. (2017). Pro- and anti-inflammatory responses of peripheral blood mononuclear cells induced by *Staphylococcus aureus* and *Pseudomonas aeruginosa* phages. *Sci Rep* 7, 1–13.
- Benler, S., Cobián-Güemes, A.G., McNair, K., Hung, S.-H., Levi, K., Edwards, R., and Rohwer, F. (2018). A diversity-generating retroelement encoded by a globally ubiquitous *Bacteroides* phage. *Microbiome* 6, 191.

Blacher, E., Bashiardes, S., Shapiro, H., Rothschild, D., Mor, U., Dori-Bachash, M., Kleimeyer, C., Moresi, C., Harnik, Y., Zur, M., et al. (2019). Potential roles of gut microbiome and metabolites in modulating ALS in mice. *Nature* 572, 474–480.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.

Booth, S.J., Van Tassell, R.L., Johnson, J.L., and Wilkins, T.D. (1979). Bacteriophages of *Bacteroides*. *Rev Infect Dis* 1, 325–336.

Breitbart, M., Bonnain, C., Malki, K., and Sawaya, N.A. (2018). Phage puppet masters of the marine microbial realm. *Nat Microbiol* 3, 754–766.

Bryan, M.J., Burroughs, N.J., Spence, E.M., Clokie, M.R.J., Mann, N.H., and Bryan, S.J. (2008). Evidence for the Intense Exchange of MazG in Marine Cyanophages by Horizontal Gene Transfer. *PLoS One* 3.

Cekanaviciute, E., Yoo, B.B., Runia, T.F., Debelius, J.W., Singh, S., Nelson, C.A., Kanner, R., Bencosme, Y., Lee, Y.K., Hauser, S.L., et al. (2017). Gut bacteria from multiple sclerosis patients modulate human T cells and exacerbate symptoms in mouse models. *PNAS* 114, 10713–10718.

Chibani, C.M., Farr, A., Klama, S., Dietrich, S., and Liesegang, H. (2019). Classifying the Unclassified: A Phage Classification Method. *Viruses* 11.

Clooney, A.G., Sutton, T.D.S., Shkoporov, A.N., Holohan, R.K., Daly, K.M., O'Regan, O., Ryan, F.J., Draper, L.A., Plevy, S.E., Ross, R.P., et al. (2019). Whole-Virome Analysis Sheds Light on Viral Dark Matter in Inflammatory Bowel Disease. *Cell Host Microbe*.

Coyne, M.J., and Comstock, L.E. (2008). Niche-Specific Features of the Intestinal Bacteroidales. *J Bacteriol* *190*, 736–742.

Creasy, A., Rosario, K., Leigh, B.A., Dishaw, L.J., and Breitbart, M. (2018). Unprecedented Diversity of ssDNA Phages from the Family Microviridae Detected within the Gut of a Protochordate Model Organism (*Ciona robusta*). *Viruses* *10*.

Dutilh, B.E., Cassman, N., McNair, K., Sanchez, S.E., Silva, G.G.Z., Boling, L., Barr, J.J., Speth, D.R., Seguritan, V., Aziz, R.K., et al. (2014). A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat Commun* *5*, 4498.

Dziarski, R., Park, S.Y., Kashyap, D.R., Dowd, S.E., and Gupta, D. (2016). Pglyrp-Regulated Gut Microflora *Prevotella falsenii*, *Parabacteroides distasonis* and *Bacteroides eggerthii* Enhance and *Alistipes finegoldii* Attenuates Colitis in Mice. *PLOS ONE* *11*, e0146162.

Eggerth, A.H., and Gagnon, B.H. (1933). The Bacteroides of Human Feces. *Journal of Bacteriology* *25*, 389–413.

Fletcher, C.M., Coyne, M.J., Bentley, D.L., Villa, O.F., and Comstock, L.E. (2007). Phase-variable expression of a family of glycoproteins imparts a dynamic surface to a symbiont in its human intestinal ecosystem. *PNAS* *104*, 2413–2418.

García-Bayona, L., and Comstock, L.E. (2019). Streamlined Genetic Manipulation of Diverse *Bacteroides* and *Parabacteroides* Isolates from the Human Gut Microbiota. *MBio* *10*, e01762-19.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5, R80.

Gilbert, R.A., Kelly, W.J., Altermann, E., Leahy, S.C., Minchin, C., Ouwerkerk, D., and Klieve, A.V. (2017). Toward Understanding Phage:Host Interactions in the Rumen; Complete Genome Sequences of Lytic Phages Infecting Rumen Bacteria. *Front. Microbiol.* 8.

Gogokhia, L., Buhrke, K., Bell, R., Hoffman, B., Brown, D.G., Hanke-Gogokhia, C., Ajami, N.J., Wong, M.C., Ghazaryan, A., Valentine, J.F., et al. (2019). Expansion of Bacteriophages Is Linked to Aggravated Intestinal Inflammation and Colitis. *Cell Host Microbe* 25, 285-299.e8.

González-Tortuero, E., Sutton, T.D.S., Velayudhan, V., Shkoporov, A.N., Draper, L.A., Stockdale, S.R., Ross, R.P., and Hill, C. (2018). VIGA: a sensitive, precise and automatic de novo Viral Genome Annotator. *BioRxiv* 277509.

Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize Implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812.

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849.

Guerin, E., Shkoporov, A., Stockdale, S.R., Clooney, A.G., Ryan, F.J., Sutton, T.D.S., Draper, L.A., Gonzalez-Tortuero, E., Ross, R.P., and Hill, C. (2018). Biology and Taxonomy of crAss-like Bacteriophages, the Most Abundant Virus in the Human Gut. *Cell Host Microbe* 24, 653-664.e6.

Hargreaves, K.R., Kropinski, A.M., and Clokie, M.R.J. (2014). What Does the Talking?: Quorum Sensing Signalling Genes Discovered in a Bacteriophage Genome. *PLOS ONE* 9, e85131.

Hawkins, S.A., Layton, A.C., Ripp, S., Williams, D., and Sayler, G.S. (2008). Genome sequence of the *Bacteroides fragilis* phage ATCC 51477-B1. *Virology Journal* 5, 97.

Hendrix, H., Kogadeeva, M., Zimmermann, M., Sauer, U., Smet, J.D., Muchez, L., Lissens, M., Staes, I., Voet, M., Wagemans, J., et al. (2019). Host metabolic reprogramming of *Pseudomonas aeruginosa* by phage-based quorum sensing modulation. *BioRxiv* 577908.

d'Hérelle, F. (1917). Sur un microbe invisible antagoniste des bacilles dysentériques. *CR Acad. Sci. Paris* 165, 373–375.

Hryckowian, A.J., Merrill, B.D., Porter, N.T., Treuren, W.V., Nelson, E.J., Garlena, R.A., Russell, D.A., Martens, E.C., and Sonnenburg, J.L. (2020). *Bacteroides thetaiotaomicron*-infecting bacteriophage isolates inform sequence-based host range predictions. *BioRxiv* 2020.03.04.977157.

Hyman, P. (2019). Phages for Phage Therapy: Isolation, Characterization, and Host Range Breadth. *Pharmaceuticals (Basel)* 12.

Jiang, X., Hall, A.B., Arthur, T.D., Plichta, D.R., Covington, C.T., Poyet, M., Crothers, J., Moses, P.L., Tolonen, A.C., Vlamakis, H., et al. (2019). Invertible promoters mediate bacterial phase variation, antibiotic resistance, and host adaptation in the gut. *Science* 363, 181–187.

Kai, M., Watanabe, S., Furuse, K., and Ozawa, A. (1985). *Bacteroides* bacteriophages isolated from human feces. *Microbiol. Immunol.* 29, 895–899.

Kang, I., Oh, H.-M., Kang, D., and Cho, J.-C. (2013). Genome of a SAR116 bacteriophage shows the prevalence of this phage type in the oceans. *Proc Natl Acad Sci U S A* *110*, 12343–12348.

Khan Mirzaei, M., Khan, M.A.A., Ghosh, P., Taranu, Z.E., Taguer, M., Ru, J., Chowdhury, R., Kabir, M.M., Deng, L., Mondal, D., et al. (2020). Bacteriophages Isolated from Stunted Children Can Regulate Gut Bacterial Communities in an Age-Specific Manner. *Cell Host Microbe* *27*, 199-212.e5.

Kieser, S., Sarker, S.A., Sakwinska, O., Foata, F., Sultana, S., Khan, Z., Islam, S., Porta, N., Combremont, S., Betrisey, B., et al. (2018). Bangladeshi children with acute diarrhoea show faecal microbiomes with increased *Streptococcus* abundance, irrespective of diarrhoea aetiology. *Environmental Microbiology* *20*, 2256–2269.

Klieve, A.V., Gregg, K., and Bauchop, T. (1991). Isolation and characterization of lytic phages from *Bacterioides ruminicola* ssbrevis. *Current Microbiology* *23*, 183–187.

Koh, G.Y., Kane, A., Lee, K., Xu, Q., Wu, X., Roper, J., Mason, J.B., and Crott, J.W. (2018). *Parabacteroides distasonis* attenuates toll-like receptor 4 signaling and Akt activation and blocks colon tumor formation in high-fat diet-fed azoxymethane-treated mice. *Int. J. Cancer*.

Kverka, M., Zakostelska, Z., Klimesova, K., Sokol, D., Hudcovic, T., Hrnčir, T., Rossmann, P., Mrazek, J., Kopečný, J., Verdu, E.F., et al. (2011). Oral administration of *Parabacteroides distasonis* antigens attenuates experimental murine colitis through modulation of immunity and microbiota composition. *Clinical & Experimental Immunology* *163*, 250–259.

Leighton Pritchard, Peter Cock, and YT (2017). widdowquinn/pyani v0.2.3 (Zenodo).

Lewis, J.D., Chen, E.Z., Baldassano, R.N., Otley, A.R., Griffiths, A.M., Lee, D., Bittinger, K., Bailey, A., Friedman, E.S., Hoffmann, C., et al. (2015). Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease. *Cell Host Microbe* 18, 489–500.

Lim, E.S., Zhou, Y., Zhao, G., Bauer, I.K., Droit, L., Ndao, I.M., Warner, B.B., Tarr, P.I., Wang, D., and Holtz, L.R. (2015). Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nature Medicine* 21, 1228–1234.

Lopetuso, L.R., Petito, V., Graziani, C., Schiavoni, E., Paroni Sterbini, F., Poscia, A., Gaetani, E., Franceschi, F., Cammarota, G., Sanguinetti, M., et al. (2018). Gut Microbiota in Health, Diverticular Disease, Irritable Bowel Syndrome, and Inflammatory Bowel Diseases: Time for Microbial Marker of Gastrointestinal Disorders. *Dig Dis* 36, 56–65.

Luo, E., Aylward, F.O., Mende, D.R., and DeLong, E.F. (2017). Bacteriophage Distributions and Temporal Variability in the Ocean's Interior. *MBio* 8.

Ma, Y., You, X., Mai, G., Tokuyasu, T., and Liu, C. (2018). A human gut phage catalog correlates the gut phageome with type 2 diabetes. *Microbiome* 6, 24.

Manrique, P., Bolduc, B., Walk, S.T., van der Oost, J., de Vos, W.M., and Young, M.J. (2016). Healthy human gut phageome. *Proceedings of the National Academy of Sciences* 113, 10400–10405.

Minot, S., Bryson, A., Chehoud, C., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2013). Rapid evolution of the human gut virome. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12450–12455.

- Ng, W.-L., and Bassler, B.L. (2009). Bacterial Quorum-Sensing Network Architectures. *Annual Review of Genetics* 43, 197–222.
- Nguyen, S., Baker, K., Padman, B.S., Patwa, R., Dunstan, R.A., Weston, T.A., Schlosser, K., Bailey, B., Lithgow, T., Lazarou, M., et al. (2017). Bacteriophage Transcytosis Provides a Mechanism To Cross Epithelial Cell Layers. *MBio* 8, e01874-17.
- Norman, J.M., Handley, S.A., Baldrige, M.T., Droit, L., Liu, C.Y., Keller, B.C., Kambal, A., Monaco, C.L., Zhao, G., and Fleshner, P. (2015). Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* 160, 447–460.
- Nurk, S., Bankevich, A., Antipov, D., Gurevich, A., Korobeynikov, A., Lapidus, A., Prjibelsky, A., Pyshkin, A., Sirotkin, A., Sirotkin, Y., et al. (2013). Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. In *Research in Computational Molecular Biology*, M. Deng, R. Jiang, F. Sun, and X. Zhang, eds. (Springer Berlin Heidelberg), pp. 158–170.
- Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Research* 27, 824–834.
- O'Donnell, M.M., Rea, M.C., O'Sullivan, Ó., Flynn, C., Jones, B., McQuaid, A., Shanahan, F., and Ross, R.P. (2016). Preparation of a standardised faecal slurry for ex-vivo microbiota studies which reduces inter-individual donor bias. *Journal of Microbiological Methods* 129, 109–116.
- Ogilvie, L.A., Caplin, J., Dedi, C., Diston, D., Cheek, E., Bowler, L., Taylor, H., Ebdon, J., and Jones, B.V. (2012). Comparative (Meta)genomic Analysis and

Ecological Profiling of Human Gut-Specific Bacteriophage ϕ B124-14. *PLOS ONE* 7, e35053.

Olbjørn, C., Cvancarova Småstuen, M., Thiis-Evensen, E., Nakstad, B., Vatn, M.H., Jahnsen, J., Ricanek, P., Vatn, S., Moen, A.E.F., Tannæs, T.M., et al. (2019). Fecal microbiota profiles in treatment-naïve pediatric inflammatory bowel disease – associations with disease phenotype, treatment, and outcome. *Clin Exp Gastroenterol* 12, 37–49.

Paepe, M.D., Hutinet, G., Son, O., Amarir-Bouhram, J., Schbath, S., and Petit, M.-A. (2014). Temperate Phages Acquire DNA from Defective Prophages by Relaxed Homologous Recombination: The Role of Rad52-Like Recombinases. *PLOS Genetics* 10, e1004181.

Papenfort, K., and Bassler, B.L. (2016). Quorum sensing signal–response systems in Gram-negative bacteria. *Nature Reviews Microbiology* 14, 576–588.

Parveen, N., and Cornell, K.A. (2011). Methylthioadenosine/S-adenosylhomocysteine nucleosidase, a critical enzyme for bacterial metabolism. *Mol Microbiol* 79, 7–20.

Petkau, A., Stuart-Edwards, M., Stothard, P., and Van Domselaar, G. (2010). Interactive microbial genome visualization with GView. *Bioinformatics* 26, 3125–3126.

Porter, N.T., Hryckowian, A.J., Merrill, B.D., Fuentes, J.J., Gardner, J.O., Glowacki, R.W.P., Singh, S., Crawford, R.D., Snitkin, E.S., Sonnenburg, J.L., et al. (2020). Multiple phase-variable mechanisms, including capsular polysaccharides, modify bacteriophage susceptibility in *Bacteroides thetaiotaomicron*. *BioRxiv* 521070.

- Puig, M., and Gironés, R. (1999). Genomic structure of phage B40-8 of *Bacteroides fragilis*. *Microbiology (Reading, Engl.)* 145 (Pt 7), 1661–1670.
- Quaiser, A., Dufresne, A., Ballaud, F., Roux, S., Zivanovic, Y., Colombet, J., Sime-
Ngando, T., and Francez, A.-J. (2015). Diversity and comparative genomics of
Microviridae in Sphagnum- dominated peatlands. *Front Microbiol* 6.
- Reyes, A., Blanton, L.V., Cao, S., Zhao, G., Manary, M., Trehan, I., Smith, M.I.,
Wang, D., Virgin, H.W., Rohwer, F., et al. (2015). Gut DNA viromes of Malawian
twins discordant for severe acute malnutrition. *PNAS* 112, 11941–11946.
- Roux, S., Enault, F., Hurwitz, B.L., and Sullivan, M.B. (2015). VirSorter: mining viral
signal from microbial genomic data. *PeerJ* 3.
- Sabiston, C.B., and Cohl, M.E. (1969). Bacteriophage virulent for species of the genus
Bacteroides. *J. Dent. Res.* 48, 599.
- Sakamoto, M., and Benno, Y. (2006). Reclassification of *Bacteroides distasonis*,
Bacteroides goldsteinii and *Bacteroides merdae* as *Parabacteroides distasonis* gen.
nov., comb. nov., *Parabacteroides goldsteinii* comb. nov. and *Parabacteroides merdae*
comb. nov. *Int. J. Syst. Evol. Microbiol.* 56, 1599–1605.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning: a laboratory
manual. *Molecular Cloning: A Laboratory Manual*.
- Shkoporov, A.N., and Hill, C. (2019). Bacteriophages of the Human Gut: The “Known
Unknown” of the Microbiome. *Cell Host & Microbe* 25, 195–209.
- Shkoporov, A.N., Khokhlova, E.V., Fitzgerald, C.B., Stockdale, S.R., Draper, L.A.,
Ross, R.P., and Hill, C. (2018a). ΦCrAss001 represents the most abundant

bacteriophage family in the human gut and infects *Bacteroides intestinalis*. *Nat Commun* 9, 1–8.

Shkoporov, A.N., Ryan, F.J., Draper, L.A., Forde, A., Stockdale, S.R., Daly, K.M., McDonnell, S.A., Nolan, J.A., Sutton, T.D.S., Dalmasso, M., et al. (2018b). Reproducible protocols for metagenomic analysis of human faecal phageomes. *Microbiome* 6, 68.

Shkoporov, A.N., Clooney, A.G., Sutton, T.D.S., Ryan, F.J., Daly, K.M., Nolan, J.A., McDonnell, S.A., Khokhlova, E.V., Draper, L.A., Forde, A., et al. (2019). The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific. *Cell Host & Microbe* 26, 527-541.e5.

Silpe, J.E., and Bassler, B.L. (2019). A Host-Produced Quorum-Sensing Autoinducer Controls a Phage Lysis-Lysogeny Decision. *Cell* 176, 268-280.e13.

Siringan, P., Connerton, P.L., Cummings, N.J., and Connerton, I.F. (2014). Alternative bacteriophage life cycles: the carrier state of *Campylobacter jejuni*. *Open Biol* 4, 130200–130200.

Sutton, T.D.S., and Hill, C. (2019). Gut bacteriophage: Current understanding and challenges. *Front. Endocrinol.* 10.

Tang, F., Bossers, A., Harders, F., Lu, C., and Smith, H. (2013). Comparative genomic analysis of twelve *Streptococcus suis* (pro)phages. *Genomics* 101, 336–344.

Tarkowski, T.A., Mooney, D., Thomason, L.C., and Stahl, F.W. (2002). Gene products encoded in the *ninR* region of phage λ participate in Red-mediated recombination. *Genes to Cells* 7, 351–363.

- Tartera, C., and Jofre, J. (1987). Bacteriophages active against *Bacteroides fragilis* in sewage-polluted waters. *Appl. Environ. Microbiol.* 53, 1632–1637.
- Tetz, G., and Tetz, V. (2018). Bacteriophages as New Human Viral Pathogens. *Microorganisms* 6.
- Turovskiy, Y., Kashtanov, D., Paskhover, B., and Chikindas, M.L. (2007). Quorum Sensing: Fact, Fiction, and Everything in Between. *Adv Appl Microbiol* 62, 191–234.
- Twort, F.W. (1915). An investigation on the nature of ultra-microscopic viruses. *The Lancet* 186, 1241–1243.
- Wang, K., Liao, M., Zhou, N., Bao, L., Ma, K., Zheng, Z., Wang, Y., Liu, C., Wang, W., Wang, J., et al. (2019). *Parabacteroides distasonis* Alleviates Obesity and Metabolic Dysfunctions via Production of Succinate and Secondary Bile Acids. *Cell Reports* 26, 222-235.e5.
- Warwick-Dugdale, J., Buchholz, H.H., Allen, M.J., and Temperton, B. (2019). Host-hijacking and planktonic piracy: how phages command the microbial high seas. *Virology* 16, 15.
- Whitehead, N.A., Barnard, A.M.L., Slater, H., Simpson, N.J.L., and Salmond, G.P.C. (2001). Quorum-sensing in Gram-negative bacteria. *FEMS Microbiol Rev* 25, 365–404.
- Yang, F., Kumar, A., Davenport, K.W., Kelliher, J.M., Ezeji, J.C., Good, C.E., Jacobs, M.R., Conger, M., West, G., Fiocchi, C., et al. (2019). Complete Genome Sequence of a *Parabacteroides distasonis* Strain (CavFT hAR46) Isolated from a Gut Wall-Cavitating Microlesion in a Patient with Severe Crohn's Disease. *Microbiol Resour Announc* 8.

Zhao, G., Droit, L., Gilbert, M.H., Schiro, F.R., Didier, P.J., Si, X., Paredes, A., Handley, S.A., Virgin, H.W., Bohm, R.P., et al. (2019). Virome biogeography in the lower gastrointestinal tract of rhesus macaques with chronic diarrhea. *Virology* 527, 77–88.

Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A.N., and Alva, V. (2017). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology*.

Zitomersky, N.L., Coyne, M.J., and Comstock, L.E. (2011). Longitudinal analysis of the prevalence, maintenance, and IgA response to species of the order Bacteroidales in the human gut. *Infect. Immun.* 79, 2012–2020.

Zuo, T., Lu, X.-J., Zhang, Y., Cheung, C.P., Lam, S., Zhang, F., Tang, W., Ching, J.Y.L., Zhao, R., Chan, P.K.S., et al. (2019). Gut mucosal virome alterations in ulcerative colitis. *Gut* 68, 1169–1179.

Thesis summary and future work

Phages form an important and abundant component of the human gut. It has become apparent that their interactions with bacteria within the complex gut ecosystem have important implications for homeostasis and disease. In the past few decades, significant progress has been made towards improving our understanding of the bacterial residents of our gut, however, our knowledge on the role played by phages is significantly lacking in comparison. Essential to expanding this is the *in vitro* and *in vivo* isolation and characterisation of novel gut phages and their hosts.

Chapter I provides a review of our current understanding of the human gut phageome. Current methodologies and hurdles in the context of both bioinformatic and laboratory analyses and possible solutions to these issues are discussed. A case study examining the crAss-like phage family highlights the importance of novel phage discovery, mining and isolation in expanding our insights into the gut phageome. In addition, the merits of phage research are discussed.

Chapter II examines the diversity of crAssphage variants, the most abundant phage type in the human gut. This was performed through bioinformatic analyses of human faecal phageome datasets leading to the development of a taxonomic classification scheme for the crAss-like phage family. The crAss-like phages were detected in the analysed datasets using the prototypical crAssphage DNA polymerase and terminase proteins as genetic signatures. The crAss-like phages were then clustered based on percentage of shared orthologous proteins. This led to the *de novo* assembly of 244 crAss-like phages. A total of 249 crAss-like phages were resolved into four sub-families and 10 candidate genera. This work also demonstrates the first

example of *ex vivo* propagation of prototypical crAssphage in a fermenter which was inoculated with faeces from a donor that persistently carried crAssphage at high abundances. Sequencing revealed the presence of six additional crAss-like phages in the fermenter that resolved into five of the new genera as defined in this study. It is likely that further lineages of this family will be identified in years to come.

Chapter III focuses on the *in vitro* isolation of crAss-like phages and associated Bacteroidales hosts using the same faecal donor as recruited in chapter II. Antibiotic enrichment was successfully implemented in driving the expansion of Bacteroidales and parallel crAss-like phage propagation in a faecal fermenter. The crAss-rich fermentate was sequenced confirming this expansion and revealed the presence of seven crAss-like phages. Primers were developed to target these phages with the goal of detecting their expansion on a specific host *in vitro* by qPCR. The crAss-rich fermentate was co-cultured with individual Bacteroidales cultures isolated from the crAssphage-rich faeces. Propagation of a candidate genus IV crAss-like phage, Φ crAss002, was detected on *Bacteroides xylanisolvens* by qPCR. Interestingly, this phage did not form plaques nor lyse liquid cultures of its host despite attaining high titres. Therefore, focusing on solely traditional properties may have hindered the detection of this phage.

Chapter IV applies the same approach for phage-host pair enrichment with a particular focus on isolating novel Bacteroidales phages due to the importance of this bacterial order in the human gut. However, here a more traditional screening approach was implemented using host lysis by spot assays as an indication of phage-host pairing. Successive enrichment and sequencing revealed the detection of a novel virulent siphovirus that infects *Parabacteroides distasonis*, Φ PDS1. Although Φ PDS1

could form visible but incomplete zones of clearing on lawns of its host and periodically formed pinpoint plaques, it also failed to lyse host cultures at high titres.

Both phage isolates do not possess lysogeny modules or are not maintained within in their hosts but yet they are able to co-exist with their host over time despite being virulent in nature. This is consistent with the observed persistence of the virulent phage core in the healthy human gut. With qPCR showing both can efficiently propagate without lysing host cultures, this could be indicative of phenotypic heterogeneity within the host population that may possibly be mediated by phase variation. Future work would focus on gaining a better insight into the dynamics that allow these phage-host interactions to occur, using transcriptomics and more complex *ex vivo* or *in vivo* models to further extend the above findings. This may help provide important insights into mechanisms that allow the co-existence of virulent phages and their hosts. This thesis also shows the importance of methodology in the screening and isolation of novel phage-host pairs and highlights the value of incorporating both traditional and non-traditional approaches. This can also be greatly aided by metagenomics which allows a more targeted experimental design. With further isolation of phages from the human gut and examination of the dynamics involved in bacterial host interactions, it is hoped that we will better understand how they influence our microbiome and health and in turn how we can manipulate them for therapeutic applications.

Acknowledgments

Firstly, I would like to thank my supervisor Professor Colin Hill and co-supervisor Professor Paul Ross for this research opportunity in the Gut Phageomics Lab. A special thanks to Colin for all your support, advice and guidance throughout my PhD.

I would like to extend my sincerest gratitude to Lorraine, both the best project manager and a great friend. I am so grateful to you for your kindness, support, encouragement and advice throughout the PhD. No problem ever seemed too big to solve thanks to you.

Next, I would like to thank all the members of the Gut Phageomics Lab for being the kindest colleagues and always so willing to help. In addition to all the work help, thank you for all the fun and banter throughout the years. Those memories are some of the highlights of my PhD. My future colleagues have a lot to live up to! I would like to thank each of you individually but that might add another hundred pages to this thesis. I would like to thank Andrey and Steve in particular. I am sincerely grateful to you both. Thank you, Andrey, for really getting me started in the lab, for everything you have taught me and for your guidance and patience throughout these past few years. Thank you, Steve, my desk neighbour, for always being so willing to help me no matter how busy you were. Thanks for sparing the time to answer my questions, teach me so much, especially bioinformatics, and for always being so patient with me.

Thank you to Ashley, Jenna and Angeliki for your friendship as well as your advice and support. Our chats over coffee (which I will miss) were always a great

boost on the tougher days. Thanks to Tom, with whom I started the PhD and somehow, we managed to make it. Thanks for always providing great advice in the most stressful moments, for the bioinformatic help and for all the laughs. Thanks to Katia and Karen for always helping me out with even the smallest of things as well as always being so caring and up for a laugh. I want to thank Julie for being so considerate, thoughtful and for our lovely afternoon walks. Thanks to Ciara and Joan for their patience with me when working with the fermenters. A warm thanks to the technical staff of the microbiology department. Furthermore, I would like to thank all my flatmates who have always been so supportive of me, in particular Bianca.

A very warm thank you to everyone at Sicilian Delights, especially Marco, for keeping me energised with your amazing coffee, pistachio biscuits and delicious food! My little daily escape from the PhD, grazie mille a voi.

A very special thanks to my parents, brother, sister and cousin Caroline for their support, optimism and enthusiasm throughout my PhD. Thanks for your interest and efforts to understand my work.

Grazie mille dal profondo del cuore a te Ale. Ti ringrazio per la pazienza e per avermi supportata sempre durante questi anni. Ti sono grata tantissimo. Aspetto con ansia il prossimo capitolo per noi!

“Non ci sono scorciatoie verso qualsiasi posto in cui valga la pena di andare”

“Det är med goda idéer som med svamp; där man finner en finns det oftast flera”.