Published in final edited form as:
Nat Rev Microbiol. 2012 September ; 10(9): 607–617. doi:10.1038/nrmicro2853.

Going viral: next generation sequencing applied to human gut phage populations

Alejandro Reyes[1], **Nicholas P. Semenkovich**[1], **Katrine Whiteson**[2], **Forest Rohwer**[2], and **Jeffrey I. Gordon**[1]

[1]Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis MO 63108

[2]Department of Biology, San Diego State University, San Diego, CA 92182

Abstract

Over the past decade researchers have begun to characterize viral diversity using metagenomic methods. These studies have shown that viruses, the majority of which infect bacteria (bacteriophages), are likely the most genetically diverse components of the biosphere. Here we briefly review the incipient rise of a phage biology renaissance catalyzed by recent advances in next generation sequencing. We explore how work characterizing phage diversity and their lifestyles in the gut is changing our view of ourselves as supra-organisms. Finally, we discuss how a new appreciation of phage dynamics may yield new applications for phage therapies designed to manipulate the structure and functions of our gut microbiomes.

Introduction

From Alfred Hershey and Martha Chase's studies indicating that DNA was the genetic material [1], to Francis Crick and Sydney Brenner's experiment establishing the triplet nature of the genetic code [2], bacteriophages ("phages") have helped define fundamental components of modern biology. Most of the tools for early molecular biology arose from the work of phage biologists [3]. The first genomes sequenced were from phages and other viruses. The first comparisons of multiple genomes were carried out on *Lactobacillus* and *Mycobacteria* phages. These early studies showed that there was extensive diversity in essentially every phage community. Now it is clear that viruses are the most diverse and uncharacterized components of the major ecosystems on Earth [4], and that viruses have intricate roles in ecosystem function, far beyond simple predator-prey dynamics [5] (Box 1).

Box 1

Phage-bacterial host cell dynamic: lessons learned from environmental ecosystems

Most of life on Earth exists as Bacteria and Archaea in the ocean, sediments, land, and potentially the deep biosphere [117]. In the early 1980s, people became aware that there are literally millions of actively growing microbes per milliliter of seawater. Marine phages were subsequently rediscovered in 1989 and efforts to characterize the impact of phage lifecycles on planetary-scale biogeochemistry were initiated [118–121]. As an example of how dramatic phage effects can be, consider the widespread cyanobacteria clades *Prochlorococcus* and *Synechococcus*. These two unicellular algae carry out about half of the world oceans' primary production. They are infected by cyanophages: variations in

Address correspondence to: jgordon@wustl.edu.

cyanophage and bacterial concentrations are tied to daily and seasonal cycles [120]. In the 1990s, isolation and characterization of phages infecting *Synechococcus* and later *Prochlorococus* not only revealed that cyanophages are widespread, often infecting 40–50% of cyanobacteria, but that they kill 10–50% of their hosts daily, [119, 120] rapidly driving the diversification of their hosts as they co-evolve resistance [122] and simultaneously driving carbon into a dissolved form when bacterial cells lyse. Under lower resource conditions, cyanophages may enter a lysogenic state. A growing list of genes that are important for bacterial host metabolism and function have been found in marine phages, including photosystem genes that can increase photosynthetic output and maintain energy production during infection so that phages can lyse their host cells [51, 123–125]. Lysogeny may be an important lifestyle under a number of suboptimal conditions, including when host abundance or nutrient abundance is low [126].

Bacteria can use temperate phage to enable invasion of new habitats by sacrificing part of the population through phage lysis. The released phage will target competitors but allow bacterial kin harboring the prophage to survive since they are resistant to attack by a process called superinfection exclusion[127–130].

The concept that a virus can have a beneficial effect beyond that experienced by its host cells is not conceptually novel. Three-way symbioses have been well described in macro- and micro-ecosystems. For example, a symbiotic bacterium that inhabits pea aphids protects them from a wasp that can otherwise lay eggs in the haemocoel of the aphids; a phage-encoded toxin expressed by the bacteria confers this protection [131, 132]. Drought, heat and cold tolerance are conferred to plants through viruses [127, 133]. In Yellowstone National Park, a fungal endophyte infecting panic grass confers thermal tolerance, allowing the grass to grow in hot geothermal soils. The fungus alone is not heat tolerant without the virus that infects it [134]. In a more phage-related example, there is a reported case of phage-associated corynetoxin synthesis in the bacterium *Rathayibacter toxicus* (formerly *Clavibacter toxicus*) that colonizes ryegrass plants; the toxin makes the grass toxic to grazing animals such as sheep [135].

At the same time, the clinical world has become increasingly interested in phage-based therapeutics because of the increased prevalence of antibiotic resistant bacteria [6]. The idea of using phages as therapeutic tools is not new. Félix d'Herelle, co-discoverer of phage, recognized the potential medical applications nearly a century ago [7] and his first phage therapies were tested as early as 1919 [8]. However, a rudimentary understanding of the composition and dynamic operations of the human microbiome, a lack of knowledge of phage biology, and poor quality control during production of phages made this therapeutic approach unreliable [9].

This Review details recent advances on the rising field of viral metagenomics with an emphasis on our current views of phage communities associated with the human gut. It does not discuss many of the resistance mechanisms such as Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs), or the extremely large field of eukaryotic viral discovery that is being propelled by metagenomics, since these topics have been reviewed elsewhere [10–14]. Instead, we focus on viruses that infect bacteria. Box 2 provides examples of some challenges facing phage metagenomics at the present time, some of which are discussed below.

Box 2

Fundamental technical challenges in viral metagenomics

- New and better tools for recovery of VLPs from small amounts of starting microbial community biomass, and methods for less biased amplification of extracted nucleic acids before shotgun sequencing.
- Improved methods for deep-draft assemblies of full-length viral genomes. Particularly problematic are the ends of phage genomes, which can be blocked, permuted, or have hairpins.
- Automation of methods (e.g., MaxiPhi) for performing comparative metagenomics and estimating beta and gamma-diversity, in order to describe the pan-virome in a given environment.
- New and better tools for defining the host specificity of known and novel phages from either assembled genomes or VLP-derived short read sequences, and for identifying the determinants of microbial host cell range.
- Improved methods of experimentally and computationally assigning functions to 'conserved' viral genes with no known functions (illuminating the 'genetic dark matter' represented by conserved hypothetical genes).
- Better *in vitro* and *in vivo* models for determining the phage-bacterial host dynamics and its impact on energy availability and niche partitioning in a microbiota.
- Experimental and computational methods, and related visualization tools, for efficient analyses of temporal variation in model viral-bacterial communities and for measuring the effects of perturbations.
- Models for predicting the cost/benefit of having prophage present in a candidate probiotic species; e.g., weighing invasiveness and persistence in a targeted microbiota.
- Methods to test the utility of phages to directly perturb a targeted microbiota in ways that facilitate invasion by a probiotic species or species consortium.

Methods for viral metagenomics

The introduction of small subunit (SSU) ribosomal RNA (e.g. 16S rRNA) as a reliable prokaryotic phylogenetic marker [15–17] opened the door to remarkable insights about microbial community diversity and dynamics. Phylogenetic marker 'envy' rapidly 'infected' the psyche of phage biologists: they did not have an SSU rRNA equivalent and there was, and is, no conserved protein or gene enabling a similar characterization of all or the majority of phages present in a sample. Efforts to characterize phage diversity focused on characterizing partially conserved fragments of phage genes such as polymerases. However, this method was useful only within certain viral families [18–20]. Horizontal gene transfer further complicates the use of marker genes in phages. For example, most Caudovirales have identifiable conserved genes including terminases, portal proteins, capsid proteins, but horizontal transfer and recombination events generate extensive genome mosaicism that challenges phylogenic characterization [21]. The arrival of next generation sequencing (NGS) together with methods for purifying virus-like particles (VLPs) set the stage for defining viral diversity based on shotgun sequencing.

Purification of VLPs

Although viral particles outnumber microbial cells 10:1 in most environments, viral DNA represents 2–5% of total community DNA [22–24]. For this reason, it is often desirable to separate viruses from microbial cells. If sample volume is large and viral density low (such as in ocean environments), tangential flow filtration can be used to remove large particles and concentrate VLPs. For solid samples with high viral density e.g., feces, a common approach is to resuspend the material in an osmotically neutral buffer followed by one or more steps designed to remove large particles (e.g., cesium chloride density gradient ultracentrifugation [25] and subsequent filtration) (Fig. 1). This procedure has been successfully applied to fecal material that had been stored at −80°C for several years, without any pre-processing of the sample, indicating that VLP structures are quite stable to freezing and thawing [22].

Amplification of VLP-derived DNA

After VLPs are purified, non-encapsulated free nucleic acids are removed by treatment with DNase and RNase, and VLP nucleic acids are then isolated. The methods chosen will determine purity, influence DNA and RNA yields, and represent a selection step that can bias interpretation of virotype abundance and viral community diversity [26]. Unfortunately, the yield of DNA following extraction of nucleic acids from purified VLPs is often below the required minimum for sequencing. Therefore, a variety of amplification methods have been developed such as random amplified shotgun library RASL [27], linker-amplified shotgun library LASL [28] and others [29–32]. Caveats concerning random PCR amplification of viral DNA include inherent bias due to exponential amplification of mixed templates, uneven coverage of viral genomes, and its limitation to dsDNA templates. Another common method uses the phage-derived phi29 polymerase for multiple displacement amplification (MDA) [33]. MDA takes advantage of the high processivity of this DNA polymerase (>70,000 nucleotides) and its strong strand-displacement capability, which permits amplification of complete viral genomes. The result is a very fast method that can efficiently amplify minute amounts of both ssDNA and dsDNA. While fast, the method is not without flaws, including over-amplification of small circular ssDNA viruses [34] and potential chimera formation [35, 36]. Procedures for avoiding some of these limitations continue to be developed [37, 38], including a novel transposon-based method for rapidly generating DNA libraries from small quantities of dsDNA [39]. RNA viruses can be sequenced by reverse transcription followed by application of the protocols described above. Alternatively, Whole Transcriptome Amplification (WTA) approaches can also be used [40].

Sequencing strategies

Although sequencing costs are falling at an astonishingly rapid rate as newer technologies offer higher degrees of parallelism (greater numbers of reads per run; multiplex sequencing using sample-specific DNA barcodes), read length matters [41]. When characterizing a viral community where most of the sequences obtained are novel and enriched in regions of low complexity repeats, accurate assembly and taxonomic assignment benefit from the longest reads possible [41, 42]. The earliest NGS analyses were powered by the 454 GS20 instrument with ~100 nt reads. Advances in pyrosequencing technology, including today's FLX+ platform, have produced average read lengths that exceed 800 nt [22, 23, 43–46]. Most recently, analysis of metagenomic datasets generated by deep sequencing of total microbial community DNA with highly parallel Illumina instruments showed that the percentage of reads with similarity to known viral sequences was generally less than 0.01% [30, 47, 48]. This low value is in part due to the short read length (≤100 nt). However, the percentage increases when VLPs are purified [32]. Other studies have obtained better assignments by pre-assembling the short reads [49, 50].

In summary, viral metagenomics has opted for technologies that prioritize long read length over platforms with short read length and higher throughput. However, as the latter platforms approach 250nt read lengths and 250 million reads per lane (Illumina HiSeq 2500), and the cost per read falls, we will undoubtedly see a very rapid transition to these types of sequencers — as long as improvements in assembly algorithms keep pace (Fig. 1).

Computational approaches for characterizing sequenced viromes

To address the question of viral community composition, shotgun metagenomic sequences are typically compared to individual viral genomes. Although public sequence databases have expanded considerably — from 500 viral genomes as of 2007 to over 3000 full viral genomes to date — the number of deposited genomes is far less than the expected number of virotypes present in 100 liters of seawater [51]. Compounding the problem, existing databases include very few viral proteins in their training sets, meaning many taxonomic assignments are based on proteins transferred between a virus and a microbial host or that are present in prophages and described as part of a microbial genome. Databases with a particular focus on viruses are under development and include a CLAssification of Mobile genetic Elements (ACLAME) [52] and the Phage SEED [53]. Novel data analysis pipelines are also being constructed to improve the accuracy and efficiency of homology searches [54–61].

Once taxonomic and functional assignments have been made for a given sample, a viral community profile can be created that characterizes the diversity present in that sample. Multi-dimensional reduction methods such as Principal Coordinate Analysis (PCA) and hierarchical clustering have been used to visualize similarities among viral communities, and methods such as supervised learning can help to identify discriminatory features.

Given that most of the viral metagenomic data lack similarity to entries in databases, similarity-independent methods have been developed to better understand viral community structure. One example, PHACCS (Phage Communities from Contig Spectrum), was designed to quantify virotypes [62, 63] based on the assumption that if a virotype is present in high abundance in a VLP sample, it is more likely to be assembled into a large contig. Moreover, we can posit that if assembly of a single sample dataset allows prediction of community structure and diversity, pooling two samples together and performing a cross-assembly analysis can determine inter-sample diversity (e.g., using MaxiPhi [64]). Another alternative for identifying shared viruses among different samples comes from crAss [65], an algorithm that allows for the simultaneous cross-assembly of all the samples in a dataset as opposed to the pairwise assemblies used in MaxiPhi. As more tools are developed, special attention should be given to the assembly parameters to prevent mixed assemblies and chimeras between viral genomes (Fig. 1).

Phages in the human gut

The gut provides an enticing place to characterize the role of phages in community assembly and dynamics. Assembly of the human gut microbiota begins at birth with evolution toward an adult-like configuration during the first three years of life [66]. The importance of early environmental exposures is emphasized by the fact that the overall phylogenetic composition of the gut microbiota of adult monozygotic twins is not significantly greater than that of dizygotic twins, and family members share a higher degree of similarity than unrelated individuals living in different households. These patterns are robust to different cultural traditions, and the observations about mono- versus dizygotic twins apply to infants and children as well as teenagers and adults [66, 67]. Microbes in this densely populated ecosystem are engaged in a constant fight for nutrients and survival. Peristalsis moves an ephemeral menu of dietary components along the cephalocaudal axis of the intestine and microbial members face the omnipresent threat of washout from the gut ‘bioreactor’.

Maintaining a foothold in this ecosystem depends not only on physical interactions with the perpetually renewing mucus layer and partially digested food particles, but also on functional interactions with other community members. Preserving functional redundancy contributes to community resilience, with horizontal gene transfer providing an opportunity to constantly refashion the genomes (and by extension the pan-genomes) of species-level phylotypes. Each adult appears to harbor a persistent collection of one or at most a few hundred species in their intestines, although strain level diversity is great [24, 68,69]. While the proportional representation of taxa changes as the community responds to various environmental perturbations, intrapersonal variation in species content is considerably less than interpersonal differences [66, 67, 70, 71].

As our knowledge of inter- and intrapersonal variations in the microbiota expands, a lagging question has been the role of phages in shaping community properties. Although a number of individual phages have been extensively characterized (providing an important genomic context against which metagenomic data can be interpreted), recently much more attention has been given to phage dynamics at a microbial community level. In 2003, the first report of a human-associated gut virome was published; it described the results of shotgun (Sanger) sequencing of VLPs isolated from a fecal sample obtained from a single healthy adult. The identifiable fraction of the virome was dominated by phages, including temperate phages ('prophages' are defined here as temperate phages in their host-incorporated state). This report estimated that there were 1200 different virotypes in the single sample analyzed, with the majority assigned to the Siphoviridae family [28]. Siphoviridae and temperate phages have subsequently been reported to be the most abundant identifiable viruses in other sampled fecal viromes, followed by members of the Podoviridae [22, 23, 72].

The prominence of Microviridae in adult human gut microbiota was initially dismissed as an artifact of the MDA method, which has a preference for ssDNA. However, a novel branch of the Microviridae has been identified recently from prophages present in the genomes from members of two genera in the Bacteroidetes: *Bacteroides* and *Prevotella* [73]. These two genera are prominently represented in the microbiota of adult human populations living in a number of diverse geographic areas [66, 74]. Another study characterizing Microviridae from healthy human donors also clustered these novel viruses with Bacteroides and Prevotella prophages [72], suggesting that Microviridae could be an important viral family in the human gut and that what was previously considered to be an exclusively lytic phage can integrate into bacterial hosts in an environment that encourages a temperate (lysogenic) viral-host lifestyle (see Fig. 2 and below).

Marine environments can contain millions of different virotypes in a single sample [51]. None of the human fecal samples characterized thus far has had greater than 1500 virotypes. Moreover, the ratio of virotypes to species-level bacterial phylotypes in the ocean is 10:1 but closer to 1:1 in the gut [22]. Microscopy counts have further validated these estimated ratios, demonstrating an average of 10^8–10^9 VLPs per gram of feces compared to ~10^9 bacterial cells per gram of feces[72]. These findings also support the notion that phage exhibit a more temperate lifestyle in the gut, in contrast to the active kill-the-winner viral-bacterial dynamic manifest in marine environments.

The temperate lifestyle of phages observed in the gut, along with bacterial CRISPR elements involved in conferring immunity to infection with foreign DNA (including phage), have facilitated bioinformatics efforts to discover new phages in this body habitat. Recently, Stern et al, [75] used datasets obtained from deep shotgun sequencing of human fecal community DNA[24] to extract CRISPR spacers present in bacterial genomes. These spacers were then used to query datasets of pyrosequencer reads produced by shotgun sequencing of fecal VLP-derived DNA[22, 23] as well as contigs assembled from VLP and total fecal community

DNAs. The results led to identification of contigs in shotgun sequencing datasets of fecal microbiomes that were not previously recognized as viral, assignment of these sequences to bacterial hosts, and an appreciation of the wide distribution of some of these novels phages across human gut communities.

Temporal variation

To date, only three reported metagenomic studies of the gut DNA virome have characterized temporal variation [22, 23, 76]. One of these studies used VLPs isolated from frozen fecal samples collected from four adult female monozygotic twin pairs and their mothers at three time points over a 12-month period [22]. VLP-derived virome datasets were compared to datasets of sequenced bacterial 16S rRNA genes and shotgun reads from total fecal community DNA generated from the same fecal samples used to prepare VLPs. The results disclosed that the viromes of co-twins and their mothers exhibited a significantly greater degree of interpersonal variation than did the corresponding bacterial communities. Despite the marked interpersonal variation in gut viromes and their encoded gene functions, intrapersonal diversity was extremely low, with >95% of virotypes retained over the period surveyed, and with DNA viromes dominated by a few temperate phages that exhibited remarkable genetic stability (>99% sequence conservation). These observations suggested that a temperate viral lifestyle is more prevalent in the distal intestine than a kill-the-winner dynamic (see Fig. 2).

Another study of temporal variation involved adults subjected to a defined diet for a period of eight days [23]. During this time, both fecal bacterial and viral communities changed in a comparable manner. Importantly, interpersonal variation at the late time points was reduced among individuals consuming the same diet, suggesting that diet has an important effect in shaping both bacterial and viral communities.

A third study examining temporal variation characterized the DNA virome of a one week old healthy infant and used DNA microarrays to compare relative viral abundances in the fecal microbiota between postnatal weeks one and two[76]. The results showed that at early stages of life, the viral population changes drastically: over half of the virotypes present at one week were undetectable at two weeks. While contrasting with the stability documented in the fecal DNA viromes of healthy adults, these results are consistent with the dynamic and rapid nature of assembly of the infant bacterial microbiota [66, 77].

Functions encoded in phage genomes

There are a number of examples of known phage-encoded host fitness factors in gut bacteria (e.g., lambda *bor* and *lom*, 933W Stx2, [78, 79]), but most of these appear to be virulence determinants of one kind or another. When comparing purified VLP viromes and fecal microbiomes in the study of monozygotic twins, phage exhibit enrichment for genes involved in anaerobic nucleotide synthesis, as well as cell wall biosynthesis and degradation [22]. Other distinctive features of phage genomes include genes that can alter bacterial receptors and prevent superinfection [80]. Interestingly, many phage receptors may be involved in carbohydrate transport and utilization. In an environment such as the gut, where carbohydrate utilization is an important fitness factor, mobilization of these genes by phages could endow their bacterial host with benefits (Fig. 2). There are likely a great number of bona-fide (metabolic) fitness factors encoded in phages that have yet to be characterized.

Intriguingly, new evidence suggests that carbohydrate-binding components of the human gut virome may change at an extremely high rate. A recent metagenomic study examining VLPs purified from fecal samples collected from 12 humans identified 51 hypervariable loci, areas

with mutation rates that are much higher than the rest of their viral genomes [50]. Protein structural predictions revealed that some do not have homology to known folds, some have similarity to Ig-superfamily proteins, and others have similarity to C-type lectin folds, which play a key role in carbohydrate binding. Moreover, these loci appear to be specifically targeted for mutation by a reverse transcriptase-based mechanism, perhaps suggesting a critical functional advantage provided by these hypervariable loci. It is tempting to speculate that these loci confer a selective advantage to phages, enabling immune evasion through IgA binding, or improving the chances of infecting a host cell in the rapidly changing conditions of the gut through adaptable binding to relevant environmental materials or bacterial surface receptors. There is a well-documented precedent for hypervariable loci conferring a fitness advantage in Bordetella phage by allowing tropism switching in the phage receptor-binding protein [81]. Is important to emphasize that the hypothesis that these loci may allow a phage to bind IgA or environmental ligands is speculative and needs experimental validation; non-receptor phage structural proteins have been found to contain Ig-like domains, which may aid in host binding by weakly interacting with the cell surface [82].

RNA virome

The RNA virome of two healthy adults has been characterized by purifying VLPs [83]. This study showed that most RNA viruses appeared to be consumed together with food. A pepper-associated virus (PMMV), comprised over 80% of the identifiable gut viruses. The only animal RNA virus observed was a picobirnavirus that had previously been found in the feces of healthy individuals as well as in patients with diarrhea: it has not been associated with any particular disease.

Comparative studies of gut viromes in other mammals

Comparative studies of different mammalian species represent a source of information about the effects of environmental factors, including diet [84, 85], and various host factors on phage diversity in the gastrointestinal tract. Extensive surveys have been conducted of viruses associated with different mammalian species in search of potential sources of zoonosis, the etiology of animal diseases, and to identify common mammalian viruses [13]. These studies, which includes mammals occupying very distinct habitats [29, 31, 32, 86–91], identified viruses from the same families identified in the human gut virome, further underscoring the prevalence and long-standing nature of the evolved viral-mammalian host relationship (Box 3). In all these studies, ssDNA viruses were ubiquitous, accompanied in some cases by positive-sense ssRNA enteric viruses [43, 90, 91].

Box 3

Characterizing the eukaryotic gut virome in healthy individuals

The eukaryotic virome can be considered from at least three different perspectives: viruses associated with the eukaryotic component of a gut microbiota, viruses associated with various human cell populations exposed to this microbiota, and viruses associated with ingested food. Metagenomic studies of healthy individuals are dominated by bacterial viruses with eukaryotic viruses either absent [23] or present at very low abundance [22, 28, 50, 72, 76]. RNA viruses are an apparent exception: while abundant, they appear to be largely derived from ingested food [83]. Our limited view of the eukaryotic gut virome is largely derived from studies that apply viral metagenomics to identify agents associated with gastrointestinal diseases [136–142]. These studies identified known enteric viruses (Adenovirus, Rotavirus, Enterovirus and Norovirus), novel members of Bocavirus, Picobirnavirus, Cosavirus, and Anellovirus that are potential human pathogens, as well as novel viruses that may be related to diet (Gyrovirus, Nodavirus and

members of the Dicistroviridae, Vigaviridae and Partitiviridae families). The high prevalence of eukaryotic ssDNA viruses in metagenomic datasets has led to new perspectives about the potential importance and the diversity of these small viruses [143]. Although identification of these viruses has come from symptomatic individuals, they have also been identified, at similar prevalence, in asymptomatic contacts of those with disease. The broad representation of these eukaryotic viruses in the human gut as well as other body habitats has prompted a call to consider the functional significance of the eukaryotic human virome [12].The almost ubiquitous presence of human viruses that are not phages and that have not been associated with any disease suggests that viruses, especially those acquired in early childhood, may be essential for proper immune system development, and that particular host genotypes or immunologic constraints may cause normally benign viruses to induce disease states.

An early survey of coliphages in cows, pigs, and humans [92] showed that they were present in titers of up to 10^7 VLPs per gram of feces and that temperate phages were the most common. Interestingly, humans and pigs (omnivores with simple guts) had higher counts of temperate coliphages than cows (herbivores with foregut fermentation chambers). In an independent study, estimates of phage diversity from bovine rumen fluid [93] suggested that up to 28,000 different virotypes could be present in titers as high as 10^9 VLPs per ml of sample, hinting at strikingly higher diversity and abundance compared to humans. In contrast to the large interpersonal variation observed in human gut viromes [22, 23], this latter study showed a high degree of similarity between the phage communities of cohabitating animals on a similar diet. Metagenomic studies in horses (herbivores containing a hindgut fermentative chamber) revealed an intermediate level of phage diversity between that documented in herbivorous foregut-fermenting ruminants and omnivorous humans with simple guts.

Together, these studies suggest that diet, gut physiology, and potentially the transit time of food, play important roles in determining the lifecycle (and diversity) of phages in the mammalian gut. Further dissection of these relationships requires manipulable, representative, and defined *in vivo* models. Moving in this direction, Maura and colleagues used mice to study the effects of a lytic enteric phage, observing stable long-term replication over three weeks [94]. As noted below, gnotobiotic mouse models may also be very informative.

Phage Therapy

Much has already been written about the history, successes, and failings of phage therapy. Most of the studies have focused on the use of lytic phages to destroy pathogenic bacteria [9596] (Fig. 3). Clinically oriented phage research began very soon after the discovery of phages, with Felix d'Herelle using phages to treat bacillary dysentery in a number of human patients [8]. This optimistic start, however, led to a number of misconceptions and missteps, both scientific and political, regarding the use of phages. d'Herelle incorrectly assumed that there was only one universally efficacious strain of lytic phage [97], though we now know phages exhibit exquisite host cell specificity. In the 1930s, pharmaceutical companies began distributing enormous amounts of lytic phages as generic antibacterial therapies, but in part because of the perceived universality of phages, they had very little knowledge of their product's components.

In retrospect, many of the commonly used phage preparations were destroyed by the organomercury preservatives added to the vials that contained them, or were contaminated with bacterial exotoxins secreted by the cultures used to generate them [98]. Inevitably these problems, along with manufacturing inconsistencies (the supposedly standardized strains of

phages would change from batch-to-batch) led to distrust among the medical and scientific community.

The recent resurgence of phages as possible therapeutics has been driven by a number of factors. The alarming prevalence of antibiotic-resistant strains of pathogenic bacteria, combined with the inexorable spread of antibiotic-degrading enzymes, such as the New Delhi metallo-beta-lactamase (NDM-1), have led to calls for new therapeutic strategies [99]. From a practical standpoint, antibiotic discovery efforts have produced few novel compounds over the past decade [100]. Phages are a promising tool as they are easy to manufacture, have good host specificity, and can be readily genetically manipulated. Moreover, resistance to phages may develop more slowly than to antibiotics, though the reasons for this are multifaceted [101]. Phage resistance can occur spontaneously in cultures (as frequently as 1 in 10^5 cells), but there can be fitness costs associated with resistance. In contrast, many forms of antibiotic resistance cannot occur spontaneously, but instead require introduction of a foreign DNA element. In many ways, addressing bacterial resistance is much easier with phages than with antibiotics because one can isolate different phages, or phages may spontaneously mutate to overcome host resistance.

Perhaps a more interesting question, in the context of community dynamics and our growing understanding of the virome and microbiome, is whether we can produce more subtle phenotypic shifts in an ecological niche. Rather than destroy a single pathogenic member of a community, lysogenic phages could be introduced to promote a community structure that is beneficial to both the human host and microbial community members (Fig. 3). For example, one could expand the capacity of the gut microbiome to degrade dietary components[102]. Similarly, phage could be used to introduce novel, beneficial traits to community members, such as those involving nutrient biosynthesis. In the latter circumstance, it may be difficult to introduce traits that are not purely beneficial to a lysogenic phage's bacterial host, as the energetic effects of synthesizing an unnecessary protein impose a selection pressure.

Given the potential power and replicating nature of phages, a number of questions must be addressed before they can be more widely adopted including issues related to bio-containment [101]. Although phages are frequently sold as viruses that 'can only infect bacteria', their safety has yet to be completely defined. The intravenous administration of phage (e.g., in the case of bacterial sepsis) is particularly complex given the immunogenicity of some preparations and rapid clearance of phage particles by the reticuloendothelial system of the spleen [103]. It is tempting to assume that other routes of administration, such as oral cocktails of phage to target the human gut microbiome, would not have such effects. However, phage DNA is detectable by PCR and FISH in serum shortly after oral consumption [104]. Other studies have provided evidence of trans-placental passage of phage [105]. There is data suggesting that enzymes transcribed from phage DNA can be expressed in mammalian cells [106], this finding has even led to attempts to use phage as gene therapy vectors [107, 108].

Despite these concerns, we are exposed to millions of phages every day, including the ones from our own microbiota, without significant observable harm. In this spirit, it is interesting to consider the potential 'therapeutic' use of phages in the context of current efforts to apply microbiome-directed therapies [109]. Questions that can be asked include whether it is beneficial or detrimental for bacterial taxa being considered as candidate probiotics to possess or lack prophages or whether phages should be deliberately administered coincident with or preceding introduction of a probiotic consortium to help create niches that promote successful invasion and engraftment of the consortium.

Future directions

There has been little experimental work done on the ecology of phages *in vivo*. Germ-free mice and mice mono-colonized with different strains of *E. coli*, including strains isolated from children with diarrhea, have been used to examine replication of T4 and T7 phages [110, 111]. Gnotobiotic mouse models of the human gut microbiota may not only provide better understanding of phage-host dynamics, but may also represent a potentially valuable tool for establishing a preclinical pipeline designed to evaluate the feasibility of phage therapy. Recent work has shown that transplanting intact uncultured human gut (fecal) microbial communities into gnotobiotic mice is efficient, capturing the majority of microbial diversity and microbiome-encoded functions present in the human donor's community in the recipient animals [112, 113]. Mice with replicated human gut microbiomes can be fed diets resembling those of the human donor to explore diet-microbiome-phage interactions. An additional refinement to this approach is to transplant sequenced collections of bacteria (some containing prophage) cultured from a given donor's fecal sample into recipient mice [113]. The effects of various perturbations of gnotobiotic mice harboring these microbiota on phage-bacterial dynamics can be studied over time under highly controlled conditions.

Yet another envisioned approach is to assemble defined communities of sequenced members of the human gut microbiota in formerly germ-free mice and then to deliberately stage phage attacks. VLPs prepared from human fecal samples and introduced into mice containing these sequenced collections of cultured microbes from human donors would allow investigators to directly determine the bacterial host specificity of the VLP-associated phage, as well as the effects of (i) the presence or absence of prophage in community members, (ii) the diet administered to the animals, and (iii) the phage's contribution to mammalian host physiology, including immune function. The impact of a staged phage attack on community structure and functions can also be defined in gnotobiotic animal models over time and as a function of location along the length of the gut, where available nutrient and energy resources vary considerably.

These model systems, coupled with observational data in human fecal samples, should help us understand how phages influence metabolism in the gut and possibly how to manipulate it. Obtaining answers will almost certainly demand new and more efficient methods for deliberately curing bacterial hosts of their prophage. It may also require the application of whole genome transposon mutagenesis methods that are married to next generation sequencing platforms [114] to identify the functional contributions of genes in a given prophage. While the journey ahead will certainly be demanding, approaches are in hand or can be envisioned that will help propel the 'new age of phage' forward so that long standing questions can be addressed and new insights can be obtained.

Acknowledgments

Work from the authors' labs described in this report was supported by grants from the NIH (DK78669, DK30292, DK70977 to JIG and GM095384 to FLR), and the Crohn's and Colitis Foundation of America. A.R. is the recipient of an International Fulbright Science and Technology Award. N.P.S is a member of Washington University's Medical Scientist Training Program (MSTP) funded by NIH grant GM007200. Due to space limits we were not able to cite many wonderful studies relevant to the topics covered.

Glossary of terms

Lytic phage cycle State whereby fast exponential viral replication is achieved by using the bacterial host's DNA replication machinery independently of

	bacterial replication. This leads to the synthesis of multiple viral particles per cell and eventually lysis of the bacterium.
Lysogenic phage	A state where linear (1:1) replication is achieved by a temperate phage through integration of its genome into the bacterial host chromosome (more rarely the phage exists as a plasmid). The integrated phage transcribes gene(s) that repress lytic action, and in some cases expresses genes that promote the fitness of the bacterial host.
Prophage	temperate phage in a host-incorporated state.
Transduction	transfer of DNA from one bacterial host to another by a phage; a common route for horizontal gene transfer.
Superinfection immunity	the ability of a prophage to block superinfection of its bacterial host from another phage due to expression of genes that directly modify the phage receptor or proteins that block the receptor preventing attachment from other phage with similar specificity.
CRISPR	A widespread genetic system in bacteria and archaea that consists of multiple copies of palindromic repeats flanking short spacers of viral or plasmid origin. CRISPR elements are believed to provide acquired resistance to foreign DNA.
Kill-the-winner dynamics	A model for the population dynamics of phage–bacteria interactions that postulates that an increase in a host population (the winner) is followed by an increase in its corresponding phage predator, resulting in an increase in the rate at which the winner is killed.
Pan-genome	The global gene repertoire defined by sequencing the genomes of multiple isolates of a given bacterial species-level phylotype. These isolates can be from a single habitat or multiple habitats.
Coliphage	A bacteriophage that infects coliform bacteria, in particular *Escherichia coli*.
Virus-like particle	Particles recovered from microbial communities using physical separation methods such as density gradient ultracentrifugation and/or filtration. Purified VLPs have physical characteristics resembling viruses although their capacity for infection has to be subsequently defined.
Multiple displacement amplification	A method for exponential isothermal amplification of a DNA template using Phi29 DNA polymerase and random primers. Exponential amplification is achieved by attachment of the polymerase to newly elongated fragments coupled to strong displacement activity of the enzyme upon extension.
Virotype	Taxonomic classification typically based on a selected percent identity threshold among viral reads, rather than on phylogenetic markers.
Bacterial phylotype	Taxonomy based on phylogenetic markers, classically, the 16S rRNA gene. Isolates can be arbitrarily assigned to a species-level phylotype if they share ≥97% sequence identity among their 16S rRNA genes.
Deep biosphere	The deepest oceanic regions where life is supported.

References

1. Hershey AD, Chase M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. J Gen Physiol. 1952; 36:39–56. [PubMed: 12981234]
2. Crick FH, Barnett L, Brenner S, Watts-Tobin RJ. General nature of the genetic code for proteins. Nature. 1961; 192:1227–1232. [PubMed: 13882203]
3. Carins, J.; Stent, GS.; Watson, JD. Phage and the Origins of Molecular Biology. Cold Spring Harbor Laboratory Press; Plainview, NY: 1992.
4. Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus discovery. Curr Opin Virol. 2012; 2:63–77. [PubMed: 22440968]
5. Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? Trends Microbiol. 2005; 13:278–284. [PubMed: 15936660]
6. Fernandes P. Antibacterial discovery and development--the failure of success? Nat Biotechnol. 2006; 24:1497–1503. [PubMed: 17160049]
7. d'Herelle F. Sur un microbe invisible antagoniste des bacilles dysenteriques. Comptes rendus Acad Sci Paris. 1917; 165:373–375.
8. Sulakvelidze A, Alavidze Z, Morris JG Jr. Bacteriophage therapy. Antimicrob Agents Chemother. 2001; 45:649–659. [PubMed: 11181338]
9. Levin BR, Bull JJ. Population and evolutionary dynamics of phage therapy. Nat Rev Microbiol. 2004; 2:166–173. [PubMed: 15040264]
10. Marraffini LA, Sontheimer EJ. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet. 2010; 11:181–190. [PubMed: 20125085]
11. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. Science. 2010; 327:167–170. An overview of CRISPR defense mechanisms against phage attack. [PubMed: 20056882]
12. Virgin HW, Wherry EJ, Ahmed R. Redefining chronic viral infection. Cell. 2009; 138:30–50. A discussion of the nuanced role of the immune system in chronic viral infections. [PubMed: 19596234]
13. Delwart E. Animal virus discovery: improving animal health, understanding zoonoses, and opportunities for vaccine development. Curr Opin Virol. 2012; 2:1–9. [PubMed: 22440959]
14. Haynes, M.; Rohwer, F. The Human Virome in Metagenomics of the Human Body. Nelson, KE., editor. 2011. p. 63-77.
15. Fox GE, et al. The phylogeny of prokaryotes. Science. 1980; 209:457–463. [PubMed: 6771870]
16. Lane DJ, et al. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci U S A. 1985; 82:6955–6959. [PubMed: 2413450]
17. Hugenholtz P, Goebel BM, Pace NR. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol. 1998; 180:4765–4774. [PubMed: 9733676]
18. Culley AI, Lang AS, Suttle CA. High diversity of unknown picorna-like viruses in the sea. Nature. 2003; 424:1054–1057. [PubMed: 12944967]
19. Breitbart M, Miyake JH, Rohwer F. Global distribution of nearly identical phage-encoded DNA sequences. FEMS Microbiol Lett. 2004; 236:249–256. [PubMed: 15251204]
20. Hambly E, et al. A conserved genetic module that encodes the major virion components in both the coliphage T4 and the marine cyanophage S-PM2. Proc Natl Acad Sci U S A. 2001; 98:11411–11416. [PubMed: 11553768]
21. Casjens SR. Comparative genomics and evolution of the tailed-bacteriophages. Curr Opin Microbiol. 2005; 8:451–8. [PubMed: 16019256]
22. Reyes A, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature. 2010; 466:334–338. The human fecal virome in healthy individuals is not highly shared between family members, and exhibits surprising stability over the course of a year. [PubMed: 20631792]
23. Minot S, et al. The human gut virome: inter-individual variation and dynamic response to diet. Genome Res. 2011; 21:1616–1625. A longitudinal study of the impact of controlled diet changes on the human gut virome. [PubMed: 21880779]

24. Qin J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010; 464:59–65. [PubMed: 20203603]
25. Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F. Laboratory procedures to generate viral metagenomes. Nat Protoc. 2009; 4:470–483. A protocol for isolating virus-like particles for subsequent metagenome characterization. [PubMed: 19300441]
26. Willner D, et al. Metagenomic detection of phage-encoded platelet-binding factors in the human oral cavity. Proc Natl Acad Sci U S A. 2011; 108 (Suppl 1):4547–4553. [PubMed: 20547834]
27. Rohwer F, Seguritan V, Choi DH, Segall AM, Azam F. Production of shotgun libraries using random amplification. BioTechniques. 2001; 31:108–112. [PubMed: 11464504]
28. Breitbart M, et al. Metagenomic analyses of an uncultured viral community from human feces. J Bacteriol. 2003; 185:6220–6223. [PubMed: 14526037]
29. Shan T, et al. The fecal virome of pigs on a high-density farm. J Virol. 2011; 85:11697–11708. [PubMed: 21900163]
30. Yozwiak NL, et al. Virus identification in unknown tropical febrile illness cases using deep sequencing. PLoS Negl Trop Dis. 2012; 6:e1485. [PubMed: 22347512]
31. Li L, et al. Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. J Virol. 2010; 84:6955–6965. [PubMed: 20463061]
32. Ge X, et al. Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in china. J Virol. 2012; 86:4620–4630. [PubMed: 22345464]
33. Hutchison CA, Smith HO, Pfannkoch C, Venter JC. Cell-free cloning using f29 DNA polymerase. Proc Natl Acad Sci U S A. 2005; 102:17332. [PubMed: 16286637]
34. Kim KH, et al. Amplification of uncultured single-stranded DNA viruses from rice paddy soil. Appl Environ Microbiol. 2008; 74:5975–5985. [PubMed: 18708511]
35. Lasken RS, Stockwell TB. Mechanism of chimera formation during the Multiple Displacement Amplification reaction. BMC Biotechnol. 2007; 7
36. Kim KH, Bae JW. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. Appl Environ Microbiol. 2011; 77:7663–7668. [PubMed: 21926223]
37. Andrews-Pfannkoch C, Fadrosh DW, Thorpe J, Williamson SJ. Hydroxyapatite-mediated separation of double-stranded DNA, single-stranded DNA, and RNA genomes from natural viral assemblages. Appl Environ Microbiol. 2010; 76:5039–5045. [PubMed: 20543058]
38. Fadrosh DW, Andrews-Pfannkoch C, Williamson SJ. Separation of single-stranded DNA, double-stranded DNA and RNA from an environmental viral community using hydroxyapatite chromatography. J Vis Exp. 2011
39. Marine R, et al. Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA. Appl Environ Microbiol. 2011; 77:8071–8079. [PubMed: 21948828]
40. Nakamura S, et al. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. PLoS ONE. 2009; 4:e4219. [PubMed: 19156205]
41. Wommack KE, Bhavsar J, Ravel J. Metagenomics: read length matters. Appl Environ Microbiol. 2008; 74:1453–1463. [PubMed: 18192407]
42. Bibby K, Viau E, Peccia J. Viral metagenome analysis to guide human pathogen monitoring in environmental samples. Lett Appl Microbiol. 2011; 52:386–392. [PubMed: 21272046]
43. Ng TF, et al. Broad surveys of DNA viral diversity obtained through viral metagenomics of mosquitoes. PLoS ONE. 2011; 6:e20579. [PubMed: 21674005]
44. Pasic L, et al. Metagenomic islands of hyperhalophiles: the case of Salinibacter ruber. BMC Genomics. 2009; 10:570. [PubMed: 19951421]
45. Vega Thurber RL, et al. Metagenomic analysis indicates that stressors induce production of herpes-like viruses in the coral Porites compressa. Proc Natl Acad Sci U S A. 2008; 105:18413–18418. [PubMed: 19017800]
46. Dinsdale EA, et al. Functional metagenomic profiling of nine biomes. Nature. 2008; 452:629–632. [PubMed: 18337718]

47. Yang J, et al. Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. J Clin Microbiol. 2011; 49:3463–3469. [PubMed: 21813714]
48. Xu B, et al. Metagenomic analysis of fever, thrombocytopenia and leukopenia syndrome (FTLS) in Henan Province, China: discovery of a new bunyavirus. PLoS Pathog. 2011; 7:e1002369. [PubMed: 22114553]
49. Coetzee B, et al. Deep sequencing analysis of viruses infecting grapevines: Virome of a vineyard. Virology. 2010; 400:157–163. [PubMed: 20172578]
50. Minot S, Grunberg S, Wu GD, Lewis JD, Bushman FD. Hypervariable loci in the human gut virome. Proc Natl Acad Sci U S A. 2012; 109:3962–3966. Hypervariable loci, targeted for mutation by a reverse transcriptase-based mechanism, are predicted to encode Ig-superfamily and C-type lectin folds. [PubMed: 22355105]
51. Rohwer F, Thurber RV. Viruses manipulate the marine environment. Nature. 2009; 459:207–212. Overview of the interactions between marine viruses and their hosts. [PubMed: 19444207]
52. Leplae R, Lima-Mendez G, Toussaint A. ACLAME: a CLAssification of Mobile genetic Elements, update 2010. Nucleic Acids Res. 2010; 38:D57–61. [PubMed: 19933762]
53. Edwards, R. Phages in the SEED Viewer. 2012. Accessed: http://www.phantome.org/PhageSeed/Phage.cgi
54. Sharon I, et al. Comparative metagenomics of microbial traits within oceanic viral communities. ISME J. 2011; 5:1178–1190. [PubMed: 21307954]
55. Ghosh TS, Mohammed MH, Komanduri D, Mande SS. ProViDE: A software tool for accurate estimation of viral diversity in metagenomic samples. Bioinformation. 2011; 6:91–94. [PubMed: 21544173]
56. Lorenzi HA, et al. TheViral MetaGenome Annotation Pipeline(VMGAP):an automated tool for the functional annotation of viral Metagenomic shotgun sequencing data. Stand Genomic Sci. 2011; 4:418–429. [PubMed: 21886867]
57. Meyer F, et al. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics. 2008; 9:386. [PubMed: 18803844]
58. Edwards, R.; Breitbart, M.; Elhai, J.; Sullivan, M. PHANTOME, PHage ANnotation TOols and MEthods. 2011. Accessed: http://www.phantome.org/
59. Bhavsar, J.; Polson, SW.; Wommack, KE. VIROME: An informatics resource for viral metagenome exploration. 2009. Accessed: http://virome.dbi.udel.edu
60. Roux S, et al. Metavir: a web server dedicated to virome analysis. Bioinformatics. 2011; 27:3074–3075. [PubMed: 21911332]
61. Sun S, et al. Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. Nucleic Acids Res. 2011; 39:D546–551. [PubMed: 21045053]
62. Breitbart M, et al. Genomic analysis of uncultured marine viral communities. Proc Natl Acad Sci U S A. 2002; 99:14250–14255. [PubMed: 12384570]
63. Angly F, et al. PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. BMC Bioinformatics. 2005; 6:41. [PubMed: 15743531]
64. Angly FE, et al. The marine viromes of four oceanic regions. PLoS Biol. 2006; 4:e368. [PubMed: 17090214]
65. Dutilh, BE., et al. Reference-independent comparative metagenomics using cross-assembly: crAss. 2012. Accessed: http://edwards.sdsu.edu/crass/
66. Yatsunenko T, et al. Human gut microbiome viewed across age and geography. Nature. 2012; 486:222–227. [PubMed: 22699611]
67. Turnbaugh PJ, et al. A core gut microbiome in obese and lean twins. Nature. 2009; 457:480–484. [PubMed: 19043404]
68. Turnbaugh PJ, et al. Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. Proc Natl Acad Sci U S A. 2010; 107:7503–7508. [PubMed: 20363958]

69. Hansen EE, et al. Pan-genome of the dominant human gut-associated archaeon, Methanobrevibacter smithii, studied in twins. Proc Natl Acad Sci U S A. 2011; 108 (Suppl 1): 4599–4606. [PubMed: 21317366]
70. Caporaso JG, et al. Moving pictures of the human microbiome. Genome Biol. 2011; 12:R50. [PubMed: 21624126]
71. Costello EK, et al. Bacterial community variation in human body habitats across space and time. Science. 2009; 326:1694–1697. [PubMed: 19892944]
72. Kim MS, Park EJ, Roh SW, Bae JW. Diversity and abundance of single-stranded DNA viruses in human feces. Appl Environ Microbiol. 2011; 77:8062–8070. [PubMed: 21948823]
73. Krupovic M, Forterre P. Microviridae goes temperate: microvirus-related proviruses reside in the genomes of Bacteroidetes. PLoS ONE. 2011; 6:e19893. [PubMed: 21572966]
74. Arumugam M, et al. Enterotypes of the human gut microbiome. Nature. 2011; 473:174–180. [PubMed: 21508958]
75. Stern A, Mick E, Tirosh I, Sagy O, Sorek R. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. Genome Res. 2012
76. Breitbart M, et al. Viral diversity and dynamics in an infant gut. Res Microbiol. 2008; 159:367–373. Describes the rapid assembly and unstable features of the gut virome following birth. [PubMed: 18541415]
77. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. PLoS Biol. 2007; 5:e177. [PubMed: 17594176]
78. Barondess JJ, Beckwith J. A bacterial virulence determinant encoded by lysogenic coliphage lambda. Nature. 1990; 346:871–4. [PubMed: 2144037]
79. Plunkett G 3rd, Rose DJ, Durfee TJ, Blattner FR. Sequence of Shiga toxin 2 phage 933W from Escherichia coli O157:H7: Shiga toxin as a phage late-gene product. J Bacteriol. 1999; 181:1767–78. [PubMed: 10074068]
80. Markine-Goriaynoff N, et al. Glycosyltransferases encoded by viruses. J Gen Virol. 2004; 85:2741–2754. [PubMed: 15448335]
81. Liu M, et al. Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. Science. 2002; 295:2091–4. [PubMed: 11896279]
82. Fraser JS, Yu Z, Maxwell KL, Davidson AR. Ig-like domains on bacteriophages: a tale of promiscuity and deceit. J Mol Biol. 2006; 359:496–507. [PubMed: 16631788]
83. Zhang T, et al. RNA viral community in human feces: prevalence of plant pathogenic viruses. PLoS Biol. 2006; 4:e3. [PubMed: 16336043]
84. Ley RE, et al. Evolution of mammals and their gut microbes. Science. 2008; 320:1647–1651. [PubMed: 18497261]
85. Muegge BD, et al. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science. 2011; 332:970–974. [PubMed: 21596990]
86. Cann AJ, Fandrich SE, Heaphy S. Analysis of the virus population present in equine faeces indicates the presence of hundreds of uncharacterized virus genomes. Virus Genes. 2005; 30:151–156. [PubMed: 15744573]
87. Donaldson EF, et al. Metagenomic analysis of the viromes of three North American bat species: viral diversity among different bat species that share a common habitat. J Virol. 2010; 84:13004–13018. [PubMed: 20926577]
88. Blinkova O, et al. Novel circular DNA viruses in stool samples of wild-living chimpanzees. J Gen Virol. 2010; 91:74–86. [PubMed: 19759238]
89. Ng TF, et al. Metagenomic identification of a novel anellovirus in Pacific harbor seal (Phoca vitulina richardsii) lung samples and its detection in samples from multiple years. J Gen Virol. 2011; 92:1318–1323. [PubMed: 21402596]
90. Phan TG, et al. The fecal viral flora of wild rodents. PLoS Pathog. 2011; 7:e1002218. [PubMed: 21909269]
91. van den Brand JM, et al. Metagenomic analysis of the viral flora of pine marten and European badger feces. J Virol. 2012; 86:2360–2365. [PubMed: 22171250]

92. Dhillon TS, Dhillon EK, Chau HC, Li WK, Tsang AH. Studies on bacteriophage distribution: virulent and temperate bacteriophage content of mammalian feces. Appl Environ Microbiol. 1976; 32:68–74. [PubMed: 987749]

93. Berg Miller ME, et al. Phage-bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. Environ Microbiol. 2012; 14:207–227. [PubMed: 22004549]

94. Maura D, et al. Intestinal colonization by enteroaggregative Escherichia coli supports long-term bacteriophage replication in mice. Environ Microbiol. 2011

95. Fischetti VA, Nelson D, Schuch R. Reinventing phage therapy: are the parts greater than the sum? Nat Biotechnol. 2006; 24:1508–1511. [PubMed: 17160051]

96. Lu TK, Koeris MS. The next generation of bacteriophage therapy. Curr Opin Microbiol. 2011; 14:524–531. [PubMed: 21868281]

97. van Helvoort T. The controversy between John H. Northrop and Max Delbruck on the formation of bacteriophage: bacterial synthesis or autonomous multiplication? Ann Sci. 1992; 49:545–575. [PubMed: 11616207]

98. Calendar, RL. The Bacteriophages. Oxford University Press; 2005.

99. Kumarasamy KK, et al. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. Lancet Infect Dis. 2010; 10:597–602. [PubMed: 20705517]

100. Piddock LJ. The crisis of no new antibiotics--what is the way forward? Lancet Infect Dis. 2012; 12:249–253. [PubMed: 22101066]

101. Clokie, MRJ.; Kropinski, AM. Bacteriophages: methods and protocols. Humana Press; Springer; London: 2009.

102. Hehemann JH, et al. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nature. 2010; 464:908–912. [PubMed: 20376150]

103. Geier MR, Trigg ME, Merril CR. Fate of bacteriophage lambda in non-immune germ-free mice. Nature. 1973; 246:221–223. [PubMed: 4586796]

104. Schubbert R, Renz D, Schmitz B, Doerfler W. Foreign (M13) DNA ingested by mice reaches peripheral leukocytes, spleen, and liver via the intestinal wall mucosa and can be covalently linked to mouse DNA. Proc Natl Acad Sci U S A. 1997; 94:961–966. [PubMed: 9023365]

105. Schubbert R, Hohlweg U, Renz D, Doerfler W. On the fate of orally ingested foreign DNA in mice: chromosomal association and placental transmission to the fetus. Mol Gen Genet. 1998; 259:569–576. [PubMed: 9819049]

106. Geier MR, Merril CR. Lambda phage transcription in human fibroblasts. Virology. 1972; 47:638–643. [PubMed: 4551993]

107. Barry MA, Dower WJ, Johnston SA. Toward cell-targeting gene therapy vectors: selection of cell-binding peptides from random peptide-presenting phage libraries. Nat Med. 1996; 2:299–305. [PubMed: 8612228]

108. Dunn IS. Mammalian cell binding and transfection mediated by surface-modified bacteriophage lambda. Biochimie. 1996; 78:856–861. [PubMed: 9116055]

109. Silverman MS, Davis I, Pillai DR. Success of self-administered home fecal transplantation for chronic Clostridium difficile infection. Clin Gastroenterol Hepatol. 2010; 8:471–473. [PubMed: 20117243]

110. Chibani-Chennoufi S, et al. In vitro and in vivo bacteriolytic activities of Escherichia coli phages: implications for phage therapy. Antimicrob Agents Chemother. 2004; 48:2558–69. [PubMed: 15215109]

111. Weiss M, et al. In vivo replication of T4 and T7 bacteriophages in germ-free mice colonized with Escherichia coli. Virology. 2009; 393:16–23. [PubMed: 19699505]

112. Turnbaugh PJ, et al. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci Transl Med. 2009; 1:6ra14.

113. Goodman AL, et al. Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. Proc Natl Acad Sci U S A. 2011; 108:6252–6257. [PubMed: 21436049]

114. Goodman AL, et al. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell Host Microbe. 2009; 6:279–289. [PubMed: 19748469]
115. Rodriguez-Valera F, et al. Explaining microbial population genomics through phage predation. Nat Rev Microbiol. 2009; 7:828–836. Discussion of the consequences of phage predation on microbial substrain diversity, presented as a constant-diversity dynamics model. [PubMed: 19834481]
116. Lu TK, Collins JJ. Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy. Proc Natl Acad Sci U S A. 2009; 106:4629–4634. [PubMed: 19255432]
117. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: the unseen majority. Proc Natl Acad Sci U S A. 1998; 95:6578–6583. [PubMed: 9618454]
118. Bergh O, Borsheim KY, Bratbak G, Heldal M. High abundance of viruses found in aquatic environments. Nature. 1989; 340:467–8. [PubMed: 2755508]
119. Clokie MR, Millard AD, Letarov AV, Heaphy S. Phages in nature. Bacteriophage. 2011; 1:31–45. [PubMed: 21687533]
120. Fuhrman JA. Marine viruses and their biogeochemical and ecological effects. Nature. 1999; 399:541–548. [PubMed: 10376593]
121. Azam F, et al. The ecological role of water column microbes in the sea. Mar Ecol Prog Ser. 1983; 10:257–263.
122. Marston MF, et al. Rapid diversification of coevolving marine Synechococcus and a virus. Proc Natl Acad Sci U S A. 2012; 109:4544–4549. [PubMed: 22388749]
123. Mann NH, Cook A, Millard A, Bailey S, Clokie M. Marine ecosystems: bacterial photosynthesis genes in a virus. Nature. 2003; 424:741. [PubMed: 12917674]
124. Sullivan MB, et al. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. PLoS Biol. 2006; 4:e234. [PubMed: 16802857]
125. Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW. Photosynthesis genes in marine viruses yield proteins during host infection. Nature. 2005; 438:86–89. [PubMed: 16222247]
126. Anderson RE, Brazelton WJ, Baross JA. Is the genetic landscape of the deep subsurface biosphere affected by viruses? Front Microbiol. 2011; 2:219. [PubMed: 22084639]
127. Roossinck MJ. The good viruses: viral mutualistic symbioses. Nat Rev Microbiol. 2011; 9:99–108. [PubMed: 21200397]
128. Roossinck MJ. Changes in population dynamics in mutualistic versus pathogenic viruses. Viruses. 2011; 3:12–19. Great outline of benefical virus-host interactions in a variety of species. [PubMed: 21994724]
129. Brown SP, Le Chat L, De Paepe M, Taddei F. Ecology of microbial invasions: Amplification allows virus carriers to invade more rapidly when rare. Current Biology. 2006; 16:2048–2052. [PubMed: 17055985]
130. Brown SP, Inglis RF, Tadddei F. Evolutionary ecology of microbial wars: within-host competition and (incidental) virulence. Evol Appl. 2009; 2:32–39.
131. Moran NA, Degnan PH, Santos SR, Dunbar HE, Ochman H. The players in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes. Proc Natl Acad Sci U S A. 2005; 102:16919–16926. [PubMed: 16195380]
132. Oliver KM, Degnan PH, Hunter MS, Moran NA. Bacteriophages encode factors required for protection in a symbiotic mutualism. Science. 2009; 325:992–994. [PubMed: 19696350]
133. Xu P, et al. Virus infection improves drought tolerance. New Phytol. 2008; 180:911–921. [PubMed: 18823313]
134. Marquez LM, Redman RS, Rodriguez RJ, Roossinck MJ. A virus in a fungus in a plant: three-way symbiosis required for thermal tolerance. Science. 2007; 315:513–515. [PubMed: 17255511]
135. Ophel KM, Bird AF, Kerr A. Association of Bacteriophage Particles with Toxin Production by Clavibacter toxicus, the Causal Agent of Annual Ryegrass Toxicity. Phytopathology. 1993; 83:676–681.
136. Holtz LR, Finkbeiner SR, Kirkwood CD, Wang D. Identification of a novel picornavirus related to cosaviruses in a child with acute diarrhea. Virol J. 2008; 5:159. [PubMed: 19102772]

137. Finkbeiner SR, et al. Metagenomic analysis of human diarrhea: viral detection and discovery. PLoS Pathog. 2008; 4:e1000011. [PubMed: 18398449]

138. Finkbeiner SR, et al. Human stool contains a previously unrecognized diversity of novel astroviruses. Virol J. 2009; 6:161. [PubMed: 19814825]

139. Phan TG, et al. A third gyrovirus species in human feces. J Gen Virol. 2012

140. Kapoor A, et al. Multiple novel astrovirus species in human stool. J Gen Virol. 2009; 90:2965–2972. [PubMed: 19692544]

141. Kapoor A, et al. Human bocaviruses are highly diverse, dispersed, recombination prone, and prevalent in enteric infections. J Infect Dis. 2010; 201:1633–1643. [PubMed: 20415538]

142. Victoria JG, et al. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. J Virol. 2009; 83:4642–4651. [PubMed: 19211756]

143. Rosario K, Duffy S, Breitbart M. A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. Arch Virol. 2012

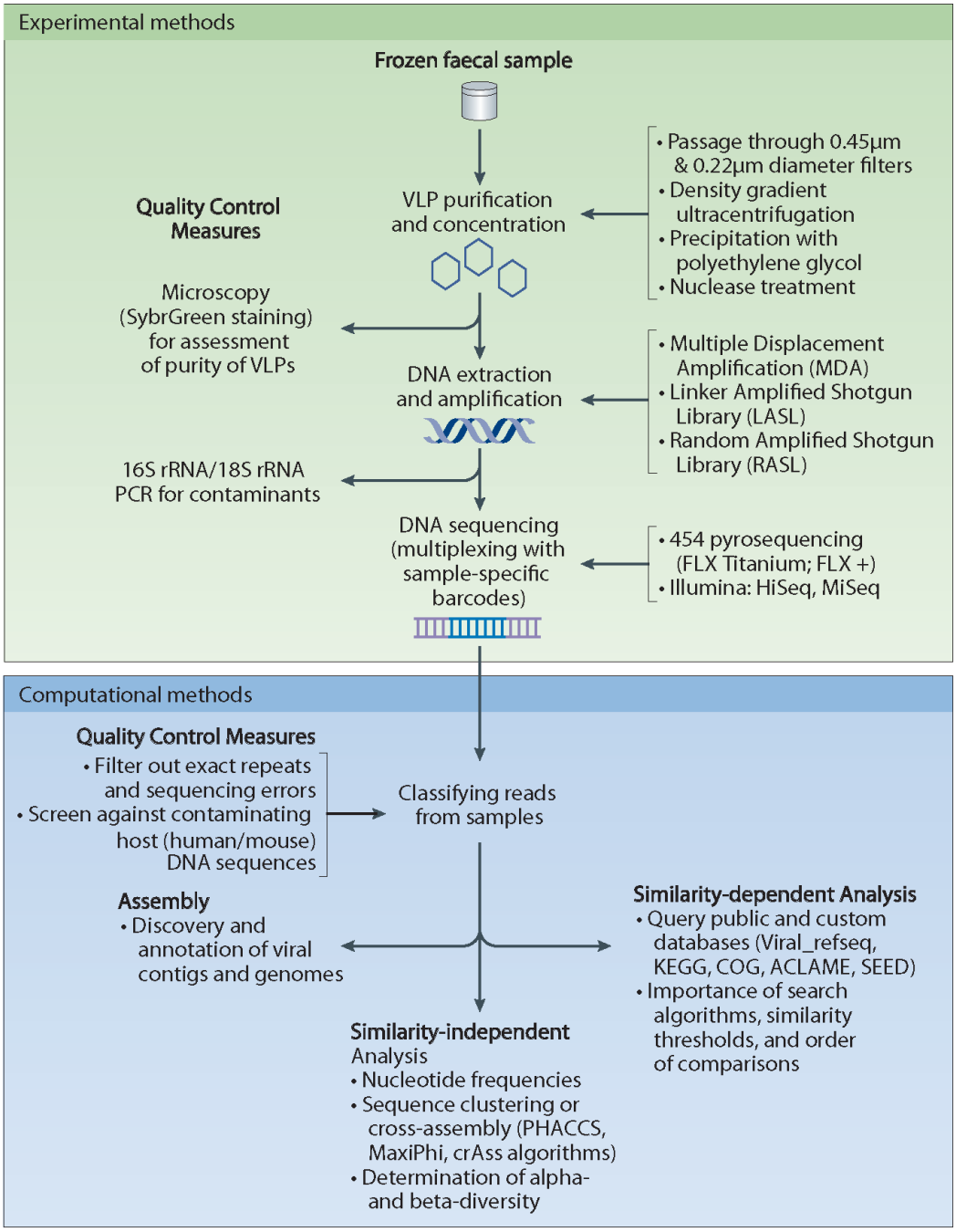

Figure 1. Experimental and computational methods for characterization of phage populations present in the human gut microbiota

See text for further discussion

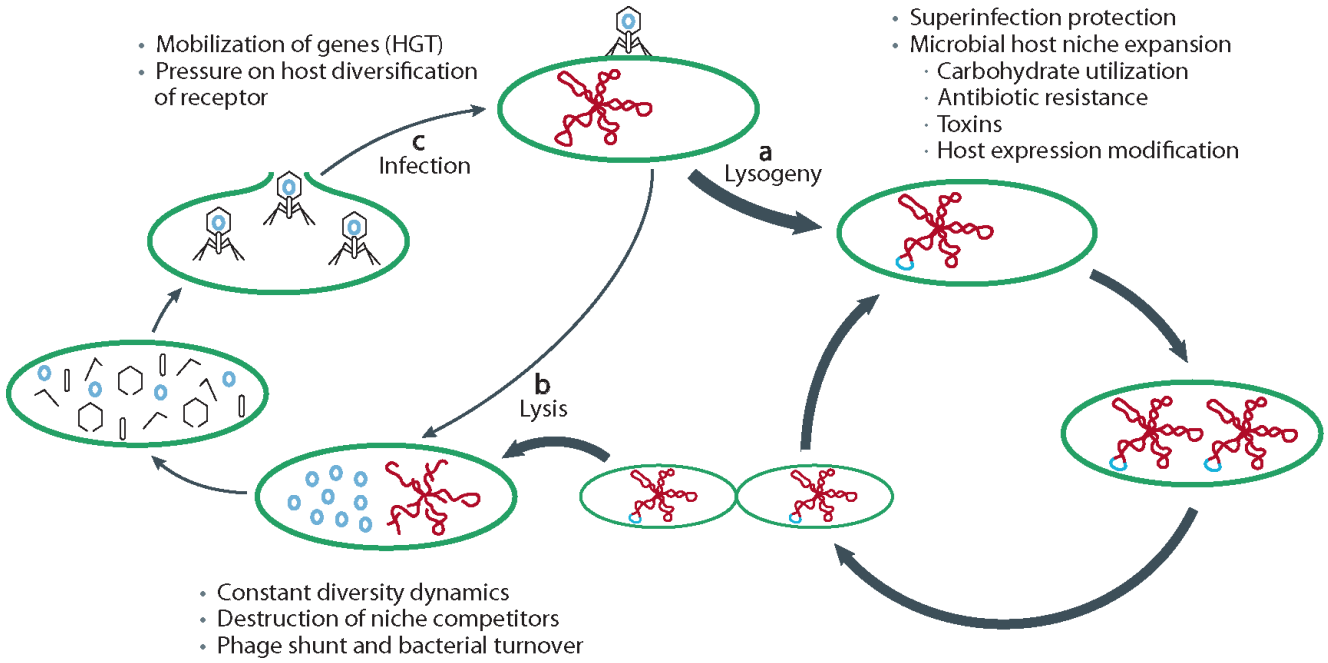

Figure 2. Potential consequences of a temperate phage lifecycle in the human gut

Viral metagenomic show that the phage population associated with the adult human gut microbiota is characterized by a relatively low number of virotypes compared to other ecosystems (e.g., soils, sediments, marine environments). These gut populations also exhibit high temporal stability of virotypes with respect to both viral community structure and nucleotide sequence conservation, and a high prevalence of temperate phages. These characteristics suggest that a temperate lifestyle is dominant in the distal human gut versus the lytic lifestyle observed in open oceans. (**a–c**) Illustration of the benefits of this temperate lifestyle on phage-host dynamics. (**a)** Integration as a prophage protects the host from superinfection, effectively 'immunizing' the bacterial host against infection from the same or closely related phages. Furthermore, the genes encoded by the viral genome may expand the niche of the bacterial host by enabling metabolism of new nutrient sources (e.g., carbohydrates), providing antibiotic resistance, conveying virulence factors, or altering host gene expression. This temperate lifecycle allows viral expansion in a 1:1 ratio with the bacterial host. If the integrated virus conveys increased fitness to its bacterial host, there will be increased prevalence of the host and phage in the microbiota. **(b)** Induction of a lytic cycle may follow a lysogenic state and can be triggered by environmental stress. As a consequence, bacterial turnover is accelerated and energy utilization optimized through 'phage shunts', where the debris remaining after lysis is used as a nutrient source by the surviving population. Furthermore, a subpopulation of bacteria that undergoes lytic induction sweeps away other sensitive species and increases the niche for survivors (i.e., bacteria that already have an integrated phage). Periodic induction of prophages can also lead to a constant diversity dynamic [115], which helps maintain community structure and functional efficiency. (**c)** Novel infections or infections of novel bacterial hosts by phages bring the benefit of horizontally transferred genes, and create selective pressure on the hosts for diversification of their phage receptors, which are often involved in carbohydrate utilization.

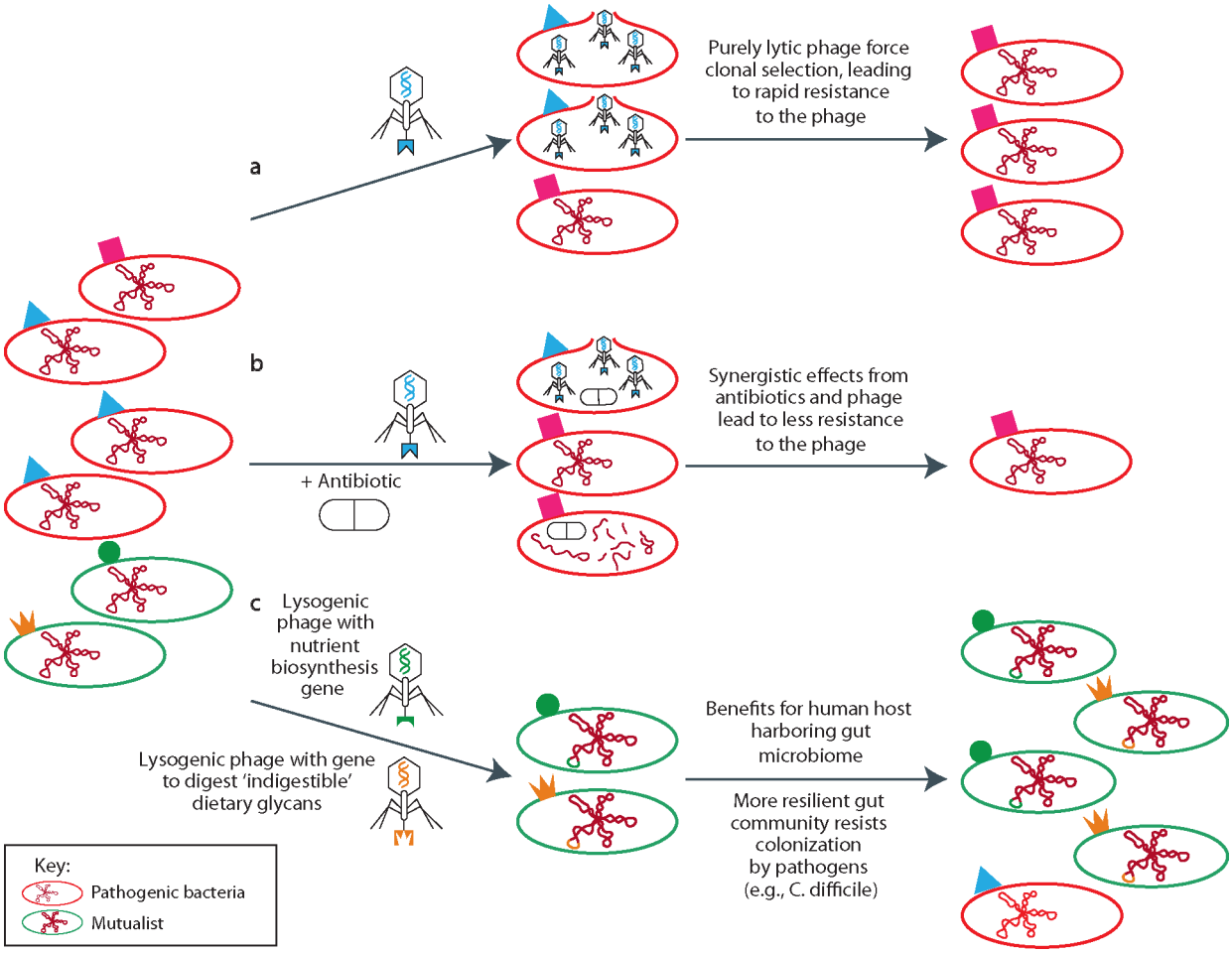

Figure 3. Potential strategies for phage therapy

(**a**) The traditional strategy has been to use a lytic phage against pathogenic bacteria. While transiently useful, this approach can lead to rapid resistance given positive selection for subpopulations ('clones') that are resistant to the lytic phage. Note that this is an antiquated approach to phage therapy that can be trivially improved by using multiple phages with non-overlapping host resistance patterns, or by selecting for phage mutants that overcome host resistance. (**b**) More recently, synergistic relationships between phage and antibiotics have been exploited, where lysogenic phages are introduced that alone do not kill the pathogen, but instead decrease its survival when used in concert with antibiotics. An example is a phage that inhibits a DNA damage repair system (SOS), which makes bacteria exquisitely sensitive to quinolone-class antibiotics [116]. (**c**) With our growing understanding of the human microbiome, it may be possible to take a more nuanced approach - selectively manipulating (enhancing) microbial community functions or clearing the way for invasion by probiotic consortia. Strategies can be envisioned to benefit both microbes and their host; for example, introducing genes into phage genomes that are involved in nutrient biosynthesis (with direct benefits to the bacterial and potentially human host), or degradation of nutrients (which may stabilize the representation and niches of beneficial microbes, especially during times of acute stress).