

# Power law rank-abundance relationships in marine phage populations

Karl Heinz Hoffmann<sup>1</sup>, Beltran Rodriguez-Brito<sup>2,3</sup>, Mya Breitbart<sup>2</sup>, David Bangor<sup>2,3</sup>, Florent Angly<sup>2</sup>, Ben Felts<sup>3</sup>, James Nulton<sup>3</sup>, Forest Rohwer<sup>2,4</sup>, Peter Salamon<sup>3</sup>

<sup>1</sup>*Physics Institute, Technical University of Chemnitz, Chemnitz, D-09107 Germany*

<sup>2</sup>*Department of Biology, San Diego State University, San Diego, CA 92182-4614*

<sup>3</sup>*Department of Mathematics and Statistics, San Diego State University, San Diego, CA 92182-7720*

<sup>4</sup>*Center for Microbial Sciences, San Diego State University, San Diego, CA 92182*

Running Title: Power law for marine phage

Keywords: phage, viruses, diversity, rank-abundance, power law, distribution, model, Lotka-Volterra, kill the winner, neutral evolution, metagenomics

**Phage are the most abundant biological entities in the biosphere, with an estimated  $10^{31}$  particles on the planet. Metagenomic analyses show that the rank-abundance curve for marine phage communities follows a power law distribution. This distribution is consistent with a proposed, modified version of Lotka-Volterra predator-prey dynamics, where blooms of a specific microbial species lead to blooms of their corresponding phage and a subsequent decrease in abundance. The model predicts that the majority of phage genotypes in a population will be rare and it is unlikely that the most abundant phage genotype will be the same at different time points. The model is based on spatial-temporal heterogeneity and a power law phage decay, which are both supported by empirical data.**

*A power law best describes the rank-abundance distribution of marine phage communities:* Six different standard mathematical and ecological distributions (reviewed in <sup>1</sup>) were tested for their goodness of fit to marine phage metagenomic data described in Breitbart et al. <sup>2</sup>. An overlap between sequences from different clones of a shotgun library (i.e., a contig) means that the same phage genotype has been re-sampled. This information was used to predict the population structure of the phage communities based on a modified Lander-Waterman equation <sup>3</sup>. To determine which distribution best described the metagenomic data, the error between the predicted and the observed contig spectrum was used in a quasi-log-likelihood function using a weighted sum of squared errors (Table 1). In both marine phage populations, the data were best described by a power law distribution (error for Scripps Pier (SP) = 1.8 and Mission Bay (MB) = 2.1), followed by the lognormal distribution (error for SP = 1.9 and MB = 2.3). In contrast, classical niche-based ecological models like Broken Stick (error for SP = 11 and MB = 15) and Niche Preemption (error for SP = 12 and MB = 16) were a very poor fit to the data.

There was a major difference between the number of phage genotypes predicted by the different distributions. For example, the lognormal distribution predicted 3 to 5 times more phage genotypes in the samples than did the power law (Table 1). This is important because recent estimations of global microbial diversity have assumed lognormal distributions and may have significantly overestimated species richness<sup>4</sup>.

*A generalized version of Lotka-Volterra dynamics predicts the power law distribution:* Determining the population structure of marine phage communities also suggests ecological models to explain how phage and their microbial hosts interact. It is assumed that different hosts interact through competition for food. This interaction is very weak because food is not a limiting resource for microbes in the marine ecosystem<sup>5,6</sup>. The second assumption is that there is a strong and specific interaction between a microbial host and its phage predators, where the most abundant microbes are killed. This relationship has been termed "kill the winner" and predicts that specific predator-prey pairs oscillate in time as blooms of a particular microbe are followed closely by blooms of its phage predator<sup>7</sup>. Using these assumptions, the problem becomes a system of many non-interacting predator-prey systems. The simplest model matching the assumptions is classical Lotka-Volterra (LV; Equation 1):

$$\begin{aligned} \frac{dB}{dt} &= \alpha B - \beta B \Phi \\ \frac{d\Phi}{dt} &= -\gamma \Phi + \delta B \Phi \end{aligned} \tag{1}$$

where  $B$  denotes the size of the microbial population for a single genotype,  $\Phi$  denotes the population size of the corresponding phage predator, and where  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  are positive constants that describe microbial growth, microbial death, phage decay and phage production, respectively. For simplicity it was assumed that all phage-host

pairs follow identical dynamics<sup>a</sup>, but bloom at independent times (i.e., they are randomly distributed along a common cycle). Note that in order to match the observed population structure, blooms must alternate with comparatively long periods of “hiding out” at concentrations several orders of magnitude lower than the bloom titers (see Fig 4 in 8). For a Lotka-Volterra model such cycles can be found by using initial states near the origin.

Figure 1a-d show the results of adopting Lotka-Volterra dynamics for a population of non-interacting phage-host pairs. Transformation of this data to a log-log plot (Fig 1d) shows a mediocre fit to the power law ( $R^2=0.83$ ). The poor fit occurs because the high abundance phage are over-represented on the log-log plot (i.e., too many ranks correspond to these values and destroy the linearity). The dynamics would fit the data much better if the blooms were shorter and more intense. Mathematically this can be accomplished by using a generalized version of Lotka-Volterra (GLV), where an exponent ( $n>1$ ) is added to the predator populations to account for cooperativity<sup>9</sup>. A GLV model with an exponent of 2 produced an almost perfect fit to the power law distribution ( $R^2=0.99$ ; Figure 1e-h).

$$\begin{aligned}\frac{dB}{dt} &= \alpha B - \beta B \Phi^2 \\ \frac{d\Phi}{dt} &= -\gamma \Phi^2 + \delta B \Phi^2\end{aligned}\tag{2}$$

We propose that in the marine environment, the cooperativity between phage in the lysis event ( $\delta B \Phi^2$ ) and microbial death ( $\beta B \Phi^2$ ) reflects spatial-temporal heterogeneity. The ocean is a gel made up of particles ranging from colloids to marine

---

<sup>a</sup> This feature of the model is for simplicity alone – the real dynamics surely varies among the types. In this neutral evolution hypothesis, our hope is that the qualitative features are captured by replacing the full complexity of the problem by one “average” type. In fact, a comparable calculation using a distribution of parameter values also gives a close fit to the power law.

snow<sup>10-12</sup>. These particles represent higher, local concentrations of nutrients that are necessary for microbial growth<sup>13</sup>. Microbes are known to chemotax to these particles<sup>14</sup>. Phage lysis of microbes on particles would create locally high concentrations of both predators and prey (i.e., a local change in mass action). This would lead to an apparent cooperativity in the predation.

One consequence of adding the cooperativity function to the lysis event ( $\delta B\Phi^2$ ) is that the decay exponent on  $\Phi$  must match the exponent in the mass action terms to keep the populations oscillating (i.e., to maintain neutral stability of the orbits). This feature of the proposed model matches data on phage decay, which shows that phage particles display rapid initial decay rates, which then decrease over time (ref <sup>15,16</sup>).

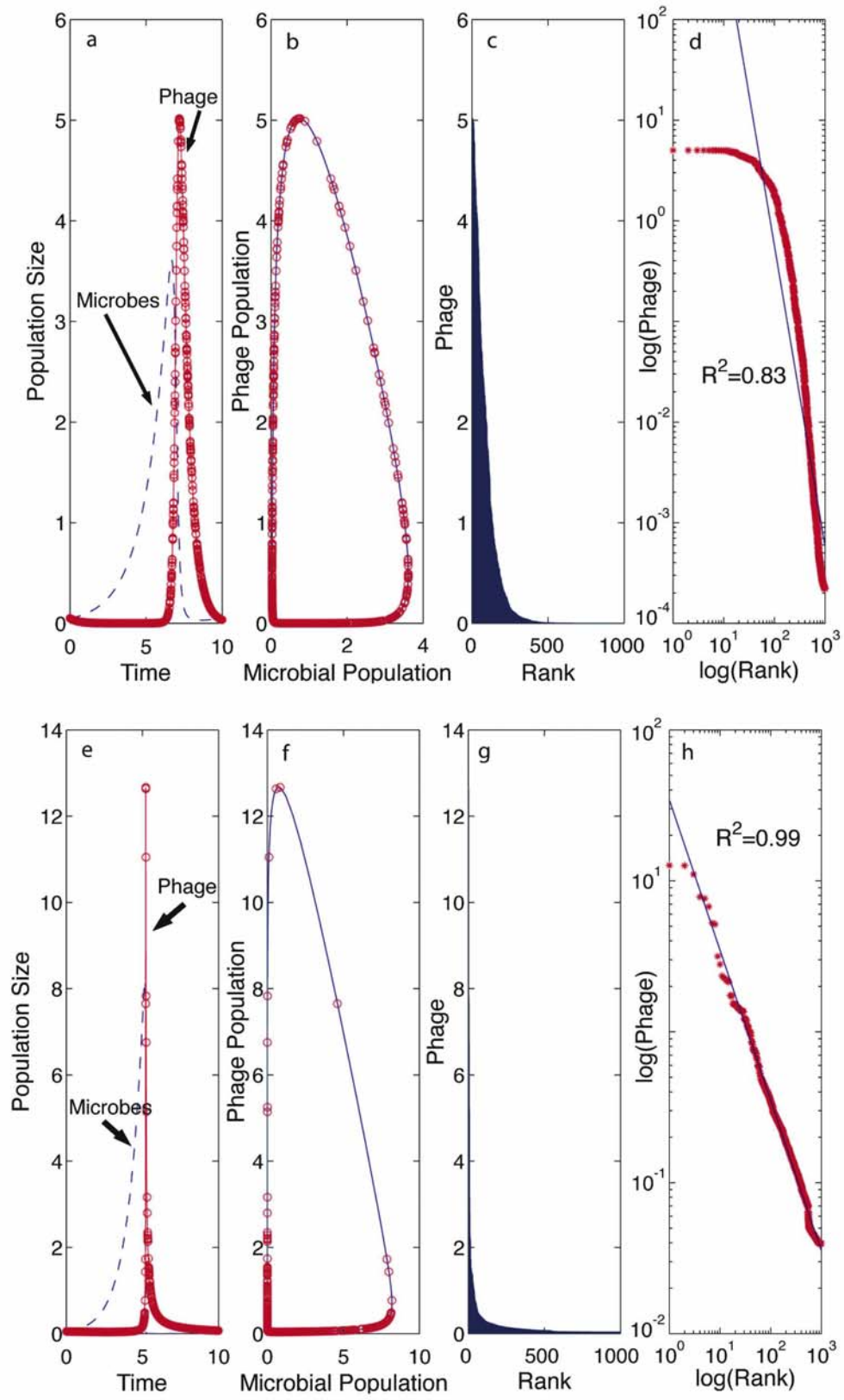
Environmental factors such as local entanglement of phage particles in microbial debris would account for this observation.

The “kill the winner” hypothesis, as interpreted through the proposed model, predicts that typical phage-host cycles involve long periods of hiding from the predators at very low numbers alternating with brief spurts of dominance. Therefore, it is unlikely that the most abundant phage genotype will be the same at different time points and there should be several orders of magnitude difference in the numbers of a specific phage present at different times. Limit cycles traversing approximately the same orbit would make similar predictions and represent a viable alternative to the proposed model. In addition to making several predictions that match empirical data, the proposed model represents a new mechanism for how a system can display power-law behavior.

Table 1. Goodness of fit for six different models to the observed contig spectra of two marine phage metagenomes. The error was calculated as the variance weighted sum squared deviation from predicted contig spectra, which gives the value of the log pseudo likelihood function as described in ref 2. The models are listed in order of increasing error for each sample. The observed contig spectra are shown in square brackets next to each of the samples, where a contig spectrum of [A, B, C, ...] represents A one contigs, B two contigs, etc.. The exponential and logarithmic models were named according to their analytic form in the frequency-rank relationship, viz. frequency proportional  $\exp(k \cdot \text{rank})$  or frequency proportional to  $1/\log(\text{rank}+1)$ , respectively.

<b>Model</b>	<b>Error</b>	<b>Predicted # of Phage Genotypes</b>
<b>Scripps Pier (SP) [1021 17 2 0]</b>		
Power Law	1.8	3400
Lognormal	1.9	13000
Logarithmic	2.5	1400
Broken Stick	11	1200
Exponential	12	10000
Niche Preemption	12	9100
<b>Mission Bay (MB) [841 13 2 0]</b>		
Power Law	2.1	7400
Lognormal	2.3	43000
Logarithmic	2.8	1600
Broken Stick	15	960
Exponential Law	16	7500
Niche Preemption	16	8200

Figure 1. Comparison of phage population structure predicted by Classical Lotka-Volterra (a-d) and the proposed Generalized Lotka-Volterra (e-h) dynamics. All population sizes and time are in arbitrary units. a) Populations of microbes (dashed blue line) and phage (solid red line) as a function of time. Red circles show 1000 random samples along the cycle. b) The relationship between the sizes of the microbial and phage populations (solid blue line). Red circles show the same 1000 random samples along the cycle. c) Rank abundance plot of the 1000 random phage samples. d) Log-log version of panel c. The solid blue line is the least-squares fit to the ranked phage populations (red asterisks). Panels e-h show the same relationships for the Generalized Lotka-Volterra dynamics. The  $R^2$  is the coefficient of determination.



## References

1. Magurran, A. *Measuring Biological Diversity* (Blackwell Publishing, Malden, MA, 2004).
2. Breitbart, M. *et al.* Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci USA* **99**, 14250-14255 (2002).
3. Lander, E. S. & Waterman, M. S. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**, 231-239 (1988).
4. Curtis, T., Sloan, W. & Scannell, J. Estimating prokaryotic diversity and its limits. *Proc Natl Acad Sci USA* **99**, 10494-10499 (2002).
5. Sano, E., Carlson, S., Wegley, L. & Rohwer, F. Movement of viruses between biomes. *Applied and Environmental Microbiology* **70**, 5842-5846 (2004).
6. Wilcox, R. M. & Fuhrman, J. A. Bacterial viruses in coastal seawater: lytic rather than lysogenic production. *Mar Ecol Prog Ser* **114**, 35-45 (1994).
7. Thingstad, T. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol Oceanogr* **45**, 1320-1328 (2000).
8. Suttle, C. & Chan, A. Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. *Applied and Environmental Microbiology* **60**, 3167-3174 (1994).
9. Dancso, A., Farkas, H., Farkas, M. & Szabo, G. Investigations on a class of generalized two-dimensional Lotka-Volterra schemes. *Acta Applicandae Mathematica* **23**, 103-127 (1990).
10. Azam, F. Microbial control of oceanic carbon flux: the plot thickens. *Science* **280**, 694-696 (1998).

11. Chin, W.-C., Orellana, M. V. & Verdugo, P. Spontaneous assembly of marine dissolved organic matter into polymer gels. *Nature* **391**, 568-572 (1998).
12. Alldredge, A. L., Cole, J. J. & Caron, D. A. Production of heterotrophic bacteria inhabiting macroscopic organic aggregates (marine snow) from surface waters. *Limnol. Oceanogr.* **31**, 68-78 (1986).
13. Martinez, J., Smith, D. C., Steward, G. F. & Azam, F. Variability in ectohydrolytic enzyme activities of pelagic marine bacteria and its significance for substrate processing in the sea. *Aquat. Microb. Ecol.* **10**, 223-230 (1996).
14. Blackburn, N., Fenchel, T. & Mitchell, J. Microscale nutrient patches in planktonic habitats shown by chemotactic bacteria. *Science* **282**, 2254-2256 (1998).
15. Mathias, C., Kirschner, A. & Velimirov, B. Seasonal variations of virus abundance and viral control of the bacterial production in a backwater system of the Danube river. *Applied and Environmental Microbiology* **61**, 3734-3740 (1994).
16. Heldal, M. & Bratbak, G. Production and decay of viruses in aquatic environments. *Mar Ecol Prog Ser* **72**, 205-212 (1991).

Acknowledgements: We thank Joseph Mahaffy and the other members of the MathPhage group at SDSU for helpful conversations. This work was supported by grants from the National Science Foundation. MB was supported by an EPA STAR Fellowship.

Correspondence should be addressed to: [salamon@sdsu.edu](mailto:salamon@sdsu.edu)