

Chapter 4

The Biophysics of Prokaryotic and Viral Diversity in Aqueous Environments

4.1 Abstract

Recent advances in techniques for enumerating viruses have led to a plethora of measurements of viral and bacterial abundances in nature that beckon for both qualitative and quantitative explanation. Here we propose a biophysical model that describes the interaction between bacteria and their lytic viruses in aqueous environments that combines both predator-prey relations and a diffusion-based transport model of viruses. In addition we postulate that the burst size is proportional to the volume ratio of the host cell and its infecting virion, for which there is empirical support for cell radii $< \sim 1\mu\text{m}$. We find that the concentration of a given bacterial species approximately scales with the radius of the cell r , as r^{-4} , suggesting that, within the context of a predator-prey model, the size of a bacterium is the most critical parameter determining its fixed point concentration. To extend our model to the community level, we postulated that there is no selection pressure on bacterial radii, i.e., *a priori*, all bacterial radii are equally probable. Given this hypothesis we predict that the size spectrum of marine bacteria follows a power law with slope -1, close to the observed average spectrum. We proceed to derive expressions for the total concentration of bacteria and viruses in the environment, reproducing for typical marine systems a virus-to-bacterium ratio (VBR) of ~ 10 . We show that the VBR is primarily determined by the average net growth rate of bacteria (growth minus predation), the average viral decay rate and, interestingly, the radius of the minimum viable bacterium. We next

derive a simple expression for the number of species in a given environment per unit volume, and predict that for offshore waters, where there are $\sim 10^5$ bacteria per ml, there should be $\sim 10^2$ to $\sim 10^3$ prokaryotic species in at most $\sim 10^2$ to $\sim 10^4$ liters of water, consistent with current empirical estimates of species richness. Thus, any given marine environment can only pack a finite degree of diversity. We use this observation to calculate an absolute lower and upper bound on the total number of active bacterial species in the ocean water column (excluding sediment), by considering the case of completely homogenous oceans and maximally heterogeneous oceans. We find that the number of species in the ocean water column should lie in the range of 10^4 – 10^{21} . We conclude by considering further experiments to test the validity of the proposed model.

4.2 Introduction

It was only in the late 1980s that the first quantitative estimates of viral abundance in the oceans using transmission electron microscopes revealed the existence of as many as millions of viral particles per milliliter of seawater [1]. Subsequently, more reliable counting methods based on epifluorescence imaging of stained nucleic acids came to the fore [1,2,3]. These methods were simple to execute even in field conditions and led to an explosion of measurements of viral and prokaryote concentrations in many environments [2,3]. In marine and fresh water ecosystems, these types of studies revealed that as a rule of thumb, viral concentrations typically exceed bacterial concentrations by one order-of-magnitude [1,2,4].

We were interested in understanding the basic processes in play that determine phage-host interactions in aqueous environments, and how these processes affect the bacterial and viral community composition. We therefore sought to identify the key variables that determine the virus and bacterium concentrations in the environment and to formulate a simple toy model that

is capable of making reasonable order of magnitude predictions and that can qualitatively explain the observed trends. Predator-prey models for host-virus interaction have earlier been examined by Campbell [5], Levin et al. [6], Lenski [7], Beretta et al. [8] and Thingstad et al. [9,10]. However, in these models the biophysical process of virus transport, which governs the contact rates between viruses and bacteria, was not considered. Stent [11] and Murray et al. [12] considered transport processes of viruses in aqueous environments but not in the context of a predator-prey model in an ecological setting. Our starting point is a simple toy model that incorporates virus transport within the context of a predator-prey model. We begin by examining the case of a particular isolated phage-host system with the goal of identifying the key variables that govern this system. We then extend our model to the community scale by hypothesizing the simplest evolutionary scenario that there is no selection pressure on bacterial radii, i.e., *a priori*, all bacterial radii are equally probable. We derive basic relations for the total concentration of bacteria and their viruses in the environment, and a basic relation for the total prokaryotic mass in the environment. Based on these results we explore questions such as, what are the critical parameters governing the system and how do variables scale with respect to these parameters? What determines the virus-to-bacterium ratio? What determines the number of species in a given environment? In what volume of water should we find this diversity? What are the bounds on the total diversity of species in Earth's oceans? Where possible we compare our predictions to observations and conclude with suggestions for experiments to further test our model.

We will further claim that the precise definition of a species lies outside the scope of our biophysical model. Consequently in Chapter 5 we will consider an evolutionary model for the generation of bacterial and viral species consistent with the definition of a species used in this

chapter. The evolutionary model is a first step in connecting the predictions of the biophysical model described in this chapter with empirical observations of diversity in the environment.

4.3. General assumptions

4.3.1 Decoupling phage-host systems

Typically a given environment will contain many species of bacteria and viruses. However, under certain assumptions, the microbial and viral communities can be treated as a set of decoupled phage-host systems [9]. Such an approximation will be valid if the following conditions are satisfied: (1) different bacterial species function independently of each other. Thus symbiotic relationships are prohibited. (2) Each viral species infects a single bacterial species and (3) each bacterial species is infected by a single viral species. The second assumption is a decent approximation given that phages characteristically exhibit species or subspecies [2,13,14] (although some exceptions, such as certain broad host range cyanophages, exist [2]). The third assumption may seem odd given that the most familiar example, *E. coli*, is known to be infected by many lytic viruses (e.g., the T-series). *E. coli*, however, is a commensal organism that lives in the intestines of animals and humans. Since the guts of animals/humans are physically separated, in principle at least, different species of phages that infect *E. coli* could have evolved in different guts. Aqueous environments on the other hand are diffusible and generally topologically connected and therefore of a different nature. Thus host range observations regarding *E. coli*, or any other gut bacterium, may not apply to marine ecosystems. That said, biogeography may play a role in marine ecosystems when considering very distant regions (e.g., the same cyanobacterium species in remote regions may be infected by different phages). Therefore both assumptions (2) and (3) may perhaps be relaxed by requiring them to be satisfied locally.

Assumption (3) is consistent with assumption (2) in the sense that two viruses cannot control the same bacterial species indefinitely, since such a system is unstable (Section 4.6). The opposite is also true, two bacterial species cannot be controlled by the same virus indefinitely, thus assumption (2) is consistent with assumption (3).

As a result, we begin our discussion by considering a simple phage-host system consisting of a single phage species infecting a single bacterial species, henceforth denoted by the index i . In Section 4.4.2 we will consider multiple independent phage-host systems.

4.3.2 Host mortality

Causes of mortality

It is generally accepted that bacterial host mortality is primarily due to either protist grazing or viral predation [4,15,16,17], both appearing to contribute about equally to microbial mortality [15,18,19]. In surface waters for example, viruses are thought to be responsible for ~10–50% of the total bacterial mortality, whereas in environments in which protists do not thrive, such as low-oxygen lake waters, viruses are thought to be responsible for 50–100% of bacterial mortality [4]. Thus it appears that the two likely fates of a bacterial cell in the ocean are either to be eaten by a protist or be lysed a virus.

Lysogenic versus lytic viruses

The process of viral predation can be mediated either through infection by lytic viruses or through induction of temperate viruses. In the case of temperate viruses, the infecting virus either enters a lytic phase and kills its host or is integrated into the genome of the host and may be induced at a later stage in response to an induction event (e.g., exposure to a mutagenic agent

[2]). In the oceans however it appears that lysogenic induction is rare [2,4,20], occurring either sporadically or at a low level [4]. Though this matter has still not been completely settled [2], it has been suggested that the majority of viruses observed in sea water are the result of successive lytic infections [4]. Other forms of infection such as chronic infection and pseudolysogeny [2] do not lead to host death and are therefore not considered to contribute to viral predation in this context. We shall therefore assume in our toy model that viral predation is exclusively the result of infection by lytic viruses.

Protists versus viruses

When comparing the effect of protist grazing to virus lysis on bacteria, there is a fundamental difference between these two predators that has to do with their host range. As a first-order approximation [10], protists can be regarded as omnivorous, i.e., they are not host selective [10,17]. On the other hand, viruses are known to be highly selective, displaying species or subspecies (strain) specificity [2,13,14]. Therefore, protists would control the total concentration of bacteria while viruses would control the individual concentration of bacterial species [10,17]. In a resource rich environment, there is evidence to suggest that because protists themselves are preyed upon, bacterial growth is determined by competition for resources and not by protist grazing [17]. Regardless of the mechanism that controls the total concentration of bacteria, in our model we simply assume that the total concentration of bacteria is fixed by some process and refer to this limiting factor as the “carrying capacity” of the environment.

Thus we will assume that every bacterial “species” is under viral control. Since grazing is thought to be a complex non-passive hydrodynamical process owing to the currents induced by the

motion of the flagella drawing its prey in [17], we account for this process by means of an effective grazing rate denoted by $\gamma_{non-viral}^{(i)}$ (see Table 4.1 for a list of notation). Another potential source for bacterial mortality is autolysis or programmed cell death in response to, for example, radiation damage [17]. Here all non-viral mediated mortality can be included effectively in $\gamma_{non-viral}^{(i)}$.

What is a bacterial species?

Note that we have not precisely defined what a bacterial “species” is or what a viral “species” is. What is the definition of a bacterial species? Similarly, what is the definition of the viral “species” that infects this bacterial “species”? We will argue that the precise definition of these concepts lies outside the scope of a biophysical model of phage-host interaction and requires a “higher” theory that probes the genetic complexity of these species (i.e., an evolutionary theory). In Chapter 5 we will propose an evolutionary mechanism that can be used to define a bacterial “species” and a viral “species”, and by which new bacterial and viral “species” co-emerge through a process of co-speciation. We will also show that when this evolutionary model is viewed in a genetic coarse-grained way, the evolutionary model converges to our current biophysical model. Our conclusion will be that while a bacterial species interacts with just one viral species, and vice versa, each of these species is comprised of strains (which are emerging new species) that are part of interaction networks with more than one viral strain. The key result that we derive is that although the species are independent, we need to multiply their concentration by (roughly) the number of strains per species to get the total concentration of a species (with a strain defined as an entity distinguishable in a consistent and clear way from all other strains). Thus, while a “strain” would have been our intuitive definition *a priori* for a

“species”, we find that strains are not independent elements (they are part of networks), and one needs to consider a more complex structure called a “species” to achieve independent phage-host systems.

4.3.3 Virus decay

Viral decay is thought to be mainly due to environmental damage from sunlight, temperature effects, and interaction with certain substances such as heat-labile colloidal dissolved organic matter [2,4,21]. These events lead to a certain rate of viral decay which we denote by $\gamma_{virus\ decay}^{(i)}$. Though protists can also potentially lead to viral removal by ingestion [21], grazing is generally not considered to be a significant factor leading to loss of viruses [2].

4.3.4 The physiological state of the host

The physiological state of bacteria in nature is generally unknown and is the subject of current research [2]. Generally speaking, bacteria appear to be growing slowly in marine environments. For example, in the cold waters of the Barents Sea in the Arctic ocean, growth rates were estimated to lie between 0.05 and 0.25 day⁻¹ [22] whereas in the warmer coastal seawater near Santa Monica growth rates were measured to be higher, ~1–3 day⁻¹ [18]. In our simple toy model we will assume that the environment is ideal in the sense that the bacteria are in a state of exponential growth. Though many environments are most likely not ideal, the notion of an “ideal environment” can be a useful construct that can at the very least serve as a null hypothesis for a given environment. In the context of our model, the growth rate of the i^{th} bacterial species is denoted by $\alpha^{(i)} = (\text{doubling rate}) \cdot \ln 2$. This growth rate is thus species specific and is determined by the availability of nutrients required by the given species.

4.3.5 Bacterial and viral abundance distribution

The virus-to-bacterium ratio (VBR) in marine systems is typically measured to be in the range of 5–25 [1,2,4,19] and in the deep waters of the Atlantic Ocean this ratio often exceeds 100 [1]. Particular phage-host systems have also been shown to exhibit VBRs as high as 8 (and locally even as high as 30 — see example discussed later on) [23]. We shall therefore assume in our model that the VBR for the i^{th} bacterial species satisfies $VBR^{(i)} \gg 1$. We will also assume that local spatial inhomogeneities in free virion concentration due to, for example, burst events [24], diffuse over time without inducing lysis in neighboring hosts. Thus, synchronized lysing (a possible mechanism for bloom termination [1,13,25]) is not accounted for by our model. Since blooms appear to be the exception rather than the rule [13], we do not expect this to affect the applicability of our model to most ecological settings. The spatial nonuniformity of viruses will be further discussed below.

4.4 A biophysical model of phage-host interaction

4.4.1 Model development part I: A single phage-host system

4.4.1.1 Viral diffusion and infection rate

We begin by estimating the infection rate of a certain bacterial species given that its viruses are freely diffusing in the medium. Let $N_{bacteria}^{(i)}$ and $N_{virus}^{(i)}$ be the number of bacteria and viruses respectively associated with the i^{th} bacterial species in a given volume V . We wish to estimate the absorption rate of the viruses to their hosts, denoted by $I_{virus}^{(i)}$ (in units of s^{-1}), given that $N_{virus}^{(i)} \gg N_{bacteria}^{(i)}$. We will assume the bacterium is stationary and is described by a simple spherical geometry with an effective radius $R_{bact}^{(i)}$. The approximation that the bacterium is stationary is supported by the following facts: Based on the Stokes-Einstein relation (see below),

the diffusion constant of a typical *E. coli*-like bacterium is expected to be roughly 30 times smaller than the diffusion constant of a typical phage particle in the same environment, thus bacteria are diffusing very slowly in comparison to their viruses. Even if a bacterium is engaged in swimming, its contact rate with viruses is relatively unaffected by the swimming motion of the bacterium [12]. Bacteria attached to marine snow may encounter enhanced viral contact rates due to the fast motion of the sinking particles [12], however these are thought to constitute a small fraction of the overall population of bacteria and should therefore not contribute much to the overall number of bacterium-viral contacts [12].

Table 4.1. Variables and parameters used in the discrete phage-host interaction model

Variables	Definition	Units
$c_{bacteria}^{(i)}$	Concentration of bacteria belonging to the i^{th} bacterial species	(number)/ m^3
$N_{bacteria}^{(i)}$	Number of bacteria belonging to the i^{th} bacterial species in volume V	dimensionless
$c_{virus}^{(i)}$	Concentration of viruses infecting the i^{th} bacterial species	(number)/ m^3
$N_{virus}^{(i)}$	Number of viruses infecting the i^{th} bacterial species in volume V	dimensionless
$I_{virus}^{(i)}$	Absorption rate of viruses onto the i^{th} bacterium	s^{-1}
$VBR^{(i)}$	Virus-to-bacterium ratio of the i^{th} phage-host system = $c_{virus}^{(i)} / c_{bacteria}^{(i)}$	dimensionless
Parameters		
i	Index of the i^{th} bacterial species. Parameters that depend on i can be interpreted as random variables drawn from a certain distribution	Dimensionless
$\alpha^{(i)}$	Specific growth rate = $\mu^{(i)} \ln 2$, where $\mu^{(i)}$ is the doubling rate	s^{-1}
$\gamma_{non-viral}^{(i)}$	Bacterial mortality rate due to non-viral mediated processes such as grazing	s^{-1}
$\gamma_{viral\ decay}^{(i)}$	Viral decay rate	s^{-1}
$R_{virus}^{(i)}$	Effective radius of the virus	m
$R_{bact}^{(i)}$	Effective radius of the bacterium	m
$b^{(i)}$	Burst size	Dimensionless
$\beta^{(i)}$	Volume fraction of host cell occupied by virions	Dimensionless
$D_{virus}^{(i)}$	Diffusion constant of the virus	m^2/s
η	Viscosity of the environment	$kg \cdot m^{-1} s^{-1}$
τ	Latency period	s
η	Viscosity of the environment	$kg \cdot m^{-1} s^{-1}$
k_B	Boltzmann constant	$kg \cdot m^2 s^{-2} K^{-1}$
T	Temperature of the environment	K

To estimate the infection rate we assume that viruses anchor to the cell surface, and that consequently the bacterium can be regarded as a perfect absorber. We then solve the diffusion equation for the virions at steady-state. We assume the bacterium is placed at the origin and that the boundary conditions are given by $c_{virus}^{(i)}(r = R_{bact}^{(i)}) = 0$ and $c_{virus}^{(i)}(r = \infty) = c_{virus}^{(i)}(\infty)$, where $c_{virus}^{(i)}(\infty) (= N_{virus}^{(i)} / V)$ is the far-field concentration of the i^{th} viral species. Solving the diffusion equation at steady-state and calculating the transport flux across the boundary of the sphere gives us the steady-state absorption rate of viruses onto the bacterium ($I_{virus}^{(i)}$) [11,26]

$$(1) \quad I_{virus}^{(i)} = 4\pi D_{virus}^{(i)} R_{bact}^{(i)} c_{virus}^{(i)}(\infty).$$

where $D_{virus}^{(i)}$ is the diffusion constant of the i^{th} viral species. Thus the average time until the i^{th} bacterial species is infected is $1/I_{virus}^{(i)}$. The assumption of a perfect absorber means that once viruses make contact with the cell, they are “absorbed” (i.e., infect the cell). Berg and Purcell [26] showed that the net flux to a cell with a small number of receptors is almost as large as the net flux into a perfectly absorbing cell. For example, fewer than 500 phage receptors are necessary for λ phage to attain half the maximum absorption rate [26,27] where *E. coli* typically has between 30 to 10,000 receptors per cell depending on the growth medium [27]. Therefore the assumption of a perfect absorber requires a small correction factor that we shall ignore in our simple toy model.

Because the perfect absorber leads to a steady-state gradient in viral concentration, the distribution of viruses is spatially nonuniform. For the case of a single absorber at the origin, the steady-state concentration of viruses is given by $c_{virus}^{(i)}(r) = c_{virus}^{(i)}(\infty)(1 - R_{bact}^{(i)}/r)$ [26], where $c_{virus}^{(i)}(\infty)$ is the far field concentration of the viruses infecting the i^{th} bacterial species. Thus, if we assume that the mean spacing between cells of a given bacterial species is significantly larger than $R_{bact}^{(i)}$ (i.e., $(c_{bact}^{(i)})^{-\frac{1}{3}} \gg R_{bact}^{(i)}$), then any given bacterial host of this species will lie in the far-field range of adjacent hosts of the same species. Thus under these conditions, to a first-order approximation, each bacterium can effectively be thought of as an isolated perfect absorber. These conditions are typically satisfied for marine ecosystems. For example, for typical marine ecosystems $c_{bact} \leq 10^6 \text{ ml}^{-1}$, or $(c_{bact})^{-\frac{1}{3}} \leq 100 \mu\text{m}$. In the extreme (and unlikely) scenario where

the entire bacterial population consists of a single species, then as long as the radius of this species is $< \sim 10 \mu\text{m}$ this condition is satisfied. Since we will see that larger bacteria are rarer (Section 4.4.2), the error incurred for larger radii will be weighted down when integrating over all radii.

4.4.1.2 Predator-prey relations

We next wish to calculate the total rate of virus infection in the population. The fraction of bacteria $\Delta N_{\text{infected}}^{(i)} / N_{\text{bacteria}}^{(i)}$ that are infected during the time Δt , where Δt satisfies $\Delta t \ll 1/I_{\text{virus}}^{(i)}$ is given by $\Delta t / (1/I_{\text{virus}}^{(i)})$. Therefore the fraction of infected cells during Δt is given by

$$\Delta N_{\text{infected}}^{(i)} / N_{\text{bacteria}}^{(i)} = \Delta t / (1/I_{\text{virus}}^{(i)}), \text{ or}$$

$$\frac{dN_{\text{infected}}^{(i)}}{dt} = I_{\text{virus}}^{(i)} N_{\text{bacteria}}^{(i)}.$$

In principle, not every virion absorption event will lead to successful infection and host lysis. However, at least in the case of T4 infecting *E. coli* this fraction appears to be close to one [11]. We will therefore assume in our toy model that each absorption event leads to host lysis. Building upon this result, we take into account bacterial growth and bacterial death due to non-viral mediated processes and obtain the following bacterial rate equation

$$\frac{dN_{\text{bacteria}}^{(i)}(t)}{dt} = \alpha^{(i)} N_{\text{bacteria}}^{(i)}(t) - \gamma_{\text{non-viral}}^{(i)} N_{\text{bacteria}}^{(i)}(t) - I_{\text{virus}}^{(i)}(t) N_{\text{bacteria}}^{(i)}(t).$$

where the first term is due to bacterial growth, the second term is due to non-viral mediated cell mortality, and the third term is due to viral predation leading to host mortality. Dividing by the system volume V and inserting Eq. 1 we obtain the following rate equation for the bacterium

$$(2) \quad \frac{dc_{bacteria}^{(i)}(t)}{dt} = (\alpha^{(i)} - \gamma_{non-viral}^{(i)})c_{bacteria}^{(i)}(t) - 4\pi D_{virus}^{(i)} R_{bact}^{(i)} c_{virus}^{(i)}(t) c_{bacteria}^{(i)}(t).$$

The corresponding rate equation for the i^{th} viral species is given by

$$\frac{dN_{virus}^{(i)}(t)}{dt} = b^{(i)} \cdot I_{virus}^{(i)}(t - \tau) N_{bacteria}^{(i)}(t - \tau) - \gamma_{virus\ decay}^{(i)} N_{virus}^{(i)}(t) - I_{virus}^{(i)}(t) N_{bacteria}^{(i)}(t).$$

where the first term is due to viral production (with τ being the latency period and $b^{(i)}$ being the average burst size of the i^{th} viral species, i.e., the number of virions released per cell into the extracellular environment), the second term is due to virion decay, and the third term is due to viral loss upon absorption (which is negligible since typically $b^{(i)} \gg 1$) [5]. Dividing by the system volume V and inserting Eq. 1 we obtain the following rate equation for the viruses

$$(3) \quad \frac{dc_{virus}^{(i)}(t)}{dt} = b^{(i)} \cdot 4\pi D_{virus}^{(i)} R_{bact}^{(i)} c_{virus}^{(i)}(t - \tau) c_{bacteria}^{(i)}(t - \tau) - \gamma_{virus\ decay}^{(i)} c_{virus}^{(i)}(t) + \\ - 4\pi D_{virus}^{(i)} R_{bact}^{(i)} c_{virus}^{(i)}(t) c_{bacteria}^{(i)}(t).$$

Equations 2 and 3 together form a predator-prey dynamical system. In the simple case where $\tau=0$, Eq. 2 and Eq. 3 form an ideal Lotka-Volterra model. This system exhibits small oscillations

with a period of $\left[(\alpha - \gamma_{non-viral}) \gamma_{virus\ decay} \right]^{-\frac{1}{2}} \sim$ hours to days around the non-trivial fixed point determined below (see Table 4.3 for typical parameters). Since the steady-state of the viral diffusion equation is achieved on the order of $t \gg R_{bact}^2 / D_{virus} \ll \sim 1$ sec, Eqs. 2 and 3 can be interpreted as describing the slow dynamics of the far-field viral concentration with the viral diffusion equation at pseudo steady-state.

We are interested in the non-trivial fixed point solutions for this system obtained by setting $d/dt = 0$. Since the steady-state solutions are time invariant we have $c_{virus}^{(i)}(t) = c_{virus}^{(i)}(t - \tau) = \text{const}$ and $c_{bacteria}^{(i)}(t) = c_{bacteria}^{(i)}(t - \tau) = \text{const}$. Solving for these two constants we find that the non-trivial fixed point solutions for this system are given by

$$(4) \quad \begin{aligned} c_{virus}^{(i)} &= \frac{\alpha^{(i)} - \gamma_{non-viral}^{(i)}}{4\pi D_{virus}^{(i)} R_{bact}^{(i)}}. \\ c_{bacteria}^{(i)} &= \frac{\gamma_{viral\ decay}^{(i)}}{(b^{(i)} - 1) \cdot 4\pi D_{virus}^{(i)} R_{bact}^{(i)}}. \end{aligned}$$

where we implicitly assume that $\alpha^{(i)} > \gamma_{non-viral}^{(i)}$. Note that equating the rate equation for bacteria to zero leads to the following condition:

$$\alpha^{(i)} \equiv I_{virus}^{(i)} + \gamma_{non-viral}^{(i)}.$$

i.e., total bacterial production equals total bacterial mortality. Though solutions to predator-prey models are typically time dependent, here we are mainly concerned with understanding the

scaling of the fixed point solutions, which we take as a proxy for the time averaged response of the system.

4.4.1.3 The virus diffusion constant

Since the shape of the virus appears to have little effect on the expected viral transport rate to the bacterium [12] we will follow Murray and Jackson and model the viruses as spheres. For a sphere of radius $R_{virus}^{(i)}$ (the effective radius for the virus) the Stokes-Einstein relation for the diffusion constant is given by

$$D_{virus}^{(i)} = k_B T / (6\pi\eta R_{virus}^{(i)})$$

where k_B is the Boltzmann coefficient, T the temperature, and η the viscosity of the medium.

Substituting this expression into the fixed point solution given in Eq. 4 we find that

$$(5A) \quad c_{virus}^{(i)} = \frac{3}{2} \frac{\eta}{k_B T} \left(\frac{R_{virus}^{(i)}}{R_{bact}^{(i)}} \right) (\alpha^{(i)} - \gamma_{non-viral}^{(i)}).$$

$$(5B) \quad c_{bacteria}^{(i)} = \frac{3}{2} \frac{\eta}{k_B T} \left(\frac{R_{virus}^{(i)}}{R_{bact}^{(i)}} \right) \frac{\gamma_{viral\ decay}^{(i)}}{(b^{(i)} - 1)}.$$

Eq. 5A makes the prediction that the larger the factor $D_{virus}^{(i)} R_{bact}^{(i)}$ in Eq. 2 (resulting in a larger viral infection rate – Eq. 1), the lower the concentration of viruses needs to be in order for the overall lysis rate (second term in Eq. 2) to match the net bacterial production rate (first term in Eq. 2). This explains the dependence of Eq. 5A on viscosity, temperature, and the virus-to-

bacterium radii ratio. Eq. 5A also predicts that the fixed point concentration of viruses does not depend on their decay rate or burst size. This paradoxical behavior is explained by the fact that viruses need to keep the net bacterial growth in check (leading to the dependence on the growth rate; first term in Eq. 2) irrespective of the viral decay rate or burst size.

Similarly, Eq. 5B predicts that the higher the factor $b^{(i)}D_{virus}^{(i)}R_{bact}^{(i)}$ in the viral production rate term (Eq. 3), the lower the fixed point concentration of bacteria needs to be in order to match viral production (first term in Eq. 3) with viral decay (second term in Eq. 3), explaining the dependence on viscosity, temperature, the virus-to-bacterium radii ratio, and the burst size. Eq. 5B also makes the intuitive prediction that the faster viruses decay, the higher the concentration of bacteria will be. This result holds because the faster viruses degrade (second term in Eq. 3) the more viruses are required to be produced (first term in Eq. 3) to sustain this degradation, and therefore more bacteria are required for viral production (since bacteria are the sources of viruses). Here too we find the paradoxical situation where the fixed point concentration of bacteria does not depend on their net growth rate. The reason for this paradoxical behavior is that as long as bacteria grow — no matter how fast — their fixed point concentration need only be high enough so that viral production (which is proportional to the bacterium concentration — first term in Eq. 3) matches viral decay (second term in Eq. 3).

4.4.1.4 The virus-to-bacterium ratio for a given phage-host system

To obtain the virus-to-bacterium ratio for the i^{th} species we divide $c_{virus}^{(i)}$ by $c_{bacteria}^{(i)}$ obtaining the simple relation

$$(6) \quad VBR^{(i)} = \frac{c_{virus}^{(i)}}{c_{bacteria}^{(i)}} \cong b^{(i)} \frac{\alpha^{(i)} - \gamma_{non-viral}^{(i)}}{\gamma_{virus\ decay}^{(i)}}.$$

where for simplicity we use the approximation $b^{(i)} - 1 \cong b^{(i)}$ since typically $b^{(i)} \gg 1$. Below we shall derive the expression for the VBR for the entire community in a given environment, i.e., for all phage-host systems.

4.4.1.5 Correlation between burst size and host/virus dimensions

Since we are interested in average scaling laws, it is worthwhile to consider the relation between burst size and the dimensions of the host and its virus, as these two quantities may be statistically highly correlated. Lytic viruses typically pack the host cytoplasm with virions upon replication, suggesting that perhaps one can make the assumption that the number of virus progeny per cell is correlated with cell volume and inversely correlated with the volume of the infecting virus. Indeed, Weinbauer et al. found that in ~50% of the visibly infected rods and spirillae and in more than 80% of the cocci found in the northern Adriatic Sea, the entire cell was occupied by mature phages (as opposed to displaying a non-uniform distribution) with the difference between cocci and other morphologies possibly explained by a shorter time span between the appearance of the first mature phages and lysis in cocci cells due to their smaller burst size [28]. Weinbauer et al. also note that almost all bacteria observed in the disruption stage were completely filled with phages [29]. That said, in 18% of the infected bacteria the phage was concentrated in two or three defined areas of the host and did not occupy the entire cell [28]. Furthermore, some bacteria may lyse prematurely [2]. Nevertheless, it has been found empirically that burst size is approximately linearly correlated with cell size for cells with a radius of ~0.2 μ m to ~1 μ m, and larger phages have been found to produce less progeny [2,19]. For example, Weinbauer et al.

[29] found a linear correlation between burst size and host cell volume, with the cell size being the only measured parameter that could account for the distribution of burst sizes [29]. In addition, Weinbauer et al. also found an inverse correlation between burst size and capsid size [28].

We will therefore assume in our toy model that to a first-order approximation the burst size is proportional to the volume ratio of the bacterium and its virus, namely, $b^{(i)} = \beta^{(i)} \cdot (R_{bact}^{(i)} / R_{virus}^{(i)})^3$, where $\beta^{(i)}$ is a positive proportionality factor ≤ 1 . Note that $\beta^{(i)}$ can be interpreted as the volume fraction of the cell occupied by viruses since: $\beta^{(i)} = b^{(i)} \cdot (R_{virus}^{(i)} / R_{bact}^{(i)})^3 = b^{(i)} \cdot V_{virus}^{(i)} / V_{bact}^{(i)}$.

Inserting this correlation into Eq. 5B and approximating $b^{(i)} - 1 \cong b^{(i)}$ we obtain

$$(7) \quad c_{bacteria}^{(i)} = \frac{3}{2} \frac{1}{\beta^{(i)}} \frac{\eta}{k_B T} \left(\frac{R_{virus}^{(i)}}{R_{bact}^{(i)}} \right)^4 \gamma_{viral\ decay}^{(i)}.$$

Further implications of the model and Eq. 7 are discussed in the following section.

We wish to estimate $\beta^{(i)}$ based on experimental observations. In Fig. 4.1 we reproduce the data of Weinbauer et al. [28], who measured in the northern Adriatic sea the burst size $b^{(i)}$ as a function of the cell volume $V_{bact}^{(i)}$ for bacterial radii ranging from ~ 0.2 to $\sim 0.9 \mu\text{m}$ and for two groups of capsid diameters: 30-60nm (group A; blue) and 60-110nm (group B; red). Note that in Fig. 4.1 the y-axis is plotted as the burst size times the average volume of a capsid for that group.

Therefore, if indeed $b^{(i)}V_{virus}^{(i)} = \beta^{(i)}V_{bact}^{(i)}$ with $\beta^{(i)} \equiv \beta \equiv \text{const}$, we would expect the slope in Fig. 4.1 to be the same for both size classes. Indeed we estimate very similar values for β for both size classes: $\beta \approx 0.005$. Furthermore, when consolidating both size groups and assuming an average capsid diameter of 60nm (the peak value found in nature) we obtain $\beta \approx 0.0049$, in agreement with the previous results. We therefore find that over a wide range of bacterial sizes, 0.5% of the cell volume is occupied with viruses upon lysis. Closer inspection of the correlation suggests however that for small cell volumes ($< \sim 1 \mu\text{m}^3$, corresponding to a radius $< \sim 0.6 \mu\text{m}$), the burst size is underestimated based on our simple linear formula given the above estimate for β . In a different work, Weinbauer et al. [29] studied the correlation between burst size and cell volume for small cells ($< 0.3 \mu\text{m}^3$) in Lake Plußsee. Based on this correlation we find that cells with a volume of $\sim 0.3 \mu\text{m}^3$ had a burst size of ~ 90 while cells with a cell volume of $\sim 0.05 \mu\text{m}^3$ had a burst size of ~ 35 . Assuming a typical capsid diameter of 60nm, this corresponds to $\beta \sim 0.03$ for the former case and $\beta \sim 0.08$ for the latter case. Since we will find however that small cells tend to be much more abundant than large cells, an underestimation of the burst size at small volumes may bias results. We will therefore assume that β is bounded in the range of $\sim 0.5\%$ to $\sim 5\%$. It is less certain how well this relation will hold for bacterial radii $> \sim 1 \mu\text{m}$. For unicellular eukaryotes with radii in the range of ≈ 2 to $\approx 7 \mu\text{m}$ we indeed estimate values of β in the range of 0.1% to 3.4% (Table 4.2), consistent with the above bounds, suggesting that our empirical correlation may hold for larger cells as well. Below we will discuss the case of extremely large bacteria that have massive cell inclusions that reduce the effective cytoplasm volume (and thus the value for β). Thus, for very large cells β may behave anomalously.

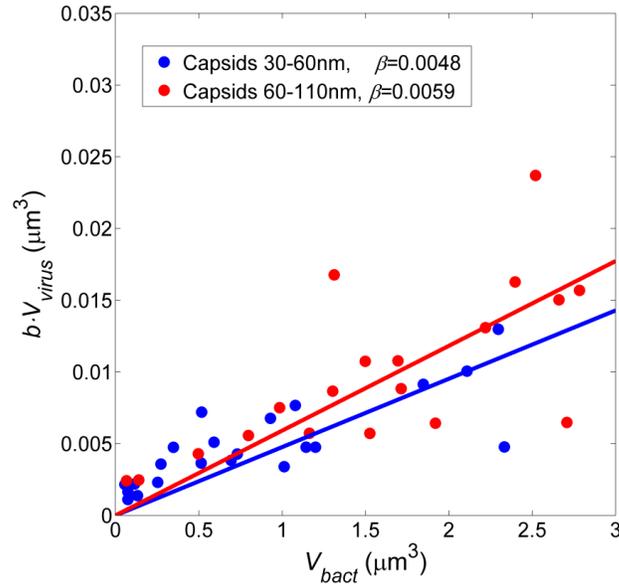


Figure 4.1. Correlation between burst size and cell volume. Here we reproduce the data from Weinbauer et al. [28] for the correlation between burst size and cell volume for two capsid diameter classes; 30-60nm (group A) and 60-110nm (group B). The y-axis was plotted as the burst size times the average capsid volume for that group. The average volume of a capsid for group A was calculated assuming a capsid diameter of $(30+60)/2 = 45\text{nm}$, while for group B the average capsid diameter was assumed to be $(60+110)/2 = 85\text{nm}$. The straight line is a least squares fit a line with a zero constant. The Pearson correlation coefficient for data points of group A was $\rho=0.79$, and for data points of group B was $\rho=0.71$.

Table 4.2. Estimation of virus volume fraction, β , for unicellular eukaryotes.

Eukaryote	Approx. radius	Burst size	Virus diameter	Ref.	β
<i>E. huxleyi</i>	$\approx 2.3\mu\text{m}$	400-1000 (mean 620)	$\approx 170\text{nm}$	[30]	$\approx 3.1\%$
<i>H. akashiwo</i>	$\sim 5-7\mu\text{m}^*$	$\sim 10^5$	$\approx 30\text{nm}$	[30,31]	$\sim 0.1-0.3\%$
<i>C. ericina</i>	$\sim 7\mu\text{m}^*$	1800-4100	$\approx 155\text{nm}$	[30,32]	$\sim 3.4\%$

* Size estimated based on different strain of this species

4.4.1.6 Dependence of host concentration on bacterium size

The most striking feature of Eq. 7 is the dependence of the concentration of the bacteria on the fourth power of the ratio $R_{\text{virus}}^{(i)} / R_{\text{bact}}^{(i)}$. Thus a bacterium that is twice as large is predicted according to this model to be $(1/2)^4 = 1/16$ times less abundant. This effect is both because larger

bacteria have a larger cross section for diffusing viruses and because larger viruses produce more virions, thus for larger cells, fewer bacteria are needed for viral production to match viral decay.

Comparing the radii of viruses and their hosts, it appears that the radii of bacteria are much more variable in natural environments. The range of the dimensions of prokaryotes in nature is tremendous, ranging from a diameter of 0.2 to 750 μm [33,34] spanning over three orders of magnitude. When raised to the power of four this variable spans an astonishing 14 orders of magnitude. On the other hand, the diameter range of heads of tailed phages (that constitute about 96% of all phages examined to date via electron microscope [35]) is very narrow and lies between 34 and 160 nm, peaking sharply at 60 nm [36]. If we use the simple rule of thumb that the dimensions of viruses is fixed at 60 nm, then for a given environment defined by η and T we can plot the fixed point concentration of bacteria as a function of the size of the bacterium (Fig. 4.2). Since *a priori* we have no reason to believe there is a correlation between $\gamma_{\text{viral decay}}^{(i)}$ and $R_{\text{bact}}^{(i)}$, we will regard $\gamma_{\text{viral decay}}^{(i)}$ as a constant. We will further approximate $\beta^{(i)}$ as a constant that is uncorrelated with $R_{\text{bact}}^{(i)}$, though we are less certain how well this assumption will hold for larger bacteria (further discussed below). From Fig. 4.2 we see that small bacteria are predicted to achieve significantly higher concentrations and lower VBRs. Thus, according to this model, the size of a bacterium appears to be the most important factor determining its fixed point concentration.

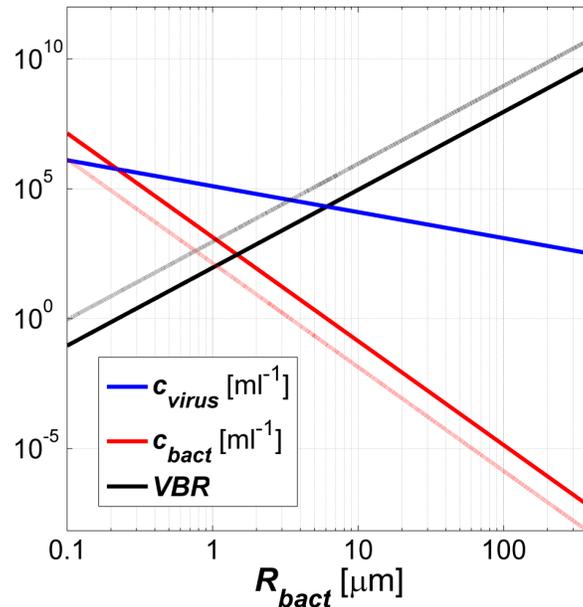


Figure 4.2. Scaling of the virus concentration, the bacterium concentration and the VBR with the radius of the bacterium for a single phage-host system. Virus concentration was calculated based on Eq. 5A and bacterium concentration was calculated based on Eq. 7. Parameters used for these equations, which are typical for marine systems, are given in Table 4.3, with $\gamma_{viral\ decay} = 2\text{ day}^{-1}$ chosen to represent an offshore marine environment [23]. Solid lines are for $\beta=0.005$ while dotted lines are for $\beta=0.05$.

4.4.1.7 Large bacteria are rare

Fig. 4.2 demonstrates that very large bacteria should exist at extremely low concentrations in the ocean. The largest bacterium known to date, *Thiomargarita namibiensis*, with a diameter of up to 750 μm , found in marine sediments [37], is predicted to occur at a frequency of 1 cell per $\sim 1.5 \cdot 10^4$ liters of water (Fig. 4.2). For very large bacteria however, our assumption of a constant β most likely breaks down. Many of the large bacteria harbor massive cell inclusions that reduce the volume of the metabolically active cytoplasm [38]. In the case of *T. namibiensis* for example, its cytoplasm is restricted to a thin $\sim 1\ \mu\text{m}$ outer layer that surrounds a large central vacuole [37]. This inclusion therefore leads to a reduction in $\beta^{(i)}$ by two orders of magnitude:

$\beta^{(i)} \rightarrow \beta^{(i)} \left(1 - (374 \mu\text{m}/375 \mu\text{m})^3\right) \cong 10^{-2} \beta^{(i)}$. Thus, the effective concentration of this bacterium would be predicted to be higher by two orders of magnitude. Consequently, inclusions have the beneficial effect of increasing the abundance of the host by reducing its effective burst size. However, even with this large inclusion, assuming $\beta^{(i)} \approx 5 \cdot 10^{-5}$, the host cell will still be very rare, with only one cell per ~ 100 liters of ocean water. Thus free-floating large cells can easily go undetected. It is therefore not surprising that *T. namibiensis* was discovered in sediments where it was found to be highly enriched [37] and not free floating in the ocean. Since large bacteria are predicted to be very rare in the open ocean and easily missed, it is worth noting that the viruses infecting such large cells are predicted to be relatively abundant, with several hundreds of virions per ml. Thus marine phages, according to this model, may be very sensitive proxies to rare, large bacterial cells.

Table 4.3. Typical parameters for phage-host systems in aquatic environments

Parameter	Value	Aquatic region	Reference
R_{virus}	≈ 30 nm	Many environments	[36]
R_{bact}	≈ 0.1 - 0.2 μm	Open ocean	[34]
α	~ 2 day^{-1}	Coastal	[18]
$\gamma_{non-viral}$	~ 1 day^{-1}	Coastal	Inferred
$\gamma_{viral\ decay}$	~ 0.1 to ~ 10 day^{-1}	Various marine	[19]
β	$\sim 0.5\%$ to $\sim 5\%$	Marine and lake	[28,29] (Inferred)
η	$\approx 10^{-3}$ $\text{kg}/(\text{m}\cdot\text{s})$	-	
$k_B T$ ($T=24^\circ\text{C}$)	$4.1 \cdot 10^{-21}$ $\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-2}$	-	
r_{max}	375 μm	Sediments	[34,38]
r_{min}	0.1 μm	Open ocean	[34,38]
D_{virus}	≈ 5 $(\mu\text{m})^2/\text{s}$	λ phage	[39] [39]

4.4.1.8 Application of the model to environmental systems

The *Synechococcus* phage-host system in the Gulf of Mexico

Eqs. 5A and 5B are very powerful in the sense that they predict the absolute equilibrium concentration of hosts and their viruses from basic parameters describing the environment, the bacterium, and the virus that infects it. We wish to see how the model predictions of the concentration of particular phage-host systems compare with measurements of specific phage-host systems in nature. One particular system of interest is cyanobacteria, which has been studied extensively. The concentration of cyanophages in coastal waters and off shore waters in the Gulf of Mexico infecting *Synechococcus* (1.5 μm in diameter) peaked at 10^5 ml^{-1} at the ocean surface [23] with a VBR for this phage-host system measured to be as high as 8 [23]. Based on the depth profiles in this study we computed the average concentration of *Synechococcus*, the average concentration of cyanophages infecting *Synechococcus* strain DC2 and the average VBR (Table 4.4). We wish to compare these observations to model predictions.

Virus concentration

For the *Synechococcus* case study $R_{bact}^{(i)} = 0.75 \mu\text{m}$. In a related study, capsid diameters of virions infecting a *Synechococcus* host were found to be in the range of 50–65 nm [40]. Thus we assumed that $R_{virus}^{(i)} \approx 30 \text{ nm}$ [36]. The average growth rate of bacteria in coastal waters is on the order of $\alpha^{(i)} \sim 2 \text{ day}^{-1}$ [18] (Table 4.3). At steady-state this growth rate equals the sum of the lysis rate and non-viral mediated mortality rate (see above). There is evidence to suggest that grazing and lysis contribute equally to microbial mortality [15,18,19], though this matter is still the subject of debate [19]. Nevertheless, as a first-order approximation we will assume that bacterial production is roughly halved by grazing so that $\gamma_{non-viral}^{(i)} \sim 1 \text{ day}^{-1}$. Thus, given that for

water at 24°C $\eta \approx 10^{-3}$ Pa·s, Eq. 5A predicts that $c_{virus}^{(i)} \approx 1.7 \cdot 10^5 \text{ ml}^{-1}$. This prediction is of the same order of magnitude as the measurements described above (see Table 4.4).

Alternatively, using the Stokes-Einstein relation we can calculate that a virus with an effective sphere diameter of 60 nm in water at 24°C should have a diffusion constant of $7.25 (\mu\text{m})^2/\text{s}$. This value is close to the measured diffusion constant of λ phage at the same temperature, which is $4.97 (\mu\text{m})^2/\text{s}$ [39]. Using the diffusion constant of the virus, one can also calculate the fixed point concentration of viruses using Eq. 4 and obtain the same solution.

Host concentration

To calculate the concentration of the bacteria one needs to know the viral decay rate and the burst size. The viral decay rate was measured in this study to be 0.1 day^{-1} inshore and 2 day^{-1} offshore [23]. Given our earlier estimate of $\beta=0.005$ we find a burst size of

$$b^{(i)} = \beta \left(\frac{R_{bact}^{(i)}}{R_{virus}^{(i)}} \right)^3 \cong 80. \text{ In a one-step growth experiment for a different strain of } \textit{Synechococcus},$$

the burst size was measured to be 250, in rough agreement with our simple linear model prediction. The one-step growth experiment burst size is most likely an overestimate since burst sizes of isolated phage-host systems are known to be consistently higher than those found in the environment since cells growing in culture are larger and thus produce more progeny and/or are better adapted to high nutrient concentrations [2]. Thus, assuming $\beta=0.005$, $\gamma_{virus\ decay}^{(i)} \sim 0.1 - 2 \text{ day}^{-1}$ and $R_{virus}^{(i)} \sim 30 \text{ nm}$ (see above) with the remaining parameters taken from Table 4.3, then based on Eq. 7 (or 5B) we find that $c_{bacteria}^{(i)} \sim 200 \text{ ml}^{-1}$ to $4.3 \cdot 10^3 \text{ ml}^{-1}$. These predictions are consistent with the range of observed concentrations of *Synechococcus* cells (Table 4.4). Note

however that a true test of the model predictions would require comparing with seasonal averages and not with one time measurements.

Table 4.4. Measured concentration of *Synechococcus* and the cyanobacteria infecting it in the Gulf of Mexico versus model predictions

Variable	Observed ($n=21$) ^a		Prediction ^b
	Mean \pm S.D.	Range	
C_{virus} (cyanophages)	$(5.6 \pm 8.6) \cdot 10^4 \text{ ml}^{-1}$	150 ml^{-1} to $2.5 \cdot 10^5 \text{ ml}^{-1}$	$1.7 \cdot 10^5 \text{ ml}^{-1}$
$C_{bacteria}$ (<i>Synechococcus</i>)	$(1.8 \pm 2.9) \cdot 10^4 \text{ ml}^{-1}$	3.0 ml^{-1} to $9.2 \cdot 10^4 \text{ ml}^{-1}$	200 ml^{-1} to $4.3 \cdot 10^3 \text{ ml}^{-1}$
VBR	6.3 \pm 8.1	0.2 to 30.7	40 to 780 ^c

^aMeasurements based on depth profiles measured by Suttle and Chan [23]. Virus concentration corresponds to viruses infecting *Synechococcus* strain DC2.

^bPredictions were made based on Eq. 5 and 7. See text for further details.

^cA better prediction could be made if the data for each station was analyzed separately as there were significant differences between stations.

4.4.2 Model development part II: Non-interacting phage-host systems

4.4.2.1 A stochastic interpretation of bacterial and viral parameters

Thus far we have considered the case of an isolated phage-host system and have treated quantities that depend on the species index i as deterministic quantities, i.e., every index i corresponds to a different phage-host system with a different set of parameters. When considering a natural environment, many different species – i.e., phage-host systems – co-exist. One can therefore imagine a hypothetical “species sample space” comprised of many phage-host systems, where each time we draw a phage-host system with index i we obtain a set of values for all model parameters based on some joint density function. Hence, all parameters can be thought of as random variables drawn from some joint distribution. Since the concentration of viruses and bacteria are functions of these parameters, these variables can be thought of as random variables themselves.

Of all the parameters that $c_{bacteria}^{(i)}$ depends on, $R_{bact}^{(i)}$ has the widest range of values, spanning over three orders of magnitude. Furthermore, given that $R_{bact}^{(i)}$ is raised to the fourth power, it is by far the most sensitive parameter in Eq. 7 (see above). For comparison, the distribution of phage capsid diameters $2R_{virus}^{(i)}$ peaks sharply at 60 nm (see above). $\gamma_{viral\ decay}^{(i)}$ varies by about two orders of magnitude across environments [19], however, we expect that for a given environment, where all phages are subject to the same conditions, the range of $\gamma_{viral\ decay}^{(i)}$ will be more restricted. In addition, $c_{bacteria}^{(i)}$ is only linearly dependent on $\gamma_{viral\ decay}^{(i)}$. Finally, $\beta^{(i)}$ also appears to display limited variability (see above).

Therefore, if we assume, to a first-order approximation, that the random variables $R_{virus}^{(i)}$, $\gamma_{viral\ decay}^{(i)}$, and $\beta^{(i)}$ are statistically independent of the random variable $R_{bact}^{(i)}$, then we can average out these parameters by taking their expected value. If these parameters are also statistically independent of each other then we have

$$(8) \quad c_{bacteria}(R_{bact} = r) \approx \frac{3}{2} \frac{\eta}{k_B T} E\beta^{-1} ER_{virus}^4 E\gamma_{viral\ decay} r^{-4} = \text{const} \cdot r^{-4}.$$

where E denotes the expectation operator. Thus, our hypothetical species sample space reduces to a single random variable, $R_{bacteria}^{(i)}$, drawn from some distribution $f_R(r)$, the functional form of which we do not know. Eq. 8 predicts what would be the average concentration of a particular bacterial species with radius r were it to exist in a given environment. In practice, the number of bacteria of a given radius present per ml of water in a given environment per radius,

$\rho_{environment}(r)$ (in units of (number)/ m^4 see Table 4.5), depends on which bacteria happen to be in the given environment to begin with. Let's assume that a given environment contains $N_{species}$ different bacterial species $i=1\dots N_{species}$, with each species characterized by its own radius $R_{bacteria}^{(i)}$, where the subscript i labels the species. Thus, for a given realization of this environment, the distribution of observed bacterial radii would be given by

$$(9) \quad \rho_{environment}(r) = c_{bacteria}(r) \sum_{i=1}^N \delta(r - R_{bacteria}^{(i)}).$$

where $\delta(r)$ denotes the Dirac delta function (in units of m^{-1}) and where $R_{bacteria}^{(i)}$ are $N_{species}$ i.i.d.¹ random variables drawn from a distribution $f_R(r)$. Note that $f_R(r)$ is the probability density that a bacterium with radius $R_{bact} = r$ *a priori* exists in the environment, whereas $\rho_{environment}(r)$ is the actual concentration of bacteria observed in the environment per bacterial radius. Thus $\rho_{environment}(r)$ is one realization of the distribution of bacteria in the given environment. To obtain the ensemble average of $\rho_{environment}(r)$, averaging over many realizations of the given environment (making the simplifying assumption that in each realization there are always $N_{species}$ different species) one should calculate the expectation value of $\rho_{environment}(r)$ with respect to the $N_{species}$ random variables $R_{bacteria}^{(i)}$. In Section 4.7 we show that this ensemble average is given by

$$(10) \quad \langle \rho_{environment}(r) \rangle = N_{species} c_{bact}(r) f_R(r)$$

¹ Independent and identically distributed

Table 4.5. Variables and parameters used in the continuous phage-host interaction model

Variables	Definition	Units
$\rho_{environment}(r)$	Concentration of bacteria with radius r per radius, predicted to exist in a given environment	(number)/ m^4
$N_{species}$	Total number of prokaryote species that exist in any given realization of the environment	dimensionless
V_{env}	Volume to find one cell of the largest bacterium ($r=r_{max}$), defining the effective size of the environment	m^3
$\rho_{species}$	Concentration of <i>species</i> (bacterial and viral) in the environment ($= N_{species}/V_{env}$)	(number)/ m^3
$f_R(r)$	Probability density function from which the radius r of a bacterial species is drawn (also defined as the density of bacterial species)	(probability)/ m
$f_\rho(r)$	Probability density function of radii measured in a given environment (empirically, the histogram of measured bacterial radii in a given environment)	(probability)/ m
c_{bact}^{tot}	Concentration of all prokaryotes in a given environment	(number)/ m^3
c_{virus}^{tot}	Concentration of all phages in a given environment	(number)/ m^3
$m_{bact}(r)$	Wet mass of bacterium of radius r	kg
$M_{bact}(r)$	Wet mass density of prokaryotes of radius r per radius in a given environment	kg/m^4
M_{bact}^{tot}	Wet mass density of all prokaryotes in the environment	kg/m^3
VBR	Virus-to-bacterium ratio in the environment $= c_{virus}^{tot}/c_{bact}^{tot}$	Dimensionless
Parameters		
ρ_{cell}	Wet mass density of a cell	kg/m^3
r_{min}/r_{max}	Minimum/maximum radius of viable bacterium in nature	m
m_{min}/m_{max}	Minimum/maximum wet mass of viable bacterium in nature	kg

4.4.2.2 A simple evolutionary scenario

In the simplest evolutionary scenario we assume that there is no selection pressure on bacterial radii, i.e., bacteria of all sizes are equally adapted to survive and therefore can all have equal probability to exist *a priori* in a given environment. Consequently evolution did not evolve more small bacterial species than large bacterial species, and hence the density of bacterial species per radii is constant. This hypothesis therefore implies that all bacterial radii are equally probable to exist and therefore the random variables $R_{bacteria}^{(i)}$ should be drawn from a uniform distribution:

$R_{bacteria}^{(i)} \sim U(r_{min}, r_{max})$, where r_{min} and r_{max} are the minimum and maximum radii for a viable bacterium, respectively, and where $U(a, b)$ denotes a uniform continuous distribution in the range $[a, b]$, thus

$$(11) \quad f_R(r) = \begin{cases} (r_{\max} - r_{\min})^{-1} & r_{\max} \leq r \leq r_{\min} \\ 0 & \text{otherwise} \end{cases}.$$

Thus $f_R(r)$ can be interpreted as the density of bacterial species, perhaps analogous to the density of states in statistical mechanics, and reflects the evolutionary history of bacteria in the given environment. If the radii of all bacterial species that have adapted to survive in the given environment were known, one could, in principle, calculate $f_R(r)$ directly. Given this scenario, using Eq. 8 and Eq. 10, we find that the ensemble average of the concentration of bacteria expected to exist in a given environment is given by

$$(12) \quad \langle \rho_{environment}(r) \rangle \approx \frac{3}{2} N_{species} \frac{\eta}{k_B T} E \beta^{-1} E R_{virus}^4 E \gamma_{virus\ decay} r_{\max}^{-1} r^{-4} = \text{const} \cdot r^{-4}$$

where we have assumed that $r_{\min} \ll r_{\max}$.

4.4.2.3 The size spectra of bacteria in aqueous environments

To calculate the size spectra of bacteria in the environment we first derive the probability density function (pdf) of observed radii in the environment. This function is obtained by normalizing $\langle \rho_{environment}(r) \rangle$ given in Eq. 12:

$$(13) \quad f_\rho(r) = \frac{\langle \rho_{environment}(r) \rangle}{\int_{r_{\min}}^{r_{\max}} \langle \rho_{environment}(r) \rangle dr} = 3 (r_{\min}^{-3} - r_{\max}^{-3})^{-1} r^{-4} \approx 3 r_{\min}^3 r^{-4}$$

where $r_{\min} \leq r \leq r_{\max}$ and where we assumed that $r \ll r_{\max}$. Note that $f_{\rho}(r)$ is the pdf of $\langle \rho_{environment} \rangle$ where as $f_R(r)$ is the pdf of R_{bact} . Thus $f_{\rho}(r)dr \approx 3r_{\min}^3 r^{-4} dr$ is the probability of observing bacteria with radii between r and $r+dr$ in a given environment. The probability that a bacterium of random volume V is greater than or equal to a given volume, v , would then be given by

$$(14) \quad \begin{aligned} \text{Prob}(V \geq v) &= \text{Prob}(R \geq r) = \int_r^{\infty} f_{\rho}(r') dr' = \int_r^{R_{\max}} 3(r_{\min}^{-3} - r_{\max}^{-3})^{-1} r'^{-4} dr' \\ &= \left(\frac{r_{\min}}{r_{\max}} \right)^3 \left[\left(\frac{r_{\max}}{r} \right)^3 - 1 \right] \approx \left(\frac{r_{\min}}{r} \right)^3 = \frac{v_{\min}}{v}. \end{aligned}$$

assuming that $r \ll r_{\max}$ (see Section 4.7 for further details). When plotting $\log(\text{Prob}(V \geq v))$ against $\log(v)$ one obtains a power law with slope -1. In 2001 Cavender-Bares *et al.* [41] measured the size spectra of microbes up to a diameter of $\sim 5 \mu\text{m}$ (i.e., from bacteria to nanophytoplankton) in the western north Atlantic Ocean. The researchers found that when plotting $\log(\text{Prob}(V \geq v))$ versus $\log(v)$, measurements fell on a straight line with a slope ranging between -1 and -1.4. The ensemble average of all environments was well described by a power law of slope -1.2. When expanding their dataset to include microzooplankton the slope was corrected to a value close to -1. A slope of -1 was also found earlier by Sheldon *et al.* [42].

Eq. 14 also predicts that the power law behavior with slope -1 is an intrinsic scaling property of the biophysical/biological dynamics of phages and their hosts and of $f_R(r)$, and therefore should remain unchanged under perturbations (irrespective of the functional form of $f_R(r)$). Thus

perturbations increasing the viral decay rate or increasing bacterial growth rate, etc., should not have an effect on this power law. This prediction was validated in IronEx II [41], an iron enrichment experiment in the equatorial Pacific, where it was shown that the slope of the power law for samples taken from outside and inside fertilized waters over the course of the experiment differed by little [41]. In both cases the power law was measured to be in the range of -1.1 to -1.2 [41].

4.4.2.4 Possible deviation from a uniform distribution

If we take into account that β tends to decrease with r , we would expect a weaker slope for the size spectra. This result may indicate that a more realistic evolutionary scenario would be one in which larger bacteria are less probable, i.e., the density of bacterial species is higher for small radii. Indeed, small cells may have certain advantages over larger cells. For example, since small cells are more numerous, their population explores collectively more mutations allowing them to adapt more quickly to changing environments and allows them to more easily exploit new habitats [34]. In addition, the high surface-to-volume ratio of cells with smaller radii allows them more efficient exchange of nutrients and higher specific metabolic rates [34,38] possibly giving them a selective advantage.

4.4.2.5 Total bacterial concentration

To obtain the total concentration of bacteria in a given environment we integrate $\langle \rho_{environment}(r) \rangle$ (Eq. 10) over the range of viable bacteria sizes

$$(15) \quad c_{bact}^{tot} = \int_{r_{min}}^{r_{max}} \langle \rho_{environment}(r) \rangle dr = N_{species} \int_{r_{min}}^{r_{max}} c_{bacteria}(r) \cdot f_R(r) dr.$$

Note that Eq. 15 can also be rewritten as $c_{bact}^{tot} = N_{species} \cdot Ec_{bacteria}$, that is the total concentration of bacteria in a given environment equals the total number of species in a given environment, $N_{species}$, times the mean concentration of a single bacterial species. Inserting the population average of $c_{bacteria}^{(i)}$ given in Eq. 8, and assuming again a uniform distribution $f_R(r) = (r_{max} - r_{min})^{-1}$ for $r_{min} \leq r \leq r_{max}$ we find that

$$(16A) \quad c_{bact}^{tot} \cong \frac{1}{2} N_{species} \frac{\eta}{k_B T} E\beta^{-1} ER_{virus}^4 E\gamma_{viral\ decay} r_{max}^{-1} r_{min}^{-3}.$$

4.4.2.6 Species richness

Given a known total concentration of bacteria (determined either by protist grazing or nutrient availability), Eq. 16A can be reversed to predict the number of species in the given environment:

$$(16B) \quad N_{species} \cong 2 \frac{k_B T}{\eta} \frac{1}{E\gamma_{viral\ decay} ER_{virus}^4 E\beta^{-1}} c_{bact}^{tot} r_{max}^{-1} r_{min}^3.$$

The largest bacterium found to date has a diameter of 750 μm (see above) and the smallest bacterium has a diameter of 0.2 μm (Table 4.3), close to the theoretical lower limit thought to be 0.14 μm [43]. Thus, given a typical marine scenario (such as the open ocean) in which direct observation reveals $\sim 10^5$ bacterial cells per ml [44] (i.e., $c_{bact}^{tot} \sim 10^5 \text{ ml}^{-1}$), then based on the parameters in Table 4.3, which are typical for marine systems, and assuming a viral decay rate of $\gamma_{viral\ decay} \sim 2 \text{ day}^{-1}$ for offshore ecosystems [23], we find via Eq. 16B that the total number of

bacterial species in any given realization of the environment is $N_{species} = 82$ to 820 , thus $N_{species} \sim 10^2$ to $\sim 10^3$. $N_{species}$ is thus the number of species (i.e., phage-host systems) that any realization of the environment must contain in order to reach the observed total bacterial concentration. Thingstad & Lignell [10] also calculated the number of species in the environment given a fixed total concentration of bacteria, however in their model the authors assumed that all hosts were identical, ignoring their distribution in the environment. Eq. 16B is expected to be a more realistic estimate since we take into account the distribution of species in the environment. Because many species can be very rare (i.e., have a low concentration due to a large radius — Fig. 4.2) the total predicted diversity is expected to be much higher.

4.4.2.7 What is a species?

Since our model is capable of predicting the number of species in a given environment, we should ask ourselves, what precisely are we counting? What is the definition of a “species” according to our model? This question has practical meaning because we would like to test our prediction against observation. However, there are many “cutoff” values for genetic diversity. For example, is a “species” equivalent to a “species” in biology? Is it equivalent to a “strain”? Does one mutation constitute a new “species”?

In the context of our model here, a “species” of a bacterium is defined by (a) a set of random variables (e.g., the size of the bacterium, its growth rate, etc.) (b) having a unique association with a “viral species” independent of all other phage-host systems, and finally, (c) there is an equal number of “bacterial species” as “viral species”. However, this definition is not sufficient. If, for example, two hosts have exactly the same parameters, they could still have totally

different genomes, and thus constitute distinguishable entities that should be counted separately. Thus, to say that two hosts with the same parameters are identical would be wrong. All that our biophysical model predicts is the number of independent phage-host systems that can be accommodated in a given environment. It does not define how these phage-host systems are different. Therefore a more detailed definition of what a “species” is lies outside the scope of the present model, necessitating us to dig deeper.

An analogy to physics. This paradoxical situation is often encountered in physics. To draw on a physics analogy, our biophysical model’s description of a species is analogous to nuclear physics’ description of a nucleus, which makes the abstraction that the nucleolus is comprised of protons and neutrons. In nuclear physics, protons and neutrons are regarded as point particles defined by certain quantum numbers (like our random variables describing a “species”). Within the framework of this theory though, it is meaningless to ask what is the internal structure of these particles. Likewise, within the context of this biophysical model it doesn’t make sense to ask what the structure of a “species” is. To better understand what a proton and neutron is, a more sophisticated model was required, called the standard model, which showed that protons and neutrons are made out of quarks held together by gluons. In Chapter 5 we propose the “standard model” of phage-host interaction, which allowed us to probe the “internal” structure of a “species”. The model proposed in Chapter 5 is a speciation model describing how new *species* of both bacteria and viruses are generated in nature, leading to a “world” of non-interacting phage-host systems, consistent with the present biophysical model. Drawing on evolution, the new model adds another metric to our description of these organisms, which is the evolutionary distance metric. Therefore in Chapter 5 we will be able to describe a model where a *species* is

comprised of many *strains*, and explain how *strains* evolve into *species*. Again, drawing on our physics analogy, our *strains* will be the “quarks and the gluons” that comprise our *species*. We will therefore revisit the question of “what is a species” in Section 5.2.3 after developing our evolutionary model. The final answer we will arrive at is that $N_{species}$ is (to within a small correction factor between) the total number of consistently distinguishable genomes (termed *strains*), a very intuitive result, with the twist that the “species” defined in our coarse-grained model are actually comprised of a collection of *strains*.

A second question that arises from our calculation is in what volume, according to our model, should we find these species? We will answer this question in the next section, and by answering this question we will be able to calculate the density of *species* in the ocean, from which we will be able to calculate an upper bound on the total bacterial diversity in the oceans.

4.4.2.8 Volume of diversity

The minimum volume that needs to be sampled to detect the $N_{species}$ species is determined by the lowest predicted concentration of bacteria, namely the concentration of the largest bacteria. This volume is given by

$$V_{env} = \left[\frac{3}{2} \frac{1}{\beta^{(max)}} \frac{\eta}{k_B T} \left(\frac{R_{virus}^{(max)}}{r_{max}} \right)^4 \gamma_{viral\ decay}^{(max)} \right]^{-1} .$$

where the index “max” corresponds to the bacterial species with the largest diameter. Taking the model at face value, given $\beta^{(max)}=0.005$, $\gamma_{viral\ decay}^{(max)}=2 \text{ day}^{-1}$ (offshore waters) and with the remaining parameters taken from Table 4.3 we find that at least $\sim 15,000$ liters of water are required to detect the largest known bacterium (see above). In other words, $\sim 1.5 \cdot 10^4$ liters of

water must contain $\sim 10^2$ – 10^3 species of bacteria in order to account for $\sim 10^5$ cells per ml. In practice this volume may be two orders of magnitude smaller due to an uncertainty in β for very large cells (see Section 4.4.1.7). Thus, anywhere between $\sim 10^2$ liters to $\sim 10^4$ liters of water are required to be sampled in order to observe the predicted number of prokaryotic species.

4.4.2.9 Species density

Dividing $N_{species}$ from Eq. 16B by V_{env} we obtain the following expression for the “species density”:

$$\rho_{species} \triangleq \frac{N_{species}}{V_{env}} = 3 \frac{1}{E\beta^{-1}} c_{bact}^{tot} \left(\frac{r_{min}}{r_{max}} \right)^3 \frac{1}{\beta^{(max)}} \approx 5 \cdot 10^{-9} c_{bact}^{tot}.$$

where we have assumed that $E\gamma_{viral\ decay} \approx \gamma_{viral\ decay}^{(max)} (R_{virus}^{(max)})^4 \approx ER_{virus}^4$, $r_{max}/r_{min} = 3750$ (Table 4.3), and, based on *T. namibiensis*, $\beta^{(max)} (E\beta^{-1})^{-1} \sim 10^{-2}$ (see above).

4.4.2.10 Observed species diversity in nature

4.4.2.10.1 Estimates of microbial diversity

How do these predictions compare with the measured prokaryotic diversity in marine systems? In a metagenome study of the Sargasso Sea, where the concentration of bacteria was indeed measured to be $\sim 10^5$ ml⁻¹ [45], it was estimated that each sample, consisting of 170–340 liters of ocean water, contained a minimum of 300 species per sample [45]. A model based on assembly depth coverage estimated between 1800 and 48,000 species [45]. A “species” in this study is defined as “a clustering of assemblies or unassembled reads more than 94% identical on the nucleotide level”, which is “roughly comparable to the 97% cutoff traditionally used for the

rRNA” [45]. In terms of rRNA diversity, in the combined study there were 1412 distinct small rRNA sequences spanning different prokaryotic phyla. Applying a similarity cutoff of 99% reduced this number to 643 strains and applying a 97% similarity cutoff reduced his number further to 148 phylotypes [45]. Given a bacterial concentration of $\sim 10^5 \text{ ml}^{-1}$ for the open sea observed in this study [45] and an offshore viral decay rate of 2 day^{-1} [23], our model predicts $N_{\text{species}} \sim 10^2 - 10^3$ species (see above). Although the predicted value for N_{species} is in agreement with the observed rRNA diversity/microdiversity and in rough agreement with the observed number of species, it is not entirely obvious how to compare N_{species} with the observed diversity. If a species is defined as a “distinguishable” genetic entity, it is not clear that the rRNA is the correct indicator for the number of species, as in principle two genomes can be “distinguishable” but have identical rRNAs. However, it is not clear that the number of “distinguishable” genetic entities is the correct measure to compare N_{species} with, since in Section 5.2.3 we will see that “strains”, which are defined to be “distinguishable” genetic elements, may not be under the sole control of a single viral species, and therefore “take up” less concentration. Thus, a certain similarity cutoff seems to be required. However, it is currently not clear how to translate this observation into an effective cutoff.

4.4.2.10.2 Viral diversity

Our model is constructed so that the number of viral species equals the number of bacterial species. Therefore we can also compare this estimate to estimate of viral diversity in the oceans. In another metagenome study, a viral metagenome was obtained from 200 liters collected from the surface seawater from Scripps Pier and a second sample was collected from Mission May, San Diego [46]. From these samples marine viruses were isolated using a combination of differential filtering and density-dependent gradient centrifugation. Several mathematical models

based on the observed number of contigs predicted between 374 and 7114 viral types. Assuming a concentration of $\sim 10^6 \text{ ml}^{-1}$ for coastal waters and a decay rate of 0.1 day^{-1} for inshore waters [23], our model predicts that $N_{\text{species}} = \sim 10^4$ to $\sim 10^5$, within rough agreement of these estimates. Here too, it is not clear what should be the correct “species” cutoff and an overestimation of diversity is not necessarily incorrect (see Section 4.4.2.10.1).

4.4.2.11 Bounds on global marine diversity

Given the expression for N_{species} and ρ_{species} we can attempt to estimate the minimum and maximum number of species in the Earth’s oceans.

Lower bound on diversity — the case of a homogeneous ocean

To estimate the minimum bound on diversity we will assume the ocean is completely homogenous, and therefore extrapolate from one region in the ocean (of the highest diversity) to the entire ocean (Fig. 4.3B). Assuming a low (onshore) decay rate for viruses of $\sim 0.1 \text{ day}^{-1}$, and a concentration of $\sim 10^6 \text{ ml}^{-1}$ we obtain an estimate of $N_{\text{species}} \sim 10^4 - \sim 10^5$. Though the sediment contains $\sim 10^3$ more cells per ml, and viral decay rates are comparable to the surface of the ocean [47,48], the particles in this region are probably not modeled well by free diffusion and therefore we will not use this region of the ocean to calculate our lower bound. The actual number of species must be higher than this since the ocean contains different regions with unique species adapted only to that region.

Upper bound on diversity — the case of a maximally heterogeneous ocean

An upper bound can be obtained by assuming that every volume V_{env} in the ocean contains a different sample of species, assuming each volume V_{env} contains the typical diversity found in the ocean (Fig. 4.3C). Thus $N_{species} \sim \rho_{species} V_{ocean} \sim 10^{-8} c_{bact}^{tot} V_{ocean}$. Given that $V_{ocean} \sim 10^{24}$ ml [44], and for the open sea $c_{bact}^{tot} \sim 10^5 \text{ ml}^{-1}$, the maximum number of species in the ocean would be $N_{species} \sim 10^{-8} \cdot 10^5 \cdot 10^{24} = 10^{21}$. Note that $\rho_{species}$ scales as $c_{bact}^{tot} (r_{min}/r_{max})^3$. Thus, the upper bound on the total diversity in the oceans essentially depends only on the “carrying capacity” of the ocean and on the size of the smallest and largest viable bacteria.

Thus the total number of actively replicating prokaryotic cells with a *distinguishable* genome in Earth’s oceans is predicted to lie somewhere between 10^4 to 10^{21} . Although 10^{21} is a large number, it is exceedingly smaller than the total number of possible bacterial strains, and 7 orders of magnitude lower than the total number of bacterial cells in the ocean, estimated to be $\sim 10^{29}$ [44]. This upper bound is of course a gross overestimate since adjacent volumes of water exchange cells constantly, and therefore the overlap in species between adjacent “volumes of diversity” will be very large. Note that using this approach one could obtain much tighter bounds on species diversity for smaller volume ecosystems such as lakes.

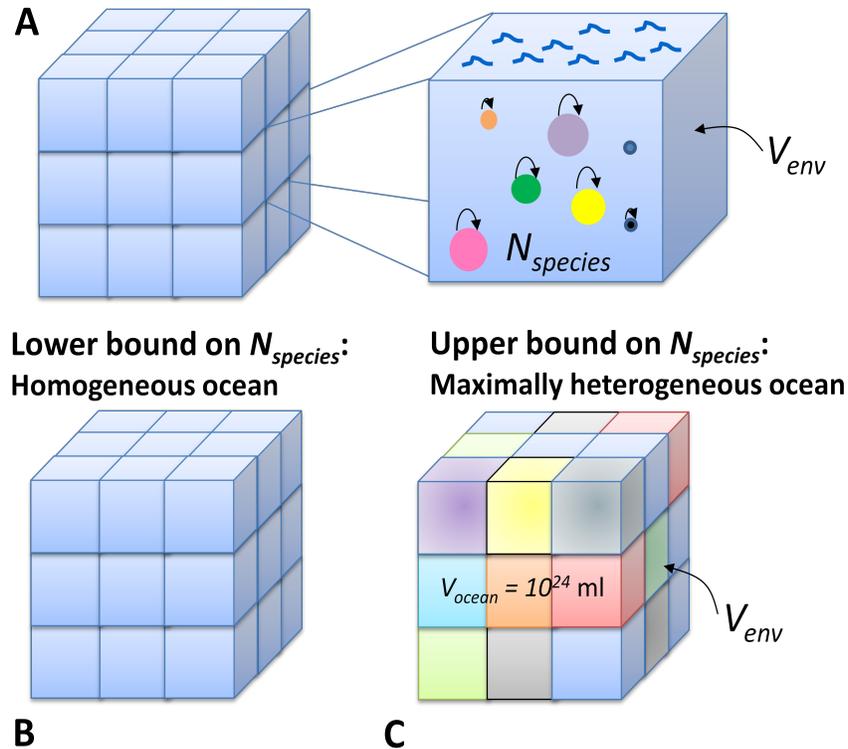


Figure 4.3. Illustration of lower and upper bounds on $N_{species}$. **A.** Each volume of diversity, V_{env} , contains $N_{species}$ species given by Eq. 16B. The concentration of each species is controlled by its lytic virus. **B.** In the case of a homogenous ocean scenario, all volumes of diversity contain the same species, resulting in a lower bound on the total diversity in the oceans **C.** In the case of a maximally heterogeneous ocean, every volume of diversity contains a different set of species, resulting in an upper bound on the total diversity in the oceans.

Current observed diversity in public databases

How do these values compare with the current estimates of diversity? If one uses the small subunit rRNA gene as a proxy for genetic diversity, then one can compare our range estimates for the total number of species in the oceans' water column with the total number of rRNA sequences that are >1% divergent. The Silva SSU Ref NR 106 [49] released in April 2011 is a non-redundant SSU rRNA database with an operational taxonomical unit (OTU) cutoff of 1%. According to this database there are 210^5 bacterial and archeal non-redundant SSU rRNA sequences. While an OTU of 1% will give us a lower bound on the number of "distinguishable"

genomes (see Chapter 5), taken at face value, this comparison suggests that the ocean appears to be more homogenous than heterogeneous.

4.4.2.12 Factors determining species richness

Nutrient availability

Eq. 16B leads to some interesting predictions regarding the diversity of species in different environments. An eutrophic environment for example, which can sustain a higher concentration of bacteria (assuming total bacterial concentration is not determined by grazers [17]), is predicted via Eq. 16B to harbor a larger number of species. Increasing the total concentration of bacteria by a factor of ten will lead via Eq. 16B to ten times the number of species and, as will be discussed below, also ten times the concentration of viruses. In fact, the increase in species diversity can even be significantly higher than a factor of ten since $N_{species}$ is proportional to $r_{min}^3 r_{max}$, and cell size often increases with growth rate, which in turn increases with nutrient availability. Thus, an increase in nutrient availability may lead to an explosion in species diversity (and also, possibility a significant increase in the VBR, discussed below). Conversely, oligotrophic environments, where the concentration of bacteria can be lower, are predicted to harbor fewer species. Thus, a direct prediction of our model is that eutrophic environments harbor a larger diversity of species compared with oligotrophic environments, given similar temperature conditions and similar viral decay rates.

Viral decay rate

Another interesting parameter that comes into play is the virus decay rate. The more viruses are allowed to thrive (i.e., decay more slowly, thus having a lower $\gamma_{viral\ decay}$), the lower the concentration of any phage-host system will be (Eq. 8), thus requiring more species to reach a

given carrying capacity (Eq. 16B). Thus viruses directly contribute to bacterial species diversity, and so “what’s good for the virus is good for the bacterium”. Since the number of bacterial species must equal the number of viral species, generating bacterial diversity also means generating viral diversity. In Section 4.5 we propose closed and open mesocosm experiments, where we show that in both cases, increasing the viral decay rate should lead to a decrease in the number of species.

The reciprocal relationship between bacterial diversity and viruses has been proposed in the past [9,10], however here we have expanded this concept by combining several ideas: (a) we have taken into account the biophysical nature of phage-host interaction, which allowed us to describe quantities in terms of physical parameters such as temperature, viscosity, the size of the virus, and the size of the host. (b) We have taken into account the observed correlation between burst size and the physical dimensions of the host and its virus. Finally, (c) we have introduced the notion of the “density of bacterial species” that was used in a statistical fashion to make the transition from a single isolated phage-host system to a community distribution. These concepts have led us to derive a realistic prediction for the number of species given in terms of physical measurable parameters and also define a physical volume associated with the predicted diversity.

Temperature

Eq. 16B predicts that given the same carrying capacity, warmer environments will contain more species. Overall this effect is not very large however. The difference between an environment just above freezing and 40°C will lead to only a 15% increase in species diversity ($(273 + 40)/273=1.15$), unless temperature will have an effect on r_{min} and r_{max} through its effect on growth rates (see above). The quantitative predictions of Eq. 16B may however be biased for

extreme temperature environments since the selection pressure in such environments may skew the bacterial radius density function, violating our assumption of uniformity.

Extreme bacteria

We have already noted that above that $N_{species} \propto r_{min}^3 r_{max}$. Thus, halving the size of the minimum viable bacterium would lead to decreasing the total diversity in the environment by about one order of magnitude (2^3), due to the great abundance of small bacteria (thus reaching the carrying capacity more quickly). On the other hand, doubling the size of a largest bacterium would only lead to a modest doubling of the total bacterial diversity in the given environment since we are adding rare species with low concentrations that do not contribute much to the total concentration, thus necessitating more species to reach a given carrying capacity. One possible way r_{min} and r_{max} may be influenced is through nutrient availability, as discussed above.

4.4.2.13 The total concentration of viruses and the VBR in the environment

Total concentration viruses

In a similar fashion we can calculate the predicted total concentration of viruses in the environment. Since the average concentration of viral species is $Ec_{virus}^{(i)}$, the total concentration of viruses is simply $c_{virus}^{tot}(r) = N_{species} Ec_{virus}^{(i)}$. Inserting Eq. 5 we find that

$$(17) \quad c_{virus}^{tot}(r) = N_{species} Ec_{virus}^{(i)} = \frac{3}{2} N_{species} \frac{\eta}{k_B T} ER_{virus}^{(i)} (E\alpha^{(i)} - E\gamma_{non-viral}^{(i)}) E\left(\frac{1}{R_{bact}^{(i)}}\right).$$

Assuming once again a uniform distribution for the bacteria (Eq. 11) we find that

$$(18) \quad E\left(\frac{1}{R_{bact}^{(i)}}\right) = \int_{r_{\min}}^{r_{\max}} r^{-1} f_R(r) dr \cong r_{\max}^{-1} \int_{r_{\min}}^{r_{\max}} r^{-1} dr = r_{\max}^{-1} \ln\left(\frac{r_{\max}}{r_{\min}}\right).$$

thus

$$(19) \quad c_{virus}^{tot} = \frac{3}{2} N_{species} \frac{\eta}{k_B T} ER_{virus}^{(i)} (E\alpha^{(i)} - E\gamma_{non-viral}^{(i)}) r_{\max}^{-1} \ln\left(\frac{r_{\max}}{r_{\min}}\right).$$

The concentration of viruses in the ocean

For an offshore ecosystem with $\gamma_{viral\ decay} \sim 2 \text{ day}^{-1}$ and $c_{bact}^{tot} \sim 10^5 \text{ ml}^{-1}$ we previously found that $N_{species} \sim 10^2$ to $\sim 10^3$ species. To calculate the total concentration of viruses in the environment we use the above estimates and the remaining parameters from Table 4.3 and find that $c_{virus}^{tot} = \sim 2 \cdot 10^5$ to $\sim 2 \cdot 10^6 \text{ ml}^{-1}$, or $\sim 10^5$ to $\sim 10^6 \text{ ml}^{-1}$. This prediction falls exactly in the range of observed concentrations: virus concentrations in offshore surface waters are typically in the range of 10^5 – 10^6 ml^{-1} [2].

The VBR in a given environment

With the total concentration of bacteria at hand we can now calculate using Eq. 16A and Eq. 19 the VBR in the environment:

$$(20) \quad VBR = \frac{c_{virus}^{tot}}{c_{bact}^{tot}} = 3 \frac{(E\alpha^{(i)} - E\gamma_{non-viral}^{(i)}) \left(\frac{r_{\min}}{R_{virus}}\right)^3 \ln\left(\frac{r_{\max}}{r_{\min}}\right)}{E\beta^{-1} E\gamma_{viral\ decay}}$$

where we have assumed that $ER_{virus}^4 \approx (ER_{virus}^{(i)})^4 \approx R_{virus}$. Given the parameters in Table 4.3 for a typical marine system, with an offshore viral decay rate of 2 day^{-1} [23], Eq. 20 predicts that VBR

~ 2 to ~ 20 , precisely as observed for typical marine systems [1]. The key observation predicted by this formula is that VBR is essentially controlled by the following parameters: (1) the net average growth of bacteria (growth minus predation), (2) the decay rate of viruses, and (3) the minimum viable bacteria (which may be related to nutrient availability). β and R_{virus} have a relatively narrow distribution and the effect of r_{max} is subdued due to the log. Thus, this basic equation can be used to predict both qualitatively and quantitatively the VBR in any aqueous environment.

Examples for environmental VBRs

VBR in nutrient-rich versus nutrient-poor environments

It has been observed that the VBR is higher for nutrient-rich, productive environments compared with nutrient-poor environments [19]. This has been attributed to the fact that “bacterioplankton host populations produce greater numbers of viruses under environmental conditions favoring fast growth and high productivity” [19]. Eq. 5 indeed predicts that — all things being equal — the higher the average growth rate of the bacterium, the higher the concentration of viruses will be, and consequently the higher the VBR. This prediction is also apparent from Eq. 20 for the VBR, where it is shown that the VBR is directly proportional to the average net growth rate of bacteria in the environment. In addition, cell size often increases with growth rate, which increases with the availability of nutrients. Since based in Eq. 20 the VBR is proportional to r_{min}^3 , even a modest increase in r_{min} would lead to a significant increase in the VBR.

VBR in oceans versus lakes

In a recent study it has been shown that the VBR is higher in marine systems than in freshwater systems [1,50]. In the surface waters of the Pacific and Arctic oceans for example, the VBRs are ~ 40 and ~ 10 respectively, while in lakes the average VBR was measured to be less than 5 [50]. Though the reasons for these differences are unknown, it has been suggested that this is related to possible higher loss rates of virus particles in freshwater environments that may be related to the presence of clays and chemicals from the terrestrial environment, which are known to contribute to viral decay [1,50]. This hypothesis is consistent with the prediction of Eq. 20, namely that the VBR should decrease with increased viral decay rate.

4.4.2.14 Total prokaryotic biomass concentration

The predicted slope of -1 (Eq. 14) for the size spectra of bacteria suggests that on average there is a tendency toward a uniform distribution of mass among size classes in aquatic ecosystems [41,42]. This result follows from Eq. 13: let $m_{bact}(r) = \frac{4}{3}\pi r^3 \rho_{cell}$ be the mass of a bacterium of radius r having a cellular mass density of ρ_{cell} . The total mass concentration per cell radius, given Eq. 12 for $\langle \rho_{environment}(r) \rangle$, scales as $M_{bact}(r) = \langle \rho_{environment}(r) \rangle m_{bact}(r) \sim r^{-1}$. Therefore the total mass concentration between radius r_1 and r_2 is given by

$$(21) \quad \text{Mass}(r_1 < r < r_2) = \int_{r_1}^{r_2} M_{bact}(r) dr \propto \int_{r_1}^{r_2} r^{-1} dr = \ln\left(\frac{r_2}{r_1}\right).$$

Thus the total mass of prokaryotes between r and $10r$ equals the total mass of prokaryotes between $10r$ and $100r$ and so on. Integrating over all viable bacterial radii (using Eq. 12) we obtain the total mass of prokaryotes per unit volume

$$(22A) \quad M_{bact}^{tot} = \int_{r_{min}}^{r_{max}} M_{bact}(r) dr \cong 2\pi\rho_{cell} N_{species} \frac{\eta}{k_B T} E\beta^{-1} ER_{virus}^4 E\gamma_{viral\ decay} r_{max}^{-1} \ln\left(\frac{r_{max}}{r_{min}}\right).$$

where we have assumed again that $r_{min} \ll r_{max}$. Combining Eq. 16A and Eq. 22A we find that

$$(22B) \quad M_{bact}^{tot} \cong c_{bact}^{tot} m_{min} \ln\left(\frac{m_{max}}{m_{min}}\right)$$

where m_{min} and m_{max} are the mass of minimum and maximum viable bacteria. Thus Eq. 22A predicts the total prokaryotes mass concentration in the ocean in terms of basic parameters such as: environmental parameters (viscosity and temperature of the water), viral parameters (average radius, average decay rate, and volume fraction within host cell) and host parameters (mass — or water — density, number of bacterial species, and minimum and maximum radii of viable bacteria). Assuming $\rho_{cell} \approx 1$ g/ml then $m_{min} = \frac{4}{3}\pi\rho_{cell}r_{min}^3 = 4.2 \cdot 10^{-15}$ g, and the total mass density of prokaryotes is $M_{bact}^{tot} = 10$ mg/m³ (including cytoplasmic water). This mass can be compared with the following simple order-of-magnitude estimate. The typical radii of bacteria in the open ocean is 0.1–0.2 μm (Table 4.3). The mass of such a bacterium is given by $m_{bact}(r \approx 0.2 \mu\text{m}) = \frac{4}{3}\pi r^3 \rho_{cell} \sim 10^{-11}$ mg. Assuming that all 10^5 cells per ml have a radius of 0.2 μm , then the total mass of cells in 1 m³ would be 10^{-11} mg \times (10^{11} cells per m³) = 1 mg. Thus, most of the mass contribution, according to Eq. 22, comes from the larger, rarer bacteria, and not the more abundant small bacteria. This can also be appreciated by noting that, whereas the total number of

prokaryotes up to radius r scales as $\sim r^{-3}$ (Eq. 14), thus decaying very fast, the total mass of prokaryotes up to radius r scales much more slowly as $\sim \ln(r)$.

4.5 Conclusions and further experiments

We developed a simple biophysical model that describes the interaction of an isolated phage-host system leading us to conclude that the single most important parameter determining the abundance of bacteria in the ocean is their size. We then extended our model to an ecological scale by making the assumption that the *a priori* distribution of bacterial radii in the environment is uniform, i.e., there is no selection pressure shaping this distribution. Given these basic ingredients we derive a model that makes reasonable predictions for the size spectra of bacteria, the VBR and the number of bacterial/viral species in the environment that largely seem to be consistent with observations. To further test our model we propose the following experiments:

4.5.1 *In vitro* investigation of phage-host systems

By choosing a particular phage-host system such as T4 and *E. coli*, one can analyze infected cultures *in vitro* as different model parameters are perturbed. To prevent total lysis of the hosts one should include an ecological factor leading to virus degradation (perhaps by introducing some organic substance that is innocuous to bacteria but would inactivate virions). Alternatively, a chemostat may be sufficient. Once a sustainable infection can be established, one can vary parameters such as growth rate, viral decay rate, temperature, and viscosity, thus testing predictions of Eqs. 5–7. Other phage-host systems can be chosen as well. Of particular interest are hosts of significantly different size. Alternatively, the growth medium of *E. coli* can be changed, thus affecting its size. The timescale of this experiment needs to be shorter than the timescale for *E. coli* and/or T4 to start evolving in a way that affects their interaction (see Section

5.4). Since the small oscillations of this system around the fixed point occur with a period of $\tau \sim [(\alpha - \gamma_{non-viral})\gamma_{virus\ decay}]^{-\frac{1}{2}}$ (assuming the latent period=0), the viral decay rate needs to be high enough to prevent large fluctuations from steady state. In addition, a high viral decay rate will prevent the fixed point bacterial concentration from becoming too low, circumventing possible bottle neck affects that can lead to *in vitro* evolution.

4.5.2 Investigating phage-host systems in nature

Our model makes many assumptions regarding viruses and their hosts. For example, we assume that bacteria are in a state of exponential growth, that radii are uniformly distributed, that the virus-host systems are independent and so on. It is therefore crucial to test our model in natural environments. One way to do this is to analyze culturable phage-host systems directly in nature, where hosts are selected to cover a wide spectrum of sizes. Of particular interest are phage-host systems involving giant bacteria. Giant bacteria are predicted by our model to have a very low density (Eq. 7), even when correcting for massive cell inclusions. However, viruses of giant bacteria are predicted by the model to be quite numerous (Eq. 5A), with as many as hundreds of virions per ml of water (see above). By designing primers against phages of giant bacteria and using quantitative assays such as quantitative PCR and/or digital PCR, one can test a direct and extreme prediction of this model, namely that phages of giant bacteria are numerous in nature (with their density predicted by Eq. 5A) and should be detected even in the absence of the host. The absence of the host can be confirmed with SSU rRNA sequencing. If the genome of a lytic phage infecting the giant bacteria cannot be obtained and the host has been sequenced, CRISPR sequences can be crossed with a viral metagenome from the environment of interest to detect phage genes for primer design.

4.5.3 Closed mesocosm experiments

Decay rate perturbation

Closed mesocosm experiments can be used to test total bacterial and viral abundances when perturbing parameters such as viral decay rate, growth rate (through nutrient availability), temperature, and viscosity. These types of experiments can be used to test the predictions of total bacterial concentration and total viral concentration (Eqs. 16 and 19). Note that in closed system experiments, the total number of species $N_{species}$ cannot increase since we cannot create species de novo. As a control, $N_{species}$ can be measured under every perturbation via a SSU rRNA library to check this assumption.

One can also test the ratio between quantities. For example, if the decay rate is changed without affecting bacterial growth (e.g., by introducing some organic chemical that decreases viral lifetime but does not affect bacterial growth or by filtering out UV bands that damage phages, assuming growth rate is not affected) then Eq. 16A makes the simple prediction that

$$(23A) \quad \frac{c_{bact}^{tot}(UV)}{c_{bact}^{tot}(no\ UV)} = \frac{N_{species}(UV)}{N_{species}(no\ UV)} \frac{E\gamma_{viral\ decay}(UV)}{E\gamma_{viral\ decay}(no\ UV)}$$

where for concreteness we designate high decay rate as UV and low decay rate as no UV. If we constrain that $N_{species} = \text{const}$ then we obtain the result that $c_{bact}^{tot}(UV) > c_{bact}^{tot}(no\ UV)$. Although this result on the one hand makes intuitive sense (viruses that degrading faster lead to more bacteria) it is counterintuitive in the sense that if the environment has the capacity to sustain a higher concentration of bacteria, then why wasn't this capacity utilized by species $i = N_{species} + 1$?

Thus a more logical alternative would be that as the viral decay rate increases, the number of species *decreases*, as some species die, allowing other species to increase in concentration (via Eq. 7) such that $c_{bact}^{tot} = \text{const}$. Thus, increasing the rate of virus degradation leads to a *decrease* in the diversity of the mesocosm by a factor of $E\gamma_{viral\ decay}(\text{no UV})/E\gamma_{viral\ decay}(\text{UV})$. This solution is pleasing in the sense that there are no undetermined degrees of freedom left. In addition, from Eq. 19 we predict that

$$(23B) \quad \frac{c_{virus}^{tot}(\text{UV})}{c_{virus}^{tot}(\text{no UV})} = \frac{N_{species}(\text{UV})}{N_{species}(\text{no UV})}.$$

If species die in the mesocosm, then when increasing the decay rate of viruses the total concentration of viruses should *decrease*. The VBR is predicted to decrease when increasing the decay rate of viruses:

$$\frac{VBR(\text{UV})}{VBR(\text{no UV})} = \frac{E\gamma_{viral\ decay}(\text{no UV})}{E\gamma_{viral\ decay}(\text{UV})} < 1.$$

Nutrient perturbation

Another critical test of the model would be an enrichment experiment on a nutrient-limited closed mesocosm. Based on Eq. 16A we have

$$(24) \quad \frac{c_{bact}^{tot}(\text{enriched})}{c_{bact}^{tot}(\text{poor})} = \frac{N_{species}(\text{enriched})}{N_{species}(\text{poor})} \frac{r_{\max}(\text{poor})}{r_{\max}(\text{enriched})} \left(\frac{r_{\min}(\text{poor})}{r_{\min}(\text{enriched})} \right)^3.$$

When adding nutrients to our mesocosm we do not expect species to die since there are more resources present in the environment. However, since $N_{species}$ also cannot grow (since there is no available reservoir for species) we conclude that $N_{species} = \text{const}$. Since bacterial size is expected to increase with nutrients, we anticipate that the total concentration of bacteria upon enrichment will *decrease*. The explanation for this paradoxical behavior is apparent from Eq. 3: as nutrients are added and the growth rate of bacteria increases, so does their radius (and thus burst size). Thus the viral production term in Eq. 3 (first term) increases, necessitating the bacterial density to decrease owing to a constant viral decay rate (second term in Eq. 3). We will see that in an open mesocosm experiment exactly the opposite response is anticipated.

Spiking approach

In another approach, a non-indigenous culturable host can be “released” into the mesocosm with its lytic virus allowing one to track host and virus concentrations upon various perturbations. The concentration of the bacterium can be monitored by a quantitative PCR (qPCR) assay targeting the SSU rRNA gene of the organism. The virus concentration can also be monitored via qPCR if there is genetic information on the virus. The advantage of this method is that one can use molecular techniques to precisely gauge the abundance of the host and its virus (instead of measuring pfus or cfus). This approach assumes however that in the time course of the experiment, primer binding sites have not mutated in the evolving viral quasispecies. This assumption can be checked by attempting to amplify plaques with the viral primers and analyzing the success rate statistically.

4.5.4 Open mesocosm experiments

Decay rate perturbation

In open mesocosm experiments the number of species is not constrained as new species can diffuse or swim into our mesocosm and existing species can diffuse or swim out. Repeating the perturbation experiment for the viral decay rate in an open mesocosm system we would predict once more Eq. 23A and 23B and, as before, there is an undetermined degree of freedom. Increasing the viral degradation rate should lead an increase in the concentration of each bacterial species (Eq. 7). However, the total concentration of bacteria should not be allowed to increase upon perturbation, since if the mesocosm could have sustained a higher concentration of bacteria, some new species would have taken advantage of this and stayed in this volume by means of chemotaxis. Thus, we conclude that upon an increase in viral decay rate the number of species will decrease, as some species will die allowing other species to increase in concentration to sustain a constant total concentration of bacteria. Thus, either in an open or closed mesocosm, it appears that increasing the decay rate of viruses should lead to a decrease in species diversity.

Note that when testing predictions of diversity, it is not sufficient to change the viral load, as this will only affect the transient response of the system. In order to observe a steady-state effect one should change the fundamental parameters governing the system, such as the viral decay rate.

Nutrient perturbation

Repeating the enrichment experiment in a nutrient limited open mesocosm Eq. 24 still holds. Here again, the concentration of any given species will decrease due to the increase in radii (Eq. 7), thus there is room for more species. Since new species entering this region can stay in the

region by means of chemotaxis, we expect the total number of species to significantly increase and with the total number of bacteria either constant or increasing.

Size spectra perturbation

Our model predicts that the size spectra of bacteria is the result of viral predation and that the slope of the resulting power law should be independent of, for example, nutrient availability, viral decay rate, temperature, medium viscosity, and so on. These predictions can be directly tested in a mesocosm, similar to the IronEx II perturbation experiments. Furthermore, removal of the lytic viral fraction should result in a certain decrease in the slope of the spectrum (more positive), with the new slope being determined presumably by nutrient availability.

Systematic mapping of prokaryotic species diversity in different aquatic zones

One of the interesting predictions of the model deals with species diversity (Eq. 16B). Species diversity changes in a very predictable manner dictated by the total bacterial concentration, viral decay rate, temperature, and so on. By systemically sampling different environments on Earth (e.g., eutrophic versus oligotrophic zones, photic versus the aphotic zones, epipelagic zones in tropical versus polar regions, marine versus freshwater ecosystems, etc.) and measuring the concentration of bacteria, the temperature, the viral decay rate, and the number of species (via SSU rRNA libraries) one can directly test the predicted number of species (Eq. 16B).

4.5.5 Investigate host range in nature

To test our assumption that a host in a given region is infected with a single viral species one can isolate different phages infecting the same host species using conventional plaque assays. Phages that appear to be morphologically different via EM can be sequenced and their genomes

$$(A3) \quad \text{viral rate equations} \rightarrow \begin{cases} -\gamma_1 + b_1 k_{11} B_{11} + \dots + b_1 k_{n1} B_{n1} = 0 \\ \vdots \\ -\gamma_n + b_n k_{1n} B_{1n} + \dots + b_n k_{nn} B_{nn} = 0 \end{cases}$$

Now let's assume we introduce bacterium $n+1$. If we write the rate equation for this bacterium, then at steady-state we will obtain the $n+1$ equation for (A2), however there are only n variables V_i $i=1..n$. The system is therefore overdetermined and therefore some species will become extinct in the transient solution. The same rational applies if we add the $n+1$ viral species. In this case we will have $n+1$ steady-state equations for the viruses (A3), yet we have only n variables for B_i $i=1..n$, again obtaining an overdetermined set of equations. If we remove one bacterial species or one viral species, we again find the same situation: the reciprocal variable will be overdetermined. Thus, the only solution which is not overdetermined is if we have n bacterial species being infected by n viral species.

For the special case of one virus species with a wide host range infecting two bacterial species the proof is the following: Let's imagine we have a closed system containing two different *distinguishable* hosts of concentration $c_{bacteria}^{(1)}$ and $c_{bacteria}^{(2)}$, both infected with the same virus of concentration c_{virus} . According to Eq. A2 and Eq. A3 the set of differential equations governing the interaction of these three species is

$$(A4) \quad \begin{cases} \frac{dc_{virus}}{dt} \cong b^{(1)}k^{(1)}c_{bacteria}^{(1)}c_{virus} + b^{(2)}k^{(2)}c_{bacteria}^{(2)}c_{virus} - \gamma_{virus \text{ decay}}c_{virus} \\ \frac{dc_{bacteria}^{(1)}}{dt} = \tilde{\alpha}^{(1)}c_{bacteria}^{(1)} - k^{(1)}c_{virus}c_{bacteria}^{(1)} \\ \frac{dc_{bacteria}^{(2)}}{dt} = \tilde{\alpha}^{(2)}c_{bacteria}^{(2)} - k^{(2)}c_{virus}c_{bacteria}^{(2)} \end{cases}$$

where $k^{(i)} = 4\pi D_{virus} R_{bact}^{(i)}$. At steady-state, this system is, however, overdetermined since the solutions $c_{virus} = \tilde{\alpha}^{(1)}/k^{(1)}$ and $c_{virus} = \tilde{\alpha}^{(2)}/k^{(2)}$ cannot be mutually satisfied. The only consistent steady-state solutions would be $c_{bacteria}^{(1)} \equiv 0$ or $c_{bacteria}^{(2)} \equiv 0$ or $c_{bacteria}^{(1)} = c_{bacteria}^{(2)} \equiv 0$, unless the two hosts have precisely the same radius and growth rate. Thus, only bacteria with the same radius and same growth rate can be infected with the same virus and sustain a population. The slightest difference and, with enough time, one species will be driven to extinction.

Similarly, if we have two viral species with a specific host range, infecting the same bacteria, we would again run into an overdetermined system of equations:

$$(A5) \quad \begin{cases} \frac{dc_{virus}^{(1)}}{dt} \cong b^{(1)}k^{(1)}c_{bacteria}c_{virus}^{(1)} - \gamma_{virus \text{ decay}}^{(1)}c_{virus}^{(1)} \\ \frac{dc_{virus}^{(2)}}{dt} \cong b^{(2)}k^{(2)}c_{bacteria}c_{virus}^{(2)} - \gamma_{virus \text{ decay}}^{(2)}c_{virus}^{(2)} \\ \frac{dc_{bacteria}}{dt} \cong \alpha_{bacteria}c_{virus}^{(1)} - k^{(1)}c_{bacteria}c_{virus}^{(1)} - k^{(2)}c_{bacteria}c_{virus}^{(2)} \end{cases} .$$

Thus at steady-state we would find that from the first equation $c_{bacteria} = \frac{\gamma_{virus\ decay}^{(1)}}{b^{(1)}k^{(1)}}$ while from the

second equation $c_{bacteria} = \frac{\gamma_{virus\ decay}^{(2)}}{b^{(2)}k^{(2)}}$, thus the system is overdetermined. It is intuitively clear

that two viruses cannot control the same species.

4.7 Power law derivation

4.7.1 The distribution of bacteria in the environment

According to Eq. 9, the concentration of bacteria of a given radius r per radius in a given realization of an environment containing $N_{species}$ bacterial species is given:

$$(B1) \quad \rho_{environment}(r) = c_{bacteria}(r) \sum_{i=1}^{N_{species}} \delta(r - R_{bacteria}^{(i)}).$$

where $R_{bacteria}^{(i)}$ are $N_{species}$ i.i.d. random variables drawn from a distribution $f_R(r)$ and where $\delta(r)$ is the Dirac delta function. To obtain the ensemble average of $\rho_{environment}(r)$, averaging over many realizations of a given environment one should calculate the expectation value of $\rho_{environment}(r)$ with respect to the N random variables $R_{bacteria}^{(i)}$:

$$\langle \rho_{environment}(r) \rangle = c_{bacteria}(r) \sum_{i=1}^{N_{species}} \int_{R^{(i)}} dR_{bacteria}^{(1)} \dots dR_{bacteria}^{(N)} f_{R^{(1)}, \dots, R^{(N_{species})}}(R_{bacteria}^{(1)}, \dots, R_{bacteria}^{(N_{species})}) \delta(r - R_{bacteria}^{(i)}).$$

Since $R_{bacteria}^{(i)}$ are i.i.d. we have

$$\begin{aligned} \langle \rho_{environment}(r) \rangle &= c_{bacteria}(r) \sum_{i=1}^{N_{species}} \int_{R_{bacteria}^{(i)}} dR_{bacteria}^{(i)} \left[f_{R^{(1)}}(R_{bacteria}^{(1)}) \dots f_{R^{(N_{species})}}(R_{bacteria}^{(N_{species})}) \right] \delta(r - R_{bacteria}^{(i)}) = \\ &= c_{bacteria}(r) \sum_{i=1}^{N_{species}} \int_{R_{bacteria}^{(i)}} dR_{bacteria}^{(i)} f_R(R_{bacteria}^{(i)}) \delta(r - R_{bacteria}^{(i)}) = \\ &= c_{bacteria}(r) \sum_{i=1}^{N_{species}} f_{R^{(i)}}(r) = N \cdot c_{bacteria}(r) \cdot f_R(r). \end{aligned}$$

Thus the average distribution of bacterium sizes in a given environment is given by Eq. B2:

$$(B2) \quad \langle \rho_{environment}(r) \rangle = N_{species} \cdot c_{bacteria}(r) \cdot f_R(r).$$

To test this equation we performed the following Monte Carlo simulation: We draw $N_{species}=100$ radii R_i ($i=1..100$) for bacteria according to a specified probability density function (pdf) $f_R(r)$. The concentration of each bacterial species as a function of its radius is given by the hypothetical distribution $c_{bacterium}(r)=r$. We then construct an empirical discrete distribution function for $\rho_{environment}(r)$ such that $\rho_{environment}(r=r_i)=c_{bacterium}(r_i)=r_i$ for $i=1..100$. Finally we average this distribution over many Monte Carlo simulations ($M=10000$), simulating many realizations of this environment to obtain $\langle \rho_{environment}(r) \rangle$. The ensemble average that we compute, $\langle \rho_{environment}(r) \rangle$, should converge according to Eq. B2 to $\langle \rho_{environment}(r) \rangle = N_{species} \cdot r \cdot f_R(r)$. Examples of two pdfs for $f_R(r)$ are shown in Fig. 4.4.

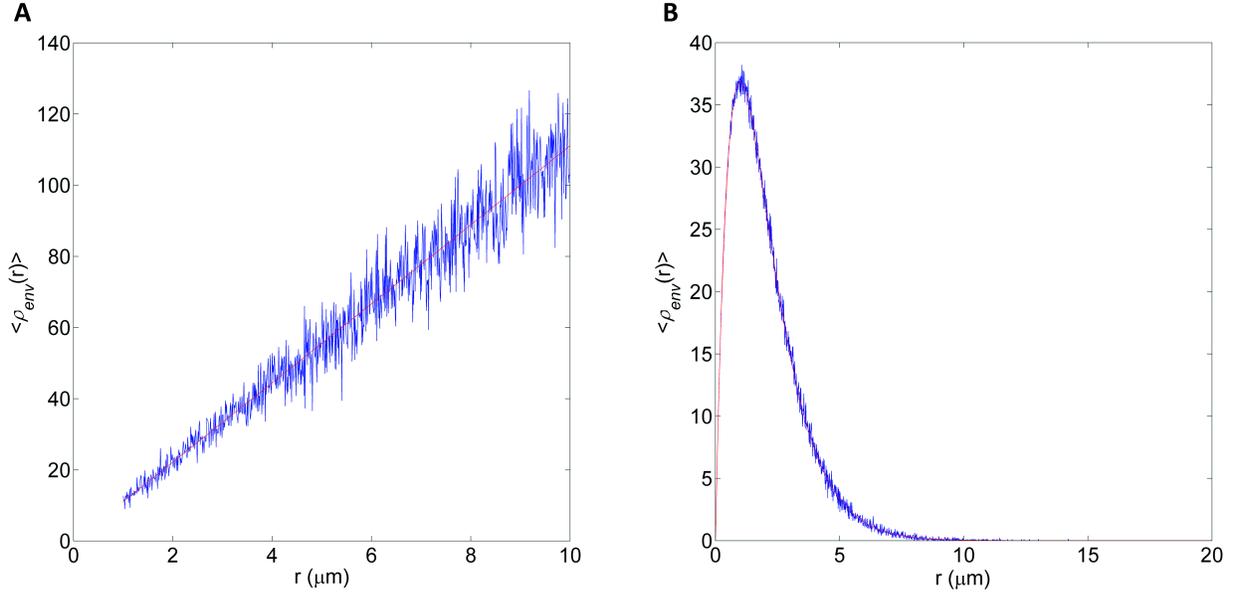


Figure 4.4. Monte Carlo simulation of a hypothetical distribution of bacteria in a given environment. In each of $M=10^4$ Monte Carlo iterations, $N_{species}=100$ bacterial radii were drawn such that in **(A)** R was exponentially distributed with rate $\lambda=1$ and in **(B)** R was uniformly distributed between $r_{min}=1$ and $r_{max}=10$. The empirical distribution of bacteria $\rho_{environment}(r)$ in both cases was calculated assuming the hypothetical relation $c_{bacterium}(r)=r$. That is, for each radius R_i drawn in a given iteration we update the empirical distribution function in the following way: $\rho_{environment}(r=r_i)=c_{bacterium}(r_i)=r_i$ (see Eq. B1). Then finally we average $M=10^4$ calculated empirical distribution functions $\rho_{environment}(r)$ to obtain the ensemble average of $\rho_{environment}(r)$, which we denote by $\langle \rho_{environment}(r) \rangle$. Based on Eq. B2 we expect that for **(A)** $\langle \rho_{environment}(r) \rangle = \lambda \cdot N_{species} \cdot r \cdot e^{-\lambda r}$ and for **(B)** $\langle \rho_{environment}(r) \rangle = N_{species} \cdot r / (r_{max} - r_{min})$. The figure demonstrates that in both cases the calculated value for $\langle \rho_{environment}(r) \rangle$ based on the Monte Carlo simulation (blue) converged precisely to the theoretical prediction (red) describe above.

4.7.2 The predicted size spectra of bacteria in the environment

In the main text we derived the probability that a bacterium of random volume V , is greater than or equal to a given volume, v (Eq. 14). Here we test Eq. 14 in the following Monte Carlo simulation: We assumed that $c_{bacterium}(r)=r^{-4}$, $R \sim U(r_{min}, r_{max})$ and we computed $\langle \rho_{environment}(r) \rangle$ as explained above (see Fig. 4.4). We then normalized the computed function $\langle \rho_{environment}(r) \rangle$ to

obtain the empirical pdf $f_\rho(r)$ and calculated $\text{Prob}(V \geq v)$. The Monte Carlo simulations should

converge to $\langle \rho_{environment}(r) \rangle = N_{species} \cdot r^{-4} / (r_{max} - r_{min})$ (following Eq. B2) and $\text{Prob}(V \geq v)$ should

converge to $\text{Prob}(V \geq v) = \left(\frac{r_{min}}{r_{max}} \right)^3 \left[\left(r_{max}/r \right)^3 - 1 \right]$ (Eq. 14). Results are shown in Fig. 4.5 and

demonstrate that the simulation converged precisely to the theoretical predictions.

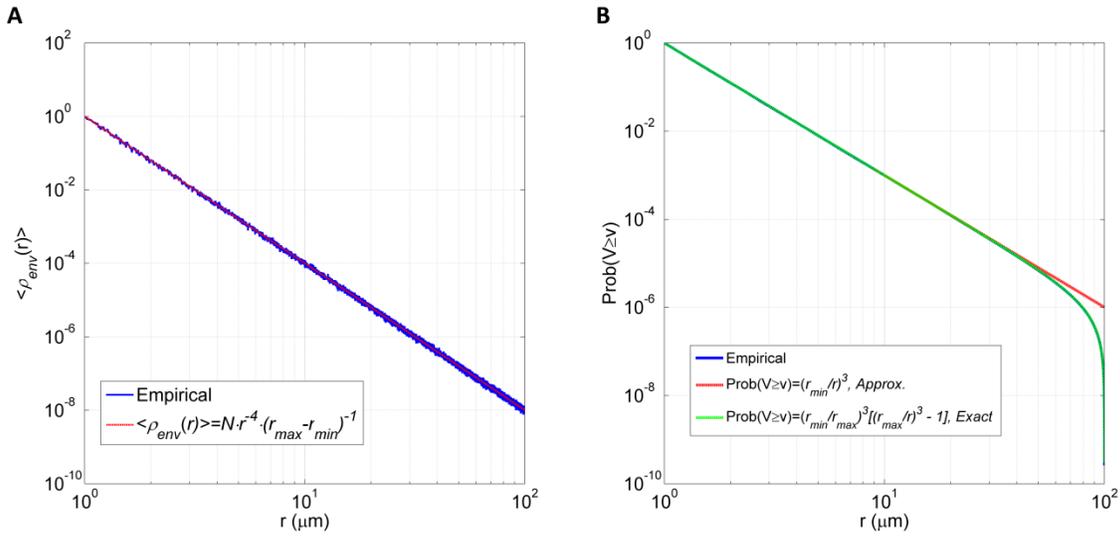


Figure 4.5. Monte Carlo simulation of the predicted size spectra of bacteria in a given environment. Monte Carlo simulation assuming the predicted concentration of a bacterium with radius r obeys $c_{bacterium}(r) = r^{-4}$ and that bacteria radii are drawn from the uniform distribution. **(A)** Theoretical prediction (red) for the ensemble average of the distribution of bacteria in the given environment $\langle \rho_{environment}(r) \rangle = N_{species} \cdot r^{-4} / (r_{max} - r_{min})$ (Eq. B2) versus Monte Carlo simulation (blue) with $M=10^4$ iterations (see caption of Fig. 4.4 for simulation details). **(B)** The numerical estimate of $\langle \rho_{environment}(r) \rangle$ was normalized to obtain an empirical pdf, which was used to calculate $\text{Prob}(V \geq v)$. The result of the Monte Carlo simulation (blue) was compared with the theoretical prediction for $\text{Prob}(V \geq v)$ (Eq. B1; red in A, green in B). The figure demonstrates that the numerical simulation converged precisely to the theoretical prediction for both (A) and (B).

4.8 References

1. Suttle C (2007) Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol* 5: 801-812.
2. Weinbauer M (2004) Ecology of prokaryotic viruses. *FEMS Microbiology Reviews* 28: 127-181.
3. Noble R, Fuhrman J (1998) Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquatic Microbial Ecology* 14: 113-118.
4. Fuhrman J (1999) Marine viruses and their biogeochemical and ecological effects. *Nature* 399: 541-548.
5. Campbell A (1961) Conditions for the existence of bacteriophage. *Evolution* 15: 153-165.
6. Levin B, Stewart F, Chao L (1977) Resource-limited growth, competition, and predation: a model and experimental studies with bacteria and bacteriophage. *American Naturalist* 111: 3-24.
7. Lenski RE (1988) Dynamics of interactions between bacteria and virulent bacteriophage. *Advances in microbial ecology* 10: 1-44.
8. Beretta E, Kuang Y (1998) Modeling and analysis of a marine bacteriophage infection. *Mathematical Biosciences* 149: 57-76.
9. Thingstad T (2000) Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography*: 1320-1328.
10. Thingstad T, Lignell R (1997) Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquatic Microbial Ecology* 13: 19-27.
11. Stent GS (1963) *Molecular Biology of Bacterial Viruses*. San Francisco: Freeman.
12. Murray A, Jackson G (1992) Viral dynamics: a model of the effects of size, shape, motion and abundance of single-celled planktonic organisms and other particles. *Marine ecology progress series* Oldendorf 89: 103-116.
13. Suttle C (2000) Ecological, evolutionary, and geochemical consequences of viral infection of cyanobacteria and eukaryotic algae. *Viral Ecology: Academic Press*. pp. 247–296.
14. Kutter E, Sulakvelidze A (2005) *Bacteriophages: biology and applications: CRC Press*.
15. Suttle C (2005) Viruses in the sea. *Nature* 437: 356-361.
16. Paul J, Kellogg C (2000) Ecology of bacteriophages in nature. *Viral Ecology*: 211–246.
17. Pernthaler J (2005) Predation on prokaryotes in the water column and its ecological implications. *Nature Reviews Microbiology* 3: 537-546.
18. Fuhrman J, Noble R (1995) Viruses and protists cause similar bacterial mortality in coastal seawater. *Limnology and Oceanography* 40: 1236-1242.
19. Wommack K, Colwell R (2000) Virioplankton: viruses in aquatic ecosystems. *Microbiology and Molecular Biology Reviews* 64: 69.
20. Wilcox R, Fuhrman J (1994) Bacterial viruses in coastal seawater: lytic rather than lysogenic production. *Marine Ecology-Progress Series* 114: 35-35.
21. Suttle CA, Chen F (1992) Mechanisms and rates of decay of marine viruses in seawater. *Applied and Environmental Microbiology* 58: 3721.
22. Howard-Jones M, Ballard V, Allen A, Frischer M, Verity P (2002) Distribution of bacterial biomass and activity in the marginal ice zone of the central Barents Sea during summer. *Journal of marine systems* 38: 77-91.

23. Suttle CA, Chan AM (1994) Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. *Applied and Environmental Microbiology* 60: 3167.
24. Seymour J, Seuront L, Doubell M, Waters R, Mitchell JG (2006) Microscale patchiness of virioplankton. *Journal of the Marine Biological Association of the UK* 86: 551-561.
25. Bratbak G, Egge J, Haldal M (1993) Viral mortality of the marine alga *Emiliania huxleyi* (Haptophyceae) and termination of algal blooms. *Marine Ecology Progress Series*.
26. Berg H, Purcell E (1977) Physics of chemoreception. *Biophysical journal* 20: 193-219.
27. Schwartz M (1976) The adsorption of coliphage lambda to its host: Effect of variations in the surface density of receptor and in phage-receptor affinity* 1. *Journal of molecular biology* 103: 521-536.
28. Weinbauer M, Peduzzi P (1994) Frequency, size and distribution of bacteriophages in different marine bacterial morphotypes. *Marine Ecology Progress Series* 108: 11-20.
29. Weinbauer M, Hoefle M (1998) Size-specific mortality of lake bacterioplankton by natural virus communities. *Aquatic Microbial Ecology* 15: 103-113.
30. Castberg T, Thyraug R, Larsen A, Sandaa RA, Haldal M, et al. (2002) Isolation and characterization of a virus that infects *Emiliania huxleyi* (Haptophyta).
31. Lawrence JE, Chan AM, Suttle CA (2001) A novel virus (HaNIV) causes lysis of the toxic bloom-forming alga *Heterosigma akashiwo* (Raphidophyceae). *Journal of Phycology* 37: 216-222.
32. Sandaa RA, Haldal M, Castberg T, Thyraug R, Bratbak G (2001) Isolation and characterization of two viruses with large genome size infecting *Chrysochromulina ericina* (Prymnesiophyceae) and *Pyramimonas orientalis* (Prasinophyceae). *Virology* 290: 272-280.
33. Schultz H, Jorgensen B (2001) Big bacteria. *Annual Review of Microbiology* 55: 105-137.
34. Madigan MT, Martinko JM (2006) *Brock biology of microorganisms*: Upper Saddle River, NJ, USA: Pearson Prentice Hall.
35. Ackermann H (2006) Classification of bacteriophages. *The bacteriophages* 2: 8-17.
36. Ackermann H (1999) Tailed bacteriophages: the order Caudovirales. *Advances in virus research* 51: 135-202.
37. Schulz H, Brinkhoff T, Ferdelman T, Mariné MH, Teske A, et al. (1999) Dense populations of a giant sulfur bacterium in Namibian shelf sediments. *Science* 284: 493.
38. Schultz H, Jorgensen B (2001) Big bacteria. *Annu Rev Microbiol* 55: 105-137.
39. Dubin S, Benedek G (1970) Molecular weights of coliphages and coliphage DNA: II. Measurement of diffusion coefficients using optical mixing spectroscopy, and measurement of sedimentation coefficients. *Journal of molecular biology* 54: 547-556.
40. Suttle CA, Chan AM (1993) Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth characteristics. *Marine Ecology-Progress Series* 92: 99-99.
41. Cavender-Bares K, Rinaldo A, Chisholm S (2001) Microbial size spectra from natural and nutrient enriched ecosystems. *Limnology and Oceanography* 46: 778-789.
42. Sheldon R, Prakash A, Sutcliffe Jr W (1972) The size distribution of particles in the ocean. *Limnology and Oceanography* 17: 327-340.
43. Maniloff J (1997) Nannobacteria: size limits and evidence. *Science* 276: 1773.
44. Whitman W, Coleman D, Wiebe W (1998) Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences* 95: 6578.

45. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, et al. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304: 66.
46. Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, et al. (2002) Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences of the United States of America* 99: 14250.
47. Mei M, Danovaro R (2004) Virus production and life strategies in aquatic sediments. *Limnology and Oceanography* 49: 459-470.
48. Danovaro R, Corinaldesi C, Luna GM, Magagnini M, Manini E, et al. (2009) Prokaryote diversity and viral production in deep-sea sediments and seamounts. *Deep Sea Research Part II: Topical Studies in Oceanography* 56: 738-747.
49. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research* 35: 7188.
50. Clasen J, Brigden S, Payet J, Suttle C (2008) Evidence that viral abundance across oceans and lakes is driven by different biological factors. *Freshwater Biology* 53: 1090-1100.