Interpreting phage ecology, theory, and models in the genomic age

Dinesh Kumar Kuppa Baskaran[1,2], Neha Kashyap[1], James C. Kosmopoulos[1,2], Marguerite V. Langwig[1,3], Kris Sankaran[4], Karthik Anantharaman[1,5*]

Author Affiliations

[1]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA

[2]Microbiology Doctoral Training Program, University of Wisconsin-Madison, Madison, WI, USA

[3]Freshwater and Marine Sciences Program, University of Wisconsin-Madison, Madison, WI, USA

[4]Department of Statistics, University of Wisconsin-Madison, Madison, WI, USA

[5]Department of Integrative Biology, University of Wisconsin-Madison, Madison, WI, USA

Correspondence to K.A. (karthik@bact.wisc.edu)

Abstract

Viruses, particularly bacteriophages, are the most abundant biological entities across nearly all ecosystems and play a central role in shaping microbial community structure, ecosystem function, and evolution. Consequently, there has been growing interest in studying phages and their interactions within microbiomes. Mathematical modeling has long provided a foundation for investigating phage–host interactions, formalizing diverse infection strategies (lytic, lysogenic and chronic infection) and their governing parameters. These approaches include resource-based trophic models that describe the flow of resources from microbes to viruses, epidemiological models that focus on the transmission of infection, stochastic models that capture environmental and demographic randomness, metacommunity models that account for spatial organization, and network-based models that characterize the topology of phage–host interactions. Together, these frameworks have given rise to influential ecological theories, such as Kill-the-Winner and Piggyback-the-Winner, which offer mechanistic explanations for phage–host dynamics. Despite these advances, theoretical insights have remained only loosely connected to environmental observations, and studies of phage communities have often been limited to descriptive characterization. Metagenomics now provides a unique opportunity to directly observe phage–host dynamics across diverse environments and timescales. The increasing availability of large-scale and longitudinal metagenomic datasets enables the application of statistical and data-driven approaches including co-occurrence and correlation analyses, network inference, dynamic modeling, and machine learning to infer interactions and ecological strategies from empirical data. In this review, we synthesize the foundations of theoretical and mathematical modeling of phage–host systems and discuss emerging approaches for integrating omics data with these frameworks. By linking metagenomic observations with mechanistic and ecological models, we highlight pathways toward moving beyond descriptive viromics and toward predictive, data-informed understanding of phage ecology and evolution in natural ecosystems.

Table 1: Glossary

Word	Meaning
Density-dependent feedback	Regulation of microbial dynamics with respect to population density. In phage-host systems, phages act on the density signals stabilizing the coexistence by increasing viral pressure on abundant host or reducing infection rates on low density hosts.
Superinfection exclusion	A mechanism by which a bacterium infected by a phage becomes resistant to subsequent infection by related phages, typically mediated by prophage-encoded immunity.
Multiplicity of Infection (MOI)	The ratio of infecting phage particles to susceptible host cells in a system. Higher MOI means more virus particles per host cell while low MOI means the opposite. MOI can determine the likelihood of infection. Weak phages require higher MOI to establish a successful infection.
Bet-hedging strategy	An evolutionary strategy in which phages prefer to follow a safer lifestyle (lysogeny over lysis) even when it has lower potential for rapid growth.
Viral shunt	The redirection of microbial biomass into dissolved organic matter through viral lysis, thereby retaining nutrients within the microbial loop rather than transferring them to higher trophic levels.
Life-history traits	Biological parameters governing phage and host dynamics, including adsorption rate, burst size, latent period, decay rate, host growth rate, and resistance costs. These traits define model parameters in mechanistic and eco-evolutionary frameworks.
Mass tranfer	The physical transport of phages, hosts, or nutrients through diffusion, advection, or mixing. Mass transfer determines encounter rates and is particularly important in spatial, chemostat, and fluid-flow models.
Permuting the dataset	A statistical procedure in which data values or labels are randomly shuffled to generate randomized null distributions. Permutation is commonly used in network inference and machine learning to assess the significance of inferred interactions or predictive performance.
Dynamic modeling approach	A modeling framework that uses a set of differential equations to explicitly represent temporal changes in population abundances or states. This can be used to study phage-host interactions over time.
Features	Quantitative variables extracted from data and used as inputs for statistical or machine learning models. In phage ecology, features may include genomic attributes, abundance trajectories, interaction metrics, or environmental variables.

Latent features	Abstract representations learned by machine learning models that capture underlying structure not directly observed in the data. In phage–host studies, latent features may correspond to infection strategies, ecological niches, or host-range determinants.
Feature space	The multidimensional space defined by all features used in a model, where each observation is represented as a vector. Comparing empirical and simulated data within a shared feature space enables quantitative evaluation of theoretical models.

Introduction

The global virosphere comprises viruses that interact with bacteria, archaea, and eukaryotes, playing a significant role in ecosystem-scale effects. Bacteriophages (viruses that infect bacteria) play a pivotal role in shaping microbial communities, regulating ecosystem processes, and influencing global biogeochemical cycles (1, 2). Phage–host interactions are inherently complex, with phages employing multiple strategies to engage bacteria, ranging from lysis to lysogeny and gene transfer. Each mode and interaction shapes microbial communities in distinct and sometimes transformative ways (3–5). For example, phage lytic infections can influence microbial diversity, while phage integration can introduce new metabolic capabilities to the microbial communities.

Modern strategies for studying viruses rely on metagenomics and viromics, along with various bioinformatics tools (6, 7), to directly identify and characterize bacteriophages in the environment (eliminating the need for cultivation). This involves annotating and identifying genes involved in different phage lifestyles. This strategy has led to a significant increase in large-scale phage studies, generating large datasets that enable the analysis of viral activity and host interactions in the environment (8, 9) and in host-associated ecosystems like the human gut (10–12). Despite the growing volumes of bioinformatics tools that utilize genomic information to study phage ecology, a significant gap remains in understanding phage lifestyles from genes and proteins, i.e., interpreting the ways in which phages interact with their hosts, such as lysis, lysogeny, or chronic infections. This is mainly due to shortcomings in viral protein annotation, which, in turn, affect our interpretation of phage ecology and evolution, including predictions of phage-host interactions, metabolic and ecological interactions, and protein function. This lack of knowledge of viral genomes and proteins is referred to as ‘viral dark matter’.

Phage interactions can be interpreted by tracking phage-host dynamics or modeling. Time-series sampling offers a powerful way to track phage-host dynamics over time, provides slightly more information on the dynamics of phage ecology in real time, but still suffers from the same bane of viral dark matter. This is mainly because the computational methods currently used to analyze both time-series and single-timepoint datasets are fundamentally similar, as they focus solely on genomic information in both cases. This approach entirely overlooks a valuable aspect of time-series data, specifically, the temporal changes in populations. There is a need for tools that integrate temporal population shifts and genomic data to improve predictions, thereby addressing

a critical gap in phage ecology research. Recognizing this gap motivates the need for integrative approaches discussed in the "*Linking metagenomics with models*" section of this review.

Simultaneously, mathematical modeling has enabled the study of phage-host interactions by establishing various infection strategies and their associated parameters. From an ecological perspective, questions about phages and their hosts are analogous to those in other predator-prey and host-parasite systems. What is the role of phage in regulating the density and diversity of the bacterial population? What are the conditions in which phages can coexist with the susceptible bacteria? How do competing phages, competing bacteria, and changes in resources influence coexistence? Theoretical models of lytic, lysogenic, and chronic infections have been used to address many of these questions. However, they often model only one or, at most, two different phage infection cycles. This does not accurately represent the heterogeneous environmental conditions that phages exist within, and therefore, theoretical insights are insufficient to explain real-time phenomena.

Table 2: Different states of phage-host system

States	Description
Host/microbe	A bacterial cell
Phage	A bacteriophage virus
Lytic	Phages infect and kills the host cells
Lysogenic	Phages infect and integrate with the host genome
Prophage	Phage genome integrated with the host genome
Resistant	Host becomes resistant to the phage
Chronic	A state that is inbetween lytic and lysogenic. Phage infects the host and buds out chronically without killing it.

In this article, we review the current state of genomic and theoretical modeling approaches available for studying phage-host lifestyles in the environment, with a focus on the disconnect between predictions from available tools and real-world observations. We also discuss gaps and scope for research in this area by proposing methods for integrating metagenomic data with theoretical models to improve predictions. In particular, we emphasize the integration of metagenomic data with theoretical models as a path toward generating predictions that more closely align with ecological observation

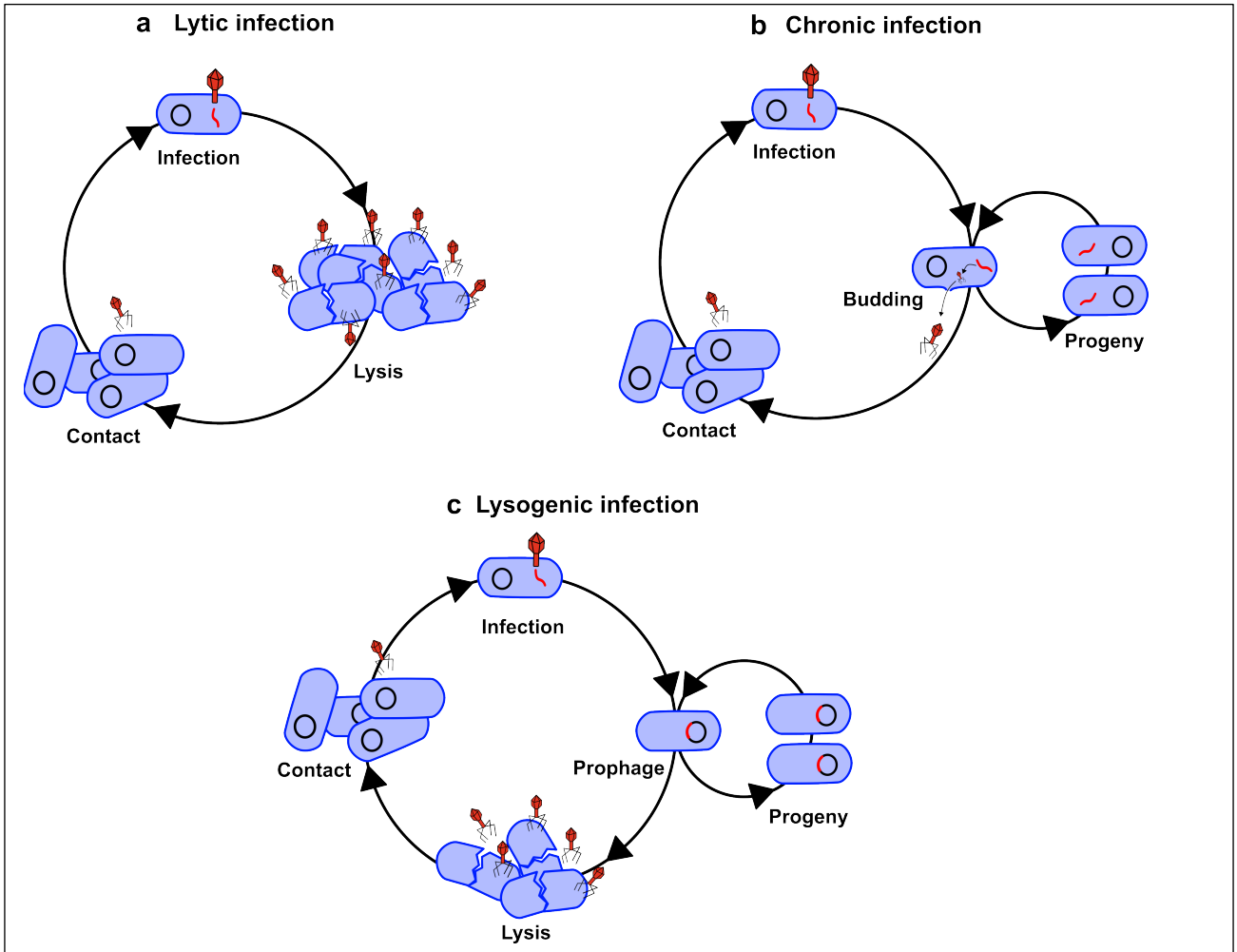

Figure 1: Phages employ three principal infection strategies: lytic (a), chronic (b), and lysogenic (c). (a) *During* ***lytic infection****, phages inject their genetic material into the host cell, hijack host machinery to replicate viral genomes and synthesize structural components (head, tail, and associated proteins) and assemble progeny virions. Completion of virion assembly is followed by host cell lysis, releasing newly produced phages.* ***(b)*** *In* ***chronic infection****, phages replicate within the host and are continuously released through budding without causing immediate host cell death; viral genetic material may also be transmitted to daughter cells during host division.* ***(c)*** *In* ***lysogeny****, phage genetic material integrates into the host chromosome as a prophage, which is passively inherited by subsequent host generations. Under appropriate environmental conditions, the prophage can be induced to exit the lysogenic state and enter the lytic cycle.*

Table 3: Summary of mathematical models

Modeling Framework	Ecological question	Components	Ecological insight	Link to Omics data	Hypothesis enabled	Applications
Resource-based trophic models	How does phages influence flow of nutrients from microbes to viruses in ecosystems?	Resources; Bacteria; Phages	Host growth and resource availability influences phage production	Bacterial abundance, Phage abundance, Virus-to-microbe ratio	Kill-the-winner	Modeling food web, microbial loop; modeling industrial bioprocesses involving phages
Epidemiological models	How do infections spread within microbial populations?	Susceptible; Infected; Phage	Viral life-history parameters determine viral lifestyles and coexistence of multiple lifestyles	Bacterial abundance, Phage abundance, Virus-to-microbe ratio	Kill-the-winner, abandon-the-sinking-ship	Designing phage therapy; modeling multiple viral lifestyles, modeling phage infections in ecosystems

Stochastic models	How does random events like evolution influence infection outcomes?	Host; Phage	Co-evolutionary arms race promotes coexistence of diverse host and phage species.	Bacterial abundance, Phage abundance, Virus-to-microbe ratio	Piggyback-the-winner, Red Queen Hypothesis	Modeling phage therapy; modeling coevolution arms race; modeling phage and host resistance systems
Network-based models	How does the phage-host interactions influence community structure? How does the topology of phage-host interactions change with environment?	Host; Phage	Phage-host interaction networks show a 'nested-modular' structure meaning that phages are specialist to niches while in a niche they can be generalist	Co-occurrence patterns estimated from host and phage abundance across time or place	NA	Detecting patterns in phage-host interactions

Meta-community models	How does spatial structure shape phage–host interactions?	Host; Phage	Phage predation can vary spatially across the community. Fast-growing peripheral regions are prone to higher phage predation than the stable basal regions	Spatial metagenomics	Peripheral Kill-the-Winner	Modeling phage interactions in biofilms

Mathematical models of phage-host interactions

There have been extensive modeling studies on how viruses shape the population dynamics and evolutionary dynamics of microbes (13–25), which often involve modeling various phage lifestyles, including lytic, lysogenic, and chronic (**Fig. 1**) (**Table 2**). Broadly, mathematical models of phage-host interactions can be classified into five major classes based on the questions answered, and the components included in the model (**Fig. 2**). Resource-based trophic models emphasize the flow and conversion of resources into microbial biomass and subsequently into viral particles. Epidemiological models focus on the transmission and spread of infection within host populations. Stochastic models incorporate randomness to capture the probabilistic nature of phage–host encounters and infection outcomes. Metacommunity models explicitly account for the spatial organization and connectivity of phage–host communities across heterogeneous environments. Finally, network-based models emphasize the topology of interactions, characterizing who infects whom, and how the structure of these interactions shapes community dynamics (**Table 3**).

Resource-based trophic models

The early phage-host models developed by *Stewart et al.* (26) were inspired by resource-based trophic models, such as predator-prey models, in which the growth of one species depends solely on its ability to take up and convert resources (**Fig. 2a**). Consequently, phages were treated only as predators (lytic agents). In the phage-host setup, there are three main components: 1) primary resources (R), 2) primary consumers or bacteria (B), and 3) predators or phages (P). Each component acts as a resource for the element in the immediately higher trophic level. The primary consumers (microbes) prey on the primary resources, while the predators (phages) prey on the primary consumers. At any given time point, microbial populations are limited by resource availability, while phage populations are limited by the availability of microbial hosts. By assuming unlimited resource availability, the model can focus solely on phage-microbe dynamics. These conditions can be observed in continuous culture setups, known as chemostats, where spatial homogeneity, abiotic variables, and nutrient supply are maintained constantly.

Starting with this setup, these models look for conditions in which all three components can coexist, to reflect real environmental conditions. On a superficial level, coexistence occurs when the concentration of resources in the environment supports the growth of primary consumers, and the density of primary consumers supports predators at all times (15). These concentrations are referred to as equilibrium concentrations. The growth characteristics of the population can provide a more intuitive explanation for the coexistence: at any point in time, the rate at which phages lyse bacteria depends on both microbial and phage densities (27). Thus, when microbial density is low, the lysis rate is also low. This effect protects the microbial population from over-predation, allowing for coexistence.

But how do these models scale in more realistic scenarios, such as the existence of multiple primary consumers? Consider a scenario with two primary consumers: a bacterial population sensitive to phages and a bacterial population resistant to phages. The chemostat models (15, 28) predicted that when the resistant bacterial population has lower competitive fitness than the sensitive population, coexistence of both bacterial populations and the phage was possible (**Fig.**

3a). Here, the sensitive bacteria persist due to their higher competitive fitness, such as a greater affinity for resources, while resource availability limits the growth of the resistant population. Again, the resistant bacteria persist because phages control sensitive bacteria. In other words, phage predation also frees up resources for consumption by resistant cells. These chemostat conditions better reflect laboratory culture conditions but do not accurately represent natural environments. However, at that time, viruses were thought to be less significant in natural environments, so these initial models focused solely on phage-host dynamics in controlled environments.

The major distinction between natural and controlled environments is the existence of a more complex food web. In this context, trophic modeling demonstrates greater adaptability in incorporating more elements into the system. This led to the formation of more elaborate trophic models with multiple resource limitations (e.g., carbon, phosphate, or nitrogen), multiple bacterial populations, multiple phage populations, and an external protozoan predator (17). This answered questions like: Can more bacterial populations coexist? What stabilizes the association between these populations? What is the role of phages in regulating the bacterial density? What are the directions and effects of selection on both bacteria and phages?

These resource-based trophic models laid out the foundations to understanding coexistence. If all the phage-host infection rates were the same, then the viruses infecting the fast-growing hosts were predicted to be the most abundant (17). This presented a reciprocal relationship between microbial diversity and phage diversity, where the host-specific phage selectively killed the most dominant populations of the community, thereby maintaining multiple bacterial populations. In contrast, the varying substrate affinities of the coexisting bacterial populations ensured the presence of multiple phage populations. This suggested that there should be one virus for every bacterial species present in the population that preys on the specific bacteria. This also satisfies a previously stated requirement of predator-prey dynamics (15): the number of distinct predator populations cannot exceed the number of distinct prey populations, and the number of distinct prey populations cannot exceed the sum of the distinct predator and resource populations in the system.

Building upon this, models next incorporated the impact of evolution on population dynamics, which allowed for the mutation of bacterial and viral species. This revealed the existence of trade-offs between susceptibility to viruses and the ability of microbes to uptake nutrients. This meant that host evolution towards phage resistance is accompanied by a cost on nutrient uptake capability (25).

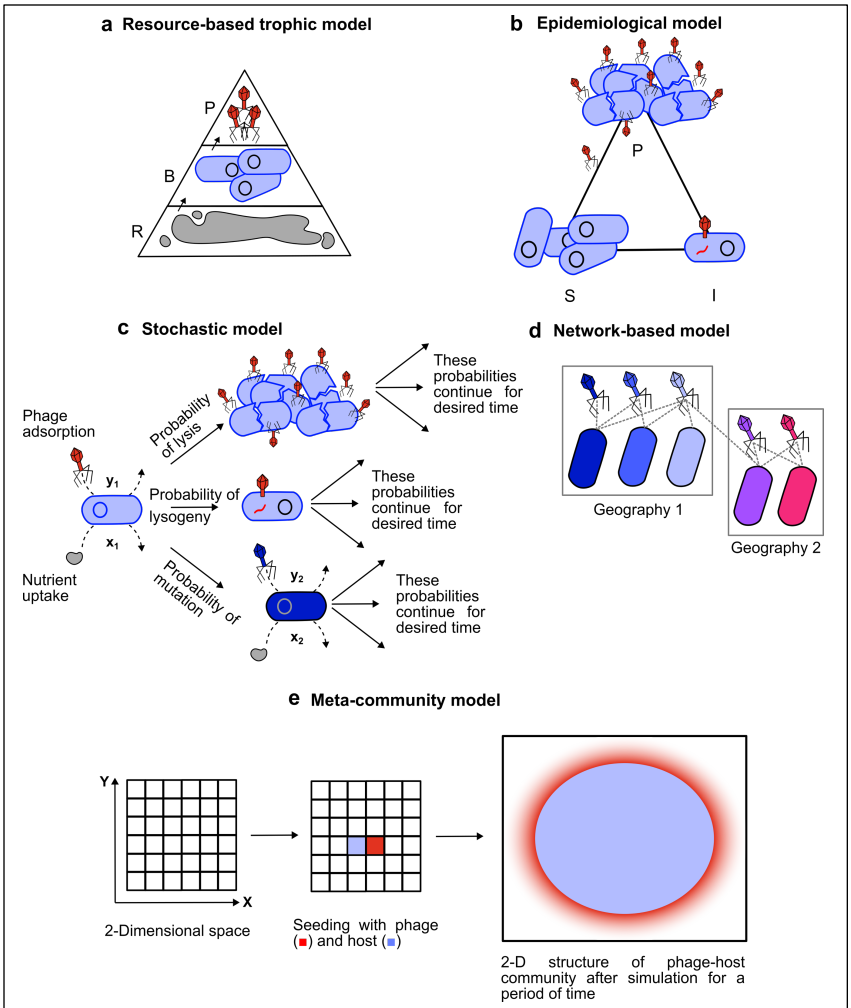

Figure 2: Mathematical models can be classified into five major classes. (a) Resource-based trophic models *are used to model the flow of resources from bacteria to viruses. The three major components of these models are resources-R, bacteria-B and phages-P.* ***(b) Epidemiological models*** *are used to model the spread and transmission of phage infections in the host population. The host population is partitioned into susceptible population-S and infected population-I, along with the phage population-P.* ***(c) Stochastic models*** *are used to model the randomness in the phage-host systems. These models represent infection outcomes as different probabilistic values.* ***(d) Network-based models*** *are used to represent the topology of phage-host interactions and are used to examine community structure and interaction patterns.* ***(e) Meta-community models*** *explicitly account for the influence of spatial organisation on phage-host dynamics.*

In chemostat conditions, bacteria are consistently maintained at an exponential/log phase. However, such conditions are not representative of natural environments, where bacteria can reach the stationary phase or grow very slowly as they near the carrying capacity of the environment. A mechanistic model developed by *Weitz et al.* (29) attempted to link the phage-induced mortality rate to the host reproduction rate. This approach only modeled two components, namely phage and bacteria, and explained the existence of two possible states: phage-host coexistence and phage extinction. The initial densities of phage and host influenced these possibilities in the system. Meaning that if too few phages were added (low multiplicity of infection, MOI), then phage infection and reproduction cannot happen sufficiently to establish the population. At the same time, if too many phages were added (high MOI), the host population was suddenly depleted, leading to extinction. Similarly, as the host population approaches the stationary phase, the timing of phage introduction also influences the outcome. Hence, an appropriate number of phages (optimization of MOI) must be added at the right time interval to ensure coexistence. Another interesting observation was that

when the phage and host coexisted, they always exhibited oscillations, whereas extinction occurred through a single boom-burst cycle of infection.

Phage-host systems deviate significantly from predator-prey dynamics when phages exhibit lysogenic characteristics. While interactions involving strictly lytic phages are relatively straightforward, lysogenic phages introduce additional layers of complexity through prophage integration, immunity, and conditional life-cycle switching. Nevertheless, similar to lytic phages, lysogenic phages can also be studied in chemostat conditions using different trophic models. The lysogenic model proposed by *Stewart et al.* (16) and *Chaudhry et al.* (24, 30) consists of 1) sensitive host cells, 2) host cells infected by lysogenic phages (lysogens), 3) resistant cells, 4) lytic phages, and 5) lysogenic phages as the main components. These models predict that coexistence can occur in a broad range of scenarios, depending on parameters such as resource concentration, adsorption rates, burst size, and the relative growth rates of sensitive, infected, and resistant bacterial populations. For example, it was seen that the resistant cells were always at a competitive disadvantage relative to the sensitive host when all three bacterial populations coexisted. Similarly, temperate and lytic phages coexisted as long as the latter did not have a net advantage in life-history traits. On the contrary, even if temperate phages had a net advantage in life-history traits, lytic phages were not eliminated. In addition to these predictions, these models identified two instances in which temperate phages were advantageous over lytic phages. First, by entertaining the idea of lysogenic phages as allelopathic agents (phages produced by lysogens as toxins for other bacterial populations) (31). This is advantageous to the host bacterial population because the prophages in the host provide superinfection immunity. Secondly, under conditions of low bacterial density, lytic phages would eventually be depleted over time. However, temperate phages can persist due to their ability to exist as prophages (**Fig. 1c**). Despite these advancements, a major information bottleneck lies in modeling the coexistence of multiple viral lifestyles as they occur in nature (32, 33). This was achieved to some extent using epidemiological models, which we discuss in the following section.

Epidemiological models

In parallel, epidemiological models, inspired by the spread of disease, have been developed to model phage-host infections (**Fig. 2b**). The three main components of a simple epidemiological model are: 1) susceptible bacterial populations (S), 2) infected bacterial populations (I), and 3) phage populations (P) (34, 35). In these models, phages can infect the susceptible population at an infection rate/adsorption rate, converting them into infected cells. New phages are released on the death of infected cells based on the burst size of the phage.

Epidemiological phage-host models (with only lytic phages) developed by *Beretta et al.* (34) showed that there were three possible stable states in this system, depending on the virus burst size (**Fig. 3b**). The first was the infection-free system, in which the infection dies down when the burst size is small. When the burst size increases and exceeds a threshold value, it transitions into a positive infection system, where all three components of the model coexist. Finally, there is the oscillatory system, where the burst size reaches a critical value, and the model exhibits sustained oscillations in the susceptible, infected, and viral populations. In all three systems, the phages do not burst out of the cell right after the infection. There is always a time required for the production of phage particles inside the host cell. When introduced as a latency period into the

model, it still retains the three states (35), but they now depend on the latency period in addition to the burst size.

Any host stress has been shown to be an important model component in influencing lytic-to-lysogeny switching (36). This has four components in the system: 1) normal hosts with a high growth rate and low mortality, 2) stressed hosts with low growth rate and high mortality, 3) free phages, and 4) lysogens. Simulation of this model showed that lysis was maximized in stressed hosts, whereas a balance between lysis and lysogeny was observed in normal hosts. This explains the 'abandon the sinking ship' hypothesis: when the host is healthy or in an exponential growth phase, maintaining a balance between lysis and lysogeny reduces the cost of losing the current host, whereas this benefit is lost in a dying host.

Different variants of epidemiological models can be constructed for lytic, lysogenic, and chronic phages, each with its respective phage population (20). Parameters such as growth rate, carrying capacity, adsorption rate, host and phage death rates, burst size, and lysis rates will remain common among all models. Lifestyle-specific parameters distinguish different viral strategies within epidemiological models. For example, lysogenic models include parameters such as the probability of genome integration and the rate of prophage induction, whereas chronic infection models include a budding rate but do not involve host lysis. These parameters define how each viral lifestyle interacts with its host and shape the resulting population dynamics. These parameters can help study whether a virus can successfully establish itself in a microbial community (called viral invasion fitness). In practical terms, invasion fitness measures how effectively an infected cell gives rise to new infections. It is defined as the average number of newly infected cells produced by a single infected cell and the virions it releases when introduced into an otherwise susceptible host population, a quantity commonly referred to as the reproduction number (20). If this value exceeds one for the given host and viral parameters, the virus can spread; if it is less than one, the virus will fail to invade. This framework also allows the mix-and-match of different lifestyles to create heterogeneous lifestyle models (14) for studying interactions between viruses with different lifestyles by determining the invasion fitness of one viral population in the presence of an established resident viral population. For instance, invasion fitness can be used to determine whether a lytic virus can invade a microbial system already dominated by a chronic virus, or whether a chronic virus can establish itself in a community where lytic infection is already prevalent.

Then, determining the life history and environmental parameters under which these reproduction numbers are feasible is the condition under which the different lifestyle strategies coexist. This model predicted a unique parameter regime where the chronic viruses require lytic viruses for survival. This is justified in a scenario where only chronic viruses are present, and the infected cells face high competition for nutrients, hence leading to eventual extinction. However, in the presence of lytic viruses, this competition is reduced by lysis, allowing for the coexistence of these organisms. This phenomenon can be extended to other latent lifestyles, such as lysogeny, in which lytic viruses reduce niche competition through cell lysis, allowing the coexistence of lysogens with low competitive fitness but resistant to lytic phages (37).

To study the competition between temperate and chronic viruses, *Clifton et al.* developed an epidemiological model (38) in which the bacterial population is susceptible to both temperate and

chronic phages. Upon infection by a temperate phage, the bacteria can either become a lysogen or lysed, producing free temperate phages. In contrast, upon infection by a chronic phage, the bacteria can become productive, producing phages without lysis, or remain in a latent state as chronic bacteria. This model observed four steady states based on the frequency of bacteria becoming latent in chronic versus temperate lifestyles: coexistence, temperate-only, chronic-only, and susceptible-only. It was observed that the frequency for temperate phages must be maintained at a level that is neither too small nor too large. If the temperate lysogen frequency is too high, chronic viruses will dominate and infect bacteria first, eventually driving temperate viruses to extinction. If too low, temperate viruses lyse bacteria quickly, allowing bacteria to reproduce and the susceptible population to grow. It was also shown that a chronic strategy benefits best by reducing the frequency of latency, as productive bacteria increase phage abundance through both horizontal (via production) and vertical (via host cell division) mechanisms.

Stochastic models

Stochastic models have been used to study the evolution of phage resistance and counter mechanisms to phages within a population (commonly referred to as an 'arms race'). In these models, the hosts and phages are allowed to mutate to gain resistance and counter resistance stochastically with a probability upon infection. Such evolutionary models often combine population dynamics models and trait-evolutionary models (**Fig. 2c**). One such model, developed by Weitz et al. (39), studies the coevolutionary arms race between bacteria and phages in a chemostat condition. The system includes two primary components: 1) bacteria and 2) phages. In this setup, a simple predator-prey model governs the population dynamics, and a phenotypic model regulates the evolution of traits in bacteria and phages. Though any life history trait can evolve in a natural environment, only the nutrient uptake rate of bacteria and the adsorption rate of phages are allowed to evolve in the controlled setup. A mutant can successfully invade a population only when its fitness is positive, even in low initial concentrations. Fitness for bacteria depends on nutrient uptake rate and the ability to escape phage. In contrast, phage fitness depends on the adsorption rate, ensuring adsorption to a host before the phage is washed out of the system. When this happens, a new set of equations corresponding to the new bacteria or phage is added to the population dynamics. The model predicts that phages drive hosts to diversify to escape infection, which in turn drives phage mutations to track the host. This phenomenon, called 'Red Queen' dynamics, leads to the coexistence of multiple species in the system (**Fig. 3c**).

Levin et al. (28) used a similar setup to study the CRISPR-based arms race. Consistent with Red queen dynamics, this model showed the emergence of phage mutants capable of evading the host CRISPR from wildtype phage that were susceptible to CRISPR. Yet this arms-race did not completely wipe out the initial wildtype bacteria as one would expect, instead, they demonstrated the persistence of both sensitive and phage-resistant hosts.

Game theory can be used to model the decision of phages to follow a temperate or a virulent lifestyle (40). This is done by simulating a phage population in a fluctuating environment. The environment is designed to randomly switch between a 'good' condition, favoring phage production, and a 'bad' condition characterized by low host availability. Here, the phage is made

to choose between lysis and lysogeny when infected, without knowing its condition in the next cycle. Phages exhibited optimal growth rates when the lysogenisation frequency was equal to the probability of encountering a hostile environment. This means that in environments prone to frequent distortions, phages will have a high lysogenisation frequency. Conversely, a higher frequency of lysis can be observed in environments with high host density, diversity, and nutrient availability.

Phage–host models have also been applied to investigate the dynamics of phage therapy in bacterial infections. Such models are used to predict key challenges in therapeutic applications, particularly the emergence of phage-resistant bacterial strains, which remain a major limitation to sustained treatment efficacy. In addition, many frameworks incorporate the combined effects of antibiotics and host immune responses, providing a more comprehensive understanding of how these factors interact to influence infection outcomes and therapeutic success.

Berryhill et al. (41) developed a system of differential equations to simulate phage therapy in bacterial populations with varying resistance profiles. The model comprised of bacterial subpopulations that were either phage-sensitive or resistant to one, two, or three distinct phage species, each with its own phage type. Transitions from phage-sensitive to resistant states occurred stochastically, reflecting spontaneous mutation or selection under phage pressure. Model simulations revealed that a three-phage cocktail provided more durable treatment efficacy than single-phage therapy, as bacteria developed resistance to one phage relatively quickly but required significantly longer to evolve simultaneous resistance to multiple phages.

A related modeling framework extended this approach to include interactions among phages, antibiotics, and bacterial resistance states. Here, bacterial populations were categorized as antibiotic- and phage-sensitive, antibiotic-resistant but phage-sensitive, antibiotic-sensitive but phage-resistant, or resistant to both. Model outcomes indicated that phage presence delayed the emergence of antibiotic resistance, and conversely, antibiotics slowed the evolution of phage resistance. However, despite these mutual delays, the combined treatment ultimately resulted in the proliferation of dual-resistant bacteria, leaving the infection unresolved. Notably, when host innate immunity was incorporated into the model, it predicted complete clearance of both the infection and the resistant population—highlighting the critical role of the immune system in determining therapeutic success.

Meta-community models

The mathematical models discussed so far have assumed that the phage-host environment is always ‘well-mixed’, meaning that every phage can meet every host. This is not always true in many natural environments, especially in soil where there are patches of densely populated areas. The species from one patch can encounter a species in another patch only through a migration event. Meta-community modeling can incorporate these spatial constraints into the setup (**Fig. 2e**). Conway’s Game of Life (42) can be considered a simple meta-community model, where each pixel represents a community within a larger community. *Ruan et al.* (43) utilized an advanced form of a cell automaton model, known as CellModeller (44), to simulate the growth of bacteria with and without phages. This approach allows modeling scenarios in which phage predation is spatially restricted, for example, occurring only at the periphery of a bacterial

community (**Fig. 2e**). The authors studied the role of phages in maintaining antibiotic resistance genes in the host population despite their reduced competitiveness. The study predicted that phage predation on wild-type, fast-growing, antibiotic-sensitive cells at the biomass periphery, through a peripheral kill-the-winner dynamic, played a significant role in maintaining the slow-growing, antibiotic-resistant cells.

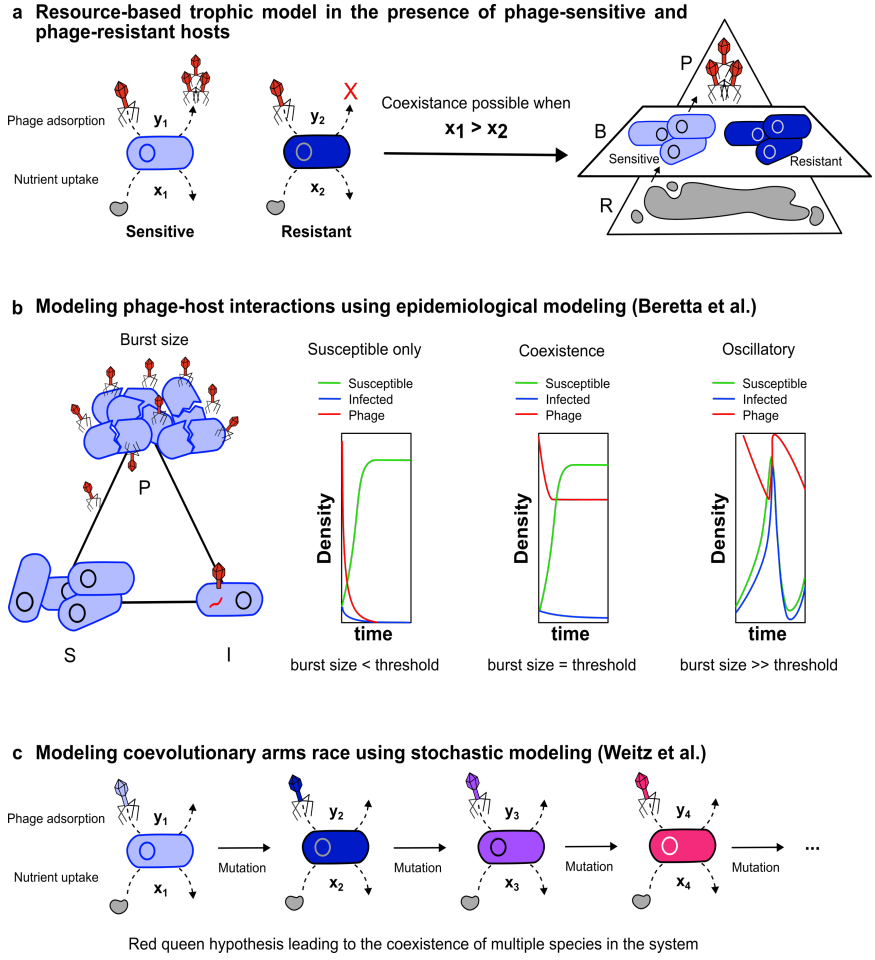

***Figure 3: Application of mathematical models in different scenarios**. **(a)** A modified resource-based trophic model incorporating phage-sensitive and phage-resistant bacterial populations predicts coexistence of both host types and phages when sensitive bacteria have higher resource competitiveness (nutrient uptake rate x_1) than resistant bacteria (x_2). **(b)** A model by Beretta et al. demonstrates that infection dynamics depend on phage burst size, with a threshold below which infection cannot be sustained; at the threshold, coexistence of hosts and phages occurs, while substantially larger burst sizes generate oscillatory population dynamics. **(c)** A stochastic model by Weitz et al. captures coevolutionary arms-race dynamics by allowing probabilistic mutation of phage adsorption rates and host nutrient uptake traits, resulting in diversified host and phage populations consistent with Red Queen dynamics.*

Network-based models

In these models, phage-host interactions can be represented as bipartite networks, with phages and bacteria as the two components (**Fig. 2d**). Edges connect any phage node to a bacterial node, representing a possible interaction. Network-based models combine empirical data with theoretical analysis to characterize the interaction patterns underlying phage-host communities. Often this involves meta-analysis of various environmental and experimental datasets to collect information on phage-host interactions. Then, the bipartite network can be constructed as described. The observations made from such analysis indicate that phage-host interaction networks often exhibit a 'nested-modular' structure (45–48). This means that, on a large

taxonomic and/or geographic scale, the interactions are modular (groups of phages specialize on distinct groups of hosts), but each sub-module shows a nested structure (with both specialist and generalist phages).

Ecological models

Phages can affect their bacterial hosts in various ways, including changes in bacterial physiology, competitive ability, and virulence (49). These effects may directly result from phage life cycles. For instance, lytic phages can reduce host population density and increase bacterial diversity, among other effects. However, they may also cause unexpected outcomes for individual bacteria and populations. For example, prophages can encode essential toxins and auxiliary metabolic genes that influence bacterial fitness. Similarly, the bacterial population can also affect the phages in several ways. The host can impart negative density-dependent feedback on the phage population (21, 37, 50), and bacteria can develop resistance to a specific phage species. These effects change with the environment (51, 52). Ecological models, such as kill-the-winner (KtW) (18), piggyback-the-winner (PtW) (21), piggyback-the-loser (PtL) (21), and cull-the-winner (CtW) (53), among others, have been used to explain this interplay between phage and bacteria.

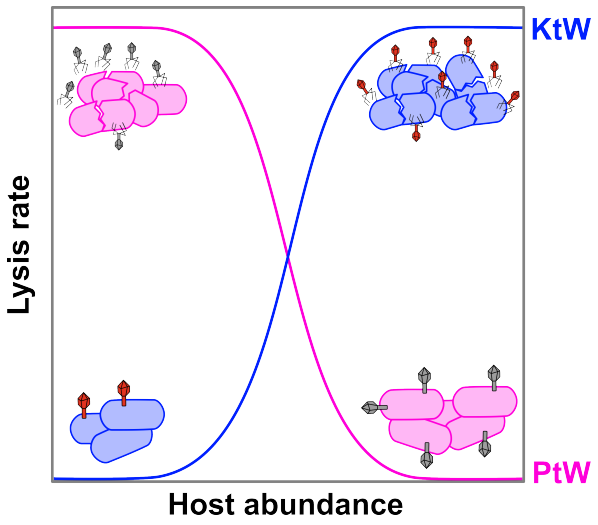

Figure 4: Host abundance–dependent viral strategies in Kill-the-Winner and Piggyback-the-Winner hypotheses. *Under the Kill-the-Winner (KtW) hypothesis, lytic activity increases with host abundance, suppressing dominant host populations, and decreases when host abundance is low. In contrast, the Piggyback-the-Winner (PtW) hypothesis proposes that phages preferentially adopt lysogeny in highly abundant host populations while lytic infection predominates in less abundant hosts.*

Each of these ecological models has been proposed in attempts to explain various environmental phenomena, such as why one fast-growing species cannot overtake a niche—KtW or CtW dynamics. *Thingstad et al.* (17) formalized the KtW model, where lytic phages preferentially target abundant, fast-growing bacterial 'winners,' preventing any single strain from dominating (**Fig. 4**). Biologically, this works on two levels. First, these models suggest that fast-growing organisms have higher lysis rates. This is because mass transfer rates limit phage infection. Thus, the phages are more likely to encounter a dominant strain (the winner) in the ecosystem, thereby lysing it more frequently. Secondly, unlike simpler predator-prey views, KtW explains coexistence in homogeneous environments, where phage-induced mortality recycles nutrients, sustaining the food web (13). The insight is that phages act as diversity enforcers, accounting for observations such as a 1:10 bacteria-to-virus ratio and viral lysis, which causes 10-50% of bacterial losses. This model interprets viruses as partitioning bacterial production: lysis releases organics for reuse, while predation channels biomass upward. Viruses increase bacterial

production relative to primary productivity by short-circuiting energy loops, without altering growth rates in nutrient-limited states. This thereby explains the coexistence of multiple competitors for the same resource in the same niche. A study (54) based on the historical Bermuda Atlantic Time Series (BATS) data revealed the role of viral shunt on formation of oxygen maxima below the sunlit photic zone of the ocean. The study showed that when the cyanobacteria population increases, feedback happens in the form of viral lysis resulting in nutrient recycling. This tight feedback supports the continual maintenance of photoautotrophs keeping the oxygen levels high in the layer. An extension to this model is the KtW with stochastic coevolution model (55), which explains the role of coevolution in maintaining within diversity in microbial populations in the presence of demographic stochasticity.

On the other hand, PtW or PtL dynamics (**Fig. 4**) can explain why phages exhibit a temperate lifestyle. The PtW model suggests that phages undergo lysogeny to eliminate phage competitors through superinfection exclusion (21). The PtL proposes a bet-hedging strategy in which temperate phages integrate with the host to survive extended periods of low host density (40). Li *et al* (37) employed mathematical modeling and evolutionary invasion fitness analysis to further elucidate the interplay between lytic and temperate phages. Their results demonstrate that when lytic phages reduce microbial cell densities, this favors the invasion of temperate phages, which subsequently protect host cells against reinfection. Strikingly, this strategy can emerge even when temperate phages confer no direct benefit or impose a fitness cost on host growth. Transforming this effect into a spatially structured environment, we can conclude that in regions of low phage-host mixing, lytic phages can run out of host cells more frequently, and hence, temperate phages may be prevalent, and vice versa, in areas of high phage-host mixing (36, 56).

Viral lysis and lysogeny are dynamic processes that respond on relatively short temporal and spatial scales to differences in environmental conditions, and particularly to changes in system productivity. Correlation and regression-based analyses in marine systems indicated that viral lysis and lysogeny were significantly coupled with chlorophyll a and the abundance, production, and growth rate of bacteria. This implies that viral lifestyles were dependent on host productivity. Therefore, when host productivity was influenced by season and temperature, it in turn influenced viral lifestyles. For example, lysis was the highest in summer when the water was warmest, stratified, and most productive. In contrast, lysogeny was highest in the spring, when the water was colder, well-mixed, and oligotrophic (50). KtW fits well with this kind of observation, as it implies that the lysis rate increases with host abundance (**Fig. 4**).

As an implication of the KtW model of phage infections, the phage-host interaction network shows a ‘nested-modular’ structure (48). Here, the network shows large-scale modularity, but each module has high local nestedness. Coevolution plays a significant role in the formation of this structure. The KtW model shows that an increase in the population size of a specific host leads to greater cell death due to phage infection. This allows a more diverse pool of less abundant hosts to thrive, promoting the spread of less dominant traits (25, 57). This results in the coevolution of phages, leading to an arms race and the emergence of a ‘nested-modular’ structure.

Furthermore, trophic models can be used to study ecological strategies, such as PtW and PtL, under resource-limiting conditions. For this, the lysis rate of infected bacteria and their growth

rates are regulated in a density-dependent manner. For example, PtW is defined by high lysogenic rates at high abundances (**Fig. 4**). Hence, the lysis rate will be suppressed. The opposite is true for the PtL strategy. These models allow us to study how different ecological strategies maintain the coexistence of various bacterial species (for example, slow growers and fast growers) under different resource conditions. Previous studies (22) have shown that PtW enabled bacterial coexistence under low-nutrient conditions via the viral shunt. Here, the viruses lyse the dying host population, leading to a sudden increase in nutrient availability that results in the persistence of lysogens. In contrast, the PtL is characterized by decreased lysis under low host conditions, resulting in increased resource competition. Hence, in such cases, bacteria can coexist only during high resource conditions.

Microcosm studies of cyanobacteria from oligotrophic and eutrophic lakes have shown that microbial populations collapse following increases in phage following nutrient enrichment (58, 59). This can be attributed to PtL and KtW dynamics, in which high bacterial abundance is followed by increased lysis, thereby enriching nutrients. In contrast, in specific environments, the microbial populations increase in conjunction with the phage population following nutrient enrichment (60). This phenomenon can be explained by PtW dynamics, in which the host replicates along with the prophage under high nutrient conditions.

Linking metagenomics with models

Though modeling provides insights into the mechanisms of different infection strategies, the interpretations are not entirely based on observed or measured features from an environmental perspective. The life history traits, such as adsorption rate, burst size, latent period, and decay rates (61), on which the model parameters rely, are derived from *in vitro* experiments. The laboratory conditions in which these traits are determined are known to be far removed from environmental conditions. Phage-host systems are known to exhibit dynamic life history traits in response to various ecological parameters, including the availability of other microbes, their own population density, and other factors (19, 61). Moreover, models study only shorter timespans. This weakens the links between theoretical predictions and ecological phenomena.

Metagenomics has emerged as a powerful tool for real-time observation of phage-host dynamics in various environments. Metagenomic samples are the most common and most feasible means for measuring microbial populations in the environment today. We argue that metagenomics should be incorporated more into modeling virus-host dynamics because metagenomic datasets can (1) generate bacterial and viral abundances, which are proxies for the bacterial and viral densities at the sampling site and time, (2) measure bacterial and viral richness and (3) measure functional diversity. However, metagenomic analysis alone cannot be used to interpret the life-history traits of microbes in the environment. This makes it challenging to directly link the theoretical predictions to measured environmental datasets. Nevertheless, it is possible to predict phage-microbe relationships and dynamics from meta-analysis of large sets of metagenomics datasets and time-series metagenomics datasets of an environment (62). In particular, three quantitative metrics: bacterial abundance, viral abundance, and the virus-to-microbe ratio (VMR) (63, 64) have emerged as powerful links between empirical observations and theoretical models, forming the basis for multiple integrative analytical approaches.

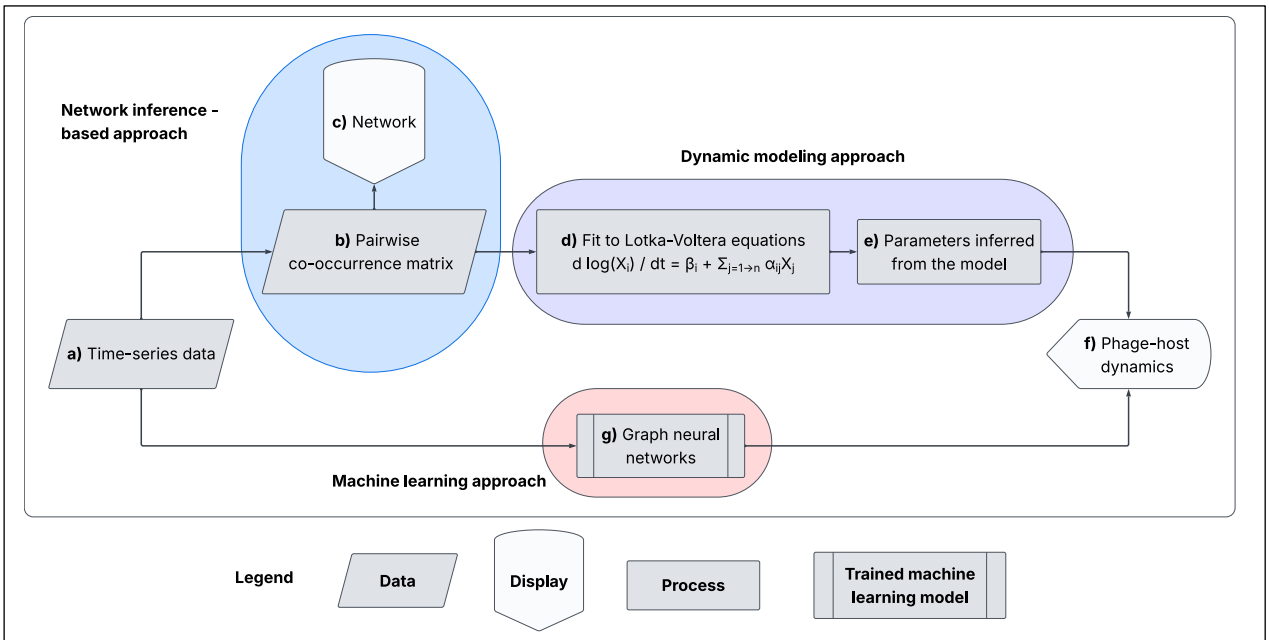

Figure 5: Linking metagenomic datasets to models. ***(a)*** *Any time-series data can be represented using an abundance matrix where the columns are different time points and the columns are abundance of different bacteria/phage population.* ***(b)*** *Applying co-occurrence/correlation analysis yields pairwise co-occurrence matrices. Here, every value represents the statistical dependence of bacteria/phage on other bacteria/phage in the environment.* ***(c)*** *The pairwise co-occurrence matrix can be used to construct a bipartite network with bacteria and phages as nodes. An edge between a bacteria and phage represents that the co-occurrence value is statistically significant denoting an interaction.* ***(d)*** *Once the information on which phage infects which bacteria is known, the time-series abundance data can be fit to Lotka-Voltera equations using linear regression.* ***(e)*** *Parameters learnt from the model can be used to* ***(f)*** *interpret the phage-host dynamics.* ***(g)*** *Alternately, graph neural networks (GNN) can also be trained directly on abundance and interaction data to predict interactions and model dynamics.*

Although most of the following methods were originally developed for studying microbial communities rather than phage systems, they can be readily applied to phage datasets by simply substituting the appropriate input data.

Network inference-based approach

Faust et al. (65) discuss one straightforward approach for predicting relationships from metagenomic data referred to as the 'checkerboard approach' (66). The basic principle behind this approach is that when two species co-occur or exhibit similar abundance patterns across multiple samples or time points, a positive relationship is predicted. Conversely, if they show mutual exclusion, a negative relationship is indicated.

An advanced form of this approach is network inference, widely used to build regulatory gene networks from gene expression datasets (67–69). These methods predict pairwise relationships based on the co-occurrence patterns and statistical dependencies in the abundance data (**Fig. 5b**). The statistical significance of these interactions can be assessed by permuting the dataset. Finally, the significant pairwise relationships are combined to form the required interaction networks (**Fig. 5c**). These methods, though potentially useful, have been applied in very few phage-host studies: *Milns et al.* (70) used a network inference approach, Bayesian networks, to infer avian ecological networks from species and habitat abundance alone. *Countinho et al.* (71) adapted a network inference based on co-occurrence associations for host prediction of phages.

However, interpreting the ecological significance of these relationships remains challenging. Multiple biological mechanisms could explain the network patterns, and time-series metagenomic data alone cannot distinguish among them. A positive correlation, for example, may reflect co-occurrence within shared habitats or biofilms, co-colonization, or overlapping ecological niches, among other possibilities. Conversely, negative correlations may arise from competitive interactions, superinfection exclusion, or similar antagonistic processes. Moreover, associations can exhibit temporal delays; an increase in one population's abundance may precede a decline in another. Therefore, rigorous statistical testing like bootstrapping datasets to determine statistical significance and experimental validation (65, 72–74) are currently needed to determine the true mechanisms underlying these observed patterns.

Dynamic modeling approaches

A dynamic model describes how the abundances of community members change over time, typically using systems of differential equations or Boolean functions. These mathematical representations capture how phage and host populations evolve over time as a function of their interactions and the underlying ecological and biological parameters. While dynamic modeling has a long and rich history in population ecology, relatively few studies have applied it directly to metagenomic datasets. *Mounier et al.* (75) studied microbial interactions in a cheese microbiome by modelling the microbial community using a multispecies Lotka-Volterra model. This dynamic model quantifies how the growth rate of one species depends linearly on the abundances of others in the community, effectively capturing the influence of interspecies interactions on population growth. The abundance data for the cheese microbiome community were generated by sampling daily for 21 days. Then the interactions between microbes can be determined by fitting this time-series abundance data to the multilinear regression equation of the multispecies Lotka-Volterra

model (**Fig. 5d-e**). This approach predicted positive and negative interactions between bacteria and yeast in the cheese community.

Hoffmann et al. (76) used a similar approach to fit the Lotka-Volterra model to rank-abundance data derived from metagenomic sampling of two marine phage communities. By integrating empirical time-series data with theoretical modeling, they uncovered dynamic phage–host population cycles. Their analysis indicated that phage and host populations alternate between prolonged periods of low abundance and short-lived bloom events (**Fig. 5f**). Phage numbers increase rapidly following host blooms, consistent with the Kill-the-Winner (KtW) hypothesis, and then decline sharply, stabilizing at very low levels until the subsequent host resurgence. This cyclical dominance of different host strains, each subsequently decimated by its corresponding phage, was later confirmed at the strain level (77).

Stein et al. (78) and *White et al.* (79) extended the use of this method to time-series metagenomics datasets. However, such applications also revealed limitations inherent to the Lotka–Volterra (LV) approach. Although more elaborate dynamic models can represent non-linear and multi-level relationships, the LV framework struggles when variables are highly correlated, making it difficult to isolate the effect of individual interactions. Moreover, the standard LV formulation assumes deterministic behavior and therefore fails to account for the stochasticity that characterizes real-world ecological systems (55). However, recent advances have addressed some of these limitations: for example, the SgLV-EKF algorithm (80) introduces stochasticity into the LV model, while LIMIT (81) employs statistical regularization techniques to mitigate the issue of correlation among variables, thereby enhancing the interpretability and robustness of inferred ecological interactions.

Machine learning approaches

An alternative approach to mechanistic mathematical models is the use of machine learning and deep learning algorithms (**Fig. 5g**). Artificial neural network frameworks have been used to study cause-and-effect relationships from abundance datasets in microbial ecology (82–84). Neural networks are advanced computational models that can automatically learn the relationship between variables in a dataset from the input data. This is typically achieved by training the model on labeled datasets, where known inputs and outputs are used to infer predictive relationships. This is called as supervised learning. In contrast, unsupervised learning approaches do not rely on labeled training data, instead, they identify intrinsic patterns, structures, or representations within the data itself. In these methods, the algorithm organizes samples based on shared features, revealing latent clusters, gradients, or interaction signatures without prior assumptions about the underlying relationships. Such approaches are particularly well-suited for metagenomic and ecological datasets, where labeled examples are scarce and system complexity is high, enabling the discovery of emergent ecological states and infection strategies.

Andersen et al. (82) used a type of artificial neural network called the graph neural networks (GNN) to model microbial community dynamics from time-series 16S rRNA sequencing data. GNNs represent microbial communities as graphs where nodes are taxa and edges capture ecological relationships, like co-occurrence patterns, then iteratively update each taxon representations by aggregating information from neighbors through graph convolution. The

authors used a dataset of 4709 samples from 24 full-scale wastewater treatment plants (WWTPs) to train, validate, and test their GNN model. This model had three neural network components: one for learning the interaction strengths between microbes in the community, another for capturing the effect of time on the microbes, and a final component that combines all learnt information. This model was able to predict species dynamics up to 10 timepoints in the future. This means that the GNN model successfully understood the community dynamics from the dataset. A key limitation of this method is that, despite its ability to capture complex dynamics, interpreting the underlying mechanisms driving those dynamics is difficult due to the neural network's inherent "black box" nature.

Studying phage-host lifestyles from time-series metagenomics data

Mathematical models have been used to describe the mechanistic foundations of phage–host interactions and the diverse lifestyles that shape their dynamics. The basis for constructing mathematical models comes from culture dependent methods like phage infection assays (46), continuous culture experiments (15, 18), one-step growth curves, and adsorption assays (19). With the increasing availability of time-series metagenomic data, an important question has emerged: Given an observed data of phage and host abundance, can we infer their lifestyles?. Addressing this question involves bringing together mathematical models and observed data. When viral lifestyles can be simulated using mechanistic models, observed time-series data can, in turn, be used to identify the model parameters and infection strategies that are most consistent with those observations. This approach is referred to as simulation-based inference (SBI). The core idea is simple: if a mathematical model can generate time-series dynamics that resemble real phage–host abundance patterns, then the parameters and assumptions underlying that simulation represent plausible explanations for the observed system (85). Rather than fitting a single model trajectory to the data, SBI explores a wide range of models and parameter combinations to determine which scenarios are consistent with the observed dynamics. It is important to note that SBI has been used in biology to infer parameters from observed data. For example, this approach was successfully used to study Batesian mimicry in butterflies. This is a phenomenon where the non-toxic butterflies evolve towards a wing-color and pattern similar to that of a toxic species to avoid predation. This evolutionary process can be captured using agent-based models that simulate how wing patterns change over time under selective pressure. Simulation-based inference serves as a bridge between such mechanistic models and observed time-series data describing the progression of mutations toward mimicry. By comparing simulated evolutionary trajectories with empirical observations, this approach has been used to infer key biological quantities, such as the mutation rates required to produce the observed patterns of color change (85).

In practice, this approach begins by defining a set of candidate phage–host models that represent different viral lifestyles, such as lytic, lysogenic, or chronic infection strategies, different objectives, such as studying the flow of resources from microbes to phages (select resource-based trophic models), studying the spread of infection (select epidemiological models), studying stochastic elements like mutation or evolution of phages and hosts (select stochastic models). Since only the models involving differential equations can be used in this approach, network-based models and metacommunity models cannot be utilized. Each model is then simulated repeatedly across

broad ranges of biologically plausible parameter values, producing an ensemble of synthetic time-series datasets. These simulated trajectories are sampled on the same temporal resolution and duration as the observed data, ensuring that differences between simulated and empirical dynamics reflect underlying mechanisms rather than mismatched timescales.

Once simulations are generated, the next step is to compare them to the observed time series. Rather than requiring an exact match, simulation-based inference evaluates similarity using summary statistics or distance metrics that capture key dynamical features, such as growth rates, oscillatory behavior, lag times, or correlations between phage and host abundances. Simulations that closely reproduce these features are retained, while those that fail to capture the observed patterns are discarded or down-weighted.

By examining the parameter values and model structures associated with successful simulations, SBI yields a set of plausible viral lifestyles and infection strategies that could have generated the observed data. Importantly, this approach does not aim to identify a single “best” model, but instead characterizes a range of mechanisms consistent with the data, naturally accounting for uncertainty and variability in both biological processes and measurements. This approach is limited by the mathematical models available in the literature. Hence when the observed data does not match with any simulated data from all candidate models, it means that the underlying viral lifestyles in the environment under study has not been established.

Available time-series datasets

This method can be applied to any time-series metagenomics datasets. Alternatively, any datasets already available in the literature can also be used. These are some of the available time-series datasets across different systems. A twenty year time-series metagenomic dataset of a Freshwater lake (8, 9), an aquatic sampling series from East Sound Fjord which consists of 133 time points sampled every 4 hours for 22 days, a longitudinal infant gut virome dataset (10) studying preterm infants over the first eleven weeks of life, the TEDDY dataset (11) studying phage-bacteria dynamics in the early life of gut and the dataset from the study comparing the human gut virome of children and their mother (12), 610 samples of marine metagenomics dataset comprising of samples collected from GEOTRACES cruise from multiple geographic locations and long-term sampling sites BATS and ALOHA (86), a metatranscriptomic time-series sampling of Sargasso sea every 4 hours for a week (87, 88).

Preprocessing

Once a time-series dataset is considered, all the phages in the dataset needs to be paired-up with a host. This step establishes phage-host linkages (which phages infect which hosts) from the observed phage and host abundance data (**Fig. 6**). For metagenomic data, the traditional, sequence-based approach is to use tools such as iPHoP (89) and others (90) that bioinformatically predict phage-host linkages based on their genomic signatures. However, these predictions are based only on genomic information from the samples and do not use the observed host and phage densities across time points, which is the most important feature of time-series datasets. Conversely, network inference-based methods, such as those used by *Milns et al.* (70) and *Coutinho et al.* (71), leverage time-series observations to predict phage-host linkages. Both approaches can be used such that the sequence-based predictions provide additional support for

the network inference-based predictions. The phage-host predictions obtained in this way form the basis for the subsequent steps.

Simulation

The first step in the approach is to define the question of interest. We then begin by identifying models that suit the objective. For example, if we want to study the phage-host system as obligately lytic then we can choose models that only model lytic infection (this applies for other lifestyles as well). Conversely if we want to study the phage-host system as a heterogeneous system with phages expressing multiple infection strategies, then a heterogenous model would need to selected as a candidate. If we are unaware of the lifestyle, which is often the case with metagenomic samples, it is best to have models from multiple categories. For each model, we can define ranges for parameters such as growth rates, adsorption rates, burst sizes, or induction probabilities, informed by prior studies or experimental constraints. The mathematical models exhibit different behavior (multiple steady states) under different parameter settings, so it is necessary to simulate across a range of parameter values. In a typical study, we would simulate an ensemble of time-series phage-host abundance data from all the candidate models. Simulations should match the temporal resolution and duration of the observed dataset. Next, we would compute the summary statistics or dynamical descriptors from both simulated and observed time series, such as trends, variability, lag structure, or correlations.

Comparison

The summary statistics from our comparisons allows us to compare simulated and observed data. We quantify similarity between observed and simulated data and retain the simulations that reproduce key aspects of the observed dynamics and analyze the associated parameter distributions to infer likely viral lifestyles. We use the inferred parameter regimes and model structures to generate mechanistic hypotheses about phage–host interactions, ecological conditions, and infection strategies.

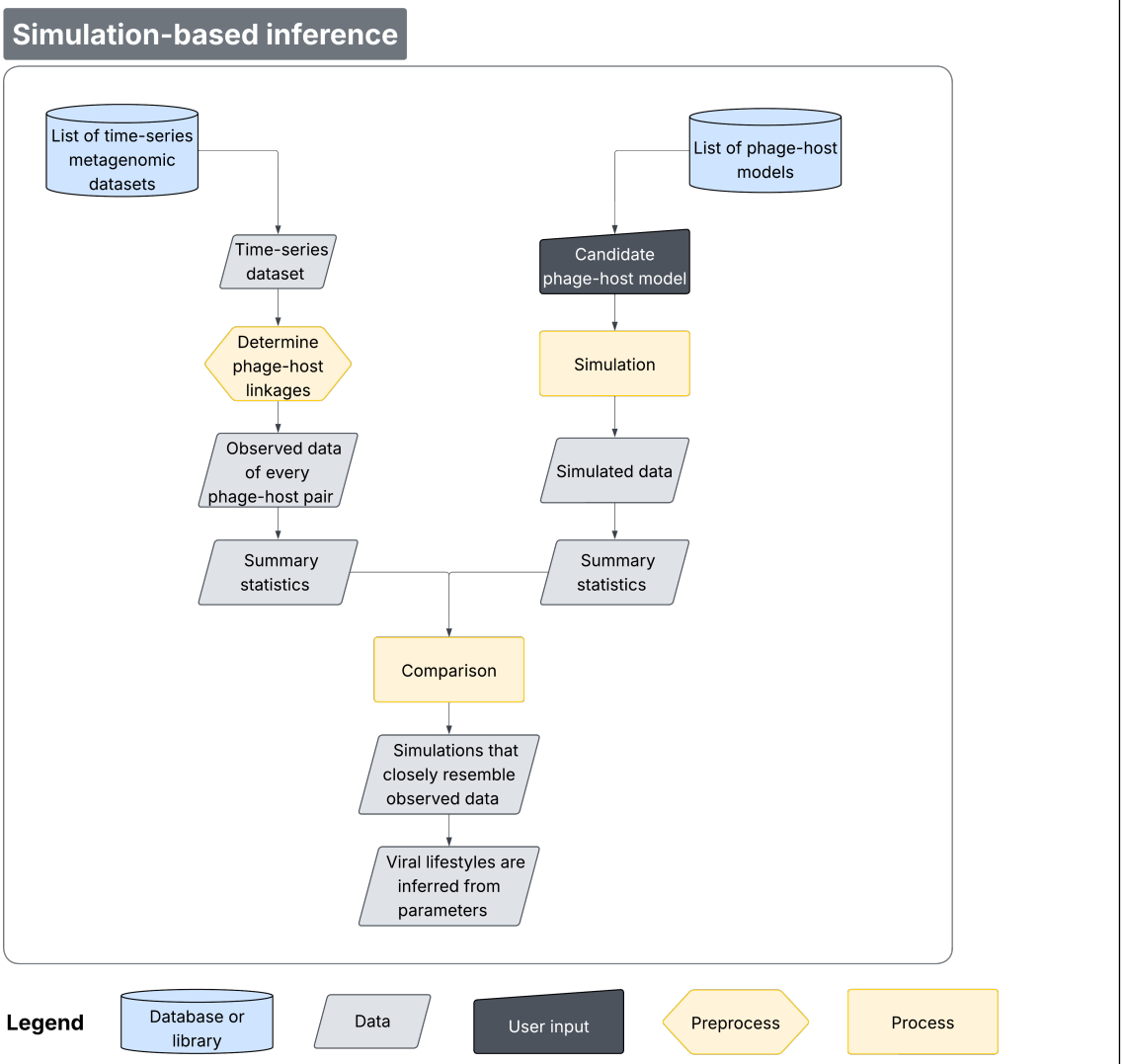

Figure 6: Conceptual workflow for inferring phage–host lifestyles from time-series metagenomic data. *Time-series metagenomic data are first organized into abundance matrices, from which co-occurrence or correlation analyses are applied to infer putative phage–host linkages and construct interaction networks. To move beyond static associations and investigate dynamic phage–host interactions, the temporal data are compared against predictions from candidate mathematical models. As an initial screening step, correlation tests are performed between model-generated abundance trajectories and observed time-series data to identify models that capture key temporal patterns. Models exhibiting strong agreement are then further evaluated using parameter sweeping or regression-based fitting to refine parameter estimates and assess explanatory power. The best-fitting models provide mechanistic interpretations of phage–host dynamics and infection strategies. When statistic-based approaches fail to capture model fits, machine learning can be employed to capture complex or nonlinear patterns present in both theoretical and observed data.*

Many of the aforementioned approaches have rarely been used to study phage lifestyles. Historically, this was due to a lack of data, but in the last 10 years, there has been a surge in metagenomics studies that have generated massive genomic datasets, with fine-scale metadata including temporal, spatial, environmental, and biogeochemical resolution. Given this high volume of data, it is time to focus research efforts on developing data science methods to obtain a deeper

743 understanding of the datasets, specifically for building hypotheses associated with predicting
744 phage lifestyles.

745 **Conclusion**

746 Mathematical models of phage–host dynamics have played a crucial role in uncovering the
747 principles governing coexistence, diversity, and ecological stability in microbial communities.
748 From chemostat-based resource–predator frameworks to stochastic, epidemiological, and
749 ecological models, such as Kill-the-Winner and Piggyback-the-Winner, each theoretical advance
750 has added layers of realism to our understanding of viral lifestyles. Yet, as George Box famously
751 remarked, “all models are wrong, but some are useful”, current models fall short of capturing the
752 full complexity of natural environments while explaining specific aspects of the dynamics.

753 The emergence of high-resolution metagenomic and time-series datasets provides a unique
754 opportunity to bridge this gap. By integrating genomic signals with dynamic modeling, researchers
755 can begin to move beyond static descriptions of viral ecology toward predictive frameworks that
756 link viral life-history traits with real ecological outcomes. Such integration is essential not only for
757 testing existing ecological theories but also for refining models to account for the fluidity of
758 infection strategies *in situ*. Looking ahead, the most fruitful progress will likely come from hybrid
759 approaches that combine ecological modeling, evolutionary theory, and multi-omics datasets to
760 resolve the ‘viral dark matter’ that still obscures much of phage ecology. In this way, phage–host
761 modeling can advance from theoretical abstraction toward an empirically grounded science that
762 illuminates the roles of viruses in shaping microbial ecosystems and global biogeochemical
763 cycles.

764

765

766

767

768

769

770 **Acknowledgements**

771 This research was supported by the National Institute of General Medical Sciences of the National
772 Institutes of Health under award (R35GM143024) and by a National Science Foundation award
773 (DBI-2047598). JCK was supported by a National Science Foundation Graduate Research
774 Fellowship under award number 2137424. MVL was supported by the Department of
775 Bacteriology's William H. Peterson Graduate Fellowship. We thank members of the
776 Anantharaman lab for project discussion and feedback on the manuscript.

777 **Competing interests**

778 The authors declare no competing interests.

References

1. Breitbart M, Bonnain C, Malki K, Sawaya NA. 2018. Phage puppet masters of the marine microbial realm. Nat Microbiol 3:754–766.

2. Wieczynski DJ, Yoshimura KM, Denison ER, Geisen S, DeBruyn JM, Shaw AJ, Weston DJ, Pelletier DA, Wilhelm SW, Gibert JP. 2023. Viral infections likely mediate microbial controls on ecosystem responses to global warming. FEMS Microbiol Ecol 99:fiad016.

3. Hunter M, Fusco D. 2022. Superinfection exclusion: A viral strategy with short-term benefits and long-term drawbacks. PLOS Comput Biol 18:e1010125.

4. Susskind MM, Wright A, Botstein D. 1971. Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*. Virology 45:638–652.

5. Díaz-Muñoz SL, Koskella B. 2014. Chapter Four - Bacteria–Phage Interactions in Natural Environments, p. 135–183. *In* Sariaslani, S, Gadd, GM (eds.), Advances in Applied Microbiology. Academic Press.

6. Roux S, Coclet C. 2025. Viromics approaches for the study of viral diversity and ecology in microbiomes. Nat Rev Genet 1–15.

7. Kosmopoulos JC, Anantharaman K. 2024. Microbial and Viral Ecology Analysis for Metagenomic Data. arXiv:2407.08858. arXiv https://doi.org/10.48550/arXiv.2407.08858.

8. Zhou Z, Tran PQ, Martin C, Rohwer RR, Baker BJ, McMahon KD, Anantharaman K. 2025. Unravelling viral ecology and evolution over 20 years in a freshwater lake. Nat Microbiol 10:231–245.

9. Rohwer RR, Hale RJ, Vander Zanden MJ, Miller TR, McMahon KD. 2023. Species invasions shift microbial phenology in a two-decade freshwater time series. Proc Natl Acad Sci 120:e2211796120.

10. Kaelin EA, Rodriguez C, Hall-Moore C, Hoffmann JA, Linneman LA, Ndao IM, Warner BB, Tarr PI, Holtz LR, Lim ES. 2022. Longitudinal gut virome analysis identifies specific viral signatures that precede necrotizing enterocolitis onset in preterm infants. Nat Microbiol 7:653–662.

11. Tisza MJ, Lloyd RE, Hoffman K, Smith DP, Rewers M, Javornik Cregeen SJ, Petrosino JF. 2025. Longitudinal phage–bacteria dynamics in the early life gut microbiome. Nat Microbiol 10:420–430.

12. Walters WA, Granados AC, Ley C, Federman S, Stryke D, Santos Y, Haggerty T, Sotomayor-Gonzalez A, Servellita V, Ley RE, Parsonnet J, Chiu CY. 2023. Longitudinal

comparison of the developing gut virome in infants and their mothers. Cell Host Microbe 31:187-198.e3.

13. Weitz JS, Stock CA, Wilhelm SW, Bourouiba L, Coleman ML, Buchan A, Follows MJ, Fuhrman JA, Jover LF, Lennon JT, Middelboe M, Sonderegger DL, Suttle CA, Taylor BP, Frede Thingstad T, Wilson WH, Eric Wommack K. 2015. A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. ISME J 9:1352–1364.

14. Gulbudak H, Weitz JS. 2019. Heterogeneous viral strategies promote coexistence in virus-microbe systems. J Theor Biol 462:65–84.

15. Levin BR, Stewart FM, Chao L. 1977. Resource-Limited Growth, Competition, and Predation: A Model and Experimental Studies with Bacteria and Bacteriophage. Am Nat 111:3–24.

16. Stewart FM, Levin BR. 1984. The population biology of bacterial viruses: Why be temperate. Theor Popul Biol 26:93–117.

17. Thingstad TF. 2000. 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.

18. Thingstad T, Lignell R. 1997. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. Aquat Microb Ecol 13:19–27.

19. Dey R, Coenen AR, Solonenko NE, Burris MN, Mackey AI, Galasso J, Sun CL, Demory D, Muratore D, Beckett SJ, Sullivan MB, Weitz JS. 2025. Emergent higher-order interactions enable coexistence in phage-bacteria community dynamics. bioRxiv https://doi.org/10.1101/2025.05.15.651590.

20. Weitz JS, Li G, Gulbudak H, Cortez MH, Whitaker RJ. 2019. Viral invasion fitness across a continuum from lysis to latency†. Virus Evol 5:vez006.

21. Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobián-Güemes AG, Coutinho FH, Dinsdale EA, Felts B, Furby KA, George EE, Green KT, Gregoracci GB, Haas AF, Haggerty JM, Hester ER, Hisakawa N, Kelly LW, Lim YW, Little M, Luque A, McDole-Somera T, McNair K, de Oliveira LS, Quistad SD, Robinett NL, Sala E, Salamon P, Sanchez SE, Sandin S, Silva GGZ, Smith J, Sullivan C, Thompson C, Vermeij MJA, Youle M, Young C, Zgliczynski B, Brainard R, Edwards RA, Nulton J, Thompson F, Rohwer F. 2016. Lytic to temperate switching of viral communities. 7595. Nature 531:466–470.

22. Voigt E, Rall BC, Chatzinotas A, Brose U, Rosenbaum B. 2021. Phage strategies facilitate bacterial coexistence under environmental variability. PeerJ 9:e12194.

23. Kerr B, West J, Bohannan BJM. 2008. Bacteriophages: models for exploring basic principles of ecology, p. 31–63. *In* Abedon, ST (ed.), Bacteriophage Ecology: Population Growth, Evolution, and Impact of Bacterial Viruses. Cambridge University Press, Cambridge.

24. Chaudhry WN, Pleška M, Shah NN, Weiss H, McCall IC, Meyer JR, Gupta A, Guet CC, Levin BR. 2018. Leaky resistance and the conditions for the existence of lytic bacteriophage. PLOS Biol 16:e2005971.

25. Thingstad TF, Våge S, Storesund JE, Sandaa R-A, Giske J. 2014. A theoretical analysis of how strain-specific viruses can control microbial species diversity. Proc Natl Acad Sci 111:7813–7818.

26. Stewart FM, Levin BR. 1973. Partitioning of Resources and the Outcome of Interspecific Competition: A Model and Some General Considerations. Am Nat 107:171–198.

27. Chao L, Levin BR, Stewart FM. 1977. A Complex Community in a Simple Habitat: An Experimental Study with Bacteria and Phage. Ecology 58:369–378.

28. Levin BR, Moineau S, Bushman M, Barrangou R. 2013. The Population and Evolutionary Dynamics of Phage and Bacteria with CRISPR–Mediated Immunity. PLoS Genet 9:e1003312.

29. Weitz JS, Dushoff J. 2008. Alternative stable states in host–phage dynamics. Theor Ecol 1:13–19.

30. Chaudhry W, Vega N, Govindan A, Garcia R, Lee E, McCall I, Levin B. 2019. The population and evolutionary dynamics of bacteriophage: Why be temperate revisited. bioRxiv https://doi.org/10.1101/824235.

31. Levin BR, Antonovics J, Sharma H, Clarke BC, Partridge L. 1997. Frequency-dependent selection in bacterial populations. Philos Trans R Soc Lond B Biol Sci 319:459–472.

32. Lwoff A. 1953. LYSOGENY1. Bacteriol Rev 17:269–337.

33. Miller RV, Ripp SA. 2002. Chapter 8 - Pseudolysogeny: A Bacteriophage Strategy for Increasing Longevity in situ, p. 81–91. *In* Syvanen, M, Kado, CI (eds.), Horizontal Gene Transfer (Second Edition). Academic Press, London.

34. Beretta E, Kuang Y. 1998. Modeling and analysis of a marine bacteriophage infection. Math Biosci 149:57–76.

35. Beretta E, Kuang Y. 2001. Modeling and analysis of a marine bacteriophage infection with latency period. Nonlinear Anal Real World Appl 2:35–74.

882 36. Gandon S. 2016. Why Be Temperate: Lessons from Bacteriophage λ. Trends
883 Microbiol 24:356–365.

884 37. Li G, Cortez MH, Dushoff J, Weitz JS. 2020. When to be temperate: on the
885 fitness benefits of lysis vs. lysogeny. Virus Evol 6:veaa042.

886 38. Clifton SM, Whitaker RJ, Rapti Z. 2021. Temperate and chronic virus competition
887 leads to low lysogen frequency. J Theor Biol 523:110710.

888 39. Weitz JS, Hartman H, Levin SA. 2005. Coevolutionary arms races between
889 bacteria and bacteriophage. Proc Natl Acad Sci 102:9535–9540.

890 40. Maslov S, Sneppen K. 2015. Well-temperate phage: optimal bet-hedging against
891 local environmental collapses. Sci Rep 5:10523.

892 41. Berryhill BA, Gil-Gil T, Levin BR. 2025. The joint action of antibiotics,
893 bacteriophage, and the innate immune response in the treatment of bacterial infections.
894 Front Microbiol 16.

895 42. 1970. Mathematical Games - The fantastic combinations of John Conway's new
896 solitaire game "life" - M. Gardner - 1970.

897 43. Ruan C, Vinod DP, Johnson DR. 2025. Phage-mediated peripheral kill-the-winner
898 facilitates the maintenance of costly antibiotic resistance. Nat Commun 16:5839.

899 44. Rudge TJ, Steiner PJ, Phillips A, Haseloff J. 2012. Computational Modeling of
900 Synthetic Microbial Biofilms. ACS Synth Biol 1:345–352.

901 45. Flores CO, Meyer JR, Valverde S, Farr L, Weitz JS. 2011. Statistical structure of
902 host–phage interactions. Proc Natl Acad Sci 108:E288–E297.

903 46. Weitz JS, Poisot T, Meyer JR, Flores CO, Valverde S, Sullivan MB, Hochberg
904 ME. 2013. Phage–bacteria infection networks. Trends Microbiol 21:82–91.

905 47. Flores CO, Valverde S, Weitz JS. 2013. Multi-scale structure and geographic
906 drivers of cross-infection within marine bacteria and phages. 3. ISME J 7:520–532.

907 48. Beckett SJ, Williams HTP. 2013. Coevolutionary diversification creates nested-
908 modular structure in phage–bacteria interaction networks. Interface Focus 3:20130033.

909 49. Rohwer F, Thurber RV. 2009. Viruses manipulate the marine environment. Nature
910 459:207–212.

911 50. Payet JP, Suttle CA. 2013. To kill or not to kill: The balance between lytic and
912 lysogenic viral infection is driven by trophic status. Limnol Oceanogr 58:465–474.

913 51. Silveira CB, Rohwer FL. 2016. Piggyback-the-Winner in host-associated
914 microbial communities. 1. Npj Biofilms Microbiomes 2:1–5.

52. Paterson JS, Smith RJ, McKerral JC, Dann LM, Launer E, Goonan P, Kleinig T, Fuhrman JA, Mitchell JG. 2019. A hydrocarbon-contaminated aquifer reveals a Piggyback-the-Persistent viral strategy. FEMS Microbiol Ecol 95:fiz116.

53. Kosmopoulos JC, Anantharaman K. 2023. To cull or kill. Nat Ecol Evol 7:1752–1753.

54. Gilbert NE, Muratore D, Gochev CS, LeCleir GR, Cagle SM, Pound HL, Sun CL, Carrillo A, Ndlovu KS, Maidanik I, Coenen AR, Chittick L, DeBruyn JM, Buchan A, Lindell D, Sullivan MB, Weitz JS, Wilhelm SW. 2025. Seasonal enhancement of the viral shunt catalyzes a subsurface oxygen maximum in the Sargasso Sea. Nat Commun 17:352.

55. Xue C, Goldenfeld N. 2017. Coevolution Maintains Diversity in the Stochastic "Kill the Winner" Model. Phys Rev Lett 119:268101.

56. Berngruber TW, Lion S, Gandon S. 2015. Spatial Structure, Transmission Modes and the Evolution of Viral Exploitation Strategies. PLOS Pathog 11:e1004810.

57. Williams HT. 2013. Phage-induced diversification improves host evolvability. BMC Evol Biol 13:17.

58. Coello-Camba A, Diaz-Rua R, Duarte CM, Irigoien X, Pearman JK, Alam IS, Agusti S. 2020. Picocyanobacteria Community and Cyanophage Infection Responses to Nutrient Enrichment in a Mesocosms Experiment in Oligotrophic Waters. Front Microbiol 11.

59. Gons HJ, Ebert J, Hoogveld HL, van den Hove L, Pel R, Takkenberg W, Woldringh CJ. 2002. Observations on cyanobacterial population collapse in eutrophic lake water. Antonie Van Leeuwenhoek 81:319–326.

60. Silveira CB, Luque A, Rohwer F. 2021. The landscape of lysogeny across microbial community density, diversity and energetics. Environ Microbiol 23:4098–4111.

61. Keen EC. 2014. Tradeoffs in bacteriophage life histories. Bacteriophage 4:e28365.

62. Faust K, Lahti L, Gonze D, de Vos WM, Raes J. 2015. Metagenomics meets time series analysis: unraveling microbial community dynamics. Curr Opin Microbiol 25:56–66.

63. López-García P, Gutiérrez-Preciado A, Krupovic M, Ciobanu M, Deschamps P, Jardillier L, López-Pérez M, Rodríguez-Valera F, Moreira D. 2023. Metagenome-derived virus-microbe ratios across ecosystems. ISME J 17:1552–1563.

64. Parikka KJ, Le Romancer M, Wauters N, Jacquet S. 2017. Deciphering the virus-to-prokaryote ratio (VPR): insights into virus–host relationships in a variety of ecosystems. Biol Rev 92:1081–1100.

65. Faust K, Raes J. 2012. Microbial interactions: from networks to models. Nat Rev Microbiol 10:538–550.

66. Cody ML, Diamond JM. 1975. Ecology and Evolution of Communities. Harvard University Press.

67. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. 2010. Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. PLOS ONE 5:e12776.

68. Larjo A, Shmulevich I, Lähdesmäki H. 2013. Structure Learning for Bayesian Networks as Models of Biological Networks, p. 35–45. *In* Mamitsuka, H, DeLisi, C, Kanehisa, M (eds.), Data Mining for Systems Biology: Methods and Protocols. Humana Press, Totowa, NJ.

69. T. Veiga DF, Dutta B, Balázsi G. 2010. Network inference and network response identification: moving genome-scale data to the next level of biological discovery. Mol Biosyst 6:469–480.

70. Milns I, Beale CM, Smith VA. 2010. Revealing ecological networks using Bayesian network inference algorithms. Ecology 91:1892–1899.

71. Coutinho FH, Silveira CB, Gregoracci GB, Thompson CC, Edwards RA, Brussaard CPD, Dutilh BE, Thompson FL. 2017. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. Nat Commun 8:15955.

72. Alrasheed H, Jin R, Weitz JS. 2019. Caution in inferring viral strategies from abundance correlations in marine metagenomes. Nat Commun 10:501.

73. Coutinho FH, Silveira CB, Gregoracci GB, Thompson CC, Edwards RA, Brussaard CPD, Dutilh BE, Thompson FL. 2019. Reply to: Caution in inferring viral strategies from abundance correlations in marine metagenomes. Nat Commun 10:502.

74. Coenen AR, Weitz JS. 2018. Limitations of Correlation-Based Inference in Complex Virus-Microbe Communities. mSystems 3:10.1128/msystems.00084-18.

75. Mounier J, Monnet C, Vallaeys T, Arditi R, Sarthou A-S, Hélias A, Irlinger F. 2008. Microbial Interactions within a Cheese Microbial Community. Appl Environ Microbiol 74:172–181.

76. Hoffmann KH, Rodriguez-Brito B, Breitbart M, Bangor D, Angly F, Felts B, Nulton J, Rohwer F, Salamon P. 2007. Power law rank–abundance models for marine phage communities. FEMS Microbiol Lett 273:224–228.

77. Rodriguez-Brito B, Li L, Wegley L, Furlan M, Angly F, Breitbart M, Buchanan J, Desnues C, Dinsdale E, Edwards R, Felts B, Haynes M, Liu H, Lipson D, Mahaffy J, Martin-Cuadrado AB, Mira A, Nulton J, Pašić L, Rayhawk S, Rodriguez-Mueller J, Rodriguez-Valera F, Salamon P, Srinagesh S, Thingstad TF, Tran T, Thurber RV, Willner D, Youle M, Rohwer F. 2010. Viral and microbial community dynamics in four aquatic environments. ISME J 4:739–751.

78. Stein RR, Bucci V, Toussaint NC, Buffie CG, Rätsch G, Pamer EG, Sander C, Xavier JB. 2013. Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota. PLOS Comput Biol 9:e1003388.

79. James Robert White. 2010. Novel Methods for Metagenomic Analysis. PhD Thesis. University of Maryland.

80. Alshawaqfeh M, Serpedin E, Younes AB. 2017. Inferring microbial interaction networks from metagenomic data using SgLV-EKF algorithm. BMC Genomics 18:228.

81. Fisher CK, Mehta P. 2014. Identifying Keystone Species in the Human Gut Microbiome from Metagenomic Timeseries Using Sparse Linear Regression. PLOS ONE 9:e102451.

82. Andersen KS, Zhao K, Agerskov A de L, Sørensen CB, Holmager TJ, Nierychlo M, Peces M, Guo C, Nielsen PH. 2025. Predicting microbial community structure and temporal dynamics by using graph neural network models. Nat Commun 16:9124.

83. Pan J, You W, Lu X, Wang S, You Z, Sun Y. 2023. GSPHI: A novel deep learning model for predicting phage-host interactions via multiple biological information. Comput Struct Biotechnol J 21:3404–3413.

84. Larsen PE, Field D, Gilbert JA. 2012. Predicting bacterial community assemblages using an artificial neural network approach. Nat Methods 9:621–625.

85. Sankaran K, Holmes SP. 2023. Generative Models: An Interdisciplinary Perspective. Annu Rev Stat Its Appl 10:325–352.

86. Biller SJ, Berube PM, Dooley K, Williams M, Satinsky BM, Hackl T, Hogle SL, Coe A, Bergauer K, Bouman HA, Browning TJ, De Corte D, Hassler C, Hulston D, Jacquot JE, Maas EW, Reinthaler T, Sintes E, Yokokawa T, Chisholm SW. 2018. Marine microbial metagenomes sampled across space and time. 1. Sci Data 5:180176.

87. Muratore D, Gilbert NE, LeCleir GR, Wilhelm SW, Weitz JS. 2025. Diel partitioning in microbial phosphorus acquisition in the Sargasso Sea. Proc Natl Acad Sci 122:e2410268122.

88. Gilbert NE, Muratore D, Gochev CS, LeCleir GR, Cagle SM, Pound HL, Sun CL, Carillo A, Ndlovu KS, Maidanik I, Coenen AR, Chittick L, DeBruyn JM, Buchan A, Lindell D, Sullivan MB, Weitz JS, Wilhelm SW. 2025. Seasonal Enhancement of the Viral Shunt Catalyzes a Subsurface Oxygen Maximum in the Sargasso Sea. bioRxiv https://doi.org/10.1101/2025.01.23.634377.

89. Roux S, Camargo AP, Coutinho FH, Dabdoub SM, Dutilh BE, Nayfach S, Tritt A. 2023. iPHoP: An integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. PLOS Biol 21:e3002083.

90. Nie W, Qiu T, Wei Y, Ding H, Guo Z, Qiu J. 2024. Advances in phage–host interaction prediction: in silico method enhances the development of phage therapies. Brief Bioinform 25:bbae117.

91. Ai D, Chen L, Xie J, Cheng L, Zhang F, Luan Y, Li Y, Hou S, Sun F, Xia LC. 2023. Identifying local associations in biological time series: algorithms, statistical significance, and applications. Brief Bioinform 24:bbad390.

92. Song C, Levine JM. 2025. Rigorous validation of ecological models against empirical time series. Nat Ecol Evol 1–14.