Monitoring Ecosystem Degradation: Bacteriophage Structural Properties as a Quantitative Environmental Indicator

PI: Antoni Luque
co-PIs: Forest Rohwer, Robert Edwards, and Peter Salamon
Viral Information Institute at San Diego State University

Abstract

Bacteriophages are viruses that infect bacteria and are present in all ecosystems, from hot springs to the human gut. DNA sequencing techniques, pioneered in the Viral Information Institute, are surveying phages in all sort of environments, showing that phages are the most abundant and sensitive biological entity on the planet. Early changes in phage properties, for example, prelude the degradation of entire ecosystems, indicating that phages are early indicators of the ecosystem dynamical state. Phage environmental data is increasing at a fast pace, but new models are necessary to obtain an insightful quantitative and mechanistic interpretation of this data. In this proposal, we hypothesize that the structural properties of phage particles, which adapt quickly to environmental changes, represent an ideal metric to characterize the state of an ecosystem and monitor its degradation. To test this, we will develop a set of complementary mathematical and physical models that estimate key structural properties of phages based on genetic information. As a proof of concept, we will focus on two important and well defined variables in marine ecosystems: Salinity and microbialization. The accomplishment of the following aims will measure the success of this proposal: (1) predict the structure of icosahedral and elongated phages based on the size of the phage genome; (2) predict the phage tail length based on the size of the tape measure protein; (3) identify the major capsid protein from the phage genome and characterize its physicochemical properties; (4 and 5) use the structural data from models (1) to (3) to obtain a mathematical phage-metric that measures the impact of (4) salinity and (5) microbialization in marine ecosystems. This research proposal builds upon the theoretical and experimental expertise of the Viral Information Institute. The phage-metric framework proposed here will eventually include other environmental variables like temperature and acidity, and will be also applied to analyze other ecosystems like the human microbiome.

Introduction

Bacteriophages, or phages, are viruses that infect bacteria. At a first glimpse, a phage might look like a marginal vestige of nature: a shell made of proteins—the capsid—protecting an infectious genetic material programmed to replicate just in bacteria. Phages, however, are ten times more abundant than their microbial hosts, and, in matter of two days, phage predation causes a turnover of the whole bacterial population in our planet. Phages keep bacterial growth under control and redistribute nutrients across ecosystems. Rather than parasites, they are a cornerstone in the stability of the biosphere [1]. A key aspect in the success of phages is their versatile genome. The phage genome is organized in functional blocks that are often interchanged among phages. This, combined with their sheer number and replication efficiency, allows phages to quickly adapt to changes in the environment. In fact, the impact of these changes in phage properties preludes more drastic consequences in the entire ecosystem, as has been observed in the degradation of coral reefs. Thus, phage properties represent a potential method to define the state of an ecosystem and monitor its degradation process. Nevertheless, a rigorous quantitative phage-metric of ecosystems for relevant environmental variables is missing. In this proposal we will use phage environmental data to implement mathematical and physical models that, based on phage genetic information [2], estimate the structural properties of phages to define a phage-metric for the salinity and microbialization of marine ecosystems (Fig. 1).

d ***Salinity.*** Salinity is the amount of salt dissolved in water and varies significantly across marine environments. The level of salinity in the sea depends on multiple factors, like rainfall, water evaporation, or inflow of rivers. Due to climate change, the weather pattern on the planet is becoming more extreme, and this is also affecting salinity; seawater salinity is increasing in dryer regions and decreasing in wetter regions [3]. Extreme levels of salinity, either too high or too low, have a devastating effect in marine ecosystems, and it is important to find early biological indicators to measure the impact of salinity in these environments. Since the average size of phage capsids is positively correlated with salinity [4], one of our main goals is to use

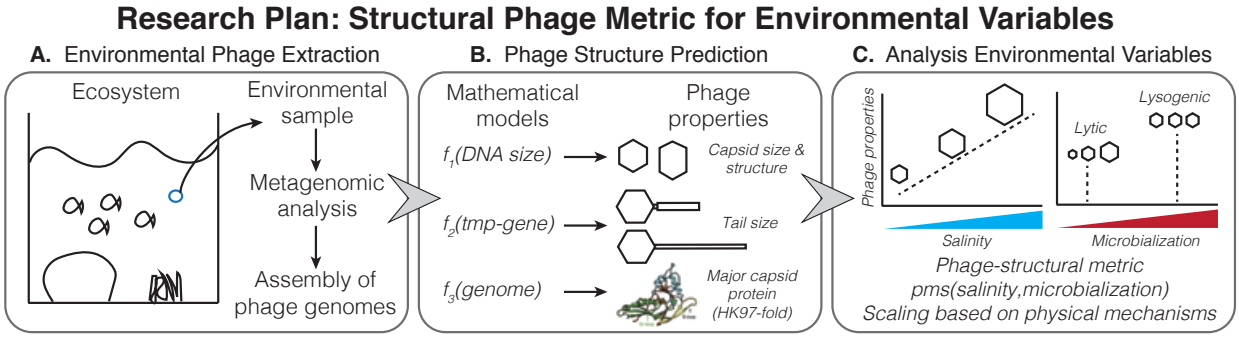

Figure 1: Research plan: Structural phage-metric for environmental variables. A) The Viral Information Institute at SDSU continuously analyzes phage samples from diverse ecosystems, where phage genomes are assembled using metagenomic analysis. B) Mathematical models aimed in this proposal: f_1, capsid size and structure; f_2, tail size; and f_3, major capsid protein. In parenthesis are the inputs for the models (*tmp* means tape measure protein). C) Phage structural properties will be analyzed based on the salinity and microbialization of the sampled environments. The environmental phage-metrics for salinity and microbialization will be based on scaling derived from statistical correlations and physical models.

our mathematical and physical models to study the distribution of phage capsids across environments with different salinity levels. We hypothesize that the distribution of capsid sizes and the electrostatic content of the major capsid protein defines a sensitive ecological metric that can monitor the impact of salinity in ecosystems (Fig. 1C).

Microbialization. In marine environments, bacteria play an important role as primary producers. But when environmental stresses disrupt the ecosystem balance, heterotrophic bacteria find an opportunity to invade new niches, increasing the level of resources used at the microbial level. This effect, called microbialization, is associated with the degradation of ecosystems. Interestingly, the level of microbialization has been recently correlated with an increment of lysogenic phages. Lysogenic phages, contrary to lytic phages, are able to integrate in the bacterial genome. Lysogens provide metabolic pathways and toxins that bacteria can use to exploit eukaryotic organisms and expand in the ecosystem. The average size of lysogenic genomes varies from the size of lytic phages, indicating that lysogenic capsids should differ from lytic ones. An important goal in this proposal is to apply our new mathematical and physical models to quantify the distribution of phage structural properties among lytic and lysogenic phages. The distance between both distributions will define the phage-metric for microbialization in ecosystems (Fig. 1C).

High throughput prediction of phage structural properties

Empirical electron microscopy studies accumulated over decades indicate that the majority of phages—up to 96%—are tailed phages and define the order of *caudovirales* [5]. Tailed phages have a polyhedral capsid with a proteic tail attached in one of the capsid vertices (Fig. 2). Their phage genome is a double stranded DNA molecule stored in the capsid at high densities [6]. The common structural properties of tailed phages set the foundations for our hypothesis and modeling approaches.

Icosahedral capsids. Approximately 80% of tailed phages have quasi-spherical capsids with icosahedral symmetry, while the remaining 20% have elongated capsids [6, 9] (Fig. 2A). The geometrical theory of viral capsids classifies icosahedral shells based on the T-number index [8]. The T-number ($T = h^2 + hk + k^2$) follows the series $T = 1, 3, 4, 7, 9, 12$, etc., and determines the number of capsid proteins ($n = 60T$) and the relative volume of viral capsids ($V/\sigma^3 = AT^{3/2}$) [10]. The packing density of the genome is very similar among tailed phages [11]. Thus, the size of the phage genome is directly related to the internal volume of the capsid, and here we propose to use the geometrical theory of viral capsids to predict the T-number for any phage based on its genome size (Figs. 2 and 3).

A preliminary application of our model estimates that 80% of phages adopt icosahedral capsids, in agree-

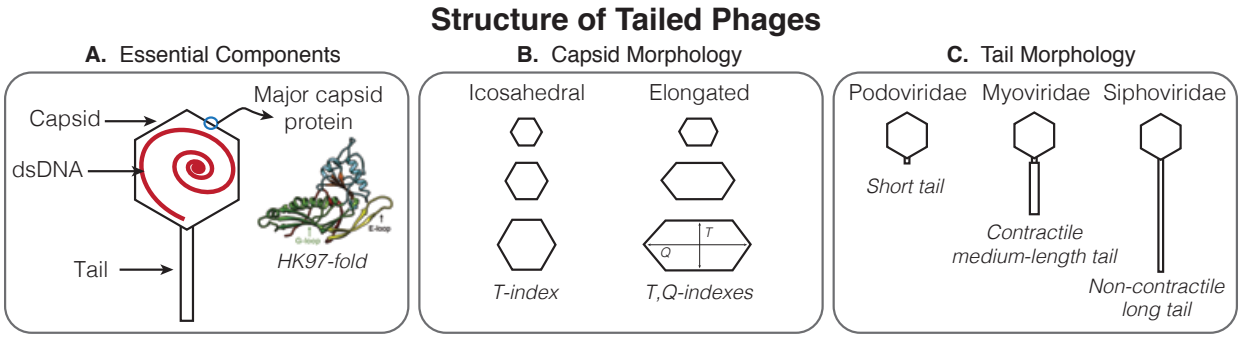

Figure 2: Structure of tailed phages. (A) Tailed phages represent 96% of total phages and have three essential structural components: capsid, genome, and tail. The capsid is made of multiple copies of the major capsid protein, which follows the HK97-fold structural lineage [7]. The phage genome is packed at high densities as double stranded DNA (dsDNA). The phage tail is a protein nanochannel responsible for triggering the infection process. (B) Tailed phage capsids can adopt icosahedral and elongated capsids. The sizes of icosahedral shells are classified with the T-index [8]; the size and length of elongated shells are determined by the T- and Q-indexes [9]. (C) Tail phages are classified in three families depending on the structure of the tail [6]: Podoviridae (short tails), myoviridae (contractile, medium-length tail), and siphoviridae (non-contractile, long tail).

ment with the literature [5]. The distribution of capsids show peaks associated to T-structures, confirming that T=7 is the most abundant architecture across *free* phages, including lytic and lysogenic (Fig. 3A). Interestingly, prophages, which are lysogenic phages integrated in bacterial genomes, show an even sharper distribution around T=7 (Fig. 3B). This result supports our hypothesis about the structural transition from lytic to lysogenic phages, a key aspect for our phage-metric to quantify the microbialization of ecosystems.

Elongated capsids. Elongated phage capsids are spherocylindrical shells with a tubular body closed by semi-icosahedral caps [12] (Fig. 2B). The geometrical theory of viral capsid has been recently extended to describe in detail the structural properties of elongated capsids [9, 13]. Our goal again is to use the genome size to identify the architecture of any elongated phage. In this case there are two structural indexes to be determined: The T-number, which is associated to the radius of the capsid, and the Q-number, which is associated to the length of the capsid [9]. In addition to provide a quantitative approach to define a phage-metric, our method will clarify if phages use elongated structures to transition between icosahedral shells of different volumes and the environmental effects that trigger this transition (Fig. 3).

***Major capsid protein.*.** The major capsid protein is the main element of the capsid (Fig. 2A). The sequence similarity among major capsid proteins is low, but their structure is conserved across tailed phages. This crucial protein adopts the *HK97-fold* and is probably the most abundant protein fold on the planet [7]. We hypothesize that the major capsid protein is the first phage element that reflects the impact of environmental changes. Salty environments, for example, reduce the electrostatic interactions that maintain the stability of the capsid; this might favor the selection of amino acids with different isoelectric points or alternative stabilization mechanisms, like hydrophobic interactions. Our goal is to study the physicochemical properties of the major capsid protein across phages to define a metric of the ecosystem state at the molecular level. The main challenge here is to identify the amino acid sequence of the major capsid protein a given a phage genome. To address this problem, we will apply an artificial neural network algorithm developed in the Viral Information Institute that identifies structural phage proteins from genomic sequences [14]; the refined algorithm will be re-trained to identify exclusively *HK97-fold* like proteins.

Tail size. The tail attached to the capsid of phages is essential to start the infection process. Phages are classified in three families based on the morphology of the tail (Fig. 2C): *Podoviridae* (short tails), *myoviridae* (medium contractile tails), and *siphoviridae* (long flexible tails). For *myoviridae* and *siphoviridae* phages, a molecular ruler, called the tape measure protein, determines the size of the tail [15]. This protein is usually the longest gene in the phage genome, which facilitates its identification from metagenomic analysis.

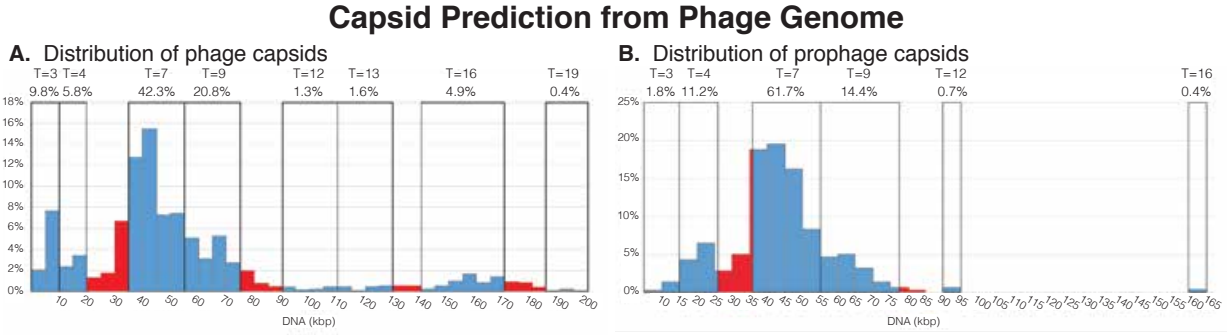

Figure 3: Capsid prediction from phage genome. We use our mathematical model based on the theory of viral capsids to identify icosahedral capsids (blue bars) and elongated capsids (red bars) as a function of the phage genome size, given in kilo base pairs (kbp). Each box groups icosahedral capsids with the same predicted T-number. We also indicate the relative presence of each T-number in the distribution. A) We plot the genome size and T-number distributions for an environmental sample of free phages (includes lysogenic and lytic phages). B) Same plot as in A) but for a sample of prophages (lysogenic phages integrated in bacteria) obtained from metagenomic analysis of bacteria.

Although the size of the tape measure protein gene is proportional to the tail length, a rigorous analysis of this relationship using a mathematical model is missing. We will implement such a model using electron microscopy and genetic data accumulated for phages [6]. Our goal is to apply the model to investigate if, similar to the capsid structure, the tail morphology varies as a function of the levels of microbialization and salinity of the environment.

Environmental phage metric

In order to develop an effective metric that captures the impact of environmental changes in ecosystems using phage structural properties, we will initially focus in two well studied variables in marine environments: Salinity and microbialization.

Salinity. Recent studies have shown that the average size of phages is positively correlated with salinity of seawater [4]. The Viral Information Institute has performed metagenomic analysis for phage sea samples with different levels of salinity. We will apply our mathematical models to these data to obtain a quantitative picture about the distribution of structural properties as a function of salinity (Fig. 1C). In order to explore potential phage-metrics using this data, we will implement a physical model to derive the mechanistic relationship, for instance, between salt concentration and phage capsid size. This model will combine continuum theory of DNA packing in phages [16, 17] and the theory of phage capsid mechanics [18, 10, 19, 20]. A preliminary analysis of these theories indicates that larger phage capsids packed with DNA are stabilized at higher salinity environments, providing a molecular mechanism to support the phage-salinity metric.

Microbialization. Increase of lysogeny is positively correlated with the microbialization and degradation of marine ecosystems. Lysogens need additional genes than lytic phages to regulate the integration of the genome in the host bacteria [21]. This suggests that lysogenic capsids should be larger than lytic capsids. On the other hand, a large genome will reduce the probability of lysogens to integrate the phage genome in the host [22]. We hypothesize that lysogens favors capsids with an intermediate size, which optimizes the length of the phage genome for integration. This hypothesis is in agreement with our preliminary results (Fig. 3), and it constitutes a molecular mechanism that justifies the use of phage structural properties as a metric of microbialization. We will investigate the phage-metric applying the mathematical models described above to a set of phage samples and environmental observations of coral reefs with different degrees of microbialization. Additionally, since salinity tends to increase the average size of phages, we will also investigate if salinity reduces lysogeny, or if, instead, it shifts the expected size of lysogens.

References

[1] F. Rohwer and R. Thurber. Viruses manipulate the marine environment. *Nature*, 459:207–212, 2009.

[2] R. Edwards and F. Rohwer. Viral metagenomics. *Nat. Rev. Microbiol*, 3:504–510, 2005.

[3] P.J. Durack, S.E. Wijffels, and R.J. Matear. Ocean salinities reveal strong global water cycle intensification during 1950 to 2000. *Science*, 336:455–458, 2012.

[4] J.R. Brum, R.O. Schenck, and M.B. Sullivan. Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. *The ISME Journal*, 7:1738–1751, 2013.

[5] H. Ackermann. 5500 Phages examined in the electron microscope. *Arch. Virol.*, 152:227–243, 2007.

[6] H.-W. Ackermann. Tailed bacteriophages: The order caudovirales. *Adv. Virus Res.*, 51:135–201, 1999.

[7] M.M. Suhanovsky and C.M. Teschke. Nature's favorite building block: Deciphering folding and capsid assembly of proteins with the HK97-fold. *Virology*, 479–480:487–97, 2015.

[8] D.L.D Caspar and A. Klug. Physical priniciples in the construction of regular viruses. *Cold Spring Harb. Sym.*, 27:1–24, 1962.

[9] A. Luque and D. Reguera. The structure of elongated viral capsids. *Biophys. J.*, 98:2993–3003, 2010.

[10] M. Aznar, A. Luque, and D. Reguera. Relevance of capsid structure in the buckling and maturation of spherical viruses. *Phys. Biol.*, 9:036003, 2012.

[11] L.F. Jover, T.C. Effler, A. Buchan, S.W. Wilhelm, and J.S. Weitz. The elemental composition of virus particles: Implications for marine biogeochemical cycles. *Nature Rev. Micriobiol.*, 12:519–528, 2014.

[12] M. Moody. Geometry of phage head construction. *J. Mol. Biol.*, 293:401–433, 1999.

[13] A. Luque, R. Roya, and D. Reguera. Optimal architectures of elongated viruses. *Proc. Nat. Acad. Sci. USA*, 107:5323–5328, 2010.

[14] V. Seguritan, N. Alves Jr., M. Arnoult, A. Raymond, D. Lorimer, A.B. Burgin Jr., P. Salamon, and A.M. Segall. Artificial neural networks trained to detect viral and phage structural proteins. *PLoS Comp. Biol.*, 8:e1002657, 2012.

[15] R. Hendrix. Tail length determination in double-stranded DNA bacteriophages. *Curr. Top. Microbiol. Immunol.*, 136:21–29, 1988.

[16] P.K. Purohit, M.M. Inamdar, P.D. Grayson, T.M. Squires, J. Kondev, and R. Phillips. Forces during bacteriophage DNA packaging and ejection. *Biophys. J*, 88:851–866, 2005.

[17] A. Luque and D. Reguera. heoretical Studies on Assembly, Physical Stability, and Dynamics of Viruses. *Subcell. Biochem. (in Mate, M.G. Ed. Structure and Physics of Viruses)*, 68:553–595, 2013.

[18] W. Roos, R. Bruinsma, and G. Wuite. Physical virology. *Nature Physics*, 10:733–743, 2010.

[19] C. Carrasco, A. Luque, M. Hernando-Pérez, R. Miranda, J. L. Carrascosa, P. A. Serena, M. de Ridder, A. Raman, J. Gómez-Herrero, I. A. T. Schaap, D. Reguera, and P. J. de Pablo. Built-in mechanical stress in viral shells. *Biophys. J.*, 100:1100–8, 2011.

[20] M. Hernando-Perez, E. Pascual, M. Aznar, A. Ionel, J.R. Caston, A. Luque, J.L. Carrascosa, P. de Pablo, and D. Reguera. The interplay between mechanics and stability of viral cages. *Nanoscale*, 6:2702–2709, 2014.

[21] L.-M. Bobay, M. Touchon, and E.P.C. Rocha. Pervasive domestication of defective prophages by bacteria. *Proc. Nat. Acad. Sci. USA*, 111:12127–12132, 20015.

[22] J.H. Paul. Prophages in marine bacteria: Dangerous molecular time bombs or the key to survival in the seas? *The ISME J.*, 2:579–589, 2009.