The elemental composition of virus particles: implications for marine biogeochemical cycles

Luis F. Jover[1], T. Chad Effler[2], Alison Buchan[2], Steven W. Wilhelm[2] and Joshua S. Weitz[1,3]

Abstract | In marine environments, virus-mediated lysis of host cells leads to the release of cellular carbon and nutrients and is hypothesized to be a major driver of carbon recycling on a global scale. However, efforts to characterize the effects of viruses on nutrient cycles have overlooked the geochemical potential of the virus particles themselves, particularly with respect to their phosphorus content. In this Analysis article, we use a biophysical scaling model of intact virus particles that has been validated using sequence and structural information to quantify differences in the elemental stoichiometry of marine viruses compared with their microbial hosts. By extrapolating particle-scale estimates to the ecosystem scale, we propose that, under certain circumstances, marine virus populations could make an important contribution to the reservoir and cycling of oceanic phosphorus.

Dissolved organic matter (DOM). Operationally defined as marine organic matter that passes through a filter with pores of 0.22 µm to 0.45 µm in diameter. DOM can be further classified on the basis of biological availability.

Particulate organic matter (POM). Operationally defined as the material in a marine environment that is retained by a filter with pores of 0.22 µm to 0.45 µm in diameter.

[1]*School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA.*
[2]*Department of Microbiology, University of Tennessee, Knoxville, Tennessee 37996, USA.*
[3]*School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332, USA.*
Correspondence to J.S.W.
e-mail: jsweitz@gatech.edu
doi:10.1038/nrmicro3289

Viruses in marine environments have received considerable attention over the past 2 decades. The lysis of host cells by marine viruses releases dissolved organic matter (DOM) and particulate organic matter (POM) back into the environment, where it can be assimilated by microorganisms or exported from surface waters to the deep ocean. This process of virus-mediated recycling of organic matter — which is known as the 'viral shunt' — is at the centre of many discussions concerning marine geochemical cycles[1–5]. Indeed, viruses are major drivers of the mortality of both phytoplankton and heterotrophic bacteria[5]. The lysis of bacterial cells by viruses (that is, by bacteriophages) and the lysis of eukaryotic host cells by viruses releases previously cell-bound organic material that might facilitate the growth of the remaining community members[6–8] (FIG. 1).

Prior efforts to characterize the viral shunt have mainly focused on the amount of organic material that is released by the lysis of microbial host cells, using carbon as a 'proxy currency'. Here we focus on the often overlooked fact that the organic matter that is released following cell lysis includes virus particles. Although previous work did consider the mass of virus particles, the focus was mainly on the carbon content; for example, Wilhelm and Suttle[1] estimated that marine viruses contained 0.2 fg carbon per virus particle, and this estimate has been carried forward in other studies[5,9]. Steward *et al.*[10] more recently estimated that marine viruses contained 0.055 fg carbon per virus particle by assuming that viruses are composed of equal parts DNA and protein[11] and leveraging the distributions of lengths of marine virus genomes. Viruses differ in capsid size and structure as well as in genome sequence, and these markedly different estimates of the carbon content of virus particles (0.055–0.2 fg carbon per particle) might reflect the inherent variability across a range of virus types within a community. However, irrespective of this variation, marine virus particles are substantially smaller — in terms of overall mass — than their bacterial hosts; for example, for cyanobacteria, the amount of carbon per cell has been estimated to be between 50 fg per cell and 250 fg per cell[12], and for heterotrophic bacteria, the estimated range is 20–100 fg per cell[13]. In both cases, a bacterial cell is predicted to contain at least two orders of magnitude more carbon than a virus particle. It is therefore not surprising that the relative contribution of virus particles to the total magnitude of the viral shunt and dissolved organic carbon (DOC) reservoirs has not been highlighted previously. However, ignoring the content and contribution of virus particles could become more problematic when we turn our attention to nitrogen and phosphorus.

The baseline for studies of the elemental stoichiometry of marine microorganisms was established more than 60 years ago by Alfred Redfield, who estimated that marine plankton (and organic marine detritus) have a

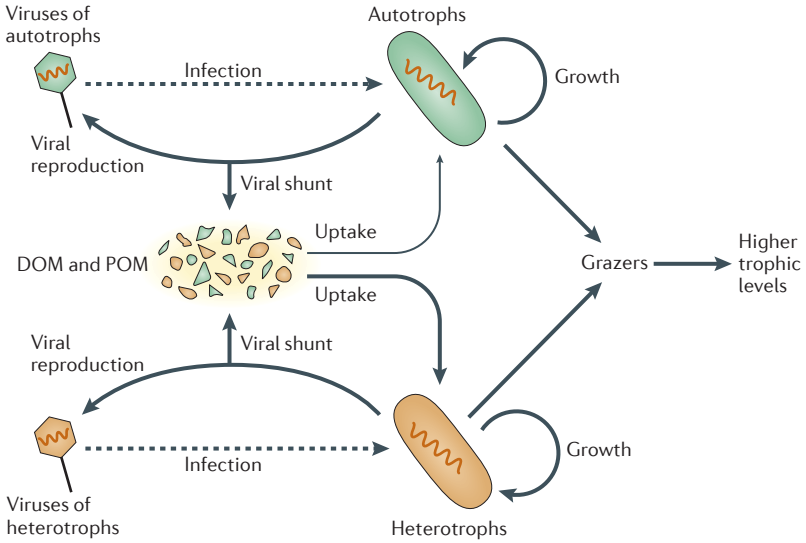

Figure 1 | **Schematic of the viral shunt.** Virus-mediated lysis of microbial cells releases dissolved organic matter (DOM) and particulate organic matter (POM) back into the microbial loop, rather than these cells being targeted by grazers (for example, nanozooplankton and microzooplankton), which can then be consumed by larger organisms in the aquatic food web. Figure adapted, with permission, from REF. 22, Faculty of 1000 Ltd.

carbon/nitrogen/phosphorus (C/N/P) ratio of 106/16/1 (REF. 14). The Redfield ratio remains the foundation for studies of the elemental composition of marine microorganisms and organic matter, particularly in the deep oceans[4,15]. Indeed, the observed elemental ratios can vary greatly with the component of marine organic biomass that is under consideration; for example, estimates of C/N/P for marine heterotrophic bacteria are reported to be around 69/16/1 (REF. 5), and measurements for individual cyanobacterial isolates include 46/10/1 (for *Prochlorococcus* sp. MED4 under phosphorus-replete conditions) and 301/49/1 (for *Synechococcus* sp. WH8013 under phosphorus-limited conditions)[12]. Thus, the elemental stoichiometry of bacterial cells varies with taxa and growth conditions and often differs substantially from the Redfield ratio (reviewed in REF. 16). Virus particles do not require the same relative investment in carbon for structural integrity and other functions as their hosts, as they are predominantly composed of proteins and nucleic acids (although many eukaryotic viruses[17] and a few bacteriophages[18,19] contain lipids). Proteins are enriched in nitrogen and nucleic acids are enriched in both nitrogen and phosphorus compared with the typical contents of a microbial cell. Hence, it should be expected that virus particles are enriched in both nitrogen and phosphorus relative to carbon when compared with microorganisms. We are aware of only one previous example in which the entire carbon, nitrogen and phosphorus content of a virus was enumerated: the elemental stoichiometry of *Paramecium bursaria Chlorella* virus 1 (PBCV1), which is a relatively large virus that infects the freshwater alga *Chlorella* NC64A, was estimated to be 17/5/1 (REF. 20). Calculating the elemental stoichiometry of PBCV1 was intended to test the consequences of stoichiometric differences between the virus and its algal host. However, neither the elemental content of virus particles nor the fraction of dissolved organic nitrogen (DON) and dissolved organic phosphorus (DOP) that is partitioned in viral populations has been systematically studied so far.

In this Analysis article, we use an *ab initio* approach to predict the elemental content of virus particles, with a focus on bacteriophages. This unified biophysical and biochemical model is generalizable and enables the prediction of the elemental content and ratios that are present in bacteriophage particles of varying sizes. We leverage this model to estimate how much of the carbon, nitrogen and phosphorus that is released following the phage-induced lysis of an individual bacterial host cell is bound in cellular debris and how much is integrated in progeny bacteriophage particles. Then, we scale up these particle-level estimates to estimate the amount of the DOC, DON and DOP that is partitioned in virus populations in marine surface waters in order to begin to quantify the relevance of virus particles to marine biogeochemical pools and fluxes.

The elemental composition of virus particles

Virus particles are commonly composed of a head and, in some cases, a tail. The two major components of the head are the capsid, which comprises a protein shell that is usually one single protein layer thick[21], and the genetic material that is packed inside it, which is either DNA or RNA. In this section, we use a geometric model of the viral head to estimate the number of macromolecules that comprises each of its components, that is, the number of proteins in the capsid and the number of nucleotides in the genetic material. We focus on double-stranded, icosahedral and spherical DNA bacteriophages that do not contain lipids, but the model could be extended to consider different viral shapes or types.

The basis for our model is an approximation of the virus head as a spherical shell with a fraction of its internal volume filled by DNA (FIG. 2a). The spherical shell can be described in terms of its external radius, which denotes the distance from the centre of the capsid to its outer boundary. The expected number of base pairs inside the capsid, is:

$$n_{bp} = fill \frac{v_{ic}}{v_{bp}} = fill \frac{4\pi}{3v_{bp}} (r_c - h)^3 \quad (1)$$

where v_{ic} is the volume inside the capsid, *fill* is the volume-filling fraction of the DNA inside the capsid, h is the thickness of the capsid and v_{bp} is the volume of a base pair.

To estimate the quantity of proteins in the capsid, each with average volume, we assume that the capsid is a spherical shell that has a uniform thickness taking up a volume. The expected number of capsid proteins is:

$$n_{pr} = \frac{v_c}{v_{pr}} = \frac{4\pi}{3v_{pr}} (r_c^3 - (r_c - h)^3) \quad (2)$$

An expression for the number of proteins as a function of the number of base pairs can be obtained using equation 1 to obtain r_c as a function of n_{bp} and substituting this result into equation 2. It can be seen from

Heterotrophic bacteria
Bacteria that use organic carbon compounds to satisfy nutritional requirements.

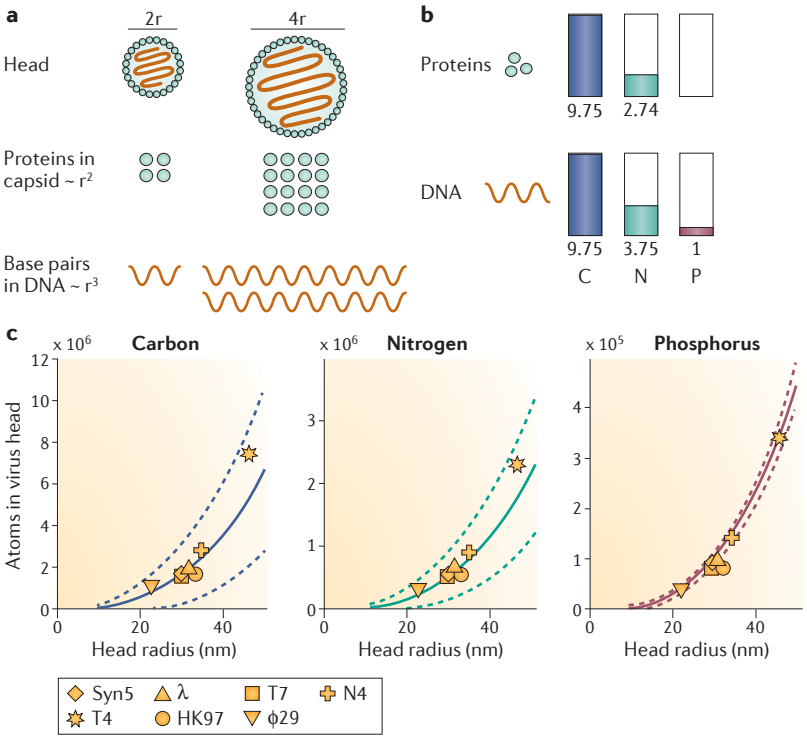

Figure 2 | **Model of the elemental stoichiometry of virus particles.** **a** | The viral head is approximated as a spherical shell of proteins with an internal core of nucleic acids. The number of proteins scales with the radius of the capsid squared (r^2) and the number of base pairs scales with the radius of the capsid cubed (r^3). **b** | The relative carbon, nitrogen and phosphorus content within proteins and nucleic acids. **c** | The carbon, nitrogen and phosphorus content of the viral head as a function of its external radius (solid lines). The data correspond to experimentally obtained carbon, nitrogen and phosphorus contents for different viral heads. The protein compositions of the capsids were obtained from the following references: T4 (REF. 84), N4 (REF. 79), Syn5 (REF. 80), λ and T7 (REF. 21), HK97 (REF. 81), Φ29 (REF. 83). The model predictions are:

$C_{head} = 41(r_c - 2.5)^3 + 130(7.5r_c^2 - 18.75r_c + 15.63)$,

$N_{head} = 16(r_c - 2.5)^3 + 36(7.5r_c^2 - 18.75r_c + 15.63)$ and

$P_{head} = 4.2\ (r_c - 2.5)^3$, where r_c is in units of nm.

The basis for model error estimates (dashed lines) is detailed in Supplementary information S1 (box). The estimate of the scaling between genome length and virus capsid radius (Supplementary information S1 (box)) is key.

equation 1 and equation 2 that the number of proteins in the capsid scales, to leading order, with the square of the viral radius — that is, the number of proteins in the capsid is proportional to the capsid surface area — and that the number of nucleotides scales, to leading order, with the cube of the viral radius — that is, the number of nucleotides is proportional to the internal volume of the capsid (FIG. 2a). This size-dependent scaling of macromolecular content is crucial to characterize the size-dependent scaling of elemental content within virus particles. To complete our biophysical scaling model of the elemental stoichiometry of virus heads, we incorporated the average elemental composition of proteins and nucleotides (proteins contain carbon and nitrogen but not phosphorus, whereas DNA (and RNA) contains carbon, nitrogen and phosphorus (FIG. 2b)). A complete list of biophysical and biochemical constants can be found in Supplementary information S1 (box); the raw data supporting the calculations in the box are available in Supplementary information S2 (table), Supplementary information S3 (table) and Supplementary information S4 (table).

Our estimates of the elemental content in virus tails follow a similar logic. The virus tail, when present, is comprised of tyre-shaped protein subunits. Hence, we approximate the tail structure as a hollow cylinder with outside radius r_t, fixed thickness h_t, and length l_t. The number of proteins in the tail is then estimated by dividing the tail volume by the volume of a protein. The abundance of carbon and nitrogen atoms in the tail is the product of the number of proteins and the average carbon and nitrogen content of proteins; unlike virus heads, virus tails should not contribute to the total phosphorus content of virus particles (for details of the calculation, see Supplementary information S1 (box)).

Elemental stoichiometry depends on virus size

The theoretical model predictions for the elemental composition of viral heads were validated using data for seven bacteriophages that infect either heterotrophic bacteria or cyanobacteria (TABLE 1). These bacteriophages were chosen as complete genome sequences, structural characterization (that is, the full complement of viral head proteins and their stoichiometry) and the amino acid sequences for all of the head protein components were publicly available. The elemental composition of each bacteriophage particle was calculated from the genome sequence (carbon, nitrogen and phosphorus) and the protein stoichiometry and amino acid content of each protein in the capsid (carbon and nitrogen only). The predictions of elemental abundance that were obtained using the biophysical scaling model strongly agreed with those that were directly measured for the seven reference phage heads (FIG. 2c; TABLE 1; Supplementary information S1 (box)), which suggests that the model can be used to predict the size-dependence of the elemental stoichiometry of intact virus particles.

To evaluate contributions from viral tails, the carbon, nitrogen and phosphorus content was calculated for three viruses (T4, N4 and Syn5) for which there is detailed structural information available about both the head and tail (Supplementary information S1 (box)). Tail inclusion modestly increased the total carbon and nitrogen content but not the phosphorus content (Supplementary information S1 (box)). Thus, considering the head size (or genome length) alone will lead to underestimates of ~2–14% for total carbon and nitrogen content in viral particles, and the degree of underestimation depends on the size of the tail relative to the head. The ratio of tail to head size can be quantified when quantitative transmission electron microscopy (qTEM) data are available (Supplementary information S1 (box)).

As is apparent, the elemental content of virus particles will scale with size. To account for this effect in the case of virus heads, our biophysical model assumes that the capsid surface layer is composed of proteins, and hence the total number of proteins scales, approximately, with the genome length raised to the power of two-thirds. Large viruses will have relatively higher proportions of

Cyanobacteria
Ubiquitous marine bacteria that fix inorganic carbon compounds into organic carbon compounds.

Quantitative transmission electron microscopy
(qTEM). Method to quantitatively estimate viral morphological characteristics (such as morphotype, capsid diameter and tail length) using transmission electron microscopy.

Table 1 | **Elemental composition, genome lengths and radii of selected viral heads**

Virus	Carbon (atoms)	Nitrogen (atoms)	Phosphorus (atoms)	C/N/P*	Genome length (bp)‡	Radius (nm)	Refs
Enterobacteria phage T4	7,339,851	2,309,461	337,806	22.3/ 6.6/ 1	168,903	46.3§	78, 84
Escherichia spp. phage N4	2,765,271	891,989	140,306	17.9/ 5.9/ 1	70,153	34.8	79
Synechococcus phage Syn5	1,664,612	563,058	92,428	17.3/ 5.9/ 1	46,214	30	80
Enterobacteria phage λ	1,897,088	621,823	97,004	19.2/6.3/1	48,502	31.5	21, 78
Enterobacteria phage HK97	1,640,059	536,465	79,464	20.6/6.8/1	39,732	33	81
Enterobacteria phage T7	1,550,638	517,474	79,874	18.1/6.1/1	39,937	30	21, 82
Bacillus spp. phage Φ29	1,058,584	326,209	38,564	23.3/7.3/1	19,282	22.8§	83

*Composition based on sequence information and experimentally validated stoichiometry of structural proteins. ‡The GenBank accession numbers of the DNA sequences are: NC_000866.4 (T4), NC_008720.1 (N4), NC_009531.1 (Syn5), NC_001416.1 (λ), NC_002167.1 (HK97), NC_001604.1 (T7), NC_011048 (φ29). §The reported radius is equivalent to that of the sphere with the same volume as an ellipsoid with the dimensions of the corresponding phage prolate head.

phosphorus, as the number of phosphorus atoms scales with the number of nucleotides. As large viruses (viruses that have capsid diameters that are greater than 100 nm) are considered, the elemental ratios of the virus particles would approach the elemental ratios of their nucleic acids asymptotically. Considering only C/N ratios initially, viruses of any size are predicted to have ratios that are similar to that of their DNA (FIG. 3). By contrast, when considering N/P ratios, virus size matters — the N/P ratios of large viruses are predicted to be similar to that of their DNA, whereas the N/P ratios of small viruses are predicted to approach the Redfield ratio (FIG. 3).

Stoichiometric mismatch between virus and host

Virus-induced host cell lysis releases virus particles and cellular debris into the environment; cellular debris is released in the form of DOM and POM. It has been suggested that the elemental stoichiometry of the lysed contents may be different from that of the original host cell[5,22]. The difference in the elemental contents of viruses and their hosts (an example of 'stoichiometric mismatch' (REF. 16)) has the potential to drive the differential release of nutrients following cell lysis. In this section, we use our scaling model for viral C/N/P to consider the role of virus-mediated lysis in transforming the elemental stoichiometry of released organic matter into two components that have potentially different stoichiometries and proportions: cellular debris and virus particles.

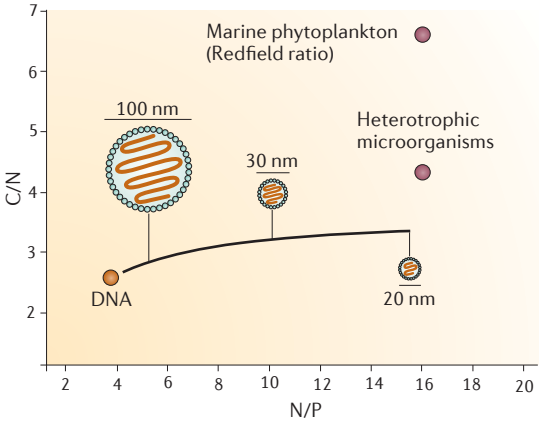

Figure 3 | **Theoretical prediction of elemental stoichiometry for viruses.** We compare the predicted C/N and N/P ratios for viruses (solid black line) with the ratios that are expected for DNA, marine phytoplankton (as per the Redfield ratio) and heterotrophic microorganisms. The theoretical curve (solid black line) denotes the predicted stoichiometry for viruses that have capsid diameters in the range of 20 nm to 300 nm. The stoichiometry of three representative viruses that have capsid diameters of 20 nm, 30 nm and 100 nm is shown.

The amount of material that is released following cell lysis can be predicted on the basis of the following principle of conservation of elements:

$$X_{\text{lysis}} = X_{\text{host}} - \beta X_{\text{virus}} + X_{\text{acquired}} \quad (3)$$

where X is the element of interest and β is the burst size (that is, the number of virus particles that is released from each lysed host cell). The subscripts denote the elemental content that is released as cellular debris (X_{lysis}), the content that is released in each virus particle (X_{virus}), the content that is found in the original host (X_{host}) and the content that is acquired (by uptake or fixation) after infection but before lysis (X_{acquired}). We expect that $X_{\text{acquired}} > 0$, as it is known that both resource uptake and fixation occur during viral infection and that viruses can repurpose and modify the uptake rates of nutrients; for example, phosphorus is a common limiting nutrient for marine cyanobacteria *Prochlorococcus* spp. and *Synechococcus* spp. (as can be assessed via the diagnostic product of the *pstS* gene (REF. 23)). Some cyanophage genomes that have been isolated from phosphorus-depleted ocean waters contain the *pstS* gene (REFS 24, 25), and laboratory experiments have shown that cyanophages can upregulate *pstS* expression to increase the acquisition of phosphorus during infection[26]. Thus, cyanophage manipulation of the host phosphorus status is crucial to the elemental stoichiometry of the virus-infected cyanobacterial host cell. However, specific predictions for a host–virus system that result from equation 3 would require measurements of X_{acquired}, and such data are not broadly available. Below, we consider a series of limit cases in order to gauge the sensitivity of our predictions to variations in post-infection uptake.

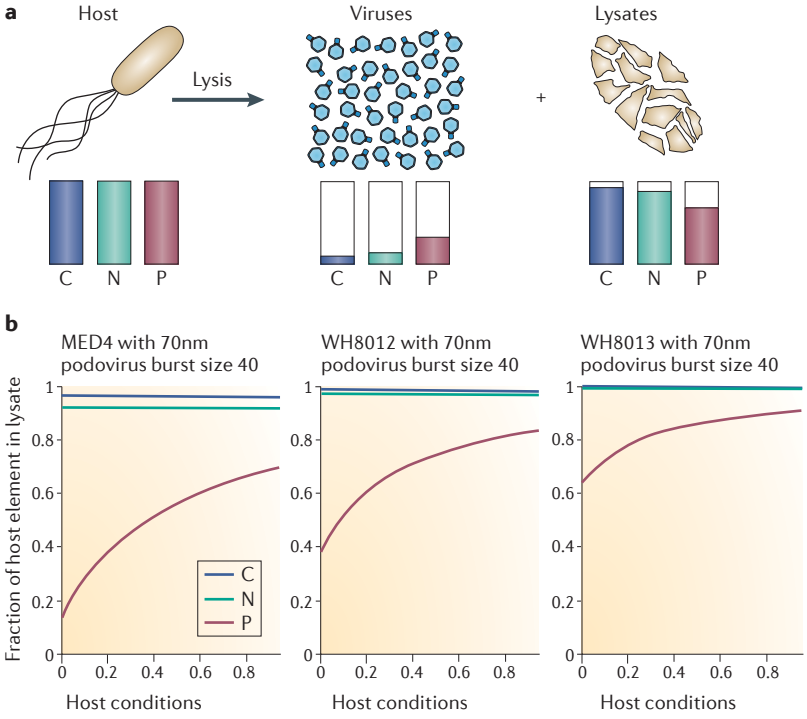

Figure 4 | **Virus-induced transformation of elemental content in cellular debris following lysis.** **a** | After lysis, the contents of a host are released as virus particles and other cellular debris (that is, the lysate). The relative carbon, nitrogen and phosphorus levels in the virus particles and the lysate are shown in bars; the bar height is normalized in each case by the total amount of the element in the host. Bar heights are based on the hypothetical infection of *Prochlorococcus* sp. MED4 host under phosphorus-replete conditions by a podovirus with 70 nm diameter head. Note that, in fact, the total carbon, nitrogen and phosphorus content differs between the host and the virus (FIG. 3). **b** | Predictions of the model of elemental transformation as applied to a viral infection of three marine cyanobacteria. The fraction of the host element in the lysate in each panel is normalized by the elemental content of the respective host. The *x* axis denotes culture conditions from phosphorus-limited (0) to phosphorus-replete (1) (REF. 12).

In the first limit case, we assume that no extracellular elements are obtained during the production of virus particles, that is, virus particles are assembled solely from organic materials that are already present in the host at the time of infection (that is, $X_{acquired}$ ~0). This condition is likely to be relevant for some marine viruses; for example, studies of marine bacteriophage genome synthesis using radioactively labelled phosphorus found that "the majority of nucleotides used during phage DNA synthesis were derived from degraded nucleic acids (that is, DNA and RNA) of the host cell" (REF. 27). Under this scenario, how much carbon, nitrogen and phosphorus would be released as cellular debris?

To answer this, we considered a hypothetical scenario in which a podovirus with a capsid diameter of 70 nm and burst size of 40 infected distinct cyanobacterial strains (FIG. 4a). We focus on three specific host systems: the cyanobacteria *Prochlorococcus* sp. MED4, *Synechococcus* sp. WH8012 and *Synechococcus* sp. WH8103, for which the total dry masses of carbon, nitrogen and phosphorus have been measured in axenic cultures using both phosphorus-replete and phosphorus-limited culture conditions[12]. We then used equation 3 (where $X_{acquired} = 0$) to estimate the abundance of carbon, nitrogen and phosphorus contained within virus particles relative to the total carbon, nitrogen and phosphorus in the lysate (FIG. 4b).

For all three virus–host pairs, the predicted relative abundance of carbon and nitrogen in the lysate is close to 1, which indicates that most host-derived organic carbon and nitrogen was released as cellular debris rather than bound in virus particles. By contrast, phosphorus was predicted to be mostly bound within virus particles. For phosphorus-limited *Prochlorococcus* sp. MED4, the model predicts that 87% of the original phosphorus content of the cells would be converted into virus particles and only 13% would be released in the non-viral component. Although under phosphorus-replete conditions *Prochlorococcus* sp. MED4 (FIG. 3a) had a C/N/P ratio of ~120.5/21.2/1, the current model conditions predict that virus particles would have a C/N/P ratio of ~17.2/5.9/1 and the remaining cellular debris would have a C/N/P ratio of ~171.9/28.8/1. Hence, in the absence of phosphorus uptake by the infected host, the cellular materials that are released as debris (and are not bound in viruses) will be depleted in phosphorus relative to carbon and nitrogen; how much so will depend on the elemental content of the host (and the elemental availability), the size of the virus and the burst size. Indeed, when virus particles are predominantly synthesized from host elements that are available before infection, burst sizes are likely to decrease in nutrient-poor hosts; for example, smaller virus burst sizes were observed when the algal virus PBCV1 infected 'poor-quality' algal hosts (that is, hosts that had a high C/P ratio)[20]. A similar phenomenon of reduced lysis (and, consequently, an increased incidence of lysogeny) has been observed during infections of phosphorus-limited *Synechococcus* sp. WH7803 hosts[28].

To contrast this scenario, we considered a second limit case in which virus particles are produced from elemental pools that are acquired post-infection (whereby $\beta X_{virus} = X_{acquired}$). In this scenario, there would be no change in the elemental stoichiometry of lysed materials that were not bound within virus particles. The reality is probably somewhere between these two limit cases; for example, formative studies of DNA synthesis in T-even phage infecting *Escherichia coli* suggested that new assimilation of phosphorus during infection is common, although host nucleic acids are also used for virus synthesis[29,30]. In cyanobacteria, the assimilation of inorganic matter that is associated with viral infection augments nutrient pools that are already available in the host cell[24–26]. The potential degree to which virus particles are formed from pre-existing or newly acquired elements can be characterized in terms of the elemental independence of the virus ($\mathcal{E}_X$), such that $\mathcal{E}_X$ approaches 0 if the virus uses only host elements and acquires all of the elements that are used in production of virus particles during infection. This index can be estimated by measuring:

$$\varepsilon_x = \frac{X_{acquired}}{X_{acquired} + X_{host}} = \frac{X_{lysis} - X_{host} + \beta X_{virus}}{X_{lysis} + \beta X_{virus}} = 1 - \frac{X_{host}}{X_{lysis} + \beta X_{virus}} \quad (4)$$

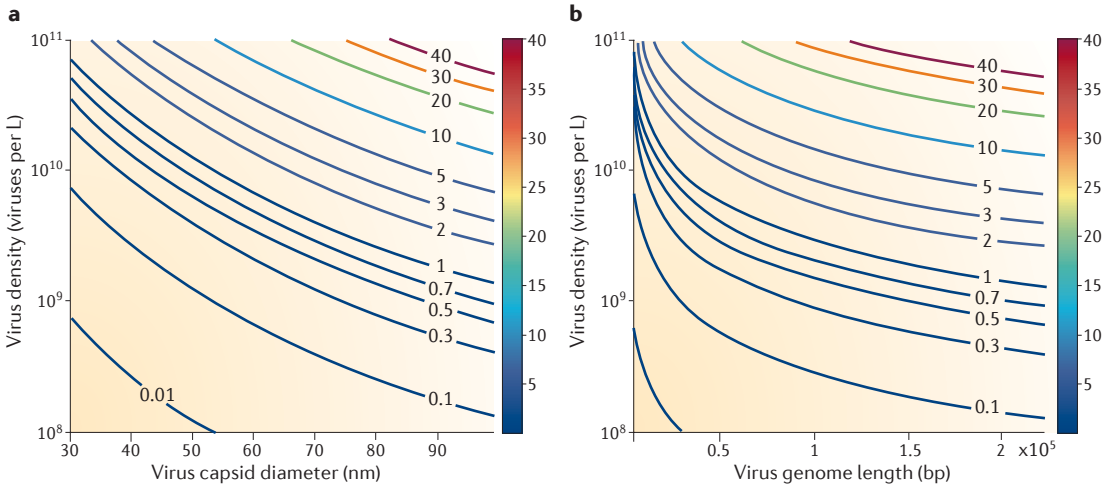

Figure 5 | **Predicted DOP concentration in viral populations as a function of viral density and virus size.** **a** | Virus size is quantified in terms of mean capsid diameter, which varies from 30 nm to 100 nm. **b** | Virus size is quantified in terms of genome length, which ranges from 4.1 kb to 220 kb. In both cases, viral density varies from 10^8 virus particles per litre to 10^{11} virus particles per litre on a logarithmically spaced axis. The contour lines denote combinations of viral density and capsid diameter, which correspond to the same predicted concentration of dissolved organic phosphorus (DOP) that is partitioned in viruses. The colour bar indicates the predicted DOP in units of nmol per litre.

If the $\mathcal{E}_X$ of a virus approaches 1, then the stoichiometry of released cellular debris should approach that of the original host cell stoichiometry.

These two limit cases show that the extent to which viruses change the stoichiometry of cellular debris depends on differences between their stoichiometry and the stoichiometry of their host as well as their elemental independence. Although previous experimental studies showed that cellular debris that is released following lysis stimulates the growth of non-targeted cells[6–8] (probably owing to the increased bioavailability of carbon (REF. 31), nitrogen (REF. 32) and iron (REF. 33)), our model predicts that the lysate tends to be depleted in nutrients relative to the host (the tendency can be characterized in terms of the parameter $\mathcal{E}_X$). This cellular debris may be a source of refractory matter that is relatively low in phosphorus and is not repurposed for viral production. These observations lay the groundwork for future hypothesis testing in ecologically important model systems (for example, in abundant viruses that infect pelagibacteria, cyanobacteria and roseobacteria[34]) to determine how viruses transform cellular material into recalcitrant organic matter[4]. Such data would be invaluable for constraining the range of predictions of our model regarding the relative elemental content of cellular debris and virus particles that are released following lysis.

Marine virus particles and elemental reservoirs

The total abundance of elements in marine viruses is a product of the abundance of viruses and their per-particle elemental content, both of which can vary across space and time. Marine virus densities range from 10^8 virus particles per litre in oligotrophic systems to more than 10^{11} virus particles per litre in highly productive environments[35–37]. Viruses (in general) and marine viruses (in particular) can vary substantially in genome length and particle size and, consequently, in elemental content (FIG. 2; Supplementary information S1 (box)). However, to extrapolate from particle-scale results to the population scale, variation in the sizes of marine virus particles must be taken into account.

We first examined the expected phosphorus content of marine virus populations. In FIG. 5a, we show the predicted DOP concentration (in nmol per litre) that is bound in virus populations if virus densities are in the range of 10^8 to 10^{11} virus particles per litre and virus capsid sizes are 30–100 nm. Likewise, in FIG. 5b, we show the predicted DOP concentration (in nmol per litre) that is bound in virus populations, using the same range of virus densities but considering ranges of virus genomes from 4.1 kb to 220 kb (as derived on the basis of our scaling model). The total DOP content of a virus population that comprises identically sized virus particles ranges from less than 0.1 nM (in the limit of low densities and small particles) to greater than 40 nM (in the limit of high densities and large particles). However, we can narrow this range on the basis of the global qTEM survey of marine viruses by Brum *et al.*[38]. This qTEM study found that most virus capsids have diameters of 50–70 nm and that individual capsid diameters range from 20 nm to 200 nm (REF. 38). Restricting our extrapolation to populations that have average capsid diameters in the 50 nm to 70 nm range, we predict a range of 0.0025 fg to 0.0074 fg phosphorus per particle. Hence, as marine virus densities range from 10^8 virus particles per litre to 10^{11} virus particles per litre, we predict that marine virus populations contain 0.0079–24 nM DOP. These predictions rely on our model of virus heads only, as any underestimates of elemental content owing to contributions from virus tails would affect carbon and nitrogen predictions but not phosphorus predictions.

By means of comparison, DOP estimates across multiple oceanic realms, including the Alaskan Gyre, North

Oligotrophic
A term used to describe an aquatic environment that has low levels of nutrients and photosynthetic production (for example, the open ocean).

Sea, North Pacific Ocean, North Pacific Gyre, Subtropical North Atlantic, Southeast Pacific and Southern Ocean, range from 30 nM to 300 nM (REFS 39–48). DOP can constitute more than 80% of the total dissolved phosphorus pool in marine surface samples[43]. The structural characterization (including solid-state ^{31}P-NMR (REFS 40, 49)) of compounds in the DOP pool does not rule out nucleic acids as important constituents. Hence, it is likely that virus particles are included in these estimates, even if their contribution to the total DOP was not specifically identified. It is evident that the upper ranges of DOP that is bound in viral populations approach the lower ranges of the total DOP pool size that was measured in marine surface waters. Considering the extremes of these ranges, we find that virus populations may potentially account for anything from <0.01% of the total DOP pool (when dividing the lower value of 0.0079 nM that is bound in viruses with the upper value of 300 nM in the total pool) to >50% of the total DOP pool (when dividing the upper value of 24 nM that is bound in viruses with the lower value of 30 nM in the total pool). However, dividing the limits of two biological ranges does not fully take into account the potential correlations between DOP and virus abundance. Hence, we use three case studies — the Sargasso Sea, Station ALOHA in the North Pacific Ocean Subtropical Gyre and a coastal site in the Southern Pacific Ocean — to illustrate circumstances in which virus populations are likely to have distinct degrees of importance with respect to the proportion of DOP that is partitioned in marine surface waters.

The Bermuda Atlantic Time-series Study (BATS) has collected oceanographic data in the Sargasso Sea, which is a well-studied open-ocean environment in the Atlantic Ocean, for 25 years. Assuming that DOP makes up 80% of the total dissolved phosphorus at BATS (REF. 43), then 95% of DOP measurements from 2000–2009 range from 25 nM to 109 nM, with an average of 58 ± 22 nM (REF. 50). Given decadal measurements (in 2000 and 2009) of seasonally varying virus abundances at BATS[51], we predict that the fraction of Sargasso Sea DOP that is partitioned in viruses is in the range of 0.10% to 8.0% (TABLE 2). Similarly, the Hawaii Ocean Time series (HOT) programme includes more than 25 years of oceanographic data, which has been collected within the North Pacific Subtropical Gyre, including a deep-water site (Station ALOHA) near to the island of Oahu, Hawaii. The data that were collected at Station ALOHA include measurements of DOP, such that 95% of the DOP measurements that were made between 1988 and 2012 range from 150 nM to 320 nM, with an average of 224 ± 46 nM (REF. 52). When these DOP ranges (and in one case, concurrent measurements of DOP) are compared with two marine virus surveys that were conducted at this site[53,54], we predict that the fractional DOP that is bound in virus populations at HOT ranges between 0.15% and 1.5%. Hence, at both the BATS and HOT sites, we predict that viruses do not generally make up a significant fraction of the DOP pool, although they can potentially exceed 5% of the total DOP pool at BATS (TABLE 2).

Table 2 | **Viral and non-viral components of DOP in marine surface waters**

Data set	Site information (for virus survey)	Virus abundance (virus particles per litre)	Estimated DOP in virus particles (nM)*	Total DOP in surface waters (nM)	Estimated % of total DOP bound in virus particles‡	Refs
Marine virus surveys		10^8–10^{11}	<0.01–24			5, 35–37
Marine DOP surveys				30–300		39–48
Atlantic Subtropical Gyre (BATS)	Late summer 2000–2009; depth: 60–100 m	6–12 x 10^9	0.48–2.9	58 ± 22 (station average)	0.61–8.0%	43, 50, 51
	Stratified summer 2000–2009; depth: 0–20 m	1–3 x 10^9	0.081–0.72	58 ± 22 (station average)	0.10–2.0%	43, 50, 51
North Pacific Subtropical Gyre (HOT), open-ocean site	September 1998; depth: 0–100 m	4.5–5.5 x 10^9	0.36–1.3	230 ± 20 (concurrent)	0.15–0.63%	52, 53
	December 2002; depth: 0–100 m	8–11.3 x 10^9	0.65–2.7	224 ± 46 (station average)	0.24–1.5%	52, 54
South Pacific Ocean and Southern Ocean (including open ocean and near-coastal sites)	South West New Zealand, Coastal; Sep–Oct 2008; depth: ~5 m	17–120 x 10^9	1.4–29	150–225§	0.61–19%	42, 55
	Australian Southern Ocean; Jan–Feb 2007; depth: 0–40 m	6.1–26 x 10^9	0.49–6.2	150–225§	0.22–4.1%	42, 58
	Drake Passage, Greenwich Meridian and Weddell Sea; Feb–Apr 2008; depth: 0–100 m	0.1–7.6 x 10^9	<0.01–1.8	150–225§	<0.01–1.2%	42, 57

BATS, Bermuda Atlantic Time-series Study; DOP, dissolved organic phosphorus; HOT, Hawaii Ocean Time-series. *The range of estimated DOP in virus particles was inferred by multiplying virus abundance by the phosphorous content per particle, using 50 nm head diameter for the lower range and 70 nm head diameter for the upper range (that is, 0.0025 fg and 0.0074 fg, respectively). ‡The percentage range of DOP bound in virus particles was estimated by dividing (lower/upper) DOP in virus particles by total DOP (upper/lower). §Comparison site of DOP measurements was separate from virus measurements (54 °0' S, 176 ° 0' W)).

By contrast, estimates of virus particles that exceed 10^{11} virus particles per litre have been reported in several productive locations (that is, during spring blooms) in the southern Pacific Ocean near New Zealand[55,56]. Given peak virus densities of 1.2×10^{11} per litre, the DOP concentration in virus populations would range from 9.7 nM to 29 nM. By contrast, peak virus abundances for open-ocean sites in the Southern Ocean are approximately four times lower[57,58]. Simultaneous measurements of DOP and virus abundances for both the coastal and open-ocean Southern Pacific Ocean and Southern Ocean sites are not available. As a proxy, we note that the total DOP in the upper 100 m of an open-ocean site in the Southern Ocean during the summer ranged between 150 nM and 225 nM (REF. 42), which is consistent with other Pacific Ocean estimates (for example, 100–200 nM (REF. 46)). Hence, we predict that the fraction of DOP that is partitioned in viruses during spring blooms would have exceeded 5% if the DOP concentration at coastal sites was similar to DOP concentrations elsewhere (TABLE 2).

We carried out a similar analysis for the contribution of virus particles to DOC and DON using our biophysical model. We predict that marine viruses have a carbon content that ranges from 0.02 fg to 0.05 fg and a nitrogen content that ranges from 0.0078 fg to 0.02 fg, assuming median capsid diameters in the range of 50–70 nm. Using estimated virus densities that can reach up to 10^{11} per litre, we predict that the total carbon content that is bound in marine viruses is probably no greater than 420 nM and that the total nitrogen content is probably no greater than 140 nM. In contrast to the case of DOP, we find that viruses are unlikely to constitute a substantial fraction of DOC, which has been estimated to be 34–80 μmol per litre in marine waters[59]. In the case of nitrogen, concentrations of approximately 2–7 μmol per litre (mean of 4.5 ± 0.4 μmol per litre) have been reported in global surface waters[60]; thus, the viral contribution to the DON pool could range from negligible amounts to up to 7%, depending on the system.

Taken together, our scaling model predicts that the DOP (and possibly DON) content that is bound in marine viruses is potentially larger than has been recognized. Our analysis shows that virus populations are likely to constitute an important fraction (that is, >5%) of the total DOP in marine surface waters when virus density exceeds 3.5×10^{10} virus particles per litre and the total DOP concentration is approximately 100 nM (FIG. 5). Note that, although coincident measurements of virus population sizes and DOP are few in number, virus densities that approach or exceed 10^{11} virus particles per litre have been reported at multiple stations, including in an estuary (Chesapeake Bay)[61], the North Sea[62] and the Coastal Northeastern Pacific Ocean[35] (reviewed in REF. 37). Efforts to quantify the relative contribution of viruses to the total DOP should also consider estimates of virus turnover rates. Virus particles in marine surface waters have lifespans on the scale of hours to days[57,58,63,64]. Hence, the DOP pool that is bound in virus particles is likely to be turned over quickly relative to other components of the DOP pool, which have been estimated to turn over in the order of days to weeks (for example, at BATS (REF. 65)). The availability of nutrient-rich virus particles at high densities suggests that virus particles may be a target for consumption by certain marine zooplankton, particularly if ingestion does not entail an infection risk. Indeed, 20 year-old observations of nanoflagellate ingestion and digestion of viruses and virus-sized particles suggest that such digestion can and does occur[66]. The importance of this target pool will depend on virus abundance and diversity, particularly in nutrient-limited regions of the ocean, and warrants renewed exploration.

Conclusions

Our analyses provide a foundation from which to evaluate the quantitative role that marine viruses have in the storage and recycling of DOC, DON and DOP. The key logical underpinning of the model is the fact that changes in the surface/volume ratio, combined with the known differential composition of proteins (which contain carbon and nitrogen but not phosphorus) and nucleotides (which contain carbon, nitrogen and phosphorus), are sufficient to make quantitative predictions regarding the elemental content of virus particles of different sizes. These particle-scale estimates can be extrapolated to ecosystem scales using marine virus survey data. Our analysis leads us to two important conclusions: first, the stoichiometric mismatch between marine viruses and their hosts provides a mechanism for the differential recycling of organic matter via the viral shunt; and second, virus particles have the potential to contribute an important component of the DOP (and, in some cases, the DON) of marine surface waters.

Although we have focused on marine bacteriophage particles, our conclusions probably apply to marine viruses generally, including viruses of archaea and eukaryotes, such as algae. Irrespective of the specific host that is involved, the elemental stoichiometry of viruses should be enriched in nitrogen and phosphorus relative to the baseline elemental stoichiometry of the host cells (as has been confirmed in the case of the algae-infecting virus PBCV1 (REF. 20)). The rationale is that the structural requirements for viruses are relatively small compared with those of their hosts (with the possible exception of 'giant' viruses, which contain carbon-rich structures, including lipids and complex carbohydrates[67,68]). We predict that there will be a stoichiometric mismatch between viruses and their hosts, leading to the depletion of nitrogen and phosphorus in cellular debris relative to host cells. Quantitative predictions of the carbon, nitrogen and phosphorus content of virus particles, broken down by virus type, may require specific extensions of the present model; this is an issue of continued interest owing to the continual discovery of novel viruses in the ocean, some of which occur at high densities (for example, single-stranded RNA (ssRNA) viruses[69]), and the discovery of virus-like particles, such as gene-transfer agents[70]. We suggest that further rigorous estimates of the total phosphorus, accompanied by characterization of the chemical nature of DOP in seawater (for an example, see REF. 71), will help to elucidate the relative

Spring blooms
Annual increases in phytoplankton abundance in response to seasonal changes, such as increased temperature and higher nutrient levels.

Gene-transfer agents
Phage-like particles that encapsulate cellular DNA that can be transferred to another bacterium.

contributions of viruses to the total DOP pool. Despite the need for simplification, the benefit of building a biophysical model is that it helps to identify potential differences in drivers of elemental content between viruses.

Our analysis of the stoichiometry of released materials also suggests that elemental acquisition by infected cells is important. Marine viruses can change the elemental assimilation rates of their infected hosts; for example, a recent study of lytic viral infections of marine *Prochlorococcus* spp. showed post-infection redirection of host metabolism towards increased assimilation of limiting elements (in this case, phosphorus) that might benefit the phage[26]. This study is consistent with the previous identification and annotation of phosphorus metabolism genes[24] and nitrogen metabolism genes[25] in viral genomes, which suggests that virus-mediated repurposing of cellular metabolism occurs during infection. Further exploration is warranted to quantify the relative rates at which viruses repurpose host elements and acquire new elements post-infection (for example, as in REFS 25, 27). We anticipate that viruses that have less elemental independence from their hosts are more likely to release cellular debris with a stoichiometry that is distinct from that of the host and are thus potentially more recalcitrant to subsequent assimilation that stimulates secondary production. Quantifying the consequences of the stoichiometric mismatch of viruses and their hosts may provide insights into the export of carbon-rich organic matter into the deep ocean[4], in much the same way that the stoichiometric mismatch between herbivores and plants and between decomposers and plant litter has improved our understanding of nutrient cycling and the storage and regeneration of organic matter in terrestrial systems[16].

In summary, this Analysis article advances long-standing efforts to quantify the relative contribution of viruses to the pools and fluxes of DOC, DON and DOP in the global oceans. In doing so, our results further highlight the importance of using classic morphological[39,72–74], biophysical[35] and biogeochemical techniques[1,3], in conjunction with modern biomolecular techniques[75–77], to provide insights into the effects of marine viruses on oceanic biogeochemistry. The fact that most marine microorganisms can be infected by viruses, including ubiquitous hosts, such as pelagibacteria, which were once thought to harbour no viruses[34], suggests that quantifying changes in nutrient assimilation by infected hosts and the partitioning of nutrients that are released from lysed hosts may have important consequences on a global scale.

1. Wilhelm, S. W. & Suttle, C. A. Viruses and nutrient cycles in the sea. *Bioscience* **49**, 781–788 (1999). **This report estimates the contribution of virus-mediated lysis of marine bacteria and algae to the biogeochemical cycling of carbon, in the process of defining the 'viral shunt'.**
2. Brussaard, C. P. D. *et al.* Global-scale processes with a nanoscale drive: the role of marine viruses. *ISME J.* **2**, 575–578 (2008).
3. Fuhrman, J. A. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**, 541–548 (1999). **This paper gives a foundational perspective on the roles of viruses in ecosystems and proposes estimates for the changes in carbon fluxes between ecosystem pools in aquatic food webs owing to virus-mediated lysis.**
4. Jiao, N. *et al.* Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nature Rev. Microbiol.* **8**, 593–599 (2010). **This review discusses the production of recalcitrant carbon in marine systems and includes the viral shunt as a central component of the hypothesized production process.**
5. Suttle, C. A. Marine viruses — major players in the global ecosystem. *Nature Rev. Microbiol.* **5**, 801–812 (2007).
6. Gobler, C. J. *et al.* Release and bioavailability of C, N, P, Se, and Fe following viral lysis of a marine Chrysophyte. *Limnol. Oceanogr.* **42**, 1492–1504 (1997).
7. Middelboe, M., Jørgensen, N. O. G. & Kroer, N. Effects of viruses on nutrient turnover and growth efficiency of noninfected marine bacterioplankton. *Appl. Environ. Microbiol.* **62**, 1991–1997 (1996).
8. Noble, R. T., Middelboe, M. & Fuhrman, J. A. The effects of viral enrichment on the mortality and growth of heterotrophic bacterioplankton. *Aquat. Microb. Ecol.* **18**, 1–13 (1999).
9. Suttle, C. A. Viruses in the Sea. *Nature* **437**, 356–361 (2005). **This paper provides the first global-scale diagram of the carbon cycle that includes marine viruses as an important component.**
10. Steward, G. F. *et al.* Microbial biomass and viral infections of heterotrophic prokaryotes in the subsurface layer of the central Arctic Ocean. *Deep-Sea Res. I* **54**, 1744–1757 (2007).
11. Steward, G. F., Montiel, J. L. & Azam, F. Genome size distributions indicate variability and similarities among marine viral assemblages from diverse environments. *Limnol. Oceanogr.* **45**, 1697–1706 (2000).
12. Bertilsson, S. *et al.* Elemental composition of marine *Prochlorococcus* and *Synechococcus*: implications for the ecological stoichiometry of the sea. *Limnol. Oceanogr.* **48**, 1721–1731 (2003). **This paper describes how phosphorus limitation shifts the subsequent elemental composition of marine cyanobacteria grown in culture.**
13. Simon, M. & Azam, F. Protein content and protein synthesis rates of planktonic marine bacteria. *Mar. Ecol. Prog. Ser.* **51**, 201–213 (1989). **This paper shows that viruses in nutrient-limited environments are predominantly synthesized from elemental pools that are already available in host cells (for example, recycled nucleotides).**
14. Redfield, A. C. Ketchum, B. H., & Richards, F. A. in *The composition of seawater. Comparative and descriptive oceanography. The sea: ideas and observations on progress in the study of the seas.* (Ed. Hill, M. N.) Vol. 2, 26–77 (Interscience Publishers,1963).
15. Jiao, N. *et al.* The microbial carbon pump and the oceanic recalcitrant dissolved organic matter pool. *Nature Rev. Microbiol.* **9**, 555 (2011).
16. Sterner, R. W. & Elser, J. J. *Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere.* (Princeton University Press, 2002).
17. Iyer, L. M., Aravind, L. & Koonin, E. V. Common origin of four diverse families of large eukaryotic DNA viruses. *J. Virol.* **75**, 11720–11734 (2001).
18. Espejo, R. T. & Canelo, E. S. Properties of bacteriophage PM2: a lipid-containing bacterial virus. *Virology* **34**, 738–747 (1968).
19. Bamford, D. H. *et al.* Constituents of SH1, a novel lipid-containing virus infecting the halophilic euryarchaeon *Haloarcula hispanica*. *J. Virol.* **79**, 9097–9107 (2005).
20. Clasen, J. L. & Elser, J. J. The effect of host *Chlorella* NC64A carbon: phosphorus ratio on the production of *Paramecium bursaria Chlorella* Virus 1. *Freshw. Biol.* **52**, 112–122 (2008). **This study shows how virus productivity during infections decreases with decreasing host quality, and tabulates the total elemental content of a single algal virus.**
21. Calendar, R. & Abedon S. T. *The Bacteriophages* (Oxford University Press, 2005).
22. Weitz, J. S. & Wilhelm, S. W. Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biol. Rep.* **4**, 17 (2012).
23. Scanlan, D. J. *et al.* An immunological approach to detect phosphate stress in populations and single cells of photosynthetic picoplankton. *Appl. Environ. Microbiol.* **63**, 2411–2420 (1997).
24. Sullivan, M. B. *et al.* Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biology* **3**, e144 (2005).
25. Sullivan, M. B. *et al.* Genomic analysis of oceanic cyanobacterial myoviruses compared with T4 like myoviruses from diverse hosts and environments. *Environ. Microbiol.* **12**, 3035–3056 (2010).
26. Zeng, Q. & Chisholm, S. W. Marine viruses exploit their host's two-component regulatory system in response to resource limitation. *Curr. Biol.* **22**, 124–128 (2012).
27. Wikner, J. *et al.* Nucleic acids from the host bacterium as a major source of nucleotides for three marine bacteriophages. *FEMS Microbiol. Ecol.* **12**, 237–248 (1993).
28. Wilson, W. H., Carr, N. G. & Mann, N. H. The effect of phosphate status on the kinetics of cyanophage infection in the oceanic cyanobacterium *Synechococcus* sp. Wh78031. *J. Phycol.* **32**, 506–516 (1996).
29. Cohen, S. S. The synthesis of nucleic acid by virus-infected bacteria. *Bacteriol. Rev.* **15**, 131–146 (1951).
30. Kozloff, L. M. & Putnam, F. W. Biochemical studies of virus reproduction: III. The origin of virus phosphorus in the *Eschericia coli* T6 bacteriophage system. *J. Biol. Chem.* **182**, 229–243 (1950).
31. Middelboe, M. *et al.* Virus-induced transfer of organic carbon between marine bacteria in a model community. *Aquat. Microb. Ecol.* **33**, 1–10 (2003).
32. Shelford, E. J. *et al.* Virus-driven nitrogen cycling enhances phytoplankton growth. *Aquat. Microb. Ecol.* **66**, 41–46 (2012).
33. Poorvin, L. *et al.* Viral release of iron and its bioavailability to marine plankton. *Limnol. Oceanogr.* **49**, 1734–1741 (2004).
34. Zhao, Y. *et al.* Abundant SAR11 viruses in the ocean. *Nature* **494**, 357–360 (2013). **This paper provides the first description of viruses that infect pelagibacteria spp., which are the most abundant heterotrophic bacterium in the oceans.**
35. Clasen, J. L. *et al.* Evidence that viral abundance across oceans and lakes is driven by different biological factors. *Freshw. Biol.* **53**, 1090–1100 (2008).
36. Wilhelm, S. W. & Matteson, A. R. Freshwater and marine virioplankton: a brief overview of commonalities and differences. *Freshw. Biol.* **53**, 1076–1089 (2008).

37. Danovaro, R. *et al.* Marine viruses and global climate change. *FEMS Microbiol. Rev.***35**, 993–1034 (2011).
38. 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).
This study provides quantitative estimates of the size and morphology of virus particles in seawater and provides constraints on the probable ranges of marine virus sizes.
39. Ridal, J. J. & Moore, R. M. Dissolved organic phosphorus concentrations in the northeast subarctic Pacific Ocean. *Limnol. Oceanogr.* **37**, 1067–1075 (1992).
40. Clark, L. L., Ingall, E. D. & Benner, R. Marine phosphorus is selectively remineralized. *Nature* **393**, 426–426 (1998).
41. Ammerman, J. W. *et al.* Phosphorus deficiency in the Atlantic: an emerging paradigm in oceanography. *Eos* **84**, 165–170 (2003).
42. Loh, A. N. & Bauer, J. E. Distribution, partitioning and fluxes of dissolved and particulate organic C, N and P in the eastern North Pacific and Southern Oceans. *Deep Sea Res. I* **47**, 2287–2316 (2000).
43. Lomas, M. W. *et al.* Sargasso Sea phosphorus biogeochemistry: an important role for dissolved organic phosphorus (DOP). *Biogeosciences* **7**, 695–710 (2010).
This study shows the importance of organic, rather than inorganic, forms of phosphorus in supporting primary production in some oceanic realms (for example, in the Sargasso Sea).
44. Nausch, M. & Nausch, G. Bioavailability of dissolved organic phosphorus in the Baltic Sea. *Mar. Ecol. Prog. Ser.* **321**, 9–17 (2006).
45. Nishimura, Y., Kim, C. & Nagata, T. Vertical and seasonal variations of bacterioplankton subgroups with different nucleic acid contents: possible regulation by phosphorus. *Appl. Environ. Microbiol.* **71**, 5828–5836 (2005).
46. Raimbault, P., Garcia, N. & Cerutti, F. Distribution of inorganic and organic nutrients in the South Pacific Ocean — evidence for long-term accumulation of organic matter in nitrogen-depleted waters. *Biogeosciences* **5**, 281–298 (2008).
47. van der Zee, C. & Chou, L. Seasonal cycling of phosphorus in the southern bight of the North Sea. *Biogeosciences Discussions* **1**, 681–707 (2004).
48. Yoshimura, T. *et al.* Distributions of particulate and dissolved organic and inorganic phosphorus in North Pacific surface waters. *Marine Chem.* **103**, 112–121 (2007).
49. Kolowith, L. C., Ingall, E. D. & Benner, R. Composition and cycling of marine organic phosphorus. *Limnol. Oceanogr.* **46**, 309–320 (2001).
50. Bermuda Institue of Ocean Sciences. Bermuda Atlantic Time-Series Study [online], http://bats.bios.edu/
51. Parsons, R. J. *et al.* Ocean time-series reveals recurring seasonal patterns of virioplankton dynamics in the northwestern Sargasso Sea. *ISME J.* **6**, 273–284 (2011).
52. Fujieki, L. A. Hawaii Ocean Time Series Data Organization and Graphical System (HOT-DOGS) [online], http://hahana.soest.hawaii.edu/hot/hot-dogs/index.html
53. Culley, A. I. & Welschmeyer, N. A. The abundance, distribution, and correlation of viruses, phytoplankton, and prokaryotes along a Pacific Ocean transect. *Limnol. Oceanogr.* **47**, 1508–1513 (2002).
54. Brum, J. R. Concentration, production and turnover of viruses and dissolved DNA pools at Stn ALOHA, North Pacific Subtropical Gyre. *Aquat. Microb. Ecol.* **41**, 103–113 (2005).
55. Matteson, A. R. *et al.* High abundances of cyanomyoviruses in marine ecosystems demonstrate ecological relevance. *FEMS Microbiol. Ecol.* **84**, 223–234 (2013).
56. Strzepek, R. F. *et al.* Spinning the 'ferrous wheel': the importance of the microbial community in an iron budget during the FeCycle experiment. *Global Biogeochem. Cycles* **19**, GB4S26 (2005).
57. Evans, C. & Brussaard, C. P. D. Regional variation in lytic and lysogenic viral infection in the Southern Ocean and its contribution to biogeochemical cycling. *Appl. Environ. Microbiol.* **78**, 6741–6748 (2012).
58. Evans, C., Pearce, I. & Brussaard, C. P. D. Viral-mediated lysis of microbes and carbon release in the sub-Antarctic and Polar Frontal zones of the Australian Southern Ocean. *Environ. Microbiol.* **11**, 2924–2934 (2009).
59. Hansell, D. A. *et al.* Dissolved organic matter in the ocean: new insights stimulated by a controversy. *Oceanography* **22**, 202–211 (2009).
60. Letscher, R. T. *et al.* Dissolved organic nitrogen in the global surface ocean: distribution and fate. *Global Biogeochem. Cycles* **27**, 141–153 (2013).
61. Winget, D. M. *et al.* Repeating patterns of virioplankton production within an estuarine ecosystem. *Proc. Natl Acad. Sci. USA* **108**, 11506–11511 (2011).
62. Winter, C. *et al.* Linking bacterial richness with viral abundance and prokaryotic activity. *Limnol. Oceanogr.* **50**, 968–977 (2005).
63. Matteson, A. R. *et al.* Production of viruses during a spring phytoplankton bloom in the South Pacific Ocean near of New Zealand. *FEMS Microbiol. Ecol.* **79**, 709–719 (2012).
64. Suttle, C. A., Chan, A. M. & Chen, F. *Cyanophages and Sunlight: a Paradox*. (eds. Guerrero, R. & Pedros-Alio, C.) 303–307 (Spanish Society for Microbiology, 1993).
65. Bjorkman, K., Thomson-Bulldis, A. L. & Karl, D. M. Phosphorus dynamics in the North Pacific subtropical gyre. *Aquat. Microb. Ecol.* **22**, 185–198 (2000).
66. Gonzalez, J. M. & Suttle, C. A. Grazing by marine nanoflagellates on viruses and virus-sized particles: ingestion and digestion. *Marine Ecol. Prog. Ser.* **94**, 1–10 (1993).
This study shows the ecological potential for viruses to be targets of grazing.
67. Fischer, M. G. *et al.* Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc. Natl Acad. Sci. USA* **107**, 19508–19513 (2010).
68. La Scola, B. *et al.* The virophage as a unique parasite of the giant mimivirus. *Nature* **455**, 100–104 (2008).
69. Steward, G. F. *et al.* Are we missing half of the viruses in the ocean? *ISME J.* **7**, 672–679 (2013).
70. Forterre, P. *et al.* Fake virus particles generated by fluorescence microscopy. *Trends Microbiol.* **21**, 1–5 (2012).
71. Dyhrman, S. T., Ammerman, J. W. & Van Mooy, B. A. S. Microbes and the marine phosphorus cycle. *Oceanography* **20**, 110–116 (2007).
72. Lindell, D. *et al.* Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**, 86–89 (2005).
This study shows that host-derived, viral-encoded protein expression during infection boosts host metabolism during viral progeny production.
73. Frank, H. & Moebus, K. An electron microscopic study of bacteriophages from marine waters. *Helgol. Mar. Res.* **41**, 385–414 (1987).
74. Ackermann, H.-W. & Heldal, M. in *Manual of Aquatic Viral Ecology*. (eds Wilhelm, S., Weinbauer, M. & Suttle C.) 182–192 (American Society of Limnology and Oceanography, 2010).
75. Angly, F. E. *et al.* The marine viromes of four oceanic regions. *PLoS Biol.* **4**, 2121–2131 (2006).
76. Williamson, S. J. *et al.* The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS ONE* **3**, e1456 (2008).
77. Deng, L. *et al.* Contrasting life strategies of viruses that infect photo- and heterotrophic bacteria, as revealed by viral tagging. *mBio* **3**, e00373-12 (2012).
78. De Paepe, M. & Taddei, F. Viruses' life history: towards a mechanistic basis of a trade-off between survival and reproduction among phages. *PLoS Biol.* **4**, e193 (2006).
79. Choi, K. H. *et al.* Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. *J. Mol. Biol.* **378**, 726–736 (2008).
80. Pope, W. H. *et al.* Genome sequence, structural proteins, and capsid organization of the cyanophage Syn5: a 'horned' bacteriophage of marine *Synechococcus*. *J. Mol. Biol.* **368**, 966–981 (2007).
81. Wikoff, W. R. *et al.* Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**, 2129–2133 (2000).
82. Ionel, A. *et al.* Molecular rearrangements involved in the capsid shell maturation of bacteriophage T7. *J. Biol. Chem.* **286**, 234–242 (2011).
83. Tao, Y. *et al.* Assembly of a tailed bacterial virus and its genome release studied in three dimensions. *Cell* **95**, 431–437 (1998).
84. Leiman, P. G. *et al.* Structure and morphogenesis of bacteriophage T4. *Cell. Mol. Life Sci.* **60**, 2356–2370 (2003).

Acknowledgements
This work was supported by US National Science Foundation (NSF) grants OCE-1233760 (to J.S.W.) and OCE-1061352 (to A.B. and S.W.W.). This work was assisted by attendance as a short-term visitor (J.S.W.) and participation (A.B., S.W.W. and J.S.W.) in the Ocean Viral Dynamics working group at the US National Institute for Mathematical and Biological Synthesis — an Institute that is sponsored by the NSF, the US Department of Homeland Security and the US Department of Agriculture through NSF Award EF-0832858, with additional support from The University of Tennessee, Knoxville, USA. J.S.W. holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund. The authors thank participants of the Ocean Viral Dynamics working group, M. Sullivan, J. Brum and three anonymous referees for their feedback and suggestions.

Competing interests statement
The authors declare no competing interests.

DATABASES
GenBank: http://www.ncbi.nlm.nih.gov
NC_000866.4 | NC_008720.1 | NC_009531.1 | NC_001416.1 | NC_002167.1 | NC_001604.1 | NC_011048

SUPPLEMENTARY INFORMATION
See online article: S1 (box) | S2 (table) | S3 (table) | S4 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

Author Biographies

Luis F. Jover is a Ph.D. candidate in physics at the Georgia Institute of Technology, Atlanta, USA. His research interests are in theoretical ecology, with a focus on modelling interactions between marine viruses and their microbial hosts.

T. Chad Effler is an undergraduate majoring in electrical engineering and computer science at the University of Tennessee, Knoxville, USA. His research interests include theoretical computing and bioinformatics.

Alison Buchan is Associate Professor of Microbiology at the University of Tennessee, Knoxville, USA. In 2001, she received a Ph.D. in marine sciences from the University of Georgia, Atlanta, USA. Her research interests are broadly in the area of marine microbial ecology, with an emphasis on bacterial transformations of lignin-derived compounds and phage–host interactions.

Steven W. Wilhelm is Professor of Microbiology at the University of Tennessee, Knoxville, USA. In 1994, he received his Ph.D. in plant sciences from the University of Western Ontario, Canada. His research interests broadly centre around aquatic biogeochemical cycles and the viruses and microbial community members that constrain them.

Joshua S. Weitz is Associate Professor of Biology at the Georgia Institute of Technology, Atlanta, USA. In 2003, he received his Ph.D. in physics from the Massachusetts Institute of Technology, Cambridge, USA. His research focuses on the application of mathematical and physical models to the life sciences, with an emphasis on virus–host interactions.

Key points

- Virus-mediated lysis of host cells results in the generation of dissolved organic carbon (DOC), dissolved organic nitrogen (DON) and dissolved organic phosphorus (DOP) via a process that is known as the 'viral shunt'.
- Previous quantitative estimates of the contribution of the viral shunt to biogeochemical cycles focused on host cellular constituents and overlooked the contribution of virus particles.
- In this Analysis article, we develop a biophysical scaling model that predicts the elemental contents and compositions of virus particles.
- This scaling model was validated using detailed sequence and structural contents of intact bacteriophage particles.
- Viruses are predicted to be enriched in phosphorus, so much so that the total phosphorus content in a burst of released viruses may approach that of the phosphorus content in an uninfected host.
- As a consequence, cellular debris may be depleted in phosphorus compared with the stoichiometry of hosts.
- Furthermore, by extrapolating the model to the ecosystem scale, marine viruses are predicted to contain an important fraction (for example, >5%) of the total DOP pool in some systems (for example, in surface waters, when virus density exceeds 3.5×10^{10} and the DOP concentration is approximately 100 nM).

000 The elemental composition of virus particles: implications for marine biogeochemical cycles

Luis F. Jover, T. Chad Effler, Alison Buchan, Steven W. Wilhelm and Joshua S. Weitz

Weitz and colleagues use a biophysical scaling model of intact virus particles to quantify differences in the elemental stoichiometry of marine viruses compared with their microbial hosts. They propose that, under certain circumstances, marine virus populations could make a previously unrecognised and important contribution to the reservoir and cycling of oceanic phosphorus.

Subject categories

Biological sciences / Microbiology / Bacteriophages [URI /631/326/1321]
Biological sciences / Microbiology / Bacteria / Marine microbiology [URI /631/326/41/2535]
Biological sciences / Biophysics [URI /631/57]

Supplementary Information for: "The elemental composition of virus particles: implications for marine biogeochemical cycles"

S.1 Parameters and ranges used in the theory of viral elemental stoichoimetry

S.1.1 Core derivation

We approximate the elemental content of the viral head as the sum of that contained in the genome and in the capsid (see Figure 2). In doing so, we utilize the notation X_{comp} to denote the number of atoms of element X in the component "comp". First, the number of atoms of element X in the genome is $X_{genome} = X_{bp} \cdot n_{bp}$ where X_{bp} denotes twice the elemental composition of a base in a single nucleic acid (i.e., representing double-stranded DNA) and the number of base pairs in a single strand is n_{bp}. The number of base pairs is calculated as described in the main text, i.e., as a fraction of the total volume in the capsid. The number of atoms of element X in the capsid is:

$$X_{capsid} = X_{pr} \cdot n_{pr} = d_X v_{pr} n_{pr} = d_X V_c \tag{S1}$$

where d_X is a constant that has units of number of atoms of element X per unit volume in a protein, X_{pr} denotes the elemental abundance in a single protein, v_{pr} is the volume of a protein and V_c is the volume of proteins in the capsid. The values of the constants used are listed in Table S1. The number of atoms of element X in the head is:

$$X_{head} = \frac{4\pi(r_c - h)^3 X_{bp} fill}{3 v_{bp}} + \frac{4\pi d_X \left(3r_c^2 h - 3h^2 r_c + h^3\right)}{3}. \tag{S2}$$

By utilizing the elemental content of nucleotides and proteins, in addition to other biochemical parameters (see Table S1), we arrive at the following scaling predictions:

$$\mathrm{C}_{head} = 41(r_c - 2.5)^3 + 130(7.5r_c^2 - 18.75r_c + 15.63) \tag{S3}$$

$$\mathrm{N}_{head} = 16(r_c - 2.5)^3 + 36(7.5r_c^2 - 18.75r_c + 15.63) \tag{S4}$$

$$\mathrm{P}_{head} = 4.2(r_c - 2.5)^3 \tag{S5}$$

where r_c is in units of nanometers and predictions are for the number of atoms of each element, respectively.

There is uncertainty related to each one of the biophysical parameters which propagates into uncertainty in X_{head}. We introduce the uncertainty in our theory, ΔX_{head}, by propagating the uncertainty of all the parameters involved. As a result, we obtain upper and lower bounds of our theory that can be written as $X_{head} \pm \Delta X_{head}$, and that are presented as dotted lines in the main text (Figure 2C). The uncertainty in the number of atoms of element X as a function of the radius is approximated as:

$$\Delta X_{head} = \sqrt{(\Delta X_{genome})^2 + (\Delta X_{capsid})^2}, \tag{S6}$$

where

$$\Delta X_{genome} = \sqrt{\left|\frac{\partial X_{genome}}{\partial h}\right|^2 (\Delta h)^2 + \left|\frac{\partial X_{genome}}{\partial X_{bp}}\right|^2 (\Delta X_{bp})^2 + \left|\frac{\partial X_{genome}}{\partial fill}\right|^2 (\Delta fill)^2} \tag{S7}$$

and

$$\Delta X_{capsid} = \sqrt{\left|\frac{\partial X_{capsid}}{\partial h}\right|^2 (\Delta h)^2 + \left|\frac{\partial X_{capsid}}{\partial d_X}\right|^2 (\Delta d_X)^2}. \quad \text{(S8)}$$

Equivalent functions for the stoichiometry of the viral head as a function of number of base pairs, n_{bp}, and the uncertainty ΔX_{head} as a function of number of base pairs were obtained using equation 1 in the main text to obtain r_c as a function of n_{bp}. Figure S1 shows the model predictions and the data corresponding to the C, N, and P content of different viral head as a function of genome size.

S.1.2 Average molecular formula of a base pair (pair of nucleotides), X_{bp}

The average molecular content of a base pair was obtained assuming that the distribution of nucleobases in the DNA is uniform. The C, N, and P composition of the nucleobases are:

$$A : C_5N_5 \qquad C : C_4N_3$$
$$T : C_5N_2 \qquad G : C_5N_5$$

The nucleotide is composed of the nucleobase, a five-carbon sugar, and a phosphate group. So, each base pair has 10 atoms of carbon, 2 of phosphorus and the atoms in the two nucleobases. Assuming a sequence comprised of nucleobases appearing with equal probability, then the average molecular formula for the base pair is:

$$C_{19.5}N_{7.5}P_2 \quad \text{(S9)}$$

However, the frequencies of nucleobases in a virus is not necessarily uniform. We calculated the frequencies of nucleobases in the DNA sequences of 776 phages (Figure S5a). From this we obtained the average C and N per nucleotide in each phage. On average the nucleobases adenine and thymine are present more frequently (mean frequency 0.52) than the nucleobases cytosine and guanine. Nonetheless, we obtain an average of 19.5 ± 0.1 atoms of C per nucleotide and 7.5 ± 0.1 atoms of N per nucleotide (see Figures S5(b),S5(c)). There is no variation in the number of P atoms per nucleotide because each nucleotide has exactly 2 P atoms. The uncertainties presented are the standard errors of the distributions ($\Delta X_{bp} = 0.1$ for both C and N).

S.1.3 Filling fraction of the genetic material inside the capsid, $fill$

The $fill$ was calculated by fitting data listed in Table S2 to the equation $n_{bp} = c \cdot (r - h)^3$ (Figure S2). We obtained $c = 2.1 \pm 0.2$ from the fit with Δc the estimated standard error. Recall that the fill is given by :

$$fill = \frac{3 v_{bp} c}{4\pi} = 0.53 \pm 0.04, \quad \text{(S10)}$$

where the uncertainty in the fill is $\Delta fill = 0.04$. Note that even when excluding the size and genome length information for phages used in the (independent) calculation of elemental composition (λ, T7, N4, Syn5, HK97), we obtain the value $fill = 0.55 \pm 0.05$, which agrees with the value being used within the margin of error. As a posterior evaluation of this key parameter value, we compiled a list of all dsDNA phages for which complete genomes (and their lengths) are available from NCBI. Then, we cross-collated that list against the catalog of phage available at the Felix d'Herelle Phage Reference Center: http://www.phage.ulaval.ca/en/phages_catalog/. Many of the phages in the reference center catalog have associated electron micrographs which have been the subject

of extensive analysis [1, 2, 3]. We cross-collated phages in the catalog that had available EMs (with scale-bars) and identified 57 candidate dsDNA phage, of which 3 were identified to be lipid-containing (phages PRD1, PM2, and AP50) and removed. Hence, our final dataset included 54 phage (see Supplementary information S2 (table) for names and accession information). We measured the capsid diameter of each phage based on the available EM image (measurements in Supplementary information S2 (table)). Linear regression of genome length against capsid size yields an estimate of the "fill" of viral capsids to be 0.52 ± 0.03 (see Figure S3). Again, this result is consistent with variation based on prior measurements and provides further posterior validation to our estimate of the fill parameter and its uncertainty.

Alternatively, one can obtain an estimate for the fill using the average concentration of DNA inside viral capsids. This concentration, ρ_{DNA} =0.5g/cm^3, is proportional to the fill and is a well conserved quantity among dsDNA phages [4]. The fill can be obtained as:

$$fill = \frac{\rho_{DNA}}{mass_{bp}/v_{bp}} = \frac{0.5 \cdot 10^{-21}\text{g/nm}^3 \cdot 1.068\text{nm}^3}{1.023 \cdot 10^{-21}\text{g}} = 0.52, \tag{S11}$$

where $mass_{bp} = 1.023 \cdot 10^{-21}$g is the average mass of one nucleotide pair [5]. We see that the estimate obtained in this way is within the uncertainty of the $fill$ used in our model.

S.1.4 Molecules of Carbon and Nitrogen per volume of protein, d_C and d_N

The number of atoms of C per unit volume of protein d_C, and the number of atoms of N per unit volume of protein d_N, were calculated using the experimental value of the average density of proteins ($0.73^{-1} \cdot 10^{-21}$g/nm^3 [6]) and the number of atoms of each element per unit of protein mass, calculated using a linear fit on 2815 viral proteins with unique RefSeq (Figure S6).

$$d_C = 0.73^{-1} \cdot 10^{-21}\text{g/nm}^3 \cdot 2.29 \cdot 10^{22}\text{atoms/g} = 31 \pm 1 \quad \text{atoms/nm}^3$$
$$d_N = 0.73^{-1} \cdot 10^{-21}\text{g/nm}^3 \cdot 6.33 \cdot 10^{21}\text{atoms/g} = 8.7 \pm 0.4 \quad \text{atoms/nm}^3$$

where Δd_C and Δd_N were obtained by propagating the uncertainties of both, the average protein density, and the the number of atoms per unit mass, which is the estimated standard error of the slope of the fit.

S.1.5 Thickness of the capsid

Bacteriophage capsids are usually one protein thick, although the exact value of the thickness varies (e.g., T7 - 2.3 nm [7], N4 - 2.6 nm [8], HK97 - 1.8 nm [9]). In our model we use an average capsid thickness of 2.5 nm and choose a conservative range for the variation in the thickness, $\Delta h = 1$ nm.

S.1.6 Model extension for viral tails

We approximate the tail structure as a hollow cylinder with outside radius r_t, fixed thickness h_t, and length ℓ_t. The number of proteins in the tail is predicted to be:

$$n_{pr} = \frac{v_{tail}}{v_{pr}} = \frac{\ell_t \pi}{v_{pr}}(r_t^2 - (r_t - h_t)^2) = \frac{\ell_t \pi}{v_{pr}}(2h_t r_t - h_t^2) \tag{S12}$$

As before, we then predict the number of atoms of element X in the tail:

$$\begin{aligned} X_{tail} &= X_{pr} \cdot n_{pr} = d_X v_{pr} n_{pr} \\ &= d_X \ell_t \pi (r_t^2 - (r_t - h_t)^2) \end{aligned} \tag{S13}$$

where the notation is the same as used in the main text. The total elemental composition of viral particles, including head and tail, for three select viruses is found in Table S3.

S.2 Virus C, N, P content distribution

The C, N, and P content of 776 different phage heads was estimated (Figure S4). The genome length of the phages were used as input to predict a capsid size and, in turn, the estimated elemental content in the head.

S.3 Elemental stoichiometry of selected phages using experimentally estimated data from the literature.

The elemental composition of phage particles was determined by considering their genetic makeup and protein content based on data available within the literature (Table 1, Table S1 and Table S3). Sequence information for every phage was retrieved from NCBI. Values of the stoichiometry of structural proteins were retrieved from empirical studies (Table S4), and the amino acid composition of each protein was retrieved from NCBI.

Concept	Variable	Value
Avg. Molecular formula of a base pair (pair of nucleotides)	X_{bp}	$C_{19.5}N_{7.5}P_2$
Avg. volume of a Base pair	v_{bp}	$(1\text{nm})^2(0.34\text{ nm})\pi$
Volume-filling fraction	$fill$	0.53 ± 0.04
Avg. thickness of the capsid	h	2.5 ± 0.3 nm
Molecules of carbon per volume of protein	d_C	31 ± 1 atoms/nm^3
Molecules of nitrogen per volume of protein	d_N	8.7 ± 0.4 atoms/nm^3

Table S1: Parameters used for the viral elemental composition calculation.

phage	genome size (kbp)	ext. diameter (nm)
λ	49 [10]	63 [11]
Mu	39 [12]	54 [12]
P1	100 [11]	85 [13]
P2	34 [14]	60 [14]
P4	12 [15]	45 [16]
$\phi 80$	45 [11]	61 [11]
$\phi X174$	5 [17]	26 [18]
T3	38 [11]	60 [11]
T5	121 [19]	90 [19]
T7	40 [20]	61 [20]
HTVC011P	40 [21]	55 [21]
HTVC019P	42 [21]	55 [21]
HTVC010P	35 [21]	50 [21]
HTVC008M	147 [21]	84 [21]
N4	71 [22]	70 [22]
Syn5	46 [23]	60 [23]
SPP1	44 [24]	60 [24]
SPO1	146 [25]	87 [26]
HK97	40 [27]	66 [9]
P22	42 [28]	61 [28]

Table S2: Genome length and capsid size for selected phages.

Phage Name	Carbon (atoms)	Nitrogen (atoms)	Phosphorus (atoms)	C:N:P (molar)	Carbon biomass (fg particle^{-1})
T4	8,508,594	2,626,186	337,800	25:8:1	0.170
N4	2,871,950	920,156	140,306	21:7:1	0.057
Syn5	1,715,042	576,459	92,428	19:6:1	0.034

Table S3: Composition of viral particles, including head and tail. Elemental composition of selected viruses based on sequence information and experimentally validated composition of structural proteins.

Phage	Function	Protein	Head	Tail	Counts	Protein id
Syn5	major capsid protein	gp39	1	0	400	YP_001285448.1
	internal virion protein	gp43	1	0	8	YP_001285452.1
	internal virion protein	gp44	1	0	8	YP_001285453.1
	internal virion protein	gp45	1	0	7	YP_001285454.1
	tail tubular protein A	gp40	0	1	6	YP_001285449.1
	tail tube B	gp41	0	1	6	YP_001285450.1
	head to tail conector	gp37	1	0	12	YP_001285446.1
	tail fiber	gp46	0	1	3	YP_001285455.1
	structural protein	gp53	1	0	6	YP_001285462.1
	structural protein	gp54	1	0	4	YP_001285463.1
	structural protein	gp58	1	0	12	YP_001285467.1
T4	Soc small outer capsid	Soc	1	0	840	NP_049644.1
	internal head	IpI	1	0	360	NP_049749.1
	internal head	IpII	1	0	360	NP_049734.1
	internal head	IpIII	1	0	370	NP_049735.1
	baseplate wedge unit	gp53	0	1	6	NP_049756.1
	baseplate hub subunit	gp5	0	1	3	NP_049757.1
	baseplate wedge subunit	gp6	0	1	12	NP_049764.1
	baseplate wedge subunit	gp7	0	1	6	NP_049765.1
	baseplate wedge subunit	gp8	0	1	12	NP_049766.1
	baseplate wedge tail fiber connector	gp9	0	1	18	NP_049767.1
	baseplate subunit and tail pin	gp10	0	1	18	NP_049768.1
	baseplagesubunit and tail pin	gp11	0	1	18	NP_049769.1
	short tail fiber	gp12	0	1	18	NP_049770.1
	fibritin	gpwac	0	1	18	NP_049771.1
	tail sheath stabilizer and completition	gp15	0	1	6	NP_049774.1
	tail sheath	gp18	0	1	144	NP_049780.1
	tail tube	gp19	0	1	144	NP_049781.1
	portal vertex	gp20	1	0	12	NP_049782.1
	major capsid	gp23	1	0	960	NP_049787.1
	capsid vertex	gp24	1	0	55	NP_049789.1
	head outer capsid	hoc	1	0	160	NP_049793.1
	baseplate wedge subunit	gp25	0	1	6	NP_049800.1
	baseplate hub subunit	gp27	0	1	3	NP_049803.1
	baseplate hub subunit	gp29	0	1	6	NP_049805.1
	baseplate subunit	gp48	0	1	6	NP_049806.1
	baseplate subunit	gp54	0	1	6	NP_049807.1
	ADP-ribosyltransferase	Alt	1	0	40	NP_049811.1
	DNA ligase	gp30	0	1	6	NP_049813.1
	synthase	dTMP	0	1	3	NP_049848.1
	dihydrofolate reductase	Frd	0	1	6	NP_049850.1
	long tail fiber, proximal subunit	gp34	0	1	18	NP_049860.1
	hinge connector of long tail fiber	gp35	0	1	6	NP_049861.1

	hinge connector of long tail fiber	gp36	0	1	18	NP_049862.1
	long tail fiber, distal subunit	gp37	0	1	18	NP_049863.1
N4	decorating protein	gp17	1	0	147	YP_950495.1
	virion RNA polymerase	gp50	1	0	4	YP_950528.1
	major coat protein	gp56	1	0	548	YP_950534.1
	portal protein	gp59	1	0	17	YP_950537.1
	non-contractil tail sheath	gp65	0	1	6	YP_950543.1
	appendage	gp66	0	1	25	YP_950544.1
lambda	major capsid decoration	gpD	1	0	405	P03712.1
	major capsid	gpE	1	0	405	YP_001700607.1
	portal	gpB	1	0	12	P03710.1
	head-tail conector	gpW	1	0	6	P68660.1
	tail attachement site	gpFII	1	0	6	P03714.1
	Protease	gpC	1	0	10	P03711.1
T7	major capsid	gp10A	1	0	395	NP_041998.1
	minor capsid	gp10B	1	0	20	NP_041997.1
	head-tail connector	gp8	1	0	12	NP_041995.1
	internal core	gp15	1	0	8	NP_042003.1
	internal core	gp16	1	0	4	NP_042004.1
	internal core	gp14	1	0	8-12	NP_042002.1
		gp6.7	1	0	18	NP_041991.1
phi29	major capsid	gp8	1	0	235	P07531.1
	head fiber	gp8.5	1	0	110	P20344.1
	upper collar	gp10	1	0	12	P04332.1
HK97	major capsid	gp5	1	0	420	YP_004123811.1

Table S4: Protein structure of virus particles.

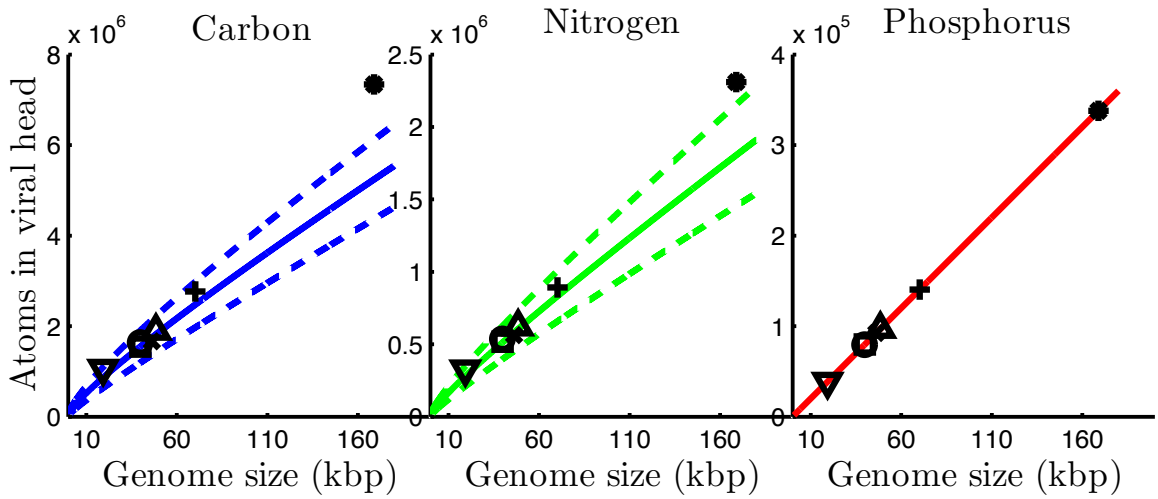

Figure S1: C, N, and P content of the viral head as a function of the number of base pairs in its DNA. The data correspond to experimentally obtained contents of C, N, and P for different viral heads: $*$ T4, $+$ N4, $\times$ Syn5, $\triangle$ λ, o HK97, $\square$ T7, $\triangledown$ ϕ29.

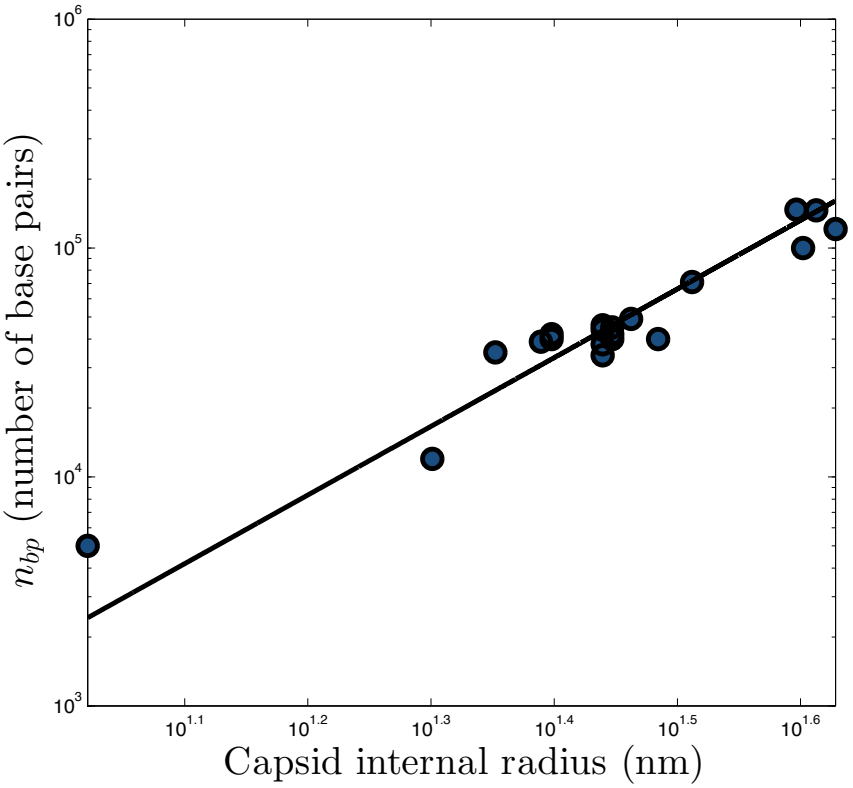

Figure S2: Number of base pairs as a function of radius of the phage capsid (data from Depaepe 2006 [11] and additional sources, see Table S2). The solid line corresponds to fitting the equation $n_{bp} = c \cdot r^3$ to the data, with c as a free parameter.

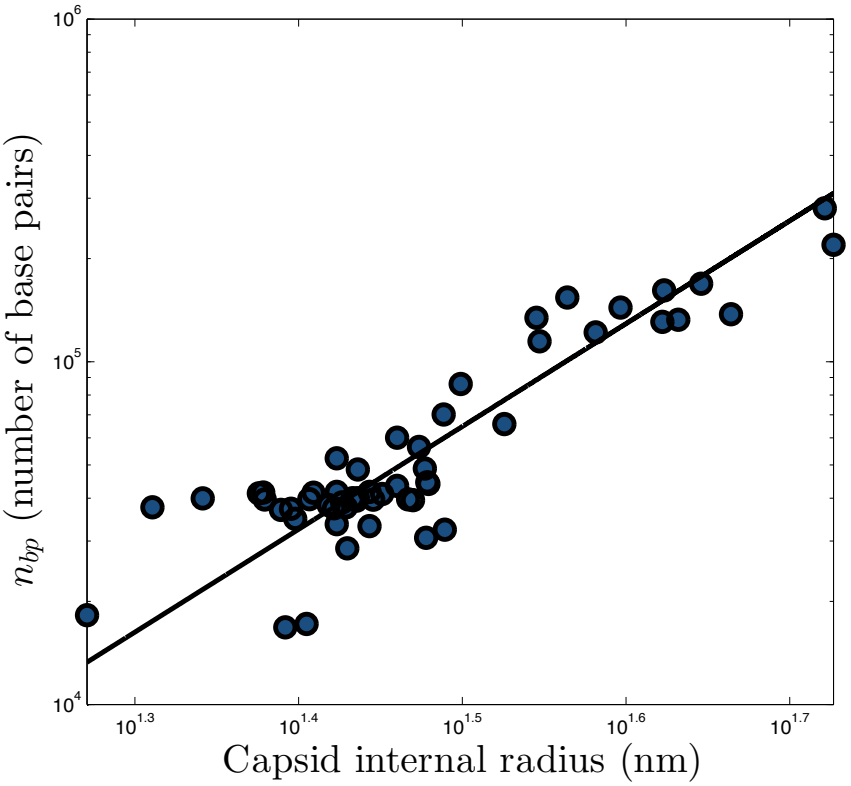

Figure S3: Number of base pairs as a function of radius of the phage capsid (posterior measurements of virus sizes from phage EM-s in the catalog of the Felix d'Herelle Phage Resource Center, see Supplementary information S2 (table) for accession information, capsid measurements and genome length information). The solid line corresponds to fitting the equation $n_{bp} = c \cdot r^3$ to the data, with c as a free parameter.

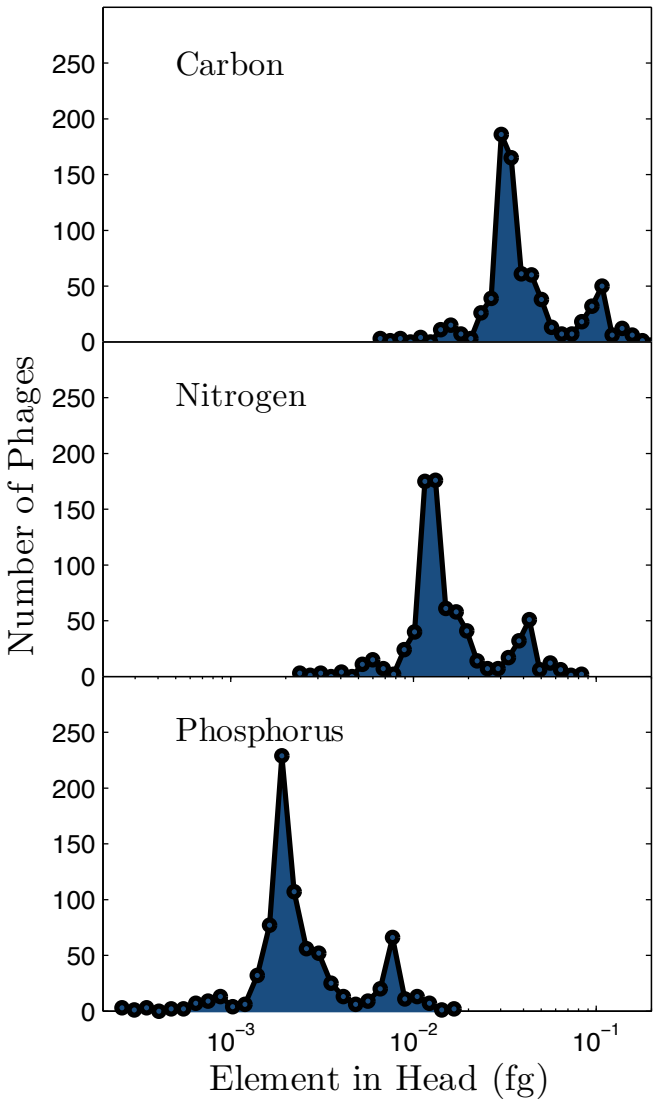

Figure S4: Estimated mass of C, N and P in virus particles. The y-axis denotes the number of phages (of a total of 776 phages analyzed, see Supplementary information S3 (table)) whose predicted elemental abundance of C, N, and P falls within a bin (denoted on the x-axis).

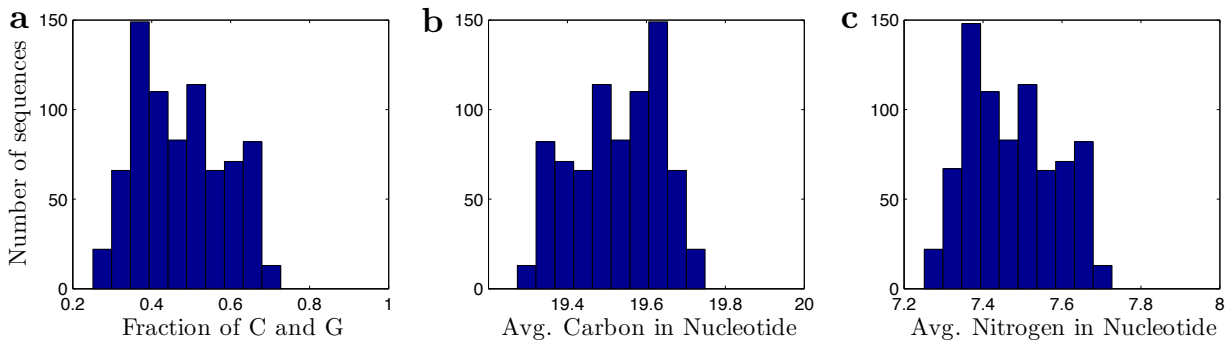

Figure S5: (a) Distribution of the fraction of nucleobases in the primary sequence of 776 phages that is either C or G (genome accession numbers in Supplementary information S3 (table)). (b) Distribution of the average carbon (C) per nucleotide in the primary DNA sequence. (c) Distribution of the average nitrogen (N) per nucleotide in the primary DNA sequence.

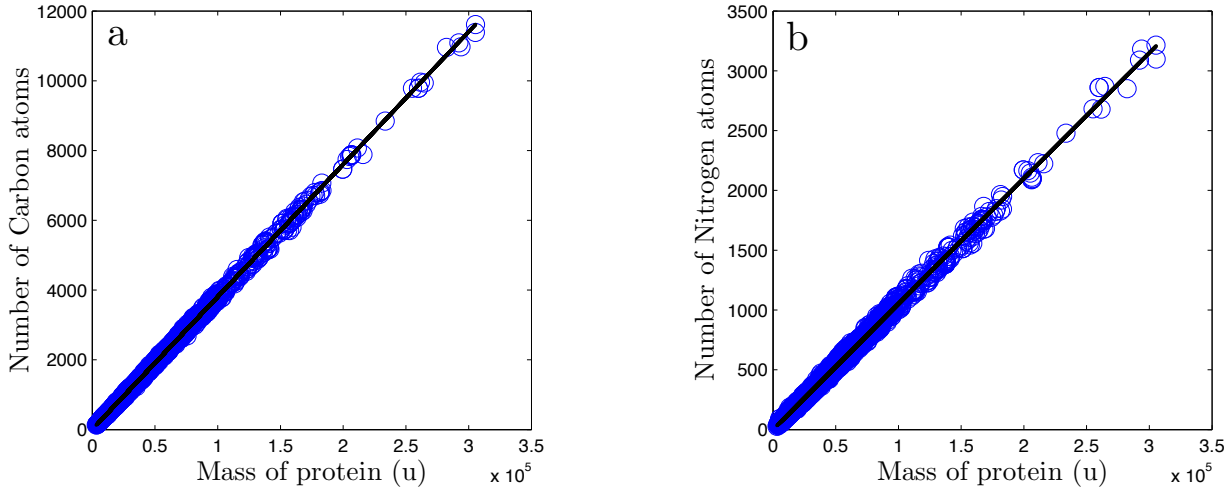

Figure S6: (a) Number of atoms of C vs. mass for 2815 viral proteins (primary structure information available in Supplementary information S4 (table)). (b) Number of atoms of N vs. mass for the same set of 2815 viral proteins.

References

[1] Ackermann, H.-W. 5500 phages examined in the electron microscope. *Archives of Virology* **152**, 227–243 (2007).

[2] Ackermann, H.-W. & Kropinski, A. M. Curated list of prokaryote viruses with fully sequenced genomes. *Research in Microbiology* **158**, 555–566 (2007).

[3] Ackermann, H.-W. & Prangishvili, D. Prokaryote viruses studied by electron microscopy. *Archives of Virology* **157**, 1843–1849 (2012).

[4] Black, L. W. & Thomas, J. A. Condensed genome structure. In *Viral Molecular Machines*, 469–487 (Springer, 2012).

[5] Dolezel, J., Bartos, J., Voglmayr, H. & Greilhuber, J. Nuclear DNA content and genome size of trout and human. *Cytometry. Part A: the journal of the International Society for Analytical Cytology* **51**, 127 (2003).

[6] Serdyuk, I. N., Zaccai, N. R. & Zaccai, J. *Methods in Molecular Biophysics: Structure, Dynamics, Function* (Cambridge University Press, 2007).

[7] Ionel, A. *et al.* Molecular rearrangements involved in the capsid shell maturation of bacteriophage T7. *Journal of Biological Chemistry* **286**, 234–242 (2011).

[8] Choi, K. H. *et al.* Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. *Journal of Molecular Biology* **378**, 726–736 (2008).

[9] Wikoff, W. R. *et al.* Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**, 2129–2133 (2000).

[10] Hendrix, R. W. & Casjens, S. Bacteriophage lambda and its genetic neighborhood. In *The Bacteriophages*, 409–447 (Oxford University Press, 2006).

[11] DePaepe, M. & Taddei, F. Viruses' life history: towards a mechanistic basis of a trade-off between survival and reproduction among phages. *PLoS Biology* **4**, e193 (2006).

[12] Paolozzi, L. & Ghelardini, P. The bacteriophage Mu. In *The Bacteriophages*, 469–496 (Oxford University Press, 2006).

[13] Yarmolinsky, M. B. & Sternberg, N. Bacteriophage P1. In *The Bacteriophages*, 291–438 (Plenum Press, 1988).

[14] Nilsson, A. S. & Haggård-Ljungquist, E. The P2-like bacteriophages. In *The Bacteriophages*, 365–390 (Oxford University Press, 2006).

[15] Dehò, G. & Ghisotti, D. The satellite phage P4. In *The Bacteriophages*, 391–408 (2006).

[16] Bertani, L. E. & Six, E. W. The P2-like phages and their parasite, P4. In *The Bacteriophages*, vol. 2, 73–143 (Plenum Press, 1988).

[17] Hayashi, M., Aoyama, A., Richardson, D. L., Jr. & Hayashi, M. N. Biology of the bacteriophage phiX174. In *The Bacteriophages*, 1–71 (Plenum Press, 1988).

[18] McKenna, R. *et al.* Atomic structure of single-stranded DNA bacteriophage phi?174 and its functional implications. *Nature* **355**, 137 (1992).

[19] McCorquodale, D. J. & Warner, H. R. Bacteriophage T5 and related phages. In *The Bacteriophages*, 439–475 (Plenum Press, 1988).

[20] Molineux, I. J. The T-7 group. In *The Bacteriophages*, 277–301 (Oxford University Press, 2006).

[21] Zhao, Y. *et al.* Abundant SAR11 viruses in the ocean. *Nature* **494**, 357–360 (2013).

[22] Kazmierczak, K. & Rothman-Denes, L. Bacteriophage N4. In *The Bacteriophages*, 302–314 (2006).

[23] Pope, W. H. *et al.* Genome sequence, structural proteins, and capsid organization of the cyanophage Syn5: A "horned" bacteriophage of marine *synechococcus*. *Journal of Molecular Biology* **368**, 966–981 (2007).

[24] Alonso, J. C., Tavarez, P., Lurz, R. & Trautner, T. A. Bacteriophage SPP1. In *The Bacteriophages*, 331–349 (2006).

[25] Stewart, C. R. *et al.* The genome of *bacillus subtilis* bacteriophage SPO1. *Journal of Molecular Biology* **388**, 48–70 (2009).

[26] Stewart, C. Bacteriophage SPO1. In *The Bacteriophages*, vol. 1, 477–515 (Plenum Press, 1988).

[27] Juhala, R. J. *et al.* Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *Journal of molecular biology* **299**, 27–51 (2000).

[28] Prevelige, P. E., Jr. Bacteriophage P22. In *The Bacteriophages*, 457–468 (2006).