



# Nutrient Loading and Viral Memory Drive Accumulation of Restriction Modification Systems in Bloom-Forming Cyanobacteria

Spiridon E. Papoulis,<sup>a</sup>  Steven W. Wilhelm,<sup>a</sup> David Talmy,<sup>a</sup> Erik R. Zinser<sup>a</sup>

<sup>a</sup>Department of Microbiology, University of Tennessee, Knoxville, Tennessee, USA

**ABSTRACT** The mechanisms driving cyanobacterial harmful algal blooms (HABs) like those caused by *Microcystis aeruginosa* remain elusive, but improved defense against viral predation has been implicated for success in eutrophic environments. Our genus-level analyses of 139,023 genomes revealed that HAB-forming cyanobacteria carry vastly more restriction modification systems per genome (RMPG) than nearly all other prokaryotic genera, suggesting that viral defense is a cornerstone of their ecological success. In contrast, picocyanobacteria that numerically dominate nutrient-poor systems have the fewest RMPG within the phylum *Cyanobacteria*. We used classic resource competition models to explore the hypothesis that nutrient enrichments drive ecological selection for high RMPG due to increased host-phage contact rate. These classic models, agnostic to the mechanism of defense, explain how nutrient loading can select for increased RMPG but, importantly, fail to explain the extreme accumulation of these defense systems. However, extreme accumulation of RMPG can be achieved in a novel “memory” model that accounts for a unique activity of restriction modification systems: the accidental methylation of viral DNA by the methyltransferase. The methylated virus “remembers” the RM defenses of its former host and can evade these defenses if they are present in the next host. This viral memory leads to continual RM system devaluation; RMs accumulate extensively because the benefit of each addition is diminished. Our modeling leads to the hypothesis that nutrient loading and virion methylation drive the extreme accumulation of RMPG in HAB-forming cyanobacteria. Finally, our models suggest that hosts with different RMPG values can coexist when hosts have unique sets of RM systems.

**IMPORTANCE** Harmful algal blooms (HABs), caused by cyanobacteria like *Microcystis aeruginosa*, are a global threat to water quality and use across the planet. Researchers have agreed that nutrient loading is a major contributor to HAB persistence. While we may understand the environmental conditions that cause HABs, we still struggle in identifying the mechanisms that explain why these organisms have a competitive edge against other, less ecologically hazardous organisms. Our interdisciplinary approach in microbiology, mathematical population modeling, and genomics allows us to use nearly 70 years of research in restriction modification systems to show that HAB-forming cyanobacteria are exceptional in their ability to defend against viruses, and this capacity is intimately tied to nutrient loading. Our hypothesis suggests that defense against viral predation is a fundamental pillar of cyanobacterial ecological strategy and an important contributor to HAB dynamics.

**KEYWORDS** *Microcystis*, harmful algal blooms, viral defense, eutrophication

The ecological success of harmful algal bloom (HAB)-forming species is not well understood despite the global threats of HABs (1). HAB-forming cyanobacteria, such as *Microcystis* spp., can produce toxins that threaten the health and safety of humans

**Citation** Papoulis SE, Wilhelm SW, Talmy D, Zinser ER. 2021. Nutrient loading and viral memory drive accumulation of restriction modification systems in bloom-forming cyanobacteria. mBio 12:e00873-21. <https://doi.org/10.1128/mBio.00873-21>.

**Editor** Mary Ann Moran, University of Georgia

**Copyright** © 2021 Papoulis et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to David Talmy, dtalmy@utk.edu, or Erik R. Zinser, ezinser@utk.edu.

This article is a direct contribution from Steven W. Wilhelm, a Fellow of the American Academy of Microbiology, who arranged for and secured reviews by Mya Breitbart, University of South Florida, and K. Eric Wommack, University of Delaware.

**Received** 9 April 2021

**Accepted** 23 April 2021

**Published** 1 June 2021

and animals interacting with contaminated water and can often damage local economies. It is well established that anthropogenic nutrient loading facilitates the proliferation of these harmful cyanobacteria, leading to the accumulation of dense phytoplankton biomass, as often witnessed in freshwater systems such as Lake Erie (North America) or Lake Taihu (China) (2, 3). Multiple studies have investigated the complex interactions between genetic potential, bottom-up (e.g., nutrient loading) and top-down (2, 4, 5) controls that lead to the ecological success of HAB-forming cyanobacteria. However, we still lack mechanistic frameworks that incorporate these different selective pressures together to elucidate the drivers of ecology and evolution of these domestic pests.

Viruses (bacteriophages or phages) are powerful and ubiquitous evolutionary drivers in the prokaryotic world (6). The lysis of microbial cells contributes to biogeochemical recycling via a process known as the “viral shunt” (7) and selects for genotypes resistant to viral infection. Antiviral innovations fall into two general classes: those that prevent virus adsorption at the cell envelope, e.g., through mutation of the virus receptor (8–10), and those that establish within the cytoplasm the ability to destroy the virus or kill the infected cell. Cytoplasmic defenses are widespread in prokaryotes (11, 12) and include CRISPR (13), argonauts (14), toxin-antitoxin systems (15), abortive infection (16), and BREX (17). While many of these cytoplasmic defenses have only recently been discovered, restriction modification (RM) systems have been investigated since the 1950s (18, 19).

RM systems galvanized the molecular biology revolution through their ability to cleave double-stranded DNA (dsDNA) at sequence-specific motifs. When expressed *in vivo*, the endonuclease (restriction enzyme) activity of the RM system can protect a potential host cell from dsDNA viruses that contain the specific sequences recognized by the RM. Individual RMs can reduce rates of infection by 2 to 6 orders of magnitude (20). Because of this antiviral effect, RMs can be thought of as innate immune systems whose targets are predetermined by the specified recognition motifs of the endonucleases. This contrasts with the “adaptive immunity” conferred by CRISPR-Cas systems that use information gathered from prior infections to provide targets for DNA cleavage (13).

Because the DNA motifs targeted by the endonuclease of RMs are short, usually spanning 6 bases (but can range from 4 to 14) (21), they are usually found in multiple sites of the phage DNA; however, they also often occur within the host’s much larger genome. To protect the cell’s genome from cleavage at these motifs, most RMs also provide DNA methyltransferase activity that methylates residues within the same target motif as the endonuclease. For type I to III RMs (see below), the endonuclease activity is specific for unmethylated DNA. Thus, the role of the methyltransferase is to block the endonuclease from cleaving host DNA, while leaving it free to attack incoming, unmethylated viral DNA. One important drawback to this defense system is that any viral DNA that escapes endonuclease attack long enough will be “immunized” by the methyltransferase (18–20, 22–25). Methylated viral progeny released from the cell will be protected from endonuclease activity if infecting a new cell with the same RM defense. Knowing the recognition sequence is critical in determining whether a methylated virus will be able to evade a host endonuclease. For example, viruses methylated at the motif GATATC would not be resistant to an endonuclease that targets GAATTC.

RMs fall into one of several classes based on protein structure and DNA target. In type II RMs, endonuclease and methyltransferase activities are in separate proteins that recognize DNA independently. Type I and III RMs involve separate proteins that complex together with or without a specificity unit, respectively. Types IIB, IIG, and IIH (collectively referred to as type IIG here) are single polypeptides with both activities covalently linked. Finally, type IV RMs are single endonucleases that cleave methylated, rather than unmethylated, DNA (26).

Previously, *Microcystis aeruginosa* was found to have the greatest fraction of its genome devoted to antiviral defenses relative to all other prokaryotes, suggesting that defense against viral predation is a cornerstone of its ecological success as a bloom-forming cyanobacterium (27). In the present study, we tested this hypothesis in an RM-specific context to determine

whether HABs are unique in their capacity to defend against foreign DNAs, such as viruses, relative to the rest of the prokaryotic world. Moreover, because of the consistency in the DNA cutting mechanism of RMs, we used mathematical models to explore the causal effect of eutrophication, a bottom-up control, on the selective pressures acting on the defense against viral predation, a top-down control. We begin with a bioinformatic assessment of the distribution of RM systems in prokaryotes. We go on to highlight interesting ecological patterns in RM systems per genome (RMPG) in diverse cyanobacteria. Finally, we introduce three mathematical population models used to evaluate ecological and molecular mechanisms of RM selection in contrasting nutrient conditions.

## RESULTS

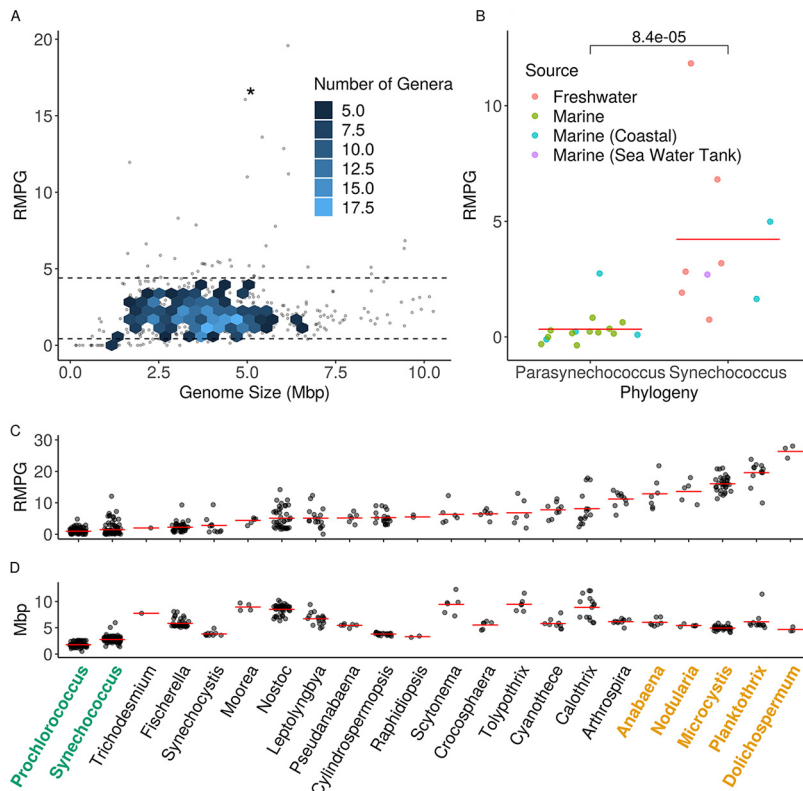
**Restriction modification distributions in prokaryotes.** We developed a custom pipeline to quantify the RMPG among the 139,023 high-quality bacterial and archaeal genomes available in the RefSeq database (see Data Set S1 in the supplemental material). To mitigate the disproportionate contribution by some taxa (e.g., genomes from overrepresented genera such as *Shewanella* and *Escherichia*) to the prokaryote mean, we aggregated data at the genus level (see Data Set S1). For the 2,522 bacterial and archaeal genera analyzed, a mean value of 2.475 and a median of 2.0 RMPG were observed, with 5th and 95th quantiles of 0 and 6 RMPG, respectively. Further restriction of the data set to include only genera with 5 or more sequenced genomes yielded mean, median, and 5th and 95th percent quantiles of 2.17, 1.91, 0.427, and 4.40 RMPG, respectively (Fig. 1A).

The low-RMPG genera (0 to 5th percent quantile,  $<0.427$  RMPG) included several organisms that are exclusively intracellular or have a large intracellular component to their lifestyle, such as *Wolbachia* and *Rickettsia* (28) (see Fig. S1A). Given that a strict intracellular lifestyle should limit contact with infectious viruses and thus reduce the pressure to maintain viral defense, it was not surprising to find these genera in the low-RM category. The high-RMPG genera (95th to 100th percent quantile,  $>4.40$  RMPG) included *Microcystis*, as well as the heterotrophic genera *Helicobacter* and *Neisseria*, both noted previously for their high number of RMPG (29, 30) (see Fig. S1B).

Previous studies reported a correlation between genome size and RMPG (31–33). To revisit these analyses, we performed both linear and negative binomial regressions on the mean RMPG of the genera with five or more sequenced genomes. Both regressions give the same result: genome size is a poor predictor of the number of RMs in prokaryotes as it can explain no more than  $\sim 2\%$  of the variation (linear: estimate = 0.145,  $R^2 = 0.0217$ ,  $P = 4.93 \times 10^{-5}$ ; negative binomial: estimate = 0.06401, McFadden Pseudo  $R^2 = 7.02 \times 10^{-3}$ ,  $P = 1.07 \times 10^{-5}$ ). Moreover, while these trends are statistically significant, the estimates from each regression suggest there would need to be a large increase in genome size for there to be an impact in RM count if genome size is the sole predictive indicator. For example, an organism with an initial genome size of 2 Mbp would need to expand its genome by an additional 6.90 or 6.50 Mbp according to linear or negative binomial regressions, respectively, to gain one RM system. However, for small genomes within the range of 0.5 to 2.5 Mbp, we observed a more pronounced scaling of RMPG as a function of genome size, a trend that is consistent with earlier studies (32, 33).

**Patterns of RM defense in cyanobacteria.** Noting an absence of strong indicators for RMPG scaling across all prokaryotes, the analysis was subsequently restricted to the phylum *Cyanobacteria*, whose members include *Microcystis* and other genera for which their ecology is well known. This restriction provided an opportunity to relate RMPG to organism lifestyle in addition to genome size.

The range of RMPG among the cyanobacterial genera was nearly as extensive as for all prokaryotes (Fig. 1C) and did not trend with genome size (Fig. 1D). The high end of the cyanobacterial RMPG distribution was dominated by the HAB-forming freshwater genera *Microcystis*, *Planktothrix*, *Nodularia*, *Dolichospermum*, and *Anabaena* (34) (Fig. 1C). This signal was robust even with more-stringent annotation calls (see Materials and Methods; see also Fig. S2 in the



**FIG 1** Distributions of RM systems. (A) 95% quantile ( $\geq 4.4$  RMPG) of the RM distribution from all prokaryotes. Error bars represent 95% confidence intervals of the genus mean, and data points represent five or more isolates. Genera were dropped if the 95% confidence interval fell below the 95th quantile. (B) Subset of *Synechococcus* genomes parsed according to Coutinho et al. (37), into either *Parasynechococcus*, which is closely related to *Prochlorococcus*, or (remaining as) *Synechococcus*. Genomes were color coordinated to their isolation environment. The Wilcoxon rank sum *P* value is  $8.4 \times 10^{-5}$ . (C and D) *Cyanobacteria* phylum, showing complete RMPG (C), and genome size (D), ordered by RMPG. Colored genera are from oligotrophic (green) or eutrophic (yellow) environments. Red bars indicate the genus mean, while data points indicate individual isolates. All genera have five or more isolates except for *Raphidiopsis* ( $n = 2$ ), *Trichodesmium* ( $n = 1$ ), and *Dolichospermum* ( $n = 3$ ).

supplemental material), suggesting a strong association between bloom formation and RM abundance.

The bottom of the RM distribution is occupied by unicellular picocyanobacteria, *Prochlorococcus* and *Synechococcus* spp. (Fig. 1C), with many *Prochlorococcus* genomes lacking RMs altogether. *Prochlorococcus* numerically dominates the low nutrient (oligotrophic) oceans and, while peaking at about  $10^5$  cells  $\text{ml}^{-1}$ , is the most abundant photosynthetic organism on Earth (35). *Synechococcus* is a picocyanobacterium with a broad habitat range: while some genotypes contribute significantly to the marine oligotrophic phytoplankton community, other genotypes can be found at high abundance in nutrient-rich coastal environments or in freshwater systems (36). Given these varied ecologies, intragenus trends in RMPG were assessed using the classification of Coutinho et al. that phylogenetically distinguishes the open ocean genotypes (e.g., *Parasynechococcus*) from other, typically freshwater, genotypes (e.g., *Synechococcus*) (37). Interestingly, genomes of *Parasynechococcus*, which are more closely related to the oligotrophic specialist *Prochlorococcus*, showed statistically fewer RMPG ( $P = 8.4 \times 10^{-5}$ , Wilcoxon rank sum; Fig. 1B). Moreover, we found that isolates from open ocean marine systems skew toward the bottom of the *Synechococcus* RMPG distribution.

The observed correlation between extremes in RMPG and cell densities of the dominant genus suggested that cell density, as a function of nutrient availability, is an important driver of RM acquisition or loss in the cyanobacterial lineage. The high end of the RMPG distribution was dominated by HAB-forming cyanobacteria, whose blooms are largely attributed to

eutrophication of water bodies from farm runoff carrying fertilizer, flooding the system with nitrogen and phosphate which promote life at high density (38). The low end was dominated by oligotrophic picocyanobacteria that are deprived of nutrients due to temperature stratification of the deep ocean and large geographic distances from coastal inputs. Given these relationships, we hypothesized that the trade-off between defense and nutrient utilization plays a central role in the evolution of RM defenses and that loss or gain is tied to nutrient availability.

**Evaluating mechanisms of RM selection.** Motivated by the observation that high nutrient input is associated with the occurrence of HAB-forming cyanobacteria with high RMPG (Fig. 1), we built upon existing ecological models (39–43) to evaluate competition in contrasting nutrient regimes. We made no attempt to encapsulate genotypic or phenotypic diversity, other than through RM-associated differences in ability to protect against viruses, which in all cases we associated with a cost to nutrient utilization ability. “Competition” specialists (low RMPG) were competed with “defense” specialists (high RMPG) in “general” and “parallel” models that encapsulate interactions within nested and parallel infection networks, respectively (see Materials and Methods for further details). We hypothesized that a low nutrient supply places strong pressure on cells to acquire nutrients, forgoing investment in RM defenses even at the cost of viral attack. We hypothesized that nutrient enrichments alleviate pressure to acquire resources and drive ecological selection for RM systems due to an increasing host-phage contact rate.

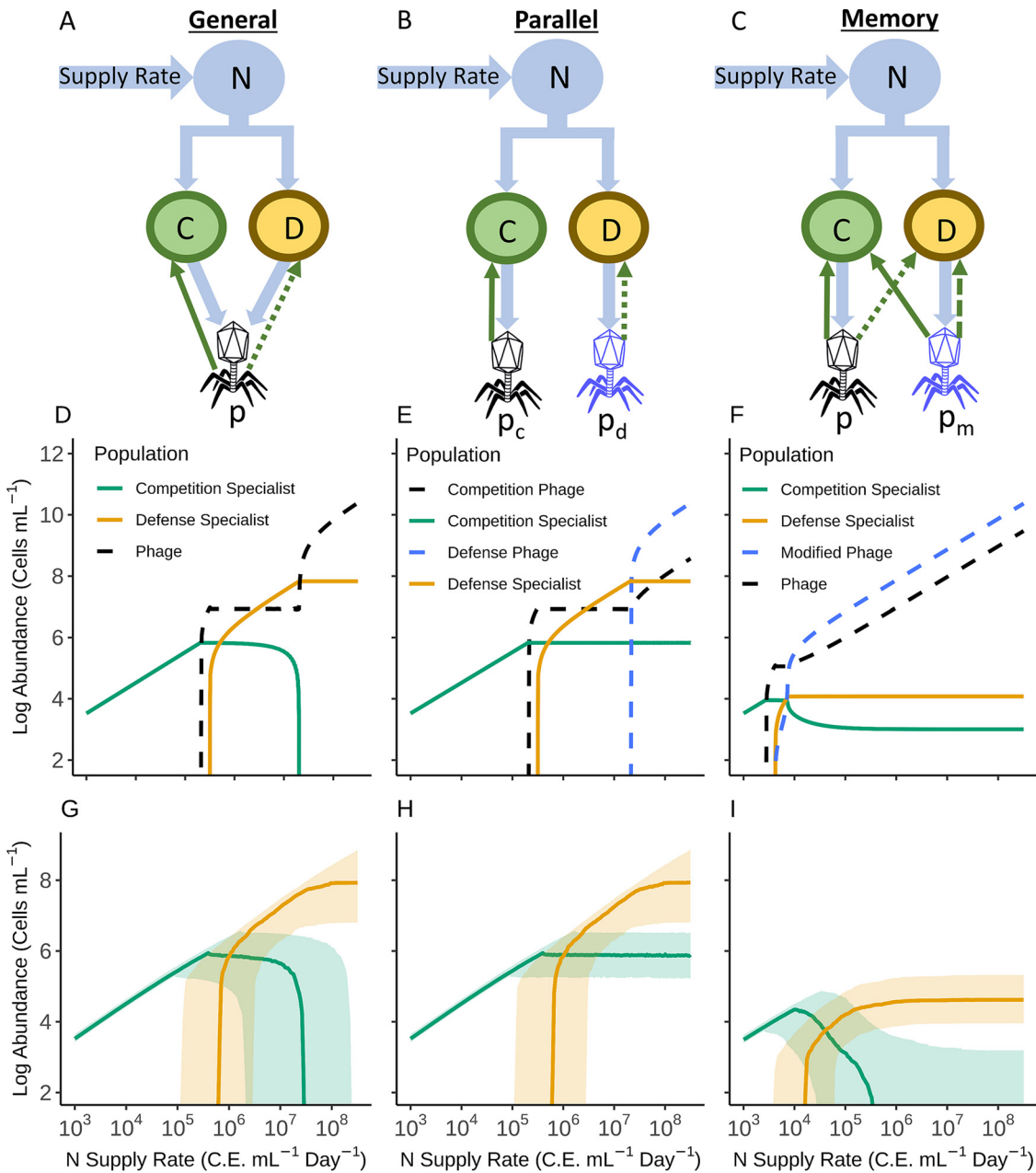
The two ecological models (Fig. 2A and B) capture the governing role of nutrient enrichment on competitive outcomes while remaining agnostic to the mechanism of defense. We developed a third, “memory” model (Fig. 2C) in which methylation of virions by host methyltransferases is captured. The parallel and general models serve as controls, allowing us to isolate the unique effects of viral methylation on investment in RM systems in contrasting nutrient conditions.

**Phage-host interaction models lacking a memory function incompletely explain RM accumulation.** We began by evaluating competition between an organism carrying 1 RMPG (competition specialist) with a newly emerged 2-RMPG subpopulation after a gain of function event (defense specialist) (Fig. 2). Outcomes were qualitatively similar when a 1-RMPG subpopulation emerged from one with 0 RMPG (data not shown). Outcomes for both general and parallel models could be binned generally by nutrient inflow as low, midrange, and high. In the “general model,” where a single generalist phage can infect both hosts, a low nutrient supply established a steady-state monoculture of the competition specialist, as both the defense specialist and the phage were eliminated from the system (Fig. 2D and G). A midrange nutrient supply led to coexisting steady-state populations of competition and defense specialists. Within this range, the defense specialist cell density scales with nutrient supply rate, whereas the competition specialist density is held in check by the virus. At the highest nutrient supply examined, the system enters a new state where the competition specialist is driven to extinction, and the defense specialist scales with nutrient supply rate until its density is held in check by phage, the latter scaling with nutrient input rate.

In the “parallel model,” outcomes for the competition and defense specialists at low to mid-range nutrient inflows are like the general model (Fig. 2E and H). In contrast to the general model, however, the parallel model predicts stable coexistence of defense and competition specialists at high nutrient supply. Both models thus indicate that carrying two RMs at a high nutrient supply confers selective advantages at high cell densities over hosts with only one.

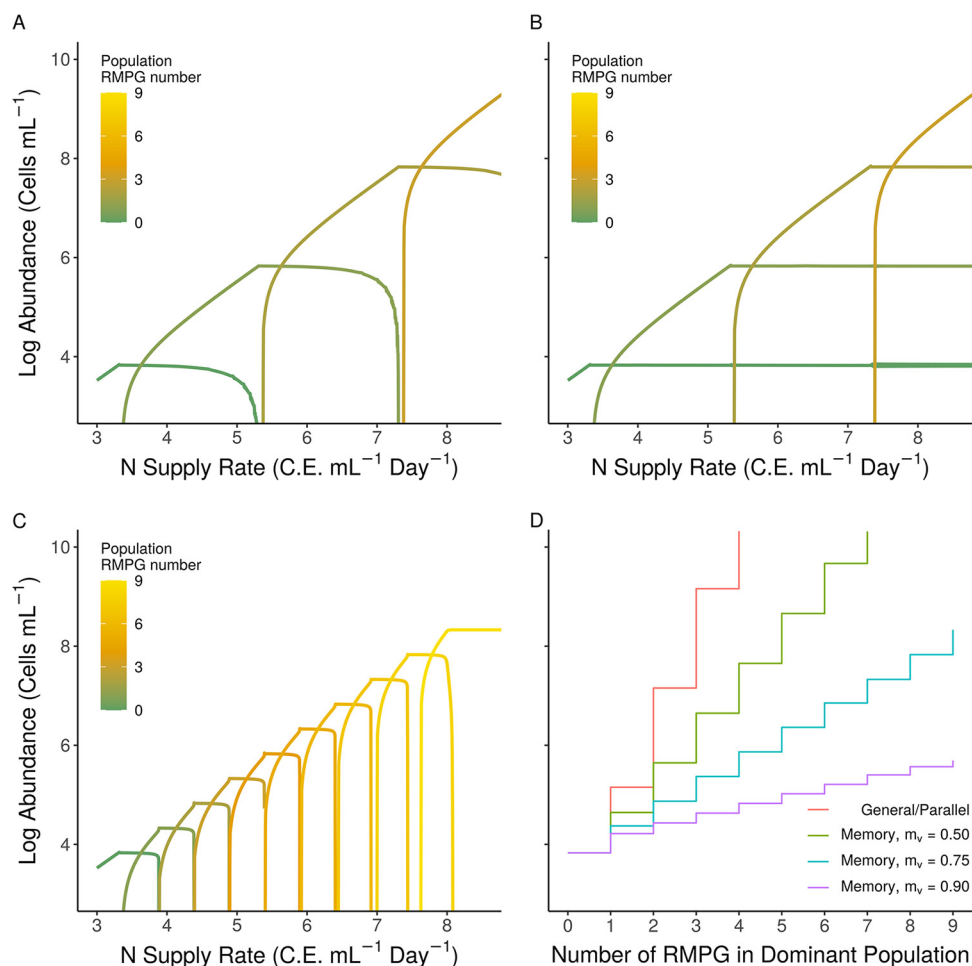
Models were next examined for their ability to explain the escalation of RM defenses such as observed in *Microcystis* (RMPG  $\approx$  16; Fig. 1C). When present in the same cell, RMs targeting different DNA sequences confer multiplicative effects on viral defense (22, 44). Assuming an equal number of restriction sites per viral genome and identical efficiencies among the endonucleases per restriction site, if a single RM system confers a moderate infection reduction of 100-fold (20), then the addition of a second RM system would reduce infections 10,000-fold.

Redundance of RM systems becomes evident in both control models when hosts with  $n = 1, 2, 3$  RMPG compete for nutrients (Fig. 3A and B). Assuming a single RM



**FIG 2** General, parallel, and memory virus-host interaction models. Figure columns correspond to general (A, D, and G), parallel (B, E, and H), and memory interaction (C, F, and I). Model structures (A to C) show the mass transfer from resource (N), to competing competition (C) and defense (D) specialist host populations, and finally into phage (p). Phage subscripts denote competition (c) or defense (d) specialist hosts (E) or phage modified by methylation (m) (F). Competition and defense specialists have 1 or 2 RMPG values, respectively. Solid green arrows represent high infection rates, dashed lines represent intermediate infection rates, while dotted lines represent low infection rates. (D to F) Steady-state abundance of each prokaryotic and viral population across a wide range of resource supply rates ( $S$  in equation 1). Solid gold and green lines are the defense and competition specialists, respectively. Dashed black lines show the abundance of phage, while the dashed blue line shows the abundance of the competition phage in the parallel model or the modified phage in the memory model. (G to I) Variation in steady-state values for the competition and defense specialists from 1,000 simulations with parameters drawn from realistic distributions (see Materials and Methods). Solid lines indicate the median value of each population, while the shaded regions show the 75th and 25th quantiles. C.E., cellular equivalents. For parameter values, see Table S1 in the supplemental material.

system has a modest reduction in infections of 100-fold (20) (resistance gained per endonuclease in our model [ $r_e$ ]=100), three RMPG is sufficient to fully protect against viruses over a range of cell densities up to roughly  $10^{10}$  cells ml<sup>-1</sup>. Hosts with >3 RMPG are not selected, even for modest assumed costs of carrying an RM (Fig. 3A and B). The selective basis for genera such as *Microcystis*, which has maximal cell densities



**FIG 3** Abundance scaling with increasing defense types in general, parallel and memory models. (A, B, and C) Steady-state abundances of populations carrying different numbers of RM systems. For simplicity, we assumed the cost and resistance of each additional RM system are identical. (C) Virion methylation ( $m_v$ ) is set to 0.75, and RMs are in a “subset” arrangement (see the text for further description). (D) Total abundance of each community plotted against the number of RM systems in the dominant subpopulation. General and parallel models are identical, while the scaling of the memory model depends on the partial resistance conferred from the degree virions are methylated. C.E., cellular equivalents. For other parameters, see Table S1.

$\sim 8 \times 10^8 \text{ ml}^{-1}$  (45), to carry more than 16 RMPG is thus poorly reconciled with the efficiency of endonucleases by these models. Clearly, our control models are unable to scale RMPG between ecological extremes when *Prochlorococcus* genomes typically carry 0 to 1 RMPG and have maximal cell densities of  $\sim 2 \times 10^5$  in the oligotrophic ocean (46).

**Viral methylation drives RM escalation and de-escalation.** Given the poor explanatory power that the two initial models provided for the observed escalation of RM defenses in *Microcystis* and other HAB-forming cyanobacteria, a new model that incorporates the viral memory conferred by the methyltransferase component of the host’s RM was developed (Fig. 2C). In this model, the (rare) surviving progeny phage of RM-expressing hosts develop immunity to that RM via methylation of its genome during infection and thus can infect both the defense and competition hosts at equal rates if the viral genome is fully methylated. We further explored relaxing this assumption by incorporating a “virion methylation” parameter ( $m_v$ ) which allows for virions to be hypomethylated. Virion methylation is a new parameter relative to previous modeling of RMs (24, 47, 48) and toggles the degree that progeny virions adopt the hosts methylation state. While some classic works suggest that most if not all virions are methylated in each viral burst ( $m_v \approx 1$ ) (23, 25), we felt the possibility of hypomethylation of virions should not go overlooked since these works focused almost exclusively on type

I RMs, with the exception of EcoRI and EcoRII which are type II RMs. Indeed, assuming plasmid conjugal transfer efficiency in *Nostoc* PCC 7120 is a proxy for resistance, we find realistic ranges with  $m_v < 1$  in experiments where cellular methyltransferase activity was experimentally manipulated (44) (see Fig. S3). To explore how hypomethylation can impact competitive outcomes, we allowed virion methylation to range from fully methylated ( $m_v = 1$ ) to half methylated ( $m_v = 0.5$ ).

Qualitatively, competitive outcomes in the memory model appear to resemble a mix of general and parallel model outcomes (Fig. 2G to I). A low nutrient supply selects monocultures of the competition specialist, and the defense specialist invades with increasing nutrient supply rates (Fig. 2F and I). However, this happens at a much lower supply rate relative to the control models because all viruses carry the methylation pattern of the competition specialist, effectively nullifying its single RM system (Fig. 2F and I). At high density, fitness and coexistence of competitive ( $n = 1$  RMPG) and defensive ( $n = 2$  RMPG) types was dependent upon parameter selection. In Fig. 2F, the defense specialist is dominant at high nutrient supply and coexists with the competitive type, resembling the parallel model when nearly all virions are fully methylated. However, the spread of the green ribbon in Fig. 3I reflects the heterogeneity of outcome in the memory model due to parameter selection. Importantly, when there was coexistence, the defense specialist dominated the community at high nutrients in nearly every simulation. The source of variation between competitive exclusion and coexistence was the virion methylation parameter. Unlike near perfect methylation of virions that establish parallel-like infections (Fig. 2F,  $m_v = 0.99$ ), decreasing virion methylation increases the effectiveness of the defense specialists RMs against viral populations and facilitates competitive exclusion at high nutrient supply (Fig. 2I).

When RM escalation or de-escalation is considered, the memory model drastically differs from the predictions of the other two models (Fig. 3C). As in the coculture (Fig. 2), methylation weakens the protective effect of each additional RM, leading to far more modest gains in cell abundance per RM along the nutrient supply gradient (Fig. 3C). Moreover, the degree of virion methylation toggles the gains between a very gradual increase in cell abundance with respect to the nutrient supply rate (Fig. 3D, purple line,  $m_v = 0.9$ ) or more drastic gains ( $m_v = 0.75$  and  $0.5$ ; Fig. 2D), approaching those of the control models (where  $m_v \approx 0$ ; Fig. 3D).

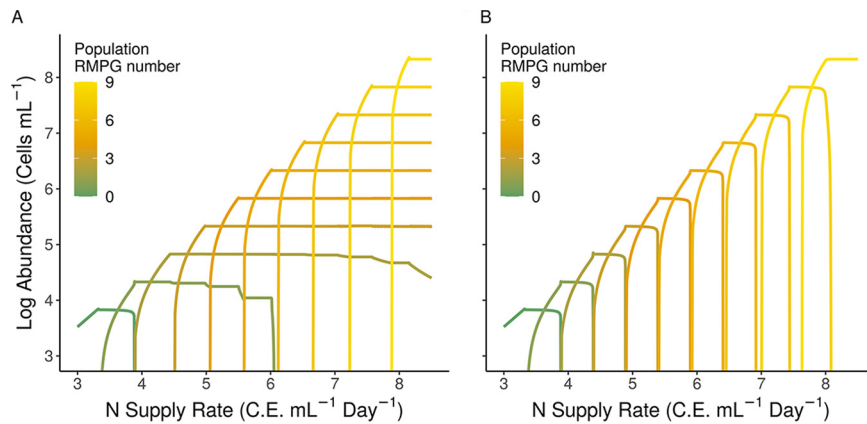
**RM identity impacts coexistence between competitive and defensive populations.** So far, we have assumed that when hosts acquire new RMs, their existing set does not change via loss or divergence. We relaxed the assumption in the memory model that all RMs exist as “subsets” by allowing all populations to have “unique sets” of RMs, representing the endpoint of a diversification scenario where multiple rounds of gene gain and loss have taken place. For both “subset” and “unique set” scenarios, increases in nutrient inflow rate led to numerical dominance by genotypes with progressively higher RMPG (Fig. 4). However, while model communities with RM subsets predict competitive exclusion of cells with fewer RMPG (Fig. 4A and Fig. 5, top), model communities with unique RM sets predict coexistence of nearly all populations (Fig. 4B and Fig. 5, bottom) at sufficiently high nutrient loads.

## DISCUSSION

This study examined the interplay between bottom-up and top-down controls in harmful algal bloom (HAB) ecology, namely, the ability of cyanobacteria to grow and sustain dense populations under constant threat by viral attack in different nutrient contexts. Specifically, we focused on RM defense systems and the possible evolutionary drivers of extensive accumulation of RM defenses in the bloom-forming cyanobacterium *Microcystis*.

Reasoning, as others have (31, 43), that the quantity of RMPG is intrinsic to ecological strategy, we focused our investigation on two ecological extremes established by resource availability: high-density eutrophic systems and low-density oligotrophic systems. Eutrophic systems promote dense blooms of not only *Microcystis* but also *Planktothrix*, *Anabaena*, and *Dolichospermum* in lakes across the planet (34, 38, 49), and

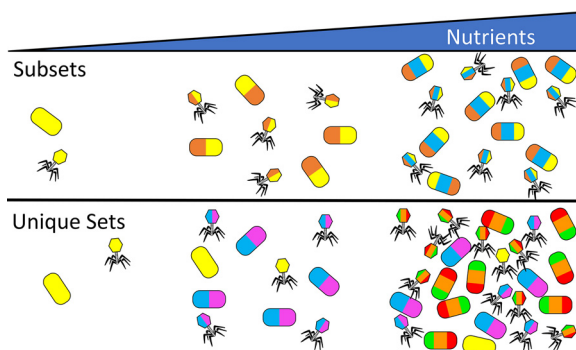




**FIG 4** Identity of RM systems among populations determine coexistence of competitive and defensive types in the memory model. (A and B) “Unique set” and “subset” RM communities, respectively, with theoretical methylated viral bursts ( $m_v = 0.75$ ). C.E., cellular equivalents.

consistently, these genera had the highest number of RMPG in cyanobacteria (Fig. 1C). Like *Prochlorococcus*, *Synechococcus* can account for a substantial fraction of the phytoplankton community in the oligotrophic ocean (46), and both genera had below-average RMPG. An intergenus difference in RMPG between oligotrophy and eutrophy could also be observed in intragenus comparisons of different *Synechococcus* isolates, further suggesting that nutrient availability drives pressure to retain RMs. In contrast to intracellular heterotrophic bacteria such as *Wolbachia* or *Chlamydia* (see Fig. S1 in the supplemental material), the low abundance or complete absence of RMs in these picocyanobacteria cannot be attributed to a lack of phages in their ecosystems, since they play host to a diverse array of viruses, and these viruses are suspected to contribute significantly to mortality of their hosts *in situ* (50, 51).

While our bioinformatics and subsequent modeling focused exclusively on RM defenses, it is important to recognize that RM represents only one form of antiviral defense available to prokaryotes. Prokaryotes can develop resistance to phage through a variety of mechanisms, including the alteration of phage receptors, the production of extracellular matrix, and the production of inhibitors (11, 12, 52). Given the disparity in RMPG between picocyanobacteria and HAB-forming cyanobacteria, we expect other



**FIG 5** Impact of RM identity on community structure. When populations of prokaryotes compete for resources in the presence of phage, the steady-state community structure depends on nutrient availability and the RM set type. Each color rendered on the bacillus hosts represents a different complete RM system, while colors rendered on phage represent the adopted methylome of the host. The number of bacilli or phage denotes relative abundance at oligotrophic, mesotrophic, and eutrophic conditions along a resource supply gradient. As the environment shifts from an oligotrophic (left) to a eutrophic (right) environment, the optimal number of RMPG increases, with either concomitant competitive exclusion of the low-RM bacilli when RMs exist as subsets or the coexistence of high- and low-RM bacilli when RMs exist as unique sets. Subsetting promotes low RM diversity, whereas unique setting promotes both higher diversity and a higher carrying capacity.

costly defensive mechanisms (e.g., CRISPR and toxin-antitoxin) to also be enriched in eutrophic environments and dispensed in oligotrophic environments over evolutionary time (53), and the observation that prokaryotes with CRISPR-Cas systems have statistically higher RMPG may reflect this reality (32). What remains unclear is whether HABs also occupy the top percentile of other defensive functional groups relative to the rest of the prokaryotic world or if RMs are specifically well suited as defense systems in these species.

Collectively, the bioinformatic results of this work strongly suggested a causal link between increased environmental nutrient availability and the investment in RM anti-predator defenses. This relationship is inconsistent with the outcome of a previous modeling study of host receptor mutations and RMs, indicating that RMs are favored at both low and high nutrient availabilities (47). It is also inconsistent with suggestions for terrestrial ecosystems that decreased resources lead to increased investments in defense (54). The reasons for the contradictions are not entirely clear, but in comparing the outcomes between our models and those of the prior RM study, we suspect that previous models assumed that receptor mutants never have a viral predator, and the costs of receptor mutations are higher than that of RM additions. We note that even if resistance from a cell surface mutation is more costly, the resistance would only apply to viruses that exploit that specific cell-surface element to gain entry, and only until a new viral variant arises that can overcome the mutation. Since RMs offer host protection regardless of viral entry, they are an evolutionary parsimonious way to increase viral resistance to all dsDNA viruses in the environment, especially when virus-host contact rates increase in eutrophic conditions.

Our mathematical population models evaluated ecological and molecular mechanisms driving enhanced RMPG in HAB-forming cyanobacteria. Our modeling builds upon prior theoretical (39–43) and experimental (55–57) investigations which have established that the fitness value of antipredator defenses can depend on nutrient supply rate. Underlying this dependency is the assumption that defense benefits come at a cost associated with nutrient utilization ability. We adopted similar approaches in our parallel and general models that served ultimately as controls for the consequences of phage methylation by RMs. While the control models—agnostic to the mechanism of defense—predict nutrients increase the selection for more RMPG, they cannot account for the extent of RMPG escalation evident in genomes of HAB-forming cyanobacteria. This prompted us to consider the special feature of RM defenses, namely, the memory conferred by the host RM system on the progeny phage. While not addressing nutrient influences *per se*, prior studies have reported (24, 47, 48) that memory can affect competition outcomes. When we introduced a memory aspect to our resource gradient model, there was a significant improvement in the ability to capture RM gain and loss events. Like the control models, the memory population model predicts that there is selection for defense against top-down controls with increased environmental nutrient loading, despite assumed costs to resource competitiveness. Unlike the control models, however, the memory model could also explain the high degree of RM accumulation observed in *Microcystis* and other HAB-forming cyanobacteria. Through the mechanism of host methylation of viral progeny, the per-RM increase in fitness is drastically reduced, requiring more RMs to achieve incremental improvements in fitness as nutrients increase.

Our final examination with the memory model evaluated the emergence of communities with differing assumptions about the structure and identity of RM systems. We found that innovation of novel RM sets tends to promote coexistence (Fig. 5). The role of RMs in promoting diversification was modeled in prior work by Sneppen et al. (48), who demonstrated that RM acquisition promotes invasion and that diverse RMs can facilitate long-term coexistence between many bacterial strains. Our results confirm this prior study and builds upon their results by assessing how virion methylation impacts selection for RMPG in contrasting resource environments.

Our modeling suggests in nutrient rich environments, RM diversity is a powerful selective force. This may help explain why RMs are so tightly associated with horizontal

transfer (58). Indeed, Oliveira et al. (58) highlighted the diversity of mobile genetic elements RMs are associated with as they are found on plasmids, prophages, transposons, integrative conjugative elements, and integrons. We note that the evolutionary history of the genus *Microcystis* is riddled with indicators of extreme genomic plasticity (59, 60), which could be the result of extreme pressure to innovate viral defense systems, contributing to their rapid genomic turnover (30, 61–64).

**Caveats and future directions.** Our hypothesis connecting RMPG with nutrient loading and viral methylation was the result of our modeling efforts to reconcile the vast number of RMPGs in HAB-forming cyanobacteria with the multiplicative efficacy of RMs in preventing phage attack. For this initial effort our models were necessarily simplified, but we recognize that additional factors will almost certainly contribute to the fitness value of RMPG, and in the following paragraphs we outline some of these factors that should be addressed in future studies.

In a prior study of a smaller set of genomes, Zhao et al. (65) noted a trend that RMPG values were higher in genomes of filamentous compared to unicellular cyanobacteria. In our larger data set, this trend was noticeably less robust, since the filamentous genera *Trichodesmium*, *Pseudanabaena*, *Leptolyngbya*, and *Moorea* or the heterocystous genus *Nostoc* (66) all had low RMPG values. Thus, while cellular organization may play a selective role in RMPG, its role is perhaps not as strong as the role that ecology, specifically nutrient loading, plays.

While the connection between high RMPG and bloom formation in cyanobacteria is striking, it is not universal: *Trichodesmium* forms blooms in marine surface waters (67, 68) and has an RMPG of 2 for the single genome available. This inconsistency may be reconciled by the actual cell densities of blooms. Global satellite monitoring shows that *Trichodesmium* rarely achieve chlorophyll *a* concentration higher than  $1 \mu\text{g liter}^{-1}$ , with the majority only showing  $0.25 \mu\text{g liter}^{-1}$  (69). While generally considered blooms, these biomass densities pale in comparison to some HAB-forming prokaryotes with higher RMPG. For example, *Microcystis*, with a mean RMPG of  $\sim 16$ , blooms in the western basin of Lake Erie or Lake Taihu at 100-fold-higher chlorophyll *a* concentrations (45, 70) than *Trichodesmium*.

In our models, we have assumed that each RM in the genome is active; however, there is reason to suspect this may not always be the case. For instance, some RM systems in *Campylobacter jejuni* (71) and *Neisseria gonorrhoeae* (72) are under phase variation control, and similar regulation may occur in cyanobacteria. While methyltransferase activity has been confirmed for some RMs of *Microcystis* (65), the restriction endonuclease components have so far not been verified to our knowledge for any RM. However, even with a conservative estimate that 50% of the RMs in any one *Microcystis* cell are active, our control models still have difficulty explaining RM escalation to this value, whereas the memory model can readily account for this accumulation.

As another caveat, we assumed that there is no redundancy in recognition sequences between RMs gained in our model, that all endonucleases cut at the same rate, and that viruses are assumed to have an equal number of restriction sites. However, natural variation in the number of recognition sequences among wild viruses and the efficiencies of different endonucleases is immense. For example, the endonuclease EcoRV could reduce the number of infections between 0 and 7 orders of magnitude among 21 different phages (20). Future experimental and laboratory work may need to account for possible diversity in cut sites among viruses, and the effects this may have on RM dynamics.

Environment and physiological state can influence RM function in ways that we did not resolve in our modeling, and these influences likely effect RM types differently. For example, Pleška et al. (73) found that there was increased fitness cost of type II RMs in minimal medium compared to rich medium, suggesting that hypomethylation can occur more frequently under these conditions. In bacterial hosts, hypomethylation of restriction sites in the chromosome can lead to DNA cleavage and induction of the SOS response and is likely a major contributor to cellular cost of carrying RMs (73). As reasoned by Pleška et al. (73), hypomethylation may be caused by stochastic gene expression (74) or protein partitioning at the cell level (75). This threat to chromosome

integrity would explain why hosts avoid restriction sites in their genome (76). Interestingly, avoidance was more intense for type II RMs than type IIG RMs, suggesting intrinsic cost differences between the different RM types resulting from hypomethylation. These observations imply RM costs could be modeled as a function of the number of host restriction sites, and the cost per site increases due to poor nutrient quality is likely higher when endonuclease and methyltransferase activities are independent, as in type II RMs. One unexplored possibility is that increased costs, resulting from increased likelihood of host genomic hypomethylation, may translate into increased protection against viruses: viruses replicating within hosts may likewise experience hypomethylation from host methyltransferases (77, 78), especially fast-replicating viruses that may fail to fully adopt host methylation prior to encapsidation.

A novel feature of our modeling was the explicit assessment of viral methylation. However, the extent to which viral methylation varies in the natural world is not well understood. Most early studies relied on plaque assays for calculating efficiency: PFU values for initial infections were compared to those for progeny phage infecting their immediate hosts; when initial PFU matched the PFU of the progeny phage, the efficiency of plating (EOP) was declared to equal 1 (23, 25). Since plaque assays cannot quantify noninfectious virions, a significant number of progeny virions could be unmethylated, even in the case of progeny phage infecting their immediate hosts. In only one of these studies was the virus measured by both particle counts and viable counts, and here the difference in counts varied from 0.31 to 0.92 (79), suggesting memory may be less than perfect. Our modeling highlights the potential importance of viral methylation on the occurrence and activity of RMs in natural conditions and points to the need for further experimental investigation to provide constraint on these parameters.

As a final consideration, it will be important to explore the interplay between RMs and CRISPR, receptor modification, and other defense innovations in contrasting environmental conditions. In some cases, such as CRISPR and RM, different mechanisms of defense can function in tandem to increase survival of the population (80, 81); however, it has also been noted that interference between mechanisms can also affect the fitness of defenses: acquiring a receptor mutation might render the intracellular RM or CRISPR defense unnecessary (27, 82). We encourage future studies assessing the complex interplay between mechanisms of defense in nutrient-rich environments and a new antiviral focus to research into the persistence of HABs.

## MATERIALS AND METHODS

**Bioinformatic search strategy.** Because of the diversity in both genomic and domain architecture of RMs, we chose a strategy that uses both BLAST 2.7.1+ (83) and HMMER 3.1b2 (84) to generate alignments to our reference database and then refined our results by using genomic context. Protein profiles are built from hidden Markov models (HMMs) and allow us to identify putative methyltransferases or endonucleases by searching for the specific functional motifs in proteins. By using profiles, we could explicitly detect functional motifs within proteins regardless of the domain architecture, a problem local alignment algorithms like BLAST cannot resolve unless there is a protein with an identical architecture capable of generating full alignments. To ensure we were not aligning multiple profiles to the same residues in each protein, we “competed” profiles that align to 75% of the same residues in a protein and select the profile with the lower E value. We used hmmscan with gathering cutoffs to collect all Pfam (release 31) HMMs (85) that represent experimentally characterized “Gold Standard” methyltransferases and endonucleases found in New England Biolabs’ REBASE (see Data Set S1, Table 3, in the supplemental material) (21). We finalized our reference HMMs after manual curation (see Fig. S4 to S6 and Data Set S1, Table 4, in the supplemental material). In curation, we found ResIII (PF04851) domains were repeatedly observed in various helicases and transcriptional regulators. Since ResIII was common in type I, type IIG, and type III RMs, we retained this domain in our reference HMMs, but added HMMs that would covary with ResIII when a protein was a not part of a RM system to flag false positives (see Data Set S1, Table 5, in the supplemental material). The Gold Standard data set from REBASE contained several endonucleases and a few methyltransferases that did not contain any HMMs from Pfam. As a way to still utilize these sequences to find homologs, all Gold Standard proteins without any identifiable Pfam HMMs were used in BLAST searches with the previously described protein profile searches to maximize our ability to identify RMs in prokaryotic genomes (see Data Set S1, Table 6, in the supplemental material). BLAST alignments were considered a match if the total alignment length was 75% of the query length and the E value was  $\leq 1 \times 10^{-5}$ . Once we generated alignments with both HMMER and BLAST, we used genomic context to count the

number of full RMs. An RM system was considered complete if there was an endonuclease  $\leq 4,000$  bp away from a methyltransferase or if both motifs were detected in one peptide.

Large proteins ( $>750$  amino acids) that contained a methyltransferase domain but did not show any additional motifs to indicate endonuclease activity were subjected to more sensitive search algorithms that are part of the HHsearch suite (86) to evaluate whether they were type IIG RMs since protein size alone can discriminate between type IIG RM and other methyltransferases (see Fig. S6D). We first preclustered these putative type IIG RMs using psi-cd-hit (87, 88) with a clustering threshold of 35% sequence identity and an alignment that covers at least 85% of each protein (parameters: -c 0.35 -aL 0.85 -aS.85 -g 1). Once the clusters were defined, representatives from each cluster were used to build profiles for HHblits. Clusters were considered type IIG proteins if the representative sequence aligned to 3S1S (89), 4PXG (90), 4XQK (91), 4ZCF (92), 5FFJ (93), or 5HR4 (94). Three iterations were used to build multiple sequence alignments with mact=0.35. Parameters for hhsearch are as follows:  $P = 20$ ,  $Z = 250$ , loc,  $z = 1$ ,  $b = 1$ ,  $B = 250$ , ssm = 2, sc = 1, seq = 1, norealign, maxres = 32000, context = context\_data.crf.

This more sensitive analysis revealed that 27,147 of the original 33,633 flagged proteins could be aligned to verified type IIG RMs in the protein data bank. While we believe there is a high likelihood these are RMs, we wanted to determine whether our initial findings (Fig. 1) depend on the veracity of our type IIG calls. When we reanalyzed our RM collection without these putative type IIG RMs, the RM distribution was qualitatively the same: *Planktothrix*, *Microcystis*, *Nodularia*, and *Anabaena* still dominated the tail end of the distribution (see Fig. S2 in the supplemental material).

**Virus-host interaction models.** To model the effects of resource on the selection for defense, we competed theoretical prokaryotic populations for a single resource in the presence of phage. Modeled prokaryotic populations differ only in the number of defense systems (i.e., RMs) they carry, where defensive types have more RMs, and thus greater resistance to phage, relative to competitive types that have fewer RMs. We further assumed a trade-off: investment in RM increases the resistance of the defense specialist to phage, but this comes at a cost to nutrient utilization (95).

We explored the influence of this trade-off on competitive outcomes within three hypothetical system structures with contrasting representations of virus-host interaction (Fig. 2A to C). In the general interaction model (Fig. 2A), the competition and defense specialists are infected by the same phage. In the parallel model (Fig. 2B), the competition and defense specialists are infected by phage that do not cross-infect the other host. In both of these first two models, the viral defense is generic, and RMs could be substituted with phage receptor modification, CRISPR, etc. We emphasize that these models are akin to controls and are meant to establish a simplified understanding of the impact of varied nutrient supply rate to host/virus biomass. In contrast, the memory model (Fig. 2C) is a variation of the control models that accounts for the unique feature of RM defenses: the “memory” bestowed upon surviving phage by the methyltransferase component of the RM defense system in the defense specialist, which renders the phage resistant to the restriction endonuclease (18–20, 23).

Our simulations were completed in a newly developed ODElib Python module (<https://github.com/SEpapoulis/ODElib>), which aims to make analysis of ecosystem ODE models more tractable by integrating several modules available in the Python ecosystem. We explicitly indicate ODElib dependencies when used for a specific computational task.

All population model simulations were computed using SciPy’s integrate module (96), and all data were managed using Pandas (97). All simulations launched with different fixed nutrient supply rates were allowed to reach an equilibrium steady state. Initial conditions were selected from the previous steady-state values when performing simulations at a range of nutrient supply rates; however, equilibrium values are independent of initial values (see Fig. S7). The final abundances of each population were plotted against the simulation resource inflow to show system state changes in R using ggplot2. A Latin Hypercube Sampling (LHS) scheme (98) from pyDOE2 module was used to randomly pick model parameters over uniform distributions in specified ranges, except for the host resistance per endonuclease ( $r_c$ ) and baseline infection rate ( $\phi$ ), which were drawn from log-uniform distributions (see Table S1 in the supplemental material). LHS is favorable over brute-force random sampling because previous samples are used to make intelligent draws for the next sample, ensuring that random draws are representative of parameter variation in multidimensional space.

**Model structure.** To explore the selective pressures on bacteria (and other prokaryotes) to increase their defense, we model the competition of  $i$  bacterial types in the presence of phage. The general model is defined with equations 1 to 4, the parallel model is defined with equations 1, 3, 5, and 6, and the memory model is defined with equation 1 and equations 7 to 9. We will first introduce all parameters and state variables through the definition of the general model and then expand to the parallel and memory models.

All models explicitly define the rate of change of nutrients in the system as follows:

$$\frac{dN}{dt} = S - \sum_{i=1}^n \alpha(1-c)B_i N \quad (1)$$

where  $N$  is a nutrient concentration, which is typically measured in molar units. Here, for notational simplicity, we quantify nutrient concentration in “cellular equivalents  $\text{ml}^{-1}$ ,” which implicitly assumes a fixed cellular nutrient quota for each competing population.  $B_i$  is the  $i$ th bacterial population ( $\text{cells ml}^{-1}$ ) at time  $t$ .  $S$  is the nutrient supply rate into the system ( $\text{cellular equivalents ml}^{-1} \text{ day}^{-1}$ ), while  $\alpha$  is the nutrient utilization rate ( $\text{ml cell}^{-1} \text{ day}^{-1}$ ), and  $c$  is the total cost of host defense, implemented as percentage of nutrient utilization

lost, and is a function of the set of RMs in bacterial population  $i$ . We assumed that RMs' costs are identical and that cost is linear; thus, if one RM system causes  $c$  to be 0.05, two RMs will cause  $c$  to be 0.10 for a respective bacterial population and will scale nutrient utilization as a percentage. Moreover, the summation of RM costs cannot be greater than 1. Growth of each bacterial population is defined as:

$$\frac{dB_i}{dt} = \alpha(1-c)B_iN - \phi\gamma_i B_i P - \delta_b B_i \quad (2)$$

where  $\phi$  is the baseline infection rate (virus<sup>-1</sup> ml<sup>-1</sup> day<sup>-1</sup>) of a bacterial population,  $\delta_b$  is the bacterial loss (day<sup>-1</sup>), and  $P$  is the abundance of phage time  $t$ . The vector  $\gamma$  stores coefficients that are calculated based on the number of RMs in each bacterial population and is defined as:

$$\gamma = \left[ \frac{1}{r_e^{b_i}} \quad \cdots \quad \frac{1}{r_e^{b_i}} \right] \quad (3)$$

where  $r_e$  is the host resistance conferred by each RM, and  $b_i$  is the total number RMs carried by bacterial population  $i$ ; thus, the total defense to phage is the multiplicative protection of all RMs in the general model. For simplicity, we assumed that all RMs in our models have the same protective value. We treat RMs in an organism as a mathematical set which imposes all RMs are unique. Vertical bars denote the cardinality of the set, where cardinality is the total number of elements in  $b_i$ . Biologically, we can think of this as each RM system targeting different recognition sequences in DNA. The assumption that RMs are multiplicative is not unfounded, since Arber and Wauters-Willems (20) were able to demonstrate that the defensive value of multiple RMs greatly reduce the efficiency of plating (EOP) of phage on host *Escherichia coli* strains (22). For example, in these experiments, one RM system decreased the EOP by  $1 \times 10^{-2}$ , another RM decreased the EOP by  $3 \times 10^{-5}$ , and together they decreased the EOP of phage to  $6 \times 10^{-7}$ . Finally, phage replication is determined by the equation:

$$\frac{dP}{dt} = \beta \sum_{i=1}^n \phi \gamma_i B_i P - \delta_p P \quad (4)$$

where  $\beta$  (virus cell<sup>-1</sup>) is the phage burst size and  $\delta_p$  is the phage loss (day<sup>-1</sup>). For the parallel model, we alter equations 2 and 4 to accommodate a unique phage population that preys on each bacterial population:

$$\frac{dB_i}{dt} = \alpha_i(1-c_i)B_iN - \phi \gamma_i B_i P_i - \delta_b B_i \quad (5)$$

$$\frac{dP_i}{dt} = \beta \phi \gamma_i B_i P_i - \delta_p P_i \quad (6)$$

In this alternate control model, phage  $i$  can only infect and reproduce on bacterium  $i$ . Altogether, equations 1 to 6 represent a diamond food-web ecosystem with predation being "keystone" to maintaining diversity within the ecosystem (39–41), where our general model is a simple diamond food web with a single phage as the predator of competing bacterial populations. Equations 1 to 6 could be extended in an infinite number of ways, for example to include separate compartments for different infection pathways (99) or nonlinear interactions (100). Equations 1 to 6 represent the most transparent and parsimonious model to explore how, to first-order, nutrient supply rate selects number of RMs. Nutrients were explicitly described in cellular equivalence, which is reported as a cell number, to acknowledge that we are remaining agnostic to which nutrient is limiting, and we neglect variability in cell nutrient quotas between organisms. Explicit representation of nonlinear interactions, additional infection pathways, and cellular quotas would introduce additional unconstrained model parameters, while also limiting the clarity with which the mechanisms driving RM selection can be presented. Nevertheless, we also considered an RM-specific ecological virus-host model that allows us to test the effects of viral methylation relative to the general and parallel models, considered control models in this study.

The "memory interaction model" addresses the biological consequences of DNA methylation by RM defenses. Host methyltransferases methylate all DNA indiscriminately, including any replicating viral DNA that evades host defenses long enough to be modified. In this manner, viral progeny adopt the host's methylation pattern, which confers immunity to the virus when infecting another bacterial cell with the same RM system(s). This adoption of host methylation patterns can be thought of as viral "memory" of its most recent prey. To incorporate the differential infection generated from the modification of viral DNA from host RMs, we create a matrix  $\Gamma$  that stores coefficients between virus-host pairs:

$$\Gamma = \begin{bmatrix} \gamma_{11} = \frac{1}{r_e^{|b_1-p_1|+(1-m_v)|b_1 \cap p_1|}} & \cdots & \gamma_{1j} = \frac{1}{r_e^{|b_1-p_j|+(1-m_v)|b_1 \cap p_j|}} \\ \vdots & \ddots & \vdots \\ \gamma_{i1} = \frac{1}{r_e^{|b_i-p_1|+(1-m_v)|b_i \cap p_1|}} & \cdots & \gamma_{ij} = \frac{1}{r_e^{|b_i-p_j|+(1-m_v)|b_i \cap p_j|}} \end{bmatrix} \quad (7)$$

$b_i$  and  $p_j$  denote the RM carried by the host and methylation state of the virus, respectively, and are treated as mathematical sets. Similar to the vector described previously (3), more RMs deteriorate viral

resistance to host defenses, but from the cardinality in the difference between  $b_i$  and  $p_j$  (RM in host  $i$  and methylation not in virus  $j$ ). The cardinality in the intersection between  $b_i$  and  $p_j$  (RM in host  $i$  and methylation in virus  $j$ ) can still reduce viral resistance of phage  $j$  to bacteria  $i$ , however, this depends on the virion methylation,  $m_v$ . Thus, methylation shared between hosts and viruses cause partial resistance if  $0 < m_v < 1$ , with 1 implying perfect methylation of progeny virions, 0 implying a complete lack of methylation, and 0.5 implying half of the restriction sites in the viral genome are methylated.

In reality, each endonuclease has an intrinsic defensive property of defense per unmodified restriction site per cell, represented by the slope of the solid lines from data in Elhai et al. (44) for three endonucleases in *Nostoc* PCC 7120, where we are using conjugal transfer efficiency as an analog for viral resistance (see Fig. S3 in the supplemental material). The positive shift in the slope is due to the reduction in unmodified sites from methyltransferase activity in the host. A slope of  $< 1$  suggest there is a small probability for each site to remain unmethylated and is likely an intrinsic property of the methyltransferase and possibly other physiological/environmental constraints (see Discussion). To simplify these cellular processes, we assumed the number of recognition sequences in our theoretical viral genomes to be identical for all RMs and simplified the viral resistance of unmodified and modified states to the idealized parameters,  $r_e$  and  $m_v$ , respectively. Moreover, we assumed  $r_e$  and  $m_v$  to be uniform for all RMs, that is, all RM endonucleases have the same efficiency and all RM methyltransferases are equally likely to methylate progeny virions. We considered the sensitivity of our main predictions, allowing inefficient/incomplete methylation due to methylase limitation (77) and unmethylated restriction sites (44). These two observations necessitate that, as long as the methylation of viral progeny is not perfect during replication, viral progenies are susceptible to RM, albeit at a much lower frequency.

To implement the memory model, we must alter the change in both bacterial and phage populations over time so that all phage can infect all bacteria, but each phage population can only emerge from a single bacterial population. This implies that there are an equal number of phage and bacterial types in our system. We accomplished this by using the following equations:

$$\frac{dB_i}{dt} = \alpha(1-c)B_iN - \sum_{j=1}^n \phi \gamma_{ij} B_i P_j - \delta_b B_i \quad (8)$$

$$\frac{dP_i}{dt} = \beta \sum_{j=1}^n \phi \gamma_{ij} B_i P_j - \delta_p P_i \quad (9)$$

In equation 8, mortality of bacterium  $i$  is now due to the collective infections of  $j$  phage within the environment. In equation 9, the reproduction of phage  $i$  can only come from bacterial population  $i$  even though all phage can infect bacterium  $i$ .

Note that the general model may be thought of as a special case of the memory model. When  $m = 0$ , we can simplify the exponents in the matrix of equation 7. This simplification leads to equation 7 becoming identical to equation 3 because  $|b_i - p_j| + |b_i \cap p_j|$  is equal to a cardinality of  $|b_i|$ . Intuitively, we can think of this as a complete lack of viral methylation which leads to the predicted outcome of the general model.

**Data availability.** All bioinformatic and mathematical modeling work was conducted in executable Jupyter Notebooks, which can be accessed, along with all source code and data, at [https://github.com/SEpapoulis/SEpapoulis/EscalationAndDe-escalationOfRM](https://github.com/SEpapoulis/EscalationAndDe-escalationOfRM). ODElib can be found at <https://github.com/SEpapoulis/ODElib>.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**FIG S1**, TIF file, 2.9 MB.

**FIG S2**, TIF file, 2.8 MB.

**FIG S3**, TIF file, 2.9 MB.

**FIG S4**, TIF file, 2.9 MB.

**FIG S5**, TIF file, 2.9 MB.

**FIG S6**, TIF file, 2.9 MB.

**FIG S7**, TIF file, 1.4 MB.

**TABLE S1**, DOCX file, 0.01 MB.

**DATA SET S1**, XLSX file, 10.9 MB.

## ACKNOWLEDGMENTS

We thank Igor Jouline and Jeffery Morris for helpful discussions involving this study and Katherine Moccia for critical review of the manuscript.

This study was supported by an NSF grant to E.R.Z. and S.W.W. (NSF IOS-1451528). This study was also supported by funds from The Great Lakes Center for Fresh Waters

and Human Health (NIH 1P01ES028939-01; NSF OCE1840715). D.T. (grant 690671) and S.W.W. (grant 735077) acknowledge the support of the Simons Foundation.

S.E.P., S.W.W., D.T., and E.R.Z. designed research. S.E.P. performed research. S.E.P., D.T., and E.R.Z. contributed analytic tools. S.E.P., D.T., and E.R.Z. analyzed data. S.E.P., S.W.W., D.T., and E.R.Z. wrote the paper.

We declare there are no conflicts of interest.

## REFERENCES

- Wilhelm SW, Bullerjahn GS, McKay RML. 2020. The complicated and confusing ecology of *Microcystis* blooms. *mBio* 11:e00529-20. <https://doi.org/10.1128/mBio.00529-20>.
- Harke MJ, Steffen MM, Gobler CJ, Otten TG, Wilhelm SW, Wood SA, Paerl HW. 2016. A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, *Microcystis* spp. *Harmful Algae* 54:4–20. <https://doi.org/10.1016/j.hal.2015.12.007>.
- Glibert PM, Maranger R, Sobota DJ, Bouwman L. 2014. The Haber-Bosch-harmful algal bloom (HB-HAB) link. *Environ Res Lett* 9:105001. <https://doi.org/10.1088/1748-9326/9/10/105001>.
- Cao H, Shimura Y, Steffen MM, Yang Z, Lu J, Joel A, Jenkins L, Kawachi M, Yin Y, Garcia-Pichel F. 2020. The trait repertoire enabling cyanobacteria to bloom assessed through comparative genomic complexity and metatranscriptomics. *mBio* 11:e01155-20. <https://doi.org/10.1128/mBio.01155-20>.
- Steffen MM, Li Z, Effler TC, Hauser LJ, Boyer GL, Wilhelm SW. 2012. Comparative metagenomics of toxic freshwater cyanobacteria bloom communities on two continents. *PLoS One* 7:e44002. <https://doi.org/10.1371/journal.pone.0044002>.
- Wommack KE, Colwell RR. 2000. Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* 64:69–114. <https://doi.org/10.1128/MMBR.64.1.69-114.2000>.
- Wilhelm SW, Suttle CA. 1999. Viruses and nutrient cycles in the sea: viruses play critical roles in the structure and function of aquatic food webs. *Bioscience* 49:781–788. <https://doi.org/10.2307/1313569>.
- Avrani S, Lindell D. 2015. Convergent evolution toward an improved growth rate and a reduced resistance range in *Prochlorococcus* strains resistant to phage. *Proc Natl Acad Sci U S A* 112:E2191–E2200. <https://doi.org/10.1073/pnas.1420347112>.
- Bohannon BJ, Kerr B, Jessup CM, Hughes JB, Sandvik G. 2002. Trade-offs and coexistence in microbial microcosms. *Antonie Van Leeuwenhoek* 81:107–115. <https://doi.org/10.1023/A:1020585711378>.
- Luria SE, Delbrück M. 1943. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 28:491–511. <https://doi.org/10.1093/genetics/28.6.491>.
- Koonin EV, Makarova KS, Wolf YI. 2017. Evolutionary genomics of defense systems in archaea and bacteria. *Annu Rev Microbiol* 71:233–261. <https://doi.org/10.1146/annurev-micro-090816-093830>.
- Labrie SJ, Samson JE, Moineau S. 2010. Bacteriophage resistance mechanisms. *Nat Rev Microbiol* 8:317–327. <https://doi.org/10.1038/nrmicro2315>.
- Westra ER, Dowling AJ, Broniewski JM, Van Houte S. 2016. Evolution and ecology of CRISPR. *Annu Rev Ecol Evol Syst* 47:307–331. <https://doi.org/10.1146/annurev-ecolsys-121415-032428>.
- Olovnikov I, Chan K, Sachidanandam R, Newman DK, Aravin AA. 2013. Bacterial Argonaute samples the transcriptome to identify foreign DNA. *Mol Cell* 51:594–605. <https://doi.org/10.1016/j.molcel.2013.08.014>.
- Yamaguchi Y, Park JH, Inouye M. 2011. Toxin-antitoxin systems in bacteria and archaea. *Annu Rev Genet* 45:61–79. <https://doi.org/10.1146/annurev-genet-110410-132412>.
- Cohen D, Melamed S, Millman A, Shulman G, Oppenheimer-Shaanan Y, Kacen A, Doron S, Amitai G, Sorek R. 2019. Cyclic GMP-AMP signaling protects bacteria against viral infection. *Nature* 574:691–695. <https://doi.org/10.1038/s41586-019-1605-5>.
- Chaudhary K. 2018. Bacteriophage EXclusion (BREX): a novel anti-phage mechanism in the arsenal of bacterial defense system. *J Cell Physiol* 233:771–773. <https://doi.org/10.1002/jcp.25973>.
- Bertani G, Weigle JJ. 1953. Host controlled variation in bacterial viruses. *J Bacteriol* 65:113–121. <https://doi.org/10.1128/JB.65.2.113-121.1953>.
- Luria SE, Human ML. 1952. A nonhereditary, host-induced variation of bacterial viruses. *J Bacteriol* 64:557–569. <https://doi.org/10.1128/JB.64.4.557-569.1952>.
- Korona R, Korona B, Levin BR. 1993. Sensitivity of naturally occurring coliphages to type I and type II restriction and modification. *J Gen Microbiol* 139:1283–1290. <https://doi.org/10.1099/00221287-139-6-1283>.
- Roberts RJ, Vincze T, Posfai J, Macelis D. 2010. REBASE—a database for DNA restriction and modification: enzymes, genes, and genomes. *Nucleic Acids Res* 38:D234–D236. <https://doi.org/10.1093/nar/gkp874>.
- Arber W, Wauters-Willems D. 1970. Host specificity of DNA produced by *Escherichia coli*. *Mol Gen Genet* 108:203–217. <https://doi.org/10.1007/BF00283350>.
- Arber W, Dussoix D. 1962. Host specificity of DNA produced by *Escherichia coli*. I. Host controlled modification of bacteriophage lambda. *J Mol Biol* 5:18–36. [https://doi.org/10.1016/S0022-2836\(62\)80058-8](https://doi.org/10.1016/S0022-2836(62)80058-8).
- Korona R, Levin BR. 1993. Phage-mediated selection and the evolution and maintenance of restriction modification. *Evolution* 47:556–575. <https://doi.org/10.2307/2410071>.
- Yoshimori R, Roulland-Dussoix D, Boyer HW. 1972. R factor-controlled restriction and modification of deoxyribonucleic acid: restriction mutants. *J Bacteriol* 112:1275–1279. <https://doi.org/10.1128/JB.112.3.1275-1279.1972>.
- Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, Blumenthal RM, Degtyarev S, Dryden DT, Dybvig K, Firman K, Gromova ES, Gumpert RI, Halford SE, Hattman S, Heitman J, Hornby DP, Janulaitis A, Jeltsch A, Josephsen J, Kiss A, Klaenhammer TR, Kobayashi I, Kong H, Krüger DH, Lacks S, Marinus MG, Miyahara M, Morgan RD, Murray NE, Nagaraja V, Piekarowicz A, Pingoud A, Raleigh E, Rao DN, Reich N, Repin VE, Selker EU, Shaw PC, Stein DC, Stoddard BL, Szybalski W, Trautner TA, Van Etten JL, Vitor JM, Wilson GG, Xu SY. 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* 31:1805–1812. <https://doi.org/10.1093/nar/gkg274>.
- Makarova KS, Wolf YI, Koonin EV. 2013. Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res* 41:4360–4377. <https://doi.org/10.1093/nar/gkt157>.
- Casadevall A. 2008. Evolution of intracellular pathogens. *Annu Rev Microbiol* 62:19–33. <https://doi.org/10.1146/annurev.micro.61.080706.093305>.
- Budroni S, Siena E, Dunning Hotopp JC, Seib KL, Serruto D, Nofroni C, Comanducci M, Riley DR, Daugherty SC, Angiuoli SV, Covacci A, Pizza M, Rappuoli R, Moxon ER, Tettelin H, Medini D. 2011. *Neisseria meningitidis* is structured in clades associated with restriction modification systems that modulate homologous recombination. *Proc Natl Acad Sci U S A* 108:4494–4499. <https://doi.org/10.1073/pnas.1019751108>.
- Nobusato A, Uchiyama I, Kobayashi I. 2000. Diversity of restriction-modification gene homologues in *Helicobacter pylori*. *Gene* 259:89–98. [https://doi.org/10.1016/S0378-1119\(00\)00455-8](https://doi.org/10.1016/S0378-1119(00)00455-8).
- Makarova KS, Wolf YI, Snir S, Koonin EV. 2011. Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J Bacteriol* 193:6039–6056. <https://doi.org/10.1128/JB.05535-11>.
- Oliveira PH, Touchon M, Rocha EP. 2014. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res* 42:10618–10631. <https://doi.org/10.1093/nar/gku734>.
- Vasu K, Nagaraja V. 2013. Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol Mol Biol Rev* 77:53–72. <https://doi.org/10.1128/MMBR.00044-12>.
- Huisman J, Codd GA, Paerl HW, Ibelings BW, Verspagen JMH, Visser PM. 2018. Cyanobacterial blooms. *Nat Rev Microbiol* 16:471–483. <https://doi.org/10.1038/s41579-018-0040-1>.
- Biller SJ, Berube PM, Lindell D, Chisholm SW. 2015. *Prochlorococcus*: the structure and function of collective diversity. *Nat Rev Microbiol* 13:13–27. <https://doi.org/10.1038/nrmicro3378>.
- Olson RJ, Chisholm SW, Zettler ER, Armbrust EV. 1990. Pigments, size, and distributions of *Synechococcus* in the North Atlantic and Pacific Oceans. *Limnol Oceanogr* 35:45–58. <https://doi.org/10.4319/lo.1990.35.1.0045>.



37. Coutinho F, Tschoeke DA, Thompson F, Thompson C. 2016. Comparative genomics of *Synechococcus* and proposal of the new genus *Parasynechococcus*. PeerJ 4:e1522. <https://doi.org/10.7717/peerj.1522>.
38. Bullerjahn GS, McKay RM, Davis TW, Baker DB, Boyer GL, D'Anglada LV, Doucette GJ, Ho JC, Irwin EG, Kling CL, Kudela RM, Kurmayer R, Michalak AM, Ortiz JD, Otten TG, Paerl HW, Qin B, Sohngen BL, Stumpf RP, Visser PM, Wilhelm SW. 2016. Global solutions to regional problems: collecting global expertise to address the problem of harmful cyanobacterial blooms. A Lake Erie case study. Harmful Algae 54:223–238. <https://doi.org/10.1016/j.hal.2016.01.003>.
39. Holt RD, Grover J, Tilman D. 1994. Simple rules for interspecific dominance in systems with exploitative and apparent competition. Am Nat 144:741–771. <https://doi.org/10.1086/285705>.
40. Leibold MA. 1996. A graphical model of keystone predators in food webs: trophic regulation of abundance, incidence, and diversity patterns in communities. Am Nat 147:784–812. <https://doi.org/10.1086/285879>.
41. McPeck MA. 1996. Trade-offs, food web structure, and the coexistence of habitat specialists and generalists. Am Nat 148:S124–S138. <https://doi.org/10.1086/285906>.
42. Våge S, Storesund JE, Giske J, Thingstad TF. 2014. Optimal defense strategies in an idealized microbial food web under trade-off between competition and defense. PLoS One 9:e101415. <https://doi.org/10.1371/journal.pone.0101415>.
43. Ruess J, Pleška M, Guet CC, Tkačik G. 2019. Molecular noise of innate immunity shapes bacteria-phage ecologies. PLoS Comput Biol 15: e1007168. <https://doi.org/10.1371/journal.pcbi.1007168>.
44. Elhai J, Veprikitskiy A, Muro-Pastor AM, Flores E, Wolk CP. 1997. Reduction of conjugal transfer efficiency by three restriction activities of *Anabaena* sp. strain PCC 7120. J Bacteriol 179:1998–2005. <https://doi.org/10.1128/JB.179.6.1998-2005.1997>.
45. Hu L, Shan K, Lin L, Shen W, Huang L, Gan N, Song L. 2016. Multi-year assessment of toxic genotypes and microcystin concentration in northern Lake Taihu, China. Toxins 8:23. <https://doi.org/10.3390/toxins8010023>.
46. Flombaum P, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, Jiao N, Karl DM, Li WK, Lomas MW, Veneziano D, Vera CS, Vrugt JA, Martiny AC. 2013. Present and future global distributions of the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. Proc Natl Acad Sci U S A 110:9824–9829. <https://doi.org/10.1073/pnas.1307701110>.
47. Frank SA. 1994. Polymorphism of bacterial restriction-modification systems: the advantage of diversity. Evolution 48:1470–1477. <https://doi.org/10.2307/2410241>.
48. Sneppen K, Semsey S, Seshasayee ASN, Krishna S. 2015. Restriction modification systems as engines of diversity. Front Microbiol 6:528–528. <https://doi.org/10.3389/fmicb.2015.00528>.
49. Heisler J, Glibert P, Burkholder J, Anderson D, Cochlan W, Dennison W, Gobler C, Dortch Q, Heil C, Humphries E, Lewitus A, Magnien R, Marshall H, Sellner K, Stockwell D, Stoecker D, Suddleson M. 2008. Eutrophication and harmful algal blooms: a scientific consensus. Harmful Algae 8:3–13. <https://doi.org/10.1016/j.hal.2008.08.006>.
50. Mann NH. 2003. Phages of the marine cyanobacterial picophytoplankton. FEMS Microbiol Rev 27:17–34. [https://doi.org/10.1016/S0168-6445\(03\)00016-0](https://doi.org/10.1016/S0168-6445(03)00016-0).
51. Sullivan MB, Waterbury JB, Chisholm SW. 2003. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. Nature 424:1047–1051. <https://doi.org/10.1038/nature01929>.
52. Forde SE, Beardmore RE, Gudelj I, Arkin SS, Thompson JN, Hurst LD. 2008. Understanding the limits to generalizability of experimental evolutionary models. Nature 455:220–223. <https://doi.org/10.1038/nature07152>.
53. Iranzo J, Cuesta JA, Manrubia S, Katsnelson MI, Koonin EV. 2017. Disentangling the effects of selection and loss bias on gene dynamics. Proc Natl Acad Sci U S A 114:E5616–E5624. <https://doi.org/10.1073/pnas.1704925114>.
54. Bryant JP, Kuropat PJ, Cooper SM, Frisby K, Owen-Smith N. 1989. Resource availability hypothesis of plant antiherbivore defence tested in a South African savanna ecosystem. Nature 340:227–229. <https://doi.org/10.1038/340227a0>.
55. Bohannan BJM, Lenski RE. 2000. The relative importance of competition and predation varies with productivity in a model community. Am Nat 156:329–340. <https://doi.org/10.2307/3079168>.
56. Gómez P, Bennie J, Gaston KJ, Buckling A. 2015. The impact of resource availability on bacterial resistance to phages in soil. PLoS One 10: e0123752. <https://doi.org/10.1371/journal.pone.0123752>.
57. 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. <https://doi.org/10.1086/283134>.
58. Oliveira PH, Touchon M, Rocha EPC. 2016. Regulation of genetic flux between bacteria by restriction-modification systems. Proc Natl Acad Sci U S A 113:5658–5663. <https://doi.org/10.1073/pnas.1603257113>.
59. Lin S, Haas S, Zemojtel T, Xiao P, Vingron M, Li R. 2011. Genome-wide comparison of cyanobacterial transposable elements, potential genetic diversity indicators. Gene 473:139–149. <https://doi.org/10.1016/j.gene.2010.11.011>.
60. Steffen MM, Dearth SP, Dill BD, Li Z, Larsen KM, Campagna SR, Wilhelm SW. 2014. Nutrients drive transcriptional changes that maintain metabolic homeostasis but alter genome architecture in *Microcystis*. ISME J 8:2080–2092. <https://doi.org/10.1038/ismej.2014.78>.
61. Jeltsch A, Pingoud A. 1996. Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. J Mol Evol 42:91–96. <https://doi.org/10.1007/BF02198833>.
62. Kobayashi I. 2001. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. Nucleic Acids Res 29:3742–3756. <https://doi.org/10.1093/nar/29.18.3742>.
63. Furuta Y, Abe K, Kobayashi I. 2010. Genome comparison and context analysis reveals putative mobile forms of restriction-modification systems and related rearrangements. Nucleic Acids Res 38:2428–2443. <https://doi.org/10.1093/nar/gkp1226>.
64. Furuta Y, Kawai M, Yahara K, Takahashi N, Handa N, Tsuru T, Oshima K, Yoshida M, Azuma T, Hattori M, Uchiyama I, Kobayashi I. 2011. Birth and death of genes linked to chromosomal inversion. Proc Natl Acad Sci U S A 108:1501–1506. <https://doi.org/10.1073/pnas.1012579108>.
65. Zhao L, Song Y, Li L, Gan N, Brand JJ, Song L. 2018. The highly heterogeneous methylated genomes and diverse restriction-modification systems of bloom-forming *Microcystis*. Harmful Algae 75:87–93. <https://doi.org/10.1016/j.hal.2018.04.005>.
66. Shih PM, Wu D, Latif A, Axen SD, Fewer DP, Talla E, Calteau A, Cai F, Tandeau de Marsac N, Rippka R, Herdman M, Sivonen K, Coursin T, Laurent T, Goodwin L, Nolan M, Davenport KW, Han CS, Ruben EM, Eisen JA, Woyke T, Gugger M, Kerfeld CA. 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. Proc Natl Acad Sci U S A 110:1053–1058. <https://doi.org/10.1073/pnas.1217107110>.
67. Bergman B, Sandh G, Lin S, Larsson J, Carpenter EJ. 2013. *Trichodesmium*: a widespread marine cyanobacterium with unusual nitrogen fixation properties. FEMS Microbiol Rev 37:286–302. <https://doi.org/10.1111/j.1574-6976.2012.00352.x>.
68. Darwin C, King PP, Fitzroy R. 1839. Narrative of the surveying voyages of His Majesty's Ships *Adventure* and *Beagle* between the years 1826 and 1836, describing their examination of the southern shores of South America, and the *Beagle's* circumnavigation of the globe, vol 3, p 14–16. Henry Colburn, London, United Kingdom.
69. Westberry TK, Siegel DA. 2006. Spatial and temporal distribution of *Trichodesmium* blooms in the world's oceans. Global Biogeochem Cycles 20:1–13. <https://doi.org/10.1029/2005GB002673>.
70. Steffen MM, Davis TW, McKay RML, Bullerjahn GS, Krausfeldt LE, Stough JMA, Neitzey ML, Gilbert NE, Boyer GL, Johengen TH, Gossiaux DC, Burtner AM, Palladino D, Rowe MD, Dick GJ, Meyer KA, Levy S, Boone BE, Stumpf RP, Wynne TT, Zimba PV, Gutierrez D, Wilhelm SW. 2017. Ecophysiological examination of the lake Erie *Microcystis* bloom in 2014: linkages between biology and the water supply shutdown of Toledo, OH. Environ Sci Technol 51:6745–6755. <https://doi.org/10.1021/acs.est.7b00856>.
71. Anjum A, Brathwaite KJ, Aidley J, Connerton PL, Cummings NJ, Parkhill J, Connerton I, Bayliss CD. 2016. Phase variation of a type IIG restriction-modification enzyme alters site-specific methylation patterns and gene expression in *Campylobacter jejuni* strain NCTC11168. Nucleic Acids Res 44:4581–4594. <https://doi.org/10.1093/nar/gkw019>.
72. Sánchez-Busó L, Golparian D, Parkhill J, Unemo M, Harris SR. 2019. Genetic variation regulates the activation and specificity of restriction-modification systems in *Neisseria gonorrhoeae*. Sci Rep 9:14685. <https://doi.org/10.1038/s41598-019-51102-2>.
73. Pleška M, Qian L, Okura R, Bergmiller T, Wakamoto Y, Kussell E, Guet CC. 2016. Bacterial autoimmunity due to a restriction-modification system. Curr Biol 26:404–409. <https://doi.org/10.1016/j.cub.2015.12.041>.
74. Elowitz MB, Levine AJ, Siggia ED, Swain PS. 2002. Stochastic gene expression in a single cell. Science 297:1183–1186. <https://doi.org/10.1126/science.1070919>.

75. Huh D, Paulsson J. 2011. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat Genet* 43:95–100. <https://doi.org/10.1038/ng.729>.
76. Gelfand MS, Koonin EV. 1997. Avoidance of palindromic words in bacterial and archaeal genomes: a close connection with restriction enzymes. *Nucleic Acids Res* 25:2430–2439. <https://doi.org/10.1093/nar/25.12.2430>.
77. Szyf M, Avraham-Haetzni K, Reifman A, Shlomai J, Kaplan F, Oppenheim A, Razin A. 1984. DNA methylation pattern is determined by the intracellular level of the methylase. *Proc Natl Acad Sci U S A* 81:3278–3282. <https://doi.org/10.1073/pnas.81.11.3278>.
78. Krüger DH, Schroeder C, Reuter M, Bogdarina IG, Buryanov YI, Bickle TA. 1985. DNA methylation of bacterial viruses T3 and T7 by different DNA methylases in *Escherichia coli* K12 cells. *Eur J Biochem* 150:323–330. <https://doi.org/10.1111/j.1432-1033.1985.tb09024.x>.
79. Kellenberger E, Arber W. 1957. Electron microscopical studies of phage multiplication. I. A method for quantitative analysis of particle suspensions. *Virology* 3:245–255. [https://doi.org/10.1016/0042-6822\(57\)90091-0](https://doi.org/10.1016/0042-6822(57)90091-0).
80. Dupuis M, Villion M, Magadán AH, Moineau S. 2013. CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. *Nat Commun* 4:2087. <https://doi.org/10.1038/ncomms3087>.
81. Hynes AP, Villion M, Moineau S. 2014. Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. *Nat Commun* 5:4399. <https://doi.org/10.1038/ncomms5399>.
82. Gurney J, Pleška M, Levin BR. 2019. Why put up with immunity when there is resistance: an excursion into the population and evolutionary dynamics of restriction-modification and CRISPR-Cas. *Philos Trans R Soc Lond B Biol Sci* 374:20180096. <https://doi.org/10.1098/rstb.2018.0096>.
83. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>.
84. Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39:W29–W37. <https://doi.org/10.1093/nar/gkr367>.
85. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the protein families database. *Nucleic Acids Res* 42:D222–D230. <https://doi.org/10.1093/nar/gkt1223>.
86. Remmert M, Biegert A, Hauser A, Söding J. 2011. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* 9:173–175. <https://doi.org/10.1038/nmeth.1818>.
87. Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28:3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>.
88. Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>.
89. Shen BW, Xu D, Chan S-H, Zheng Y, Zhu Z, Xu S-Y, Stoddard BL. 2011. Characterization and crystal structure of the type IIG restriction endonuclease RM. *Nucleic Acids Res* 39:8223–8236. <https://doi.org/10.1093/nar/gkr543>.
90. Xu CY, Yu F, Hu XJ, He JH. 2015. The 2.45 Å crystal structure of the restriction endonuclease Sau3AI suggests a self-inhibition mechanism. *Protein Data Bank* <https://doi.org/10.2210/pdb4PXG/pdb>.
91. Chand MK, Nirwan N, Diffin FM, van Aelst K, Kulkarni M, Pernstich C, Szczelkun MD, Saikrishnan K. 2015. Translocation-coupled DNA cleavage by the type ISP restriction-modification enzymes. *Nat Chem Biol* 11:870–877. <https://doi.org/10.1038/nchembio.1926>.
92. Gupta YK, Chan SH, Xu SY, Aggarwal AK. 2015. Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I. *Nat Commun* 6:7363. <https://doi.org/10.1038/ncomms8363>.
93. Kulkarni M, Nirwan N, van Aelst K, Szczelkun MD, Saikrishnan K. 2016. Structural insights into DNA sequence recognition by type ISP restriction-modification enzymes. *Nucleic Acids Res* 44:4396–4408. <https://doi.org/10.1093/nar/gkw154>.
94. Callahan SJ, Luyten YA, Gupta YK, Wilson GG, Roberts RJ, Morgan RD, Aggarwal AK. 2016. Structure of type IIL restriction-modification enzyme MmI in complex with DNA has implications for engineering new specificities. *PLoS Biol* 14:e1002442. <https://doi.org/10.1371/journal.pbio.1002442>.
95. Våge S, Storesund JE, Thingstad TF. 2013. Adding a cost of resistance description extends the ability of virus-host model to explain observed patterns in structure and function of pelagic microbial communities. *Environ Microbiol* 15:1842–1852. <https://doi.org/10.1111/1462-2920.12077>.
96. Millman KJ, Aivazis M. 2011. Python for scientists and engineers. *Comput Sci Eng* 13:9–12. <https://doi.org/10.1109/MCSE.2011.36>.
97. McKinney W. 2010. Data structures for statistical computing in Python, p 56–61. Proceedings of the 9th Python in Science Conference <https://doi.org/10.25080/Majora-92bf1922-00a>.
98. McKay MD, Beckman RJ, Conover WJ. 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21:239–245. <https://doi.org/10.2307/1268522>.
99. Thamatrakoln K, Talmy D, Haramaty L, Maniscalco C, Latham JR, Knowles B, Natale F, Coolen MJL, Follows MJ, Bidle KD. 2019. Light regulation of coccolithophore host-virus interactions. *New Phytol* 221:1289–1302. <https://doi.org/10.1111/nph.15459>.
100. Record NR, Talmy D, Våge S. 2016. Quantifying tradeoffs for marine viruses. *Front Mar Sci* 3:251. <https://doi.org/10.3389/fmars.2016.00251>.