Phage diversity, genomics and phylogeny

*Moïra B. Dion[1,2], Frank Oechslin[1,2] and Sylvain Moineau[1,2,3]**

Abstract | Recent advances in viral metagenomics have enabled the rapid discovery of an unprecedented catalogue of phages in numerous environments, from the human gut to the deep ocean. Although these advances have expanded our understanding of phage genomic diversity, they also revealed that we have only scratched the surface in the discovery of novel viruses. Yet, despite the remarkable diversity of phages at the nucleotide sequence level, the structural proteins that form viral particles show strong similarities and conservation. Phages are uniquely interconnected from an evolutionary perspective and undergo multiple events of genetic exchange in response to the selective pressure of their hosts, which drives their diversity. In this Review, we explore phage diversity at the structural, genomic and community levels as well as the complex evolutionary relationships between phages, moulded by the mosaicity of their genomes.

Lysogens
Bacterial cells containing an integrated prophage, which can be induced, excised from the chromosome and enter the lytic cycle.

Mosaicism
The observation that different regions (genes and gene blocks) of the phage genomes have distinct evolutionary histories, owing to horizontal gene transfer events.

Viral metagenomics
Sequencing genomes of the viral fraction in a sample.

[1]*Département de biochimie, de microbiologie et de bio-informatique, Faculté des sciences et de génie, Université Laval, Québec City, Québec, Canada.*

[2]*Groupe de recherche en écologie buccale, Faculté de médecine dentaire, Université Laval, Québec City, Québec, Canada.*

[3]*Félix d'Hérelle Reference Center for Bacterial Viruses, Université Laval, Québec City, Québec, Canada.*

**e-mail: sylvain.moineau@bcm.ulaval.ca*

https://doi.org/10.1038/s41579-019-0311-5

'Phages are the most abundant and diverse biological entities on the planet'. This opening statement has become a favourite of many viral ecologists. With an estimated 10^{31} on the planet[1], phages can even outnumber bacteria by approximately tenfold in some ecosystems. They are found in every explored biome, from the human gastrointestinal tract to the global ocean but also in remarkable places such as the oceanic basement[2] and a Middle Age fossilized stool specimen[3]. In aquatic environments, phages have major roles in biogeochemical cycling, by short-circuiting the flow of carbon through bacterial killing, known as the viral shunt[1]. Phages are also modulators in the human gut, where they predominantly exist in lysogens, which can affect the physiology and metabolism of their hosts[4]. In addition to their ubiquity, phages exhibit a plethora of structural morphologies, with tailed phages and their double-stranded DNA (dsDNA) genome being the most represented in public databases as of 2019. Other seemingly less common phages can package their single-stranded DNA (ssDNA), single-stranded RNA (ssRNA) or double-stranded RNA (dsRNA) genome into non-tailed virions. Despite their relatively small genomes, phages show impressive genomic diversity and complex evolutionary relationships that do not follow traditional hierarchical phylogeny, due to pervasive mosaicism.

Much of our knowledge of phage diversity has been overhauled following advances in large-scale viral metagenomics and culturing efforts. In recent years, scientists have discovered phages with a genome size nearly 10 times larger than the average[5]. Non-tailed dsDNA[6,7] and ssDNA[8,9] phages are increasingly identified and even perhaps dominant in some biomes. Thousands of viral sequences have been identified from metagenomic sequencing projects, yet many of them share no detectable homology with reference phage genomes[10,11]. One feature that was emphasized by metagenomic studies is the exceptional viral diversity at the genomic level.

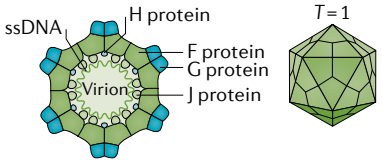
In this Review, we focus on phage diversity and explore four levels of organization. First, we present the unique morphologies of phages and the structural proteins that form viral particles. Second, we examine the genomic diversity and scarceness in gene-content similarities. From these two analyses emerges a contradiction: even when no sequence homology exists between morphologically distinct phages, some viral proteins still show conservation at the structural level. Third, we evaluate viral diversity at the community level, by comparing phage abundance and composition in various ecosystems. Recent progress in viral metagenomics has broadened our view of phage abundance and diversity, especially in marine environments. Lastly, we explore gene exchanges between phages, which generate mosaicity and drive diversification, and illustrate that they are interconnected through a complex network.

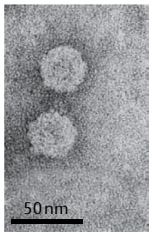
Morphological and structural diversity

Phage genomes are composed of either DNA or RNA, which may be double-stranded or single-stranded. This genetic material is packaged into a capsid that can be polyhedral (*Microviridae*, *Corticoviridae*, *Tectiviridae*, *Leviviridae* and *Cystoviridae*), filamentous (*Inoviridae*), pleomorphic (*Plasmaviridae*) or connected to a tail (*Caudovirales*)[12] (FIG. 1). To date, most isolated phages are tailed and have dsDNA genomes[13]. Taxonomic classification of phage taxa is carried out by the International

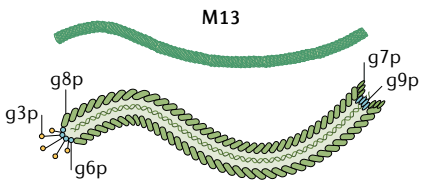
a ssDNA

***Microviridae* (phiX174)**

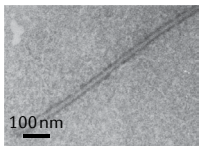




Inoviridae



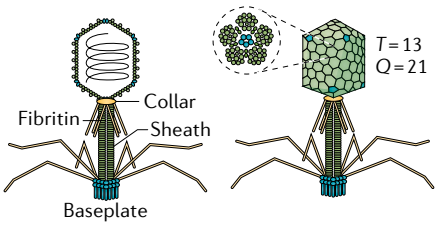
I2-2

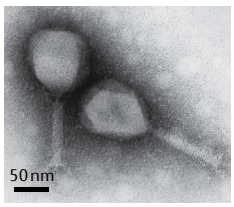


b dsDNA

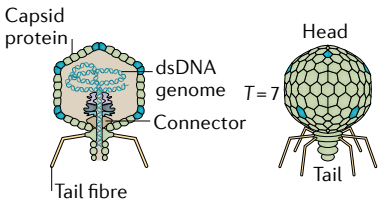
Tailed

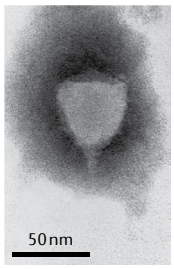
Myoviridae* (T4) and *Herelleviridae



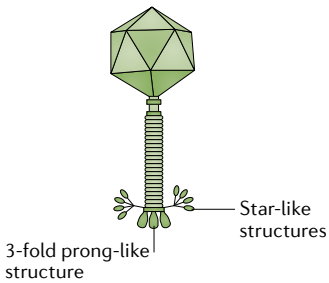


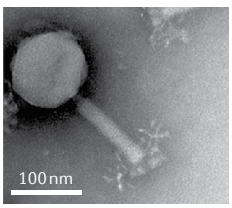
***Podoviridae* (T7)**



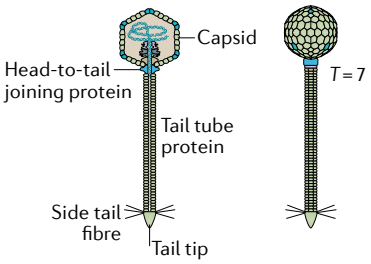


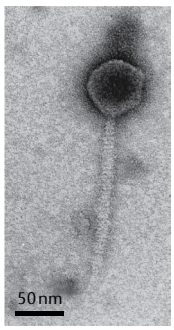
***Ackermannviridae* (AG3)**





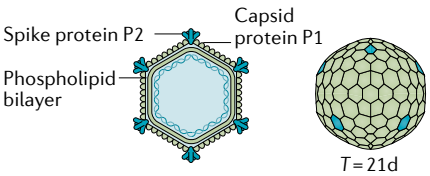
***Siphoviridae* (lambda)**

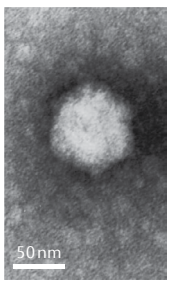




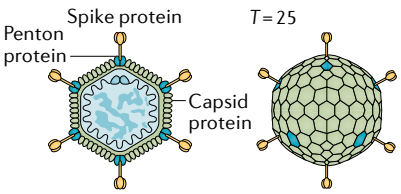
Non-tailed

***Corticoviridae* (PM2)**

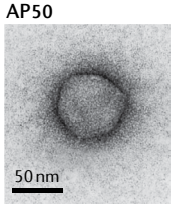




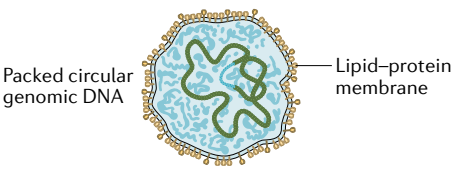
***Tectiviridae* (PRD1)**

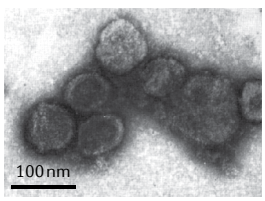


AP50



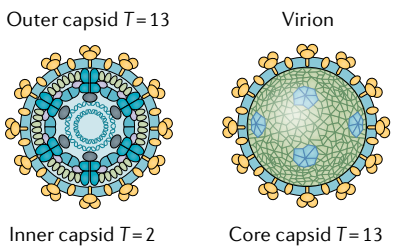
***Plasmaviridae* (MVL2)**

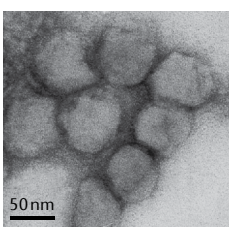




c dsRNA

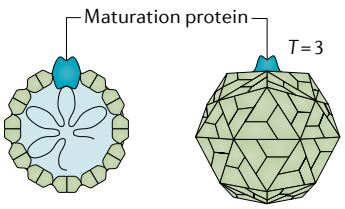
***Cystoviridae* (phi6)**



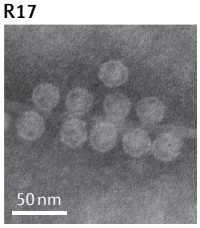


d ssRNA

***Leviviridae* (MS2)**



R17



◀ Fig. 1 | **Phage classification based on morphology and genome type.** A schematic representation (SR) and a transmission electron micrograph (TEM) are shown for each morphology. **a** | *Microviridae* have icosahedral capsids and small circular single-stranded DNA (ssDNA) genomes (SR and TEM of phiX174 are shown). The genome of filamentous phages of the *Inoviridae* family is composed of a circular supercoiled ssDNA molecule which is packaged within a long filament (>500 nm) composed of thousands of copies of the major capsid protein (MCP)[13,144] (SR of M13 and TEM of I2-2 are shown). **b** | Most of the characterized phages are tailed with double-stranded DNA (dsDNA) genomes and belong to the *Caudovirales* order. To date, five families have been described for this order: *Myoviridae* (long contractile tail; SR and TEM of T4 are shown), *Podoviridae* (short non-contractile tail; SR and TEM of T7 are shown), *Ackermannviridae* (*Myoviridae* morphology with tail spikes at the base of the tail; SR and TEM of AG3 are shown) and *Siphoviridae* (long non-contractile tail; SR and TEM of phage lambda are shown). *Herelleviridae*, although an official family, shares the same morphology as *Myoviridae*. *Corticoviridae* have a circular dsDNA genome and a capsid composed of an internal lipidic membrane surrounded by MCPs (SR and TEM of PM2 are shown). *Tectiviridae* (SR of PRD1 and TEM of AP50 are shown) have an icosahedral capsid, which contains a linear dsDNA genome and an internal lipidic membrane. Viruses belonging to the *Plasmaviridae* family (SR and TEM of MVL2 are shown) have a circular dsDNA genome surrounded by a lipidic envelope and no capsid. **c** | *Cystoviridae* have a tri-segmented double-stranded RNA (dsRNA) genome contained within a spherical capsid (SR and TEM of phi6 are shown) with three structural layers: an outer lipidic membrane and a two-layer inner capsid. **d** | *Leviviridae* have a ssRNA genome encoding only four proteins (MCP, replicase, maturation and lysis proteins) and a capsid with icosahedral and spherical geometries (SR of MS2 and TEM of R17 are shown). *T* values correspond to triangulation numbers, a measure of the complexity of the capsid shape, defined as the number of proteins per asymmetric unit. Each triangular facet of the icosahedral capsid is made of three asymmetrical units. Triangulation number has *T* for equilateral faces or *Q* if the face is an irregular triangle. Electron microscopy images are courtesy of the late Prof. Dr Hans-Wolfgang Ackermann and are available from the Félix d'Hérelle Reference Center for Bacterial Viruses. The image of *Ackermannviridae* is reprinted from REF.[17], CC-BY-4 (https://creativecommons.org/licenses/by/4.0/). The image of *Plasmaviridae* is reprinted with permission from REF.[40], Microbiology Society.

Committee on Taxonomy of Viruses (ICTV)[14]. Although phage classification was historically based on characteristics such as genome type (ssDNA, ssRNA, dsDNA or dsRNA), viral morphology and host range, it is currently undergoing a major overhaul, primarily using genomic-based methods. For example, the 1999 ICTV report classified tailed phages into three families, 16 genera and 30 species, whereas the 2018 report grouped them into five families, 26 subfamilies, 363 genera and 1,320 species (https://talk.ictvonline.org/taxonomy/p/taxonomy_releases). Comprehensive guidelines have been modernized for phage classification and it is expected that the list of virus taxa will substantially increase in the coming years[15].

Tailed phages. The large majority of phages described to date have a tailed morphology with a dsDNA genome and belong to the *Caudovirales* order[13]. This viral order, although under reclassification[16], currently comprises five families: *Myoviridae*, *Siphoviridae*, *Podoviridae*, *Ackermannviridae* and *Herelleviridae*. The last two families were created only recently because network-based approaches and meta-analyses indicated that they represented distinct groups within the *Myoviridae* family[16,17]. A large variation in capsid size can be observed among members of the *Caudovirales*, with diameters ranging from 45 to 185 nm, which is usually linked to genome size[18]. Most of the tailed phages (~75%) have icosahedral capsid structures and ~15% have an elongated capsid aligned with the axis of the tail[13]. Interestingly, members of *Caudovirales* share the same major capsid protein (MCP) fold (HK97 fold)[19]. The HK97 fold was identified following X-ray crystallography of the capsid of phage HK97. The capsid is connected to its tail through a connector complex often composed of a portal protein coupled to head completion or connector proteins. Structural studies have revealed that the portal complex is a dodecameric ring with a similar overall structure shared between most tailed phages, despite low sequence similarity[20–22]. Capsid completion or connector proteins also form dodecameric rings as observed in the *Siphoviridae* SPP1 and HK97 phages. Homologues of these proteins are found in various phages that have contractile and non-contractile tails[23]. Moreover, the head to tail connector protein gp4 of the *Podoviridae* P22 has a similar structure to those present in siphophages SPP1 and HK97, despite no observable sequence similarity[24].

The tails of *Siphoviridae* are composed of a central tape measure protein surrounded by a tail tube and ending with a terminator protein[25]. A similar architecture is observed for phages of the *Myoviridae* family, where an additional layer (the protein sheath) enables contraction for the insertion of the tail tube into the bacterial host during infection[26]. Interestingly, the capsid–tail joining protein gpFII of the *Siphoviridae* phage lambda has a similar tertiary fold to its tail tube protein gpV and adopts the same quaternary structure when assembled in the phage[27]. gpV also shares structural homology with the tail tube of *Myoviridae* phages as well as some components of the bacterial type VI secretion system such as the Hcp1 protein[28]. Moreover, the folds of proteins gpFII and gpV are similar to those of the baseplate hub of the myophages T4 and Mu, without any sequence homology[27]. These observations suggest that the tail tube-like fold adopted by the capsid–tail connector, the tail tube protein itself and the baseplate are an important building block for members of *Siphoviridae* and *Myoviridae*. Members of the *Podoviridae* family, such as *Escherichia coli* phage T7, have a very short and non-contractile tail. A tube-like extension of the tail that penetrates both cell membranes is essential for genome delivery into the host[29].

Receptor binding proteins (RBPs) present at the tip of the tail or at the baseplate were characterized at the structural level in siphophages and showed high levels of structural similarity despite low sequence homology[30–32]. Several of these RBPs were even found to be structurally related to their counterparts in some mammalian adenoviruses[33]. The RBP domains are also interchangeable between different phages, thereby providing a swift means to change hosts. Members of the *Ackermannviridae* family, formerly known as *Viunalikevirus*, have a myovirus-like morphology but differ by their complex and unique adsorption structures. Short filaments with bulbous tips that resemble an umbrella and prong-like structures are attached to the baseplate[17].

Membrane-containing phages. Phages belonging to the *Tectiviridae* (for example, phage PRD1) and *Corticoviridae* (for example, phage PM2) families comprise icosahedral non-tailed virions that have an internal lipidic membrane and linear or circular dsDNA

Polyhedral
A shape of the phage capsid, which consists of many polygonal faces and is most commonly found as an icosahedron (polyhedron with 20 faces).

Pleomorphic
Variability in shapes and sizes for phages.

HK97 fold
A 3D conformation termed after the capsid protein structure of phage HK97.

Portal complex
A dodecamer forming a central channel involved in viral DNA packaging and injection, providing a docking site for attachment of the tail machinery.

genomes, respectively. A hallmark characteristic of these two phage families is their trimeric MCP, which is composed of a double β-barrel structure[34,35]. Structural analyses of the MCP of phage PRD1 revealed an N-terminal α-helix, which interacts directly with the phage inner membrane and shares structural homology with the MCP of adenoviruses[34]. Furthermore, the RBP present at the icosahedral vertices of phage PRD1 and PM2 capsids is also structurally similar to N-terminal domains of all human adenoviruses[35–37]. Phage PRD1 does not have a tail to deliver its genome into its Gram-negative host but its membrane was observed to transform into a proteolipidic tube, which can pierce host envelopes[38]. Unlike *Corticoviridae* and *Tectiviridae* phages that have inner lipidic membranes, members of the *Cystoviridae* family, including phage phi6, have lipidic membranes that surround their icosahedral capsids[39]. *Acholeplasma* virus L2 (AVL2; also known as MVL2) is currently the only classified member of the *Plasmaviridae* family. AVL2 infects the wall-less *Acholeplasma* bacterial genus, and new virions are released by membrane budding without causing cell lysis[40]. *Plasmaviridae* phages do not possess any capsid but their genomes are enclosed in a proteinaceous lipid vesicle that has a similar composition to the outer membrane of phage phi6 (REF.[41]).

Phages with small icosahedral capsids or a filamentous morphology. *Microviridae* and its most studied member, phage phiX174, have a small icosahedral capsid (26 nm) and ssDNA genome (5,386 bp)[42] (FIG. 1). The capsid is built on a protein fold that has a 'jelly roll' β-barrel structure and has similarities with ssDNA eukaryotic viruses, including rhinoviruses[42]. *Microviridae* are currently classified into two subfamilies named *Bullavirinae* and *Gokushovirinae*. DNA delivery into the bacterial host relies on a protein, which oligomerizes to form a tube that crosses the host periplasmic space by joining the outer and inner membranes[43]. Structural differences in proteins mediating host attachment have been observed for both subfamilies. *Bullavirinae* have pentameric major spike protein complexes at the end of each capsid vertex[44], whereas *Gokushovirinae* have 'mushroom-like' protrusions that extend along the three-fold icosahedral axes of the capsid[44]. The major spike protein complexes in the *Bullavirinae* subfamily are also divergent, but as their structures are superimposable, they can be exchanged between phages[45].

Other small icosahedral viruses include members of the *Leviviridae* family, such as the phage MS2 that has a ssRNA genome (FIG. 1). The MS2 viral particle is made of only two proteins: the MCP and a single copy of the maturation protein that interacts with the genomic RNA during packaging and with the host receptor during adsorption[46]. The MCP of phage MS2 can control replication by interacting with the initiation codon of the replicase-encoding gene, which switches the replication cycle to viral assembly[47]. Recent cryo-electron microscopy reconstruction of the viral capsid also revealed that the RNA genome is highly involved in virion assembly by forming secondary structures that act as a scaffold[48]. Of note, this system of genome packing is radically different from the *Caudovirales*, in which an empty capsid is first assembled and then filled with the phage genome before the packaged capsid is connected to its tail[49].

Members of the family *Inoviridae* are drastically different in terms of morphology and lifestyle (FIG. 1). Phage particles contain a ssDNA genome that is surrounded by thousands of copies of the MCP that are assembled and then extruded from the host in a continuous manner[50]. The MCPs of filamentous phages are unique in their architecture, which consists of a long α-helix with an N-terminal signal peptide for membrane translocation[51]. The signal peptide is then cleaved before the proteins are assembled into a long cylindrical shell spiral with the C terminus interacting directly with the viral DNA[52].

Two sides of the same coin. From a morphological point of view, several diverse phages still share some commonalities. One example is the MCP fold, which is conserved at the structural level between all tailed phages but also extends to archaeal viruses and adenoviruses[53,54]. For the majority of these proteins where conservation is observed in their structure, no trace of homology can be detected at both the amino acid and the nucleic acid levels. This paradigm is explored further in BOX 1, where convergent evolution or a common ancestor are discussed as possible explanations for structural similarities between viruses infecting different domains of life.

Genomic diversity and viral metagenomics

Number of complete genomes. According to the National Center for Biotechnology Information (NCBI), as of September 2019, there are 8,437 complete phage genomes divided into 12 families (based on the ICTV classification at the time) and one unclassified group (FIG. 2a). More than half of these are members of the *Siphoviridae* family. This over-representation is owed in large part to the isolation and genome sequencing of 1,537 siphophages infecting *Mycobacterium smegmatis* by the SEA-PHAGES (Science Education Alliance–Phage Hunters Advancing Genomics and Evolutionary Science) programme[55]. *Myoviridae* and *Podoviridae* represent 17% and 12% of the total phages, respectively, rendering *Caudovirales* (also comprising *Herelleviridae* and *Ackermannviridae*) the most abundant group of phages (>85%) in public genomic databases. The disproportionate representation of tailed dsDNA phages will likely decrease in the near future with the discovery of new phages. For example, the genomic diversity within the *Microviridae* family was largely underestimated until 258 new ssDNA phages were detected in the gut of the marine invertebrate *Ciona robusta*[56]. In addition, the unclassified bacterial virus group within the NCBI database consists of phages discovered through metagenomic projects that have yet to be isolated or have been very recently propagated in a bacterial host. Part of this latter group includes 283 non-tailed dsDNA phages, infecting the ubiquitous *Vibrionaceae* bacterial family[7]. A recent study used a machine learning approach to mine microbial genomes and metagenomes for inoviruses[9]. A total of 10,295 inovirus-like sequences were found, from which 5,964 distinct species appear to have been identified. This study alone would represent a 100-fold expansion of the diversity previously described (57 genomes) within the *Inoviridae* family. The ever-increasing number

Synteny
Physical co-localization in the genome of genes with associated functions.

of complete phage genomes in the NCBI database still represents only a small fraction of the actual phage genomic diversity, because half of them infect only seven host genera (*Mycobacterium*, *Streptococcus*, *Escherichia*, *Pseudomonas*, *Gordonia*, *Lactococcus* and *Salmonella*). The number of complete phage genomes available in public databases is also certainly far greater because of the numerous unidentified prophages in bacterial genomes[57].

Range in genome size. Phages have a wide range of genome sizes (FIG. 2b). Apparently, the smallest phage genome reported to date is that of *Leuconostoc* phage L5, with only 2,435 bp. At the other end of the spectrum, an increasing number of jumbo phages (>200 kb) are being characterized and show unique genomic features[58]. Their large genome size allows jumbo phages to carry genes involved in replication and nucleotide metabolism that are absent in smaller phage genomes. The organization of these large viral genomes is also atypical because genes with associated functions do not show strong synteny and are, instead, more dispersed[58]. A new group of phages with the largest genomes ever recorded to date, called megaphages (>540 kb), were recently uncovered from human and animal gut metagenomes

Box 1 | **A common ancestor of phages?**

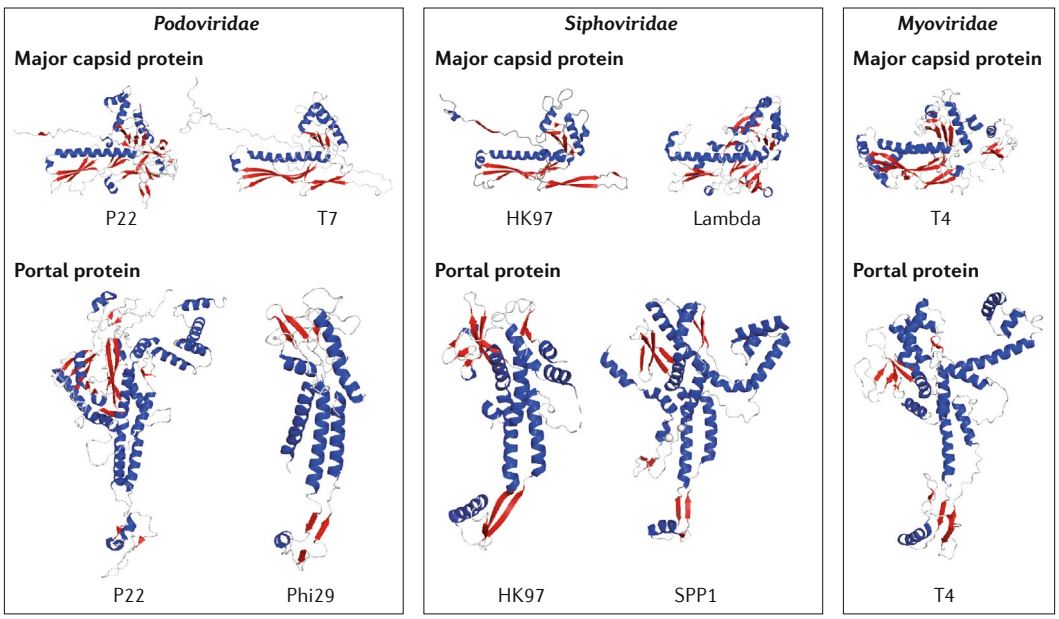
Despite extensive gene exchanges, which generate diversity, and the absence of homology at the nucleotide and amino acid levels for most phage pairs, a finite and relatively small number of different virion structures exist in nature. This raises the question of whether these structural similarities can be explained by divergent or convergent evolution. A divergent evolution would indicate that viruses share a common ancestor and have diverged beyond detectable sequence homology, while maintaining the basic architecture of their structural proteins. A convergent evolution would suggest that viruses share no common ancestors, but rather have converged towards a structure that is particularly optimal to build a virion. Although both can lead to a single common trait, the accumulation of similar structural characteristics seems to point towards the divergent evolution hypothesis and the existence of a common ancestor.

First, the *Tectiviridae* phage PRD1 major capsid protein (MCP) fold is highly similar to that of the archaeal virus STIV[134] and mammalian adenoviruses[34]. The MCP is a trimeric protein made of two eight-stranded jelly rolls (β-barrels). There are four different ways to fold such jelly rolls, but only one is seen in these viruses[135]. Other features shared between PRD1 and adenoviruses include a linear double-stranded DNA genome with inverted terminal repeats, the organization of the MCP on the capsid surface and the structure of proteins at the virion surface[136]. Other viruses have a PRD1-like structure, such as *Tectiviridae* infecting Gram-positive hosts (PRD1 infects Gram-negative hosts), *Corticoviridae*, eukaryotic and archaeal viruses[137]. This suggests a common ancestor to PRD1-like viruses.

Second, a relationship also exists between tailed double-stranded DNA phages, the archaeal virus HSTV-1 (REF. [54]) and herpesviruses[138]. The MCP of these viruses has a common fold, called the HK97 fold. Other structural similarities exist in HK97-like viruses, such as a portal on one vertex of the capsid and their capsid assembly pathways[139]. HK97-like capsid protein and portal protein tertiary structures for *Podoviridae*, *Siphoviridae* and *Myoviridae* phages are shown in the figure and illustrate the similarities between phages belonging to different families. The colour scheme used is based on secondary structures: red, β-strands; blue, α-helices; and grey, loops.

A third case is the similarity of *Cystoviridae* phages phi6 and phi8 with eukaryotic viruses belonging to *Reoviridae* (for example, blue tongue virus (BTV)) and *Totiviridae*[140]. These double-stranded RNA viruses share a similar inner coat protein[141] and have a segmented genome packaged in a double-shelled capsid[136].

Such structural similarities between viruses infecting hosts spanning different domains of life provide evidence towards understanding the origin of viruses. Based on the previous examples of common ancestors, it has been proposed that viruses form polyphyletic lineages (PRD1-like, HK97-like and BTV-like) in contrast with the monophyletic origin of cellular life[142,143].



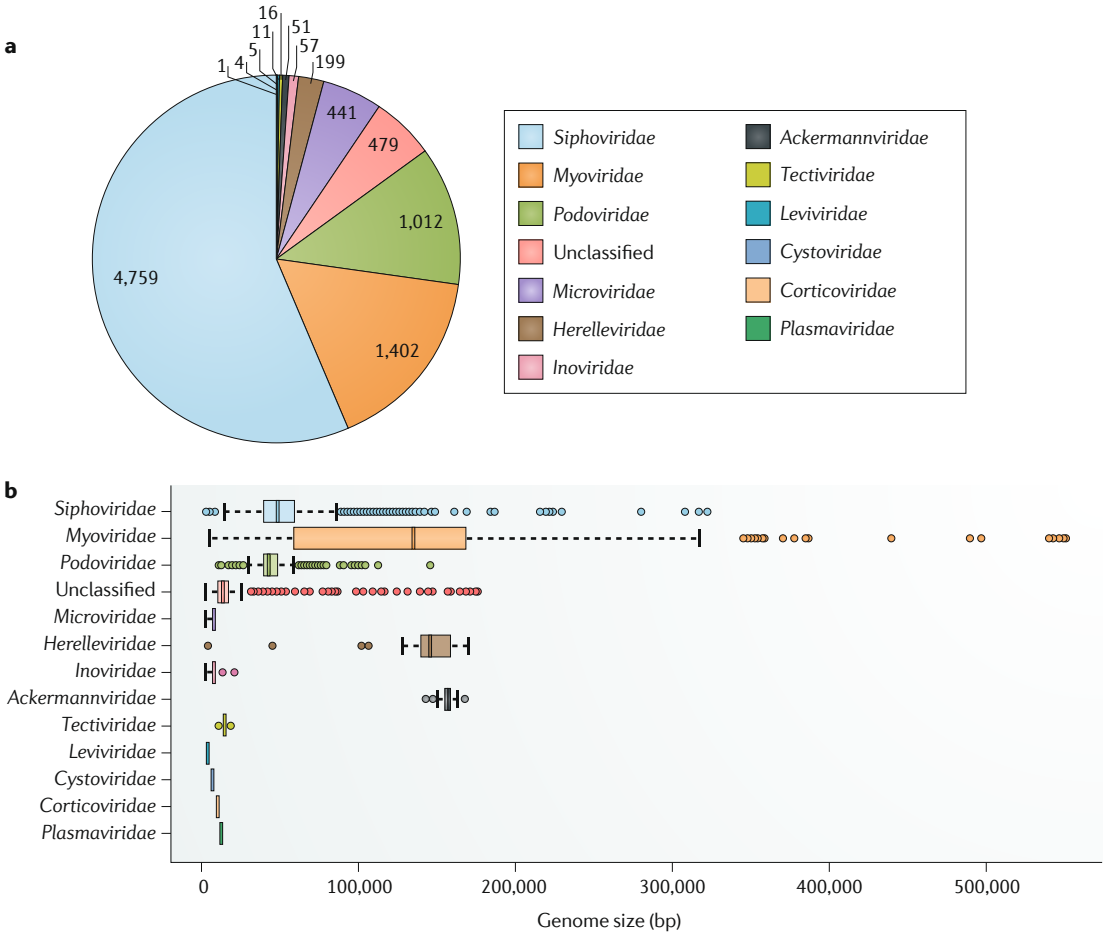


Fig. 2 | **Number of complete genomes and genome size distribution in phage families.** The number of complete genomes (part **a**) and the distribution of genome size (part **b**) in the National Center for Biotechnology Information (NCBI) nucleotide database as of September 2019 are shown. The assignment of each phage to a family was performed using the NCBI taxonomy database. The unclassified group combines 'unclassified *Caudovirales*', 'unclassified double-stranded DNA phages' and 'unclassified bacterial viruses'. This unclassified group is the fourth largest, emphasizing the increasing number of phages discovered through viral metagenomics for which no family can be assigned based on sequence information. Among the *Caudovirales* order, *Herelleviridae* and *Ackermannviridae* are the most homogeneous families in terms of genome size. This is most likely because these two families were created after genomic analyses rather than morphological similarities.

and are predicted to infect *Prevotella* species[5]. These phages seem widespread in gut microbiomes, as they were identified in humans, baboons and pigs[5]. They were overlooked owing to genome fragmentation and their use of an alternative genetic code, which consists of a repurposed stop codon[5].

Contribution of viral metagenomics to exploring phage genomic diversity. Given the absence of a conserved genetic marker and the predicted large number of phages in the biosphere[59], phage genomic diversity is difficult to comprehend. Phages infecting different hosts typically have little to no sequence similarity, and phages that infect a single host may also exhibit considerable sequence differences[60–62]. For instance, pairwise comparisons of 2,333 phages showed no detectable homology in 97% of cases when measuring the nucleotide distance and gene content[63]. Thanks to modern techniques that explore viral 'dark matter' (that is, viral genetic material that is unclassified), such as viral metagenomics, we are starting to understand the extent of phage global diversity. Viral metagenomics overcomes the challenges of culture-based approaches and single marker genes by assessing the total viral nucleic acids (mostly dsDNA) isolated from any given environment. Before the arrival of next-generation sequencing, the first viral metagenomics study was published in 2002 from surface seawater samples[64]. In recent years, the optimization of the steps required to obtain viral nucleic acids of good quality[65], the reducing costs of sequencing and an improved set of analytical tools[66] have allowed the construction of large-scale virome (that is, viral sequences obtained from viral metagenomics) datasets from viral communities, mostly from marine and human gut samples. There are now at least 90 studies describing viromes from aquatic environments[67], 38 from the human gut and eight from soil[67]. Among these, three research consortia — Tara Oceans[68], the Pacific Ocean Virome[69] and the Malaspina oceanic research expeditions[70] — have performed viral metagenomics on marine samples from various depths

Microdiversity
Intra-population genetic variation.

Viral tagging metagenomics
A high-throughput method to link a virus to its host, consisting of labelling viruses with a fluorescent dye, collecting infected cells by flow cytometry and sequencing the viral DNA.

and locations. This has led to the detailed characterization of ocean dsDNA viruses and their abundance patterns on local and global scales[71,72]. On the other hand, the first human gut virome was published in 2003 from a single healthy individual[73]. More studies of twins and their mothers[74], healthy adults[75,76] and individuals with ulcerative colitis[77] have followed to describe longitudinal and interpersonal viral variations in health and diseases. In 2014, the mining of viral metagenomic libraries (viromes) also resulted in the discovery of the most abundant and widespread phage in the human gut, called crAssphage[78]. The results of these projects are summarized in the following sections.

Beyond viral metagenomics. A major drawback in describing viral communities with metagenomics is the lack of high-enough resolution to reconstruct genomes of closely related sequences. This causes phage populations with high levels of microdiversity to be discarded from the metagenomics assembly. The detection of this microdiversity is necessary to better understand phage–host interaction dynamics[79]. Single-virus genomics overcomes this obstacle by sorting individual phages prior to sequencing. Such an approach led to the discovery of the most abundant marine phage[80], vSAG 37-F6, which infects *Pelagibacter* species[81]. Viral tagging metagenomics may also provide additional insights into phage–host interactions, as reported for cyanophages infecting *Synechococcus* species[82]. Although metagenomics does not specifically target viral DNA, a wealth of information can still be discovered about phage sequences[10]. Using an exhaustive collection of viral protein families manually identified as bait, >125,000 viral genomes were detected from 3,042 metagenomes of diverse geographical origins[10]. This study was a major contribution to our understanding of viral diversity, as it expanded the number of viral genes by 16-fold. The study also suggested that, on a global scale, phage genomic diversity still remains widely uncharacterized but that the discovery rate in marine and human samples (the most studied biomes) may be approaching saturation[10]. Yet unknown phages still consistently represent the majority of the sequences in the viral fraction of any given environmental sample, sometimes accounting for >90% of the reads[11,83]. FIG. 3 outlines how 'omics' and culturing efforts can be integrated to fully characterize entire phage communities.

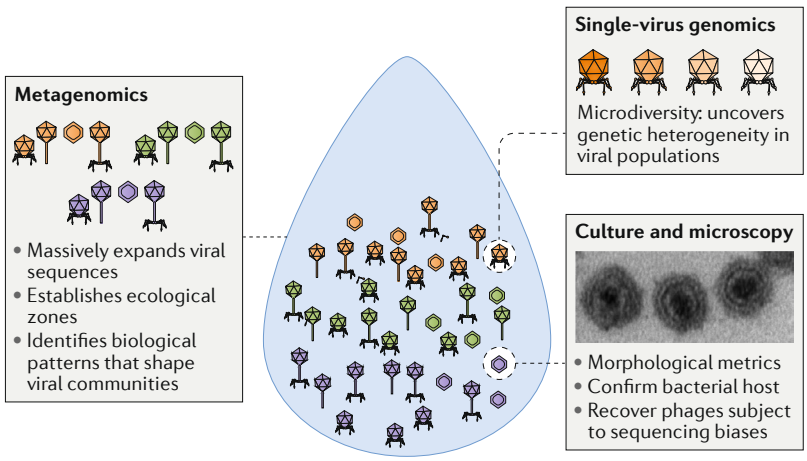


Fig. 3 | **Integrating metagenomics, single-virus genomics, culture and microscopy to uncover viral diversity.** Several techniques have been developed to characterize phage diversity in biological communities, mostly from marine samples[66]. Here, we highlight techniques that do not require previous knowledge and that can characterize the entire community. Metagenomics can characterize the diversity of phages to a great extent, with up to thousands of viral populations being identified in a single experiment[11]. Single-virus genomics enables sequencing of individual virions[80], which helps to reveal phage populations with high levels of microdiversity (represented here by different shades of orange in the podovirus), which normally impede genome assembly in metagenomics pipelines. Culturing techniques combined with observations through a transmission electron microscope permit the discovery of phages otherwise subject to sequencing biases. Reprinted from REF.[7], Springer Nature Limited.

Distribution and abundance

Phages in marine environments. Marine phages are thought to have major roles in modulating microbial communities, generating genetic diversity and influencing the nutrient cycle through bacterial mortality[84]. The crucial role of marine phages can be attributed to their tremendous abundance and diversity. In a recent analysis combining 22 distinct marine surveys, 95% of viral abundance was observed to range from 10^5 to 10^7 virus-like particles (VLPs) per millilitre, with a median virus-to-microbial cell ratio of 10:1 (REF.[85]). Analyses of samples from six global ocean regions using quantitative transmission electron microscopy[6] revealed a dominance of non-tailed viruses (79%) in the samples (FIG. 4a), followed by *Myoviridae* (14%), *Podoviridae* (6%) and *Siphoviridae* (1%). Interestingly, the morphological distribution did not vary consistently with depth or oceanic region[6].

Comparative genomic analyses of more than 100 *Synechococcus*-infecting cyanophages collected over 15 years revealed genomic clusters and sub-clusters that exhibited clear temporal and/or spatial patterns of abundance[86]. Viral tagging metagenomics indicated that phages infecting *Synechococcus* are clustered into at least 26 discrete populations with relative abundances ranging from 0.06 to 18.2% (REF.[82]). Possibly, the most abundant and well-distributed phages are those infecting *Pelagibacter* species, hosts that dominate in marine surface bacterioplankton communities[87]. Indeed, pelagiphages were among the most abundant phages in metagenomic datasets along longitudinal and depth gradients from all oceans[80].

The isolation and genome sequencing of 31 phages that infect *Cellulophaga baltica* (phylum Bacteroidetes)[88] showed that cellulophage diversity was even higher than that observed for *Synechococcus* phages and comprised non-tailed dsDNA phages. Comparisons with existing metagenomic data also revealed that cellulophages are widespread in oceans, but in low numbers. More recently, a group of dsDNA non-tailed viruses called autolykiviruses, which was missed in analyses due to multiple methodological biases, was isolated[7]. Genomic sequencing of these new phages revealed that they are present in the genome of major bacterial phyla and in metagenomic datasets from the water column and sediments[7].

Taxonomic analyses of 24 metagenomes from various sites in the Mediterranean Sea that vary by depth reported the dominance of *Caudovirales*, with *Myoviridae*

accounting for 67–96% of the viral reads (followed by *Podoviridae* and *Siphoviridae*), independent of the water depth[89]. The largest marine viral metagenomics study to date was recently published[11], which surveyed 145 samples from the Tara research expedition, including 41 samples from the polar circle. The authors identified 195,728 viral populations, 90% of which could not be taxonomically annotated, and found that *Caudovirales* dominated the known sequences. They confirmed that phages in the ocean form discrete populations and identified potential drivers of phage diversity, such as nutrients, photosynthetically active radiation (an indicator of productivity in photosynthetic organisms) and latitude.

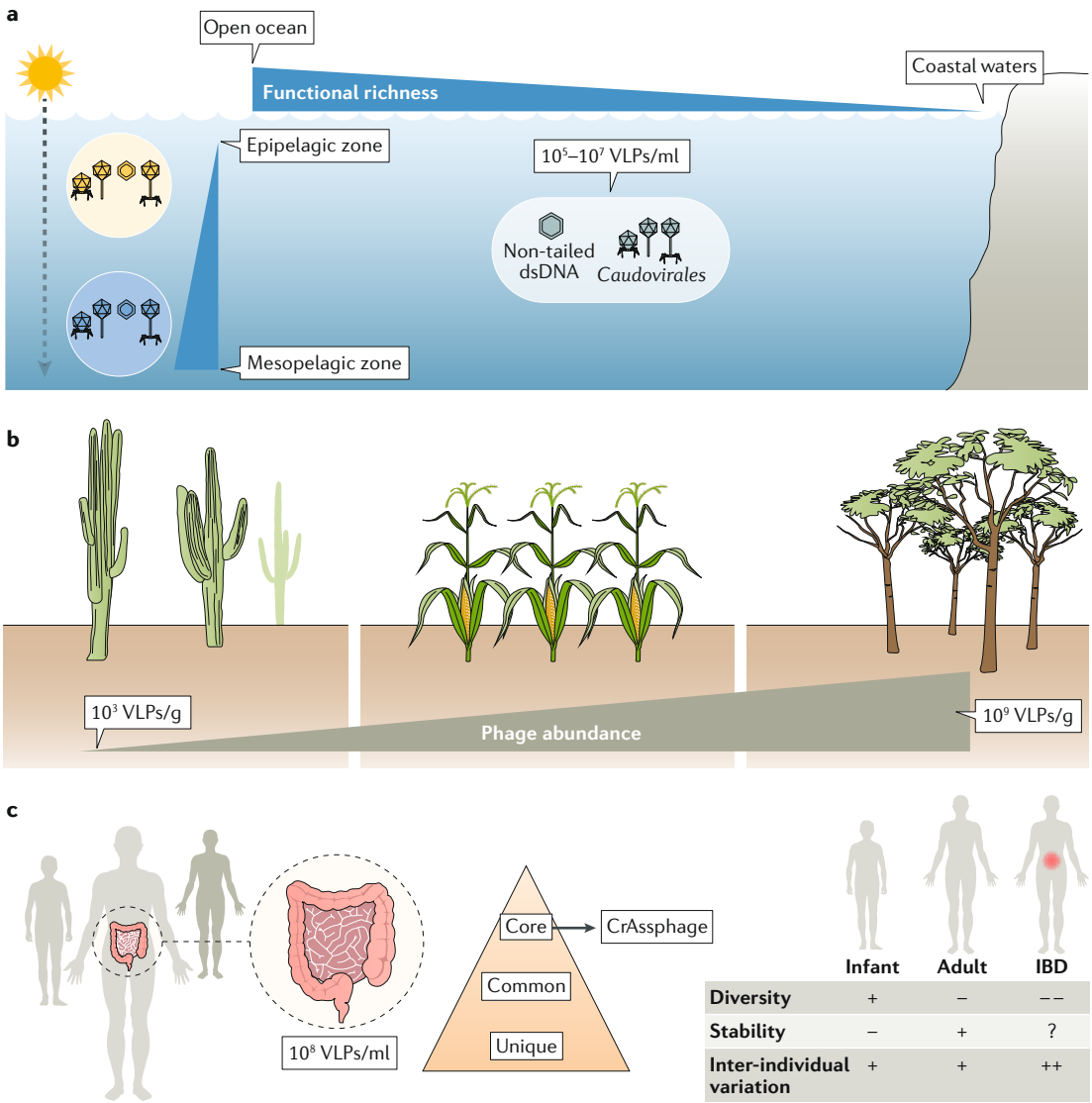


Fig. 4 | **Phage distribution and abundance in three ecosystems. a** | Phages in the marine environment are extremely abundant, with a virus-to-bacteria ratio often ranging from 1:1 to 100:1. Quantitative transmission electron microscopy of marine samples indicated that non-tailed phages are much more represented than tailed phages, which was also confirmed by metagenomic data[6,145,146]. Furthermore, phages from the mesopelagic zone were distinct from phages isolated from the epipelagic zone in terms of gene content, life history traits and temporal persistence[147]. Similarly, functional richness was found to decrease from deep to surface water and with distance from the shore for surface water only[69]. **b** | Phage abundance in the soil is highly variable and correlates with biome type (for example, desert, agricultural or forest soils), pH and bacterial abundance. Indeed, viral abundance is the lowest in hot deserts, intermediate in agricultural soils and the highest in forest and wetland soils[67]. Viral abundance also positively correlates with bacterial abundance in the soil and negatively correlates with pH, with phage counts decreasing at higher pH. **c** | The phage community in the human gut is mainly composed of members of the *Caudovirales* and *Microviridae*, and the majority of these phages remain unclassified[75,103,104]. Phage composition is unique to individuals, with global metagenomic analyses indicating that some phages are globally distributed[75,76,103,104]. The gut phage community is also stable over time, but rapid changes are observed in early life[8]. Changes in the diversity and composition of the human virome were also reported to be related to the gut health status, particularly in the case of inflammatory bowel disease (IBD)[77,148]. A set of widespread phages exists, named the core phage community, which includes crAssphage, likely the most prevalent human gut phage[76]. dsDNA, double-stranded DNA; VLP, virus-like particle.

Lytic
A replication strategy where a phage takes control of the host cell to replicate its genetic material, produce its structural components, self-assemble to form new virions and burst (lyses) the cell to release new viral particles.

In addition to exhibiting various morphological compositions, phage communities in the ocean have different replication strategies according to season. In the western Antarctic Peninsula[90] and in the Canadian Arctic Shelf[91], prophages dominate in the spring whereas lytic infections prevail in the summer. This fluctuation is likely explained by the kill-the-winner hypothesis, which states that a high bacterial abundance (caused by favourable growth conditions in the summer) is coupled with a high rate of lytic infections[92,93]. This model was further extended with the piggyback-the-winner model, in which the lysogenic lifestyle is instead dominant at high bacterial densities[94,95]. This was first observed in coral reefs, where the virus-to-host ratio is low despite substantial microbial density[94]. Abundant hosts were either killed by phages or became resistant lysogens, which in turn decreased phage titres when limited hosts were available for replication. According to a recent Review[66], the new phages that occupy the niche are more likely to be descendants of a 'royal family' (that is, they are variants of the most abundant phages that overcame host resistance). The authors proposed the 'royal family model' to describe the persistence of dominant phages in aquatic ecosystems.

Phages in the soil. Compared with marine environments, soils are intrinsically diverse due, in part, to their wide compositional spectrum and spatial heterogeneity in terms of physicochemical properties. A recent meta-analysis of 24 soils indicated that viral abundance is highly variable and correlates with soil type, ranging from approximately 10^3 VLPs per gram in desert soils to 10^9 VLPs per gram in forest soils[67] (FIG. 4b). Transmission electron microscopy observations of different soil types reported the predominance of non-tailed particles over tailed phages and higher morphological diversity in forest soils compared with agricultural soils, in some cases[96,97]. Metagenomics was also used to assess the richness and evenness of viral communities in prairie, desert and rainforest soils[98]. Similar phage sequences were observed in all of these soils but were significantly different from the dominant types found in marine or faecal samples. Metagenomic analyses of different Antarctic soils revealed that tailed phages were dominant in all samples, with the presence of *Myoviridae* and *Siphoviridae* inversely correlated[99]. Of note, samples with low and medium diversity were dominated by *Siphoviridae* genetic signatures. Abiotic factors like pH and the altitude of the sampling site appeared to be the main drivers of viral community composition[99].

Phages in the human gut. Phages are also highly abundant in the human gut microbiome, with up to 10^8 VLPs per millilitre in faecal filtrates[100] (FIG. 4c). Of note, higher phage titres were found in gut mucosal biopsies (10^9 phages per biopsy), possibly owing to the affinity of host-associated phages to bind and accumulate in the mucosal secretion[101,102]. Transmission electron microscopy visualizations demonstrated that *Caudovirales* dominate in the human gut, with striking inter-individual differences in morphologies and types[100]. As most of the bacteria residing in the gut are difficult to culture, metagenomic sequencing is highly useful for assessing the complexity and diversity of gut phage populations. Recent analyses confirmed that the majority of sequences that could be identified belonged to the *Caudovirales* order, but members of the *Microviridae* family were also detected[75,103,104]. It should be mentioned that contigs with taxonomic attribution were low, which highlights the prominence of the viral dark matter. The composition of the human gut virome appears highly specific and stable over time. The differences between individuals are the main sources of variation, despite the fact that a core set of phages was found in 20–50% of individuals[75,76,103,104]. The viral community can also evolve considerably during the first years of life, leading to an increased abundance of *Microviridae*[8]. Lastly, phage distribution is also dependent on individual health status. For example, patients with Crohn's disease and ulcerative colitis exhibit a distinct virome with a significantly increased number of *Caudovirales* phages compared with *Microviridae*[105].

Evolutionary relationships

Genetic mosaicism as the main actor in phage evolution. Defining clear evolutionary relationships is no easy task when it comes to phages. Ironically, what makes phages so diverse and unique is perhaps one of the few features that they have in common: the mosaicity of their genomes. Genetic mosaicism refers to genomes that share regions of high sequence similarity with abrupt transitions into adjacent regions with no detectable resemblance[106]. These regions are often the result of recombination between two non-identical ancestors. Such recombination events, called horizontal gene transfer (HGT), are major mediators of phage evolution, which complicates how we view their evolutionary relationships.

Horizontal gene transfer mechanisms at a glance. The molecular mechanisms leading to HGT have been well studied in model phages and consist of non-homologous, relaxed and homologous recombinations. Non-homologous (also called illegitimate) recombination occurs randomly across the genome[107,108], disrupting genes and gene blocks, likely leaving most of the phage recombinants to be eliminated by counterselection such as host barriers, including anti-phage systems. The mosaic joints (recombination sites) of the few recombinants that emerge are not located randomly. Rather, they are positioned at gene or gene block boundaries as a result of natural selection favouring phages whose biological functions remained undamaged[109]. Relaxed (also called homeologous) recombination takes place at sites of limited homology but that are somewhat related between genomes. In several phages, such as phage lambda, Rad52-like recombinases are responsible for gene shuffling. Relaxed recombination efficiency depends on sequence identity and occurs more frequently than non-homologous recombination[110]. Homologous recombination, although hard to detect[111], is presumed to be the most frequent mechanism for HGT and is promoted by the phage recombination machinery[112].

Virulent phages
Phages that can strictly undergo a lytic mode of replication.

Temperate phages
Phages that can perform either a lytic or a lysogenic mode of replication.

Cryptic
A non-functional prophage within a bacterial chromosome.

Temperate phages are the brokers in HGT. Genetic mosaicity has been studied most extensively in dsDNA phages and was first described in phage lambda[113]. In theory, all dsDNA phage genomes are mosaic because they have access to a large common gene pool through HGT[106]. However, phages do not have equal accessibility to the entire reservoir, as this depends on the number of genetic exchanges required to bring any given sequence from that pool and a particular phage together. For gene exchange to occur, two phages need to infect the same host cell. One scenario involves two virulent phages exchanging genetic material during co-infection. Co-infection appears to be prevalent in natural bacterial populations[114], and a bioinformatic analysis suggested that a recombinant between a ssRNA virus and a ssDNA virus may even arise during co-infection[115]. Because temperate phages can integrate into the host genome and become prophages, they are thought to act as viral sequence reservoirs that likely have a central role in HGT[116]. When a prophage (functional or cryptic) behaves as a sequence donor, the infecting phage (virulent or temperate) becomes the recipient of a new gene or gene block allele, as demonstrated by the emergence of phage lambda recombinants after phage lambda infection of an *E. coli* strain harbouring a cryptic prophage[110]. Similar viral recombinants were observed with dairy phages[117]. These events were made possible by recombinases in the infecting phage genome, which catalyse homologous recombination with a high tolerance to sequence divergence. Bioinformatics analyses support the idea that temperate phages (and prophages) also undergo frequent HGTs, whereas for virulent phages, which form clustered viral populations, mosaicism is still present but seems less crucial[118]. One study showed that phages have two evolutionary modes with distinct rates of HGT[63]. Virulent phages typically fall into the low gene-content flux category whereas temperate phages tend to be distributed in both low and high gene-content flux categories. Another study (discussed below) showed that if we represent phage relationships and gene exchanges as a large network, we find temperate phages at its centre[119], connecting groups of virulent phages located at the periphery. Thus, temperate phages function as banks for HGT[119].

Evolutionary relationships between phages differ by host. Along with lifestyle, the rate and the differential manner in which phages exchange genetic material depend on their hosts and the environment[63] (FIG. 5). Additionally, groups of phages infecting the same host can form discrete genotypic clusters, an uninterrupted genetic continuum or something in between (that is, discrete clusters with an intra-cluster genetic continuum)[120]. For example, despite regular exchanges of photosynthesis genes by homologous recombination, cyanophage genomes still cluster into stable discrete groups[86,121,122]. Virulent dairy phages infecting *Streptococcus thermophilus* have likely recombined with phages infecting other lactic acid bacteria species[123] and exhibit a high gene-content flux despite their lytic lifestyle[124,125]. Mycobacteriophages fall into the 'something in between' category as they are grouped in clusters and display an overall continuous spectrum of genetic diversity. However, intra-cluster diversity and discreetness are highly variable, and temperate mycobacteriophages in different clusters evolve either in the low or high gene-content flux[63,126]. More phages with different nucleic acid genomes (ssDNA and RNA) and that infect diverse bacteria still need to be characterized and sequenced. This will help to elucidate any possible universal patterns in viral evolutionary relationships, confirm the existence of discrete populations in nature and verify whether they are the result of insufficiently sampled environments[127].

A network representation of phage phylogeny. Phage phylogeny has undergone several changes in the past two decades. Classification was initially based on morphology, and traditional phylogenetic trees were used to visualize evolutionary relationships. With the rapid advance of viral metagenomics, a plethora of phage sequences were discovered without determination of the virion morphology. It also became clear that no single gene or protein exists in all phage genomes, making it difficult to build a tree based on a single, shared genomic feature[128]. In addition, phylogenetic trees cannot support the combinatorial nature of phage genomes[119]. Therefore, an alternative way to visualize phage phylogeny is to use networks, with nodes (usually drawn as points) corresponding to phage genomes and edges (connections between nodes, usually drawn as lines) representing similarities at the gene, protein or genome level. This was first shown in 2008, using a set of 306 phage genomes[119]. In this network, temperate phages were shown to be closely interconnected, whereas virulent phages were on the periphery, forming discrete clusters. The path from one virulent phage cluster to another had to pass through temperate phages in the centre of the network. Gene-sharing networks were further explored on the complete dsDNA virosphere (eukaryotic and prokaryotic viruses)[129]. A total of 19 modules (that is, clusters in the network) were identified. Most modules grouped phages according to their ICTV-classified family, although some modules contained phages belonging to different families and even viruses infecting all domains of life. The fact that most modules could be robustly segregated suggests either that, despite fluid gene exchange, phages still evolve in somewhat isolated gene pools or that more phages need to be sequenced to increase connectivity between modules. The presence of modules with viruses from different families also supports the efforts to update the viral classification. Another advance in phage phylogeny is the development of vConTACT[16,130,131], a network-based analytical tool for virus classification (FIG. 5). Already in its second version (vConTACT2 (REF. [131])), this program extracts predicted proteins from each viral genome to build viral protein clusters, which are then used to calculate genome similarities between pairs of viruses. Genome pairs with a similarity score above a given threshold become linked by an edge, and the viral cluster formation is performed by a program that can disentangle complex network relationships and delineate clusters. With this approach, the authors showed that viruses can be accurately clustered at the genus level and that the more the virosphere is sampled, the more robust the network will become.

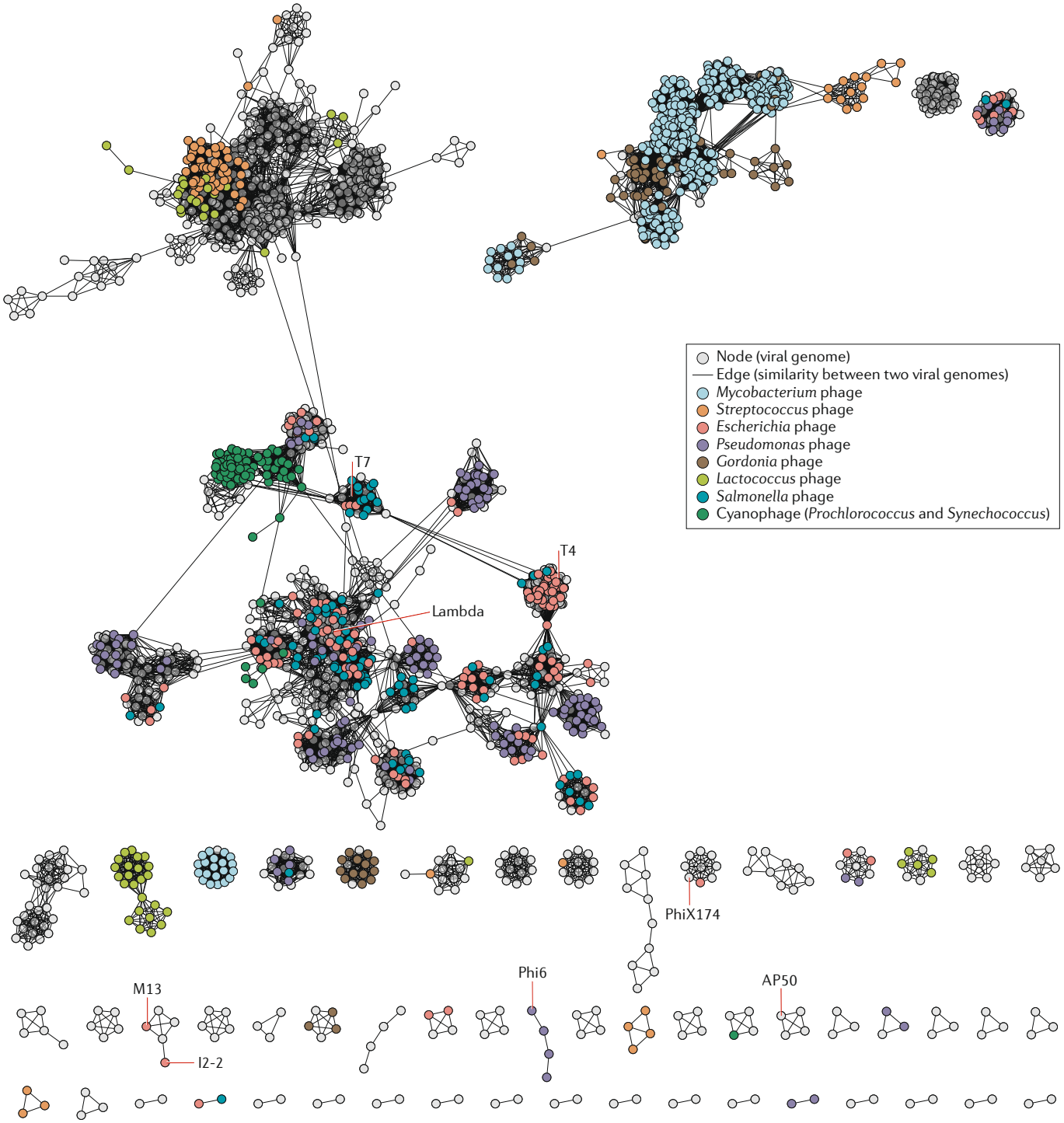


Fig. 5 | **Network representation of phage phylogeny.** Phage phylogeny is traditionally illustrated with trees, but this representation fails to account for horizontal gene transfer events. A new proposed method uses networks to show the complexity of their evolutionary relationships and how phages are interconnected. Here, vConTACT2 (a program for network-based phylogeny) was used to discover relationships between phages, and visualization of the network was performed using Cytoscape v3.7.1. This network comprises 2,617 National Center for Biotechnology Information (NCBI) RefSeq phage genomes, each represented by a node (point). An edge (line) represents a connection between two nodes (genomes) based on the number of shared protein clusters. Orange, dark green and blue nodes correspond to phages infecting *Streptococcus* species, cyanobacteria (*Synechococcus* and *Prochlorococcus* species) and *Mycobacterium* species, respectively. Mycobacteriophages dominate the cluster on the upper-right side of the figure, which has no edges connecting the super-cluster on the left. *Streptococcus* phages are located in a densely connected area of the upper part of the super-cluster, in accordance with their high genetic flux. Cyanophages are more in periphery of the lower part of the super-cluster, consistent with a low genetic flux. Traditional phylogenetic trees are used with the assumption that phages follow a linear evolution. A network-based phylogeny improves our understanding of phage evolutionary relationships, as it gives information on horizontal gene transfers, which are pervasive in phages, and is therefore more representative.

Conclusions

Phage diversity operates differently depending on which aspect of phage biology is investigated. Nucleotide and gene contents are extremely diverse in phage genomes, whereas protein structures are highly conserved even among distinct phage families. In this Review, we have highlighted recent work on viral metagenomics and how it has contributed to the discovery of perhaps the most abundant phages in marine and gut environments. Viral metagenomics also expanded our knowledge of phage diversity in diverse ecosystems and the confines of this global diversity[132]. The ever-expanding catalogue of new phage sequences called for a reflection on how to best adapt the current viral taxonomy to properly classify phages discovered through metagenomics[133]. Phages are also interconnected from an evolutionary perspective and several factors drive higher or lower rates of gene exchange. These complex phylogenetic relationships are more accurately represented by a network rather than a traditional tree, and the former may be better suited to define new phage genera, subfamilies and families[16].

We would also like to emphasize the success of the SEA-PHAGES educational programme, which contributed to the isolation, characterization and sequencing of the largest collection of phages infecting the same host. In an era when a plethora of new phages with completely new sequences are being discovered, this model highlights the importance of integrating phage research at the various teaching levels, which benefits both students and the scientific community. As more phages are discovered, the better the community will be at identifying more of them and illuminating the viral dark matter. Adding more phage sequences to reference databases will help identify a larger diversity of viral sequences from metagenomes. Resolving the structure of more viral proteins could also provide additional insights into the existence of a common ancestor, as intermediate ancestors within the lineage may be uncovered. Finally, network-based phylogenies will be improved when more phage sequences will be added, as this will help to improve the accuracy of clustering phage groups that are poorly sampled at present[130].

Published online: 03 February 2020

1. Suttle, C. A. Viruses in the sea. *Nature* **437**, 356–361 (2005).
2. Nigro, O. D. et al. Viruses in the oceanic basement. *mBio* **8**, 1–15 (2017).
3. Appelt, S. et al. Viruses in a 14th-century coprolite. *Appl. Environ. Microbiol.* **80**, 2648–2655 (2014).
4. Kim, M.-S. & Bae, J.-W. Lysogeny is prevalent and widely distributed in the murine gut microbiota. *ISME J.* **12**, 1127–1141 (2018).
5. Devoto, A. E. et al. Megaphages infect *Prevotella* and variants are widespread in gut microbiomes. *Nat. Microbiol.* **4**, 693–700 (2019).
6. Brum, J. R., Schenck, R. O. & Sullivan, M. B. Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. *ISME J.* **7**, 1738–1751 (2013).
7. Kauffman, K. M. et al. A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. *Nature* **554**, 118–122 (2018).
8. Lim, E. S. et al. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat. Med.* **21**, 1228–1234 (2015).
9. Roux, S. et al. Cryptic inoviruses are pervasive in bacteria and archaea across Earth's biomes. *Nat. Microbiol.* **4**, 1895–1906 (2019).
This study uses a machine learning approach to identify 10,295 previously uncharacterized inoviruses from microbial genomes and metagenomes.
10. Paez-Espino, D. et al. Uncovering Earth's virome. *Nature* **536**, 425–430 (2016).
11. Gregory, A. C. et al. Marine DNA viral macro- and microdiversity from pole to pole. *Cell* **177**, 1–15 (2019).
12. Ackermann, H. W. Phage classification and characterization. *Methods Mol. Biol.* **501**, 127–140 (2009).
13. Ackermann, H. W. 5500 Phages examined in the electron microscope. *Arch. Virol.* **152**, 227–243 (2007).
14. Adams, M. J. et al. 50 years of the International Committee on Taxonomy of Viruses: progress and prospects. *Arch. Virol.* **162**, 1441–1446 (2017).
15. Adriaenssens, E. & Brister, J. R. How to name and classify your phage: an informal guide. *Viruses* **9**, 70 (2017).
16. Barylski, J. et al. Analysis of Spounaviruses as a case study for the overdue reclassification of tailed phages. *Syst. Biol.* **69**, 110–123 (2019).
17. Adriaenssens, E. M. et al. A suggested new bacteriophage genus: 'Viunalikevirus'. *Arch. Virol.* **157**, 2035–2046 (2012).
18. Hua, J. et al. Capsids and genomes of jumbo-sized bacteriophages reveal the evolutionary reach of the HK97 fold. *mBio* **8**, e01579-17 (2017).
19. Duda, R. L. & Teschke, C. M. The amazing HK97 fold: versatile results of modest differences. *Curr. Opin. Virol.* **36**, 9–16 (2019).
20. Agirrezabala, X. et al. Structure of the connector of bacteriophage T7 at 8A resolution: structural homologies of a basic component of a DNA translocating machinery. *J. Mol. Biol.* **347**, 895–902 (2005).
21. Lebedev, A. A. et al. Structural framework for DNA translocation via the viral portal protein. *EMBO J.* **26**, 1984–1994 (2007).
22. Lokareddy, R. K. et al. Portal protein functions akin to a DNA-sensor that couples genome-packaging to icosahedral capsid maturation. *Nat. Commun.* **8**, 14310 (2017).
23. Cardarelli, L. et al. The crystal structure of bacteriophage HK97 gp6: defining a large family of head–tail connector proteins. *J. Mol. Biol.* **395**, 754–768 (2010).
This study shows the evolutionary relationships that can exist among diverse groups of phage proteins.
24. Olia, A. S., Prevelige Jr., P. E., Johnson, J. E. & Cingolani, G. Three-dimensional structure of a viral genome-delivery portal vertex. *Nat. Struct. Mol. Biol.* **18**, 597–603 (2011).
25. Arnaud, C.-A. et al. Bacteriophage T5 tail tube structure suggests a trigger mechanism for *Siphoviridae* DNA ejection. *Nat. Commun.* **8**, 1953 (2017).
26. Leiman, P. G., Chipman, P. R., Kostyuchenko, V. A., Mesyanzhinov, V. V. & Rossmann, M. G. Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell* **118**, 419–429 (2004).
27. Cardarelli, L. et al. Phages have adapted the same protein fold to fulfill multiple functions in virion assembly. *Proc. Natl Acad. Sci. USA* **107**, 14384–14389 (2010).
28. Pell, L. G., Kanelis, V., Donaldson, L. W., Howell, P. L. & Davidson, A. R. The phage λ major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proc. Natl Acad. Sci. USA* **106**, 4160–4165 (2009).
29. Wang, C., Tu, J., Liu, J. & Molineux, I. J. Structural dynamics of bacteriophage P22 infection initiation revealed by cryo-electron tomography. *Nat. Microbiol.* **4**, 1049–1056 (2019).
30. Legrand, P. et al. The atomic structure of the phage Tuc2009 baseplate tripod suggests that host recognition involves two different carbohydrate binding modules. *mBio* **7**, e01781–e01815 (2016).
31. Tremblay, D. M. et al. Receptor-binding protein of *Lactococcus lactis* phages: identification and characterization of the saccharide receptor-binding site. *J. Bacteriol.* **188**, 2400–2410 (2006).
32. Spinelli, S. et al. Modular structure of the receptor binding proteins of *Lactococcus lactis* phages. The RBP structure of the temperate phage TP901-1. *J. Biol. Chem.* **281**, 14256–14262 (2006).
33. Spinelli, S. et al. Lactococcal bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. *Nat. Struct. Mol. Biol.* **13**, 85–89 (2006).
34. Benson, S. D., Bamford, J. K., Bamford, D. H. & Burnett, R. M. Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell* **98**, 825–833 (1999).
35. Abrescia, N. G. et al. Insights into virus evolution and membrane biogenesis from the structure of the marine lipid-containing bacteriophage PM2. *Mol. Cell* **31**, 749–761 (2008).
36. Abrescia, N. G. et al. Insights into assembly from structural analysis of bacteriophage PRD1. *Nature* **432**, 68–74 (2004).
37. Fabry, C. M. S. et al. A quasi-atomic model of human adenovirus type 5 capsid. *EMBO J.* **24**, 1645–1654 (2005).
38. Peralta, B. et al. Mechanism of membranous tunnelling nanotube formation in viral genome delivery. *PLoS Biol.* **11**, e1001667 (2013).
39. Vidaver, A. K., Koski, R. K. & Van Etten, J. L. Bacteriophage ϕ6: a lipid-containing virus of *Pseudomonas phaseolicola*. *J. Virol.* **11**, 799–805 (1973).
40. Krupovic, M. & ICTV Report Consortium. ICTV virus taxonomy profile: *Plasmaviridae*. *J. Gen. Virol.* **99**, 617–618 (2018).
41. Greenberg, N. & Rottem, S. Composition and molecular organization of lipids and proteins in the envelope of mycoplasmavirus MVL2. *J. Virol.* **32**, 717–726 (1979).
42. McKenna, R. et al. Atomic structure of single-stranded DNA bacteriophage ϕX174 and its functional implications. *Nature* **355**, 137–143 (1992).
43. Sun, L. et al. Icosahedral bacteriophage ΦX174 forms a tail for DNA transport during infection. *Nature* **505**, 432–435 (2014).
44. Chipman, P. R., Agbandje-McKenna, M., Renaudin, J., Baker, T. S. & McKenna, R. Structural analysis of the *Spiroplasma* virus, SpV4: implications for evolutionary variation to obtain host diversity among the *Microviridae*. *Structure* **6**, 135–145 (1998).
45. Doore, S. M. & Fane, B. A. The kinetic and thermodynamic aftermath of horizontal gene transfer governs evolutionary recovery. *Mol. Biol. Evol.* **32**, 2571–2584 (2015).
46. Valegard, K., Liljas, L., Fridborg, K. & Unge, T. The three-dimensional structure of the bacterial virus MS2. *Nature* **345**, 36–41 (1990).
47. Peabody, D. S. The RNA binding site of bacteriophage MS2 coat protein. *EMBO J.* **12**, 595–600 (1993).

48. Koning, R. I. et al. Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. *Nat. Commun.* **7**, 12524 (2016).
This article reports the ability of RNA phages to adopt defined conformations that can be involved in genome packaging and virion assembly.
49. Casjens, S. R. The DNA-packaging nanomotor of tailed bacteriophages. *Nat. Rev. Microbiol.* **9**, 647–657 (2011).
50. Marvin, D. A. Filamentous phage structure, infection and assembly. *Curr. Opin. Struct. Biol.* **8**, 150–158 (1998).
51. Xu, J., Dayan, N., Goldbourt, A. & Xiang, Y. Cryo-electron microscopy structure of the filamentous bacteriophage IKe. *Proc. Natl Acad. Sci. USA* **116**, 5493 (2019).
52. Russel, M. & Model, P. A mutation downstream from the signal peptidase cleavage site affects cleavage but not membrane insertion of phage coat protein. *Proc. Natl Acad. Sci. USA* **78**, 1717–1721 (1981).
53. Suhanovsky, M. M. & Teschke, C. M. Nature's favorite building block: deciphering folding and capsid assembly of proteins with the HK97-fold. *Virology* **479–480**, 487–497 (2015).
54. Pietilä, M. K. et al. Structure of the archaeal head–tailed virus HSTV-1 completes the HK97 fold story. *Proc. Natl Acad. Sci. USA* **110**, 10604 (2013).
This article focuses on the MCP HK97 fold and its conservation at the structural level between tailed phages and archaeal and eukaryotic viruses.
55. Jordan, T. C. et al. A broadly implementable research course for first-year undergraduate students. *mBio* **5**, 1–8 (2014).
56. Creasy, A., Rosario, K., Leigh, B. A., Dishaw, L. J. & Breitbart, M. Unprecedented diversity of ssDNA phages from the family *Microviridae* detected within the gut of a protochordate model organism (*Ciona robusta*). *Viruses* **10**, 404 (2018).
57. Roux, S., Hallam, S. J., Woyke, T. & Sullivan, M. B. Viral dark matter and virus—host interactions resolved from publicly available microbial genomes. *eLife* **4**, 1–20 (2015).
58. Yuan, Y. & Gao, M. Jumbo bacteriophages: an overview. *Front. Microbiol.* **8**, 1–9 (2017).
59. Bergh, Ø., Børsheim, K. Y., Bratbak, G. & Heldal, M. High abundance of viruses found in aquatic environments. *Nature* **340**, 467–468 (1989).
60. Hatfull, G. F. Bacteriophage genomics. *Curr. Opin. Microbiol.* **11**, 447–453 (2008).
61. Krupovic, M., Prangishvili, D., Hendrix, R. W. & Bamford, D. H. Genomics of bacterial and archaeal viruses: dynamics within the Prokaryotic virosphere. *Microbiol. Mol. Biol. Rev.* **75**, 610–635 (2011).
This Review presents phage genomic diversity with a main focus on tailed dsDNA phages and an overview of the other phage families.
62. Grose, J. H. & Casjens, S. R. Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family *Enterobacteriaceae*. *Virology* **468**, 421–443 (2014).
63. Mavrich, T. N. & Hatfull, G. F. Bacteriophage evolution differs by host, lifestyle and genome. *Nat. Microbiol.* **2**, 1–9 (2017).
This study presents a large-scale bioinformatic analysis of evolutionary relationships and the rate of HGT in a dataset of more than 2,300 phages.
64. Breitbart, M. et al. Genomic analysis of uncultured marine viral communities. *Proc. Natl Acad. Sci. USA* **99**, 14250–14255 (2002).
65. Brum, J. R. & Sullivan, M. B. Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nat. Rev. Microbiol.* **13**, 147–159 (2015).
66. Breitbart, M., Bonnain, C., Malki, K. & Sawaya, N. A. Phage puppet masters of the marine microbial realm. *Nat. Microbiol.* **3**, 754–766 (2018).
67. Williamson, K. E., Fuhrmann, J. J., Wommack, K. E. & Radosevich, M. Viruses in soil ecosystems: an unknown quantity within an unexplored territory. *Annu. Rev. Virol.* **4**, 201–219 (2017).
68. Brum, J. R. et al. Patterns and ecological drivers of ocean viral communities. *Science* **348**, 1261498 (2015).
69. Hurwitz, B. L. & Sullivan, M. B. The Pacific Ocean virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS One* **8**, 1–12 (2013).
70. Duarte, C. M. Seafaring in the 21st century: the Malaspina 2010 Circumnavigation Expedition. *Limnol. Oceanogr. Bull.* **24**, 11–14 (2015).
71. Roux, S. et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537**, 689–693 (2016).
72. Coutinho, F. H. et al. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nat. Commun.* **8**, 1–12 (2017).
73. Breitbart, M. et al. Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.* **185**, 6220–6223 (2003).
74. Reyes, A. et al. Viruses in the fecal microbiota of monozygotic twins and their mothers. *Nature* **466**, 334–338 (2010).
75. Minot, S. et al. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res.* **21**, 1616–1625 (2011).
76. Manrique, P. et al. Healthy human gut phageome. *Proc. Natl Acad. Sci. USA* **113**, 201601060 (2016).
This study identifies 44 phage groups in the gut microbiota, nine of which are shared across more than one-half of individuals and are proposed to be part of a healthy gut phageome.
77. Zuo, T. et al. Gut mucosal virome alterations in ulcerative colitis. *Gut* **68**, 1169–1179 (2019).
78. Dutilh, B. E. et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* **5**, 4498 (2014).
79. Avrani, S., Wurtzel, O., Sharon, I., Sorek, R. & Lindell, D. Genomic island variability facilitates *Prochlorococcus*–virus coexistence. *Nature* **474**, 604–608 (2011).
80. Martinez-Hernandez, F. et al. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat. Commun.* **8**, 15892 (2017).
This study uses single-virus genomics to identify the most widespread phages in the ocean, which were previously overlooked in metagenomics projects because of their high microdiversity.
81. Martinez-Hernandez, F. et al. Single-cell genomics uncover *Pelagibacter* as the putative host of the extremely abundant uncultured 37-F6 viral population in the ocean. *ISME J.* **13**, 232–236 (2019).
82. Deng, L. et al. Viral tagging reveals discrete populations in *Synechococcus* viral genome sequence space. *Nature* **513**, 242–245 (2014).
This viral ecology study proposes an approach to quantitatively link phage populations and their genomes to their hosts.
83. Aggarwala, V., Liang, G. & Bushman, F. D. Viral communities of the human gut: metagenomic analysis of composition and dynamics. *Mob. DNA* **8**, 12 (2017).
84. Suttle, C. A. Marine viruses — major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
85. Wigington, C. H. et al. Re-examination of the relationship between marine virus and microbial cell abundances. *Nat. Microbiol.* **1**, 15024 (2016).
86. Marston, M. F. & Martiny, J. B. H. Genomic diversification of marine cyanophages into stable ecotypes. *Environ. Microbiol.* **18**, 4240–4253 (2016).
87. Zhao, Y. et al. Abundant SAR11 viruses in the ocean. *Nature* **494**, 357–360 (2013).
88. Holmfeldt, K. et al. Twelve previously unknown phage genera are ubiquitous in global oceans. *Proc. Natl Acad. Sci. USA* **110**, 12798 (2013).
89. López-Pérez, M., Haro-Moreno, J. M., Gonzalez-Serrano, R., Parras-Moltó, M. & Rodriguez-Valera, F. Genome diversity of marine phages recovered from Mediterranean metagenomes: size matters. *PLoS Genet.* **13**, e1007018 (2017).
90. Brum, J. R., Hurwitz, B. L., Schofield, O., Ducklow, H. W. & Sullivan, M. B. Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics. *ISME J.* **10**, 437–449 (2016).
91. Payet, J. P. & Suttle, C. A. To kill or not to kill: the balance between lytic and lysogenic viral infection is driven by trophic status. *Limnol. Oceanogr.* **58**, 465–474 (2013).
92. Thingstad, T. F. & Lignell, R. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat. Microb. Ecol.* **13**, 19–27 (1997).
93. Thingstad, T. F., Vage, S., Storesund, J. E., Sandaa, R.-A. & Giske, J. A theoretical analysis of how strain-specific viruses can control microbial species diversity. *Proc. Natl Acad. Sci. USA* **111**, 7813–7818 (2014).
94. Knowles, B. et al. Lytic to temperate switching of viral communities. *Nature* **531**, 466–470 (2016).
95. Silveira, C. B. & Rohwer, F. L. Piggyback-the-winner in host-associated microbial communities. *NPJ Biofilms Microbiomes* **2**, 16010 (2016).
96. Williamson, K. E., Radosevich, M. & Wommack, K. E. Abundance and diversity of viruses in six Delaware soils. *Appl. Environ. Microbiol.* **71**, 3119–3125 (2005).
97. Chen, L. et al. Effect of different long-term fertilization regimes on the viral community in an agricultural soil of southern China. *Eur. J. Soil. Biol.* **62**, 121–126 (2014).
98. Fierer, N. et al. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Appl. Environ. Microbiol.* **73**, 7059 (2007).
99. Adriaenssens, E. M. et al. Environmental drivers of viral community composition in Antarctic soils identified by viromics. *Microbiome* **5**, 83 (2017).
100. Hoyles, L. et al. Characterization of virus-like particles associated with the human faecal and caecal microbiota. *Res. Microbiol.* **165**, 803–812 (2014).
101. Lepage, P. et al. Dysbiosis in inflammatory bowel disease: a role for bacteriophages? *Gut* **57**, 424–425 (2008).
102. Barr, J. J. et al. Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc. Natl Acad. Sci. USA* **110**, 10771–10776 (2013).
103. Minot, S. & Bryson, A. Rapid evolution of the human gut virome. *Proc. Natl Acad. Sci. USA* **110**, 12450–12455 (2013).
104. Shkoporov, A. N. et al. Reproducible protocols for metagenomic analysis of human faecal phageomes. *Microbiome* **6**, 68 (2018).
105. Norman, J. M. et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* **160**, 447–460 (2015).
106. Hendrix, R. W., Smith, M. C. M., Burns, R. N., Ford, M. E. & Hatfull, G. F. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc. Natl Acad. Sci. USA* **96**, 2192–2197 (1999).
107. Highton, P. J., Chang, Y. & Myers, R. J. Evidence for the exchange of segments between genomes during the evolution of lambdoid bacteriophages. *Mol. Microbiol.* **4**, 1329–1340 (1990).
108. Hatfull, G. F. Dark matter of the biosphere: the amazing world of bacteriophage diversity. *J. Virol.* **89**, 8107–8110 (2015).
109. Juhala, R. J. et al. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J. Mol. Biol.* **299**, 27–51 (2000).
110. De Paepe, M. et al. Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: the role of Rad52-like recombinases. *PLoS Genet.* **10**, e1004181 (2014).
111. Nilsson, A. S. & Haggård-Ljungquist, E. Detection of homologous recombination among bacteriophage P2 relatives. *Mol. Phylogenet. Evol.* **21**, 259–269 (2001).
112. Bobay, L., Touchon, M. & Rocha, E. P. C. Manipulating or superseding host recombination functions: a dilemma that shapes phage evolvability. *PLoS Genet.* **9**, 1–9 (2013).
113. Hershey, A. D. (ed.) *The Bacteriophage Lambda* (Cold Spring Harbor Laboratory Press 1971).
114. Roux, S. et al. Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *eLife* **3**, e03125 (2014).
115. Diemer, G. S. & Stedman, K. M. A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. *Biol. Direct* **7**, 1–14 (2012).
116. Lawrence, J. G., Hatfull, G. F. & Hendrix, R. W. Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. *J. Bacteriol.* **184**, 4891–4905 (2002).
117. Labrie, S. J. & Moineau, S. Abortive infection mechanisms and prophage sequences significantly influence the genetic makeup of emerging lytic lactococcal phages. *J. Bacteriol.* **189**, 1482–1487 (2007).
118. Chopin, A., Bolotin, A., Sorokin, A., Ehrlich, S. D. & Chopin, M.-C. Analysis of six prophages in *Lactococcus lactis* IL1403: different genetic structure of temperate and virulent phage populations. *Nucleic Acids Res.* **29**, 644–651 (2001).
119. Lima-Mendez, G., Helden, J. Van, Toussaint, A. & Leplae, R. Reticulate representation of evolutionary and functional relationships between phage genomes. *Mol. Biol. Evol.* **25**, 762–777 (2008).
This study shows that phage evolutionary relationships are better represented with a reticulate network because mosaicism leads to phages belonging to multiple groups.

120. Hendrix, R. W., Hatfull, G. F. & Smith, M. C. M. Bacteriophages with tails: chasing their origins and evolution. *Res. Microbiol.* **154**, 253–257 (2003).
121. Marston, M. F. & Amrich, C. G. Recombination and microdiversity in coastal marine cyanophages. *Environ. Microbiol.* **11**, 2893–2903 (2009).
122. Gregory, A. C. et al. Genomic differentiation among wild cyanophages despite widespread horizontal gene transfer. *BMC Genomics* **17**, 930 (2016).
123. Szymczak, P., Janzen, T., Neves, R. & Kot, W. Novel variants of *Streptococcus thermophilus* bacteriophages are indicative of genetic recombination among phages from different bacterial species. *Appl. Environ. Microbiol.* **83**, 1–16 (2017).
124. Lavelle, K. et al. A decade of *Streptococcus thermophilus* phage evolution in an Irish dairy plant. *Appl. Environ. Microbiol.* **84**, 1–17 (2018).
125. Kupczok, A. et al. Rates of mutation and recombination in *Siphoviridae* phage genome evolution over three decades. *Mol. Biol. Evol.* **35**, 1147–1159 (2018).
126. Pope, W. H. et al. Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. *eLife* **4**, e06416 (2015).
This study uses the largest collection of phages infecting the same host (*M. smegmatis*) to evaluate evolutionary relationships, genomic clusters and discreteness of these clusters.
127. Hendrix, R. W. Bacteriophages: evolution of the majority. *Theor. Popul. Biol.* **61**, 471–480 (2002).
128. Rohwer, F. & Edwards, R. The phage proteomic tree: a genome-based taxonomy for phage. *J. Bacteriol.* **184**, 4529–4535 (2002).
129. Iranzo, J., Krupovic, M. & Koonin, E. V. The double-stranded DNA virosphere as a modular hierarchical network of gene sharing. *mBio* **7**, 1–21 (2016).
130. Bolduc, B. et al. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect archaea and bacteria. *PeerJ* **5**, e3243 (2017).
131. Jang, H. Bin et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **37**, 632–639 (2019).
132. Cesar Ignacio-Espinoza, J., Solonenko, S. A. & Sullivan, M. B. The global virome: not as big as we thought? *Curr. Opin. Virol.* **3**, 566–571 (2013).
133. Simmonds, P. et al. Virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol.* **15**, 161–168 (2017).
134. Khayat, R. et al. Structure of an archaeal virus capsid protein reveals a common ancestry to eukaryotic and bacterial viruses. *Proc. Natl Acad. Sci. USA* **102**, 18944–18949 (2005).
135. Benson, S. D., Bamford, J. K. H., Bamford, D. H. & Burnett, R. M. Does common architecture reveal a viral lineage spanning all three domains of life? *Mol. Cell* **16**, 673–685 (2004).
136. Hendrix, R. W. Evolution: the long evolutionary reach of viruses. *Curr. Biol.* **9**, 914–917 (1999).
137. Krupovič, M. & Bamford, D. H. Virus evolution: how far does the double β-barrel viral lineage extend? *Nat. Rev. Microbiol.* **6**, 941–948 (2008).
138. Baker, M. L., Jiang, W., Rixon, F. J. & Chiu, W. Common ancestry of herpesviruses and tailed DNA bacteriophages. *J. Virol.* **79**, 14967–14970 (2005).
139. Rixon, F. J. & Schmid, M. F. Structural similarities in DNA packaging and delivery apparatuses in herpesvirus and dsDNA bacteriophages. *Curr. Opin. Virol.* **5**, 105–110 (2014).
140. El Omari, K. et al. Plate tectonics of virus shell assembly and reorganization in phage φ8, a distant relative of mammalian reoviruses. *Structure* **21**, 1384–1395 (2013).
141. Huiskonen, J. T. et al. Structure of the bacteriophage ϕ6 nucleocapsid suggests a mechanism for sequential RNA packaging. *Structure* **14**, 1039–1048 (2006).
142. Bamford, D. H. Do viruses form lineages across different domains of life? *Res. Microbiol.* **154**, 231–236 (2003).
143. Sinclair, R., Ravantti, J. & Bamford, D. H. Nucleic and amino acid sequences support structure-based viral classification. *J. Virol.* **91**, 1–13 (2017).
144. Ackermann, H.-W. Bacteriophage electron microscopy. *Adv. Virus Res.* **82**, 1–32 (2012).
145. Hurwitz, B. L., Brum, J. R. & Sullivan, M. B. Depth-stratified functional and taxonomic niche specialization in the 'core' and 'flexible' Pacific Ocean virome. *ISME J.* **9**, 472–484 (2015).
146. Villar, E. et al. Ocean plankton. Environmental characteristics of Agulhas rings affect interocean plankton transport. *Science* **348**, 1261447 (2015).
147. Luo, E., Aylward, F. O., Mende, D. R. & DeLong, E. F. Bacteriophage distributions and temporal variability in the ocean's interior. *mBio* **8**, e01903–e01917 (2017).
148. Gogokhia, L. et al. Expansion of bacteriophages is linked to aggravated intestinal inflammation and colitis. *Cell Host Microbe* **25**, 285–299.e8 (2019).

Acknowledgements
This work is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Program and the Canadian Institutes of Health Research (team grant on Intestinal Microbiomics, Institute of Nutrition, Metabolism and Diabetes). M.B.D. is a recipient of graduate scholarships from the Fonds de Recherche du Québec — Nature et Technologies (FRQNT) as well as Sentinelle Nord, and is a recipient of the Goran-Enhorning Graduate Student Research Award from the Canadian Allergy, Asthma and Immunology Foundation. F.O. is a recipient of a fellowship from the Swiss National Science Foundation (Early Postdoc. Mobility). S.M. holds the Tier 1 Canada Research Chair in Bacteriophages and is a member of the PROTEO and Op+Lait FRQNT Networks.

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

RELATED LINKS
Félix d'Hérelle Reference Center for Bacterial Viruses:
http://www.phage.ulaval.ca