Marine viral particles reveal an expansive repertoire of phage-parasitizing mobile elements

John M. Eppley[a] 🔵, Steven J. Biller[b,1] 🔵, Elaine Luo[a,1,2], Andrew Burger[a], and Edward F. DeLong[a,3] 🔵

Phage satellites are mobile genetic elements that propagate by parasitizing bacteriophage replication. We report here the discovery of abundant and diverse phage satellites that were packaged as concatemeric repeats within naturally occurring bacteriophage particles in seawater. These same phage-parasitizing mobile elements were found integrated in the genomes of dominant co-occurring bacterioplankton species. Like known phage satellites, many marine phage satellites encoded genes for integration, DNA replication, phage interference, and capsid assembly. Many also contained distinctive gene suites indicative of unique virus hijacking, phage immunity, and mobilization mechanisms. Marine phage satellite sequences were widespread in local and global oceanic virioplankton populations, reflecting their ubiquity, abundance, and temporal persistence in marine planktonic communities worldwide. Their gene content and putative life cycles suggest they may impact host-cell phage immunity and defense, lateral gene transfer, bacteriophage-induced cell mortality and cellular host and virus productivity. Given that marine phage satellites cannot be distinguished from bona fide viral particles via commonly used microscopic techniques, their predicted numbers (~3.2×10^{26} in the ocean) may influence current estimates of virus densities, production, and virus-induced mortality. In total, the data suggest that marine phage satellites have potential to significantly impact the ecology and evolution of bacteria and their viruses throughout the oceans. We predict that any habitat that harbors bacteriophage will also harbor similar phage satellites, making them a ubiquitous feature of most microbiomes on Earth.

bacteriophage | phage satellites | mobile elements | lateral gene transfer | marine virus

Bacterial viruses (also known as bacteriophages) can impact host-cell phenotype, population structure, pathogenicity, host immunity, lateral gene transfer, and other diverse aspects of microbial physiology, ecology, and evolution (1). Viruses can themselves be parasitized by small mobile genetic elements (MGEs) called phage satellites that propagate by exploiting co-occurring viral reproduction machinery. Comprising a few known types—most notably satellite phage [e.g., the *Escherichia coli* phage P4/P2 system (2, 3)], phage-inducible chromosomal islands [PICIs; found in several gram-positive genera and two gram-negative bacterial orders (4–6)], and phage-inducible chromosomal island–like elements [PLEs; of *Vibrio cholerae* (7–9)]—these phage-parasitizing MGEs are genetically diverse. They do, however, share a common life cycle strategy, which depends upon co-opting their "helper phage" reproductive cycles to redirect phage capsid assembly toward production of phage-like particles containing phage satellite DNA instead of bacteriophage DNA (10). Known phage satellites (P4-like, PICIs, and PLEs) typically have genomes about one-third of the size of the helper phages they parasitize. They encode a set of functionally similar genes involved in replication and propagation (integrases, primases, and DNA polymerases that provide autonomous replication capabilities), regulatory genes (that interact with helper phage life cycles), specific DNA recognition sequences that help redirect packaging specificity, and specific host genome integration sites, all of which enable them to propagate and interfere with helper phage reproduction (6, 10).

Phage satellites can influence host-cell phage immunity (5, 10), the frequency of host gene transduction (9, 11–14), host-cell pathogenicity (4, 8, 14, 15), and host-cell and helper phage genome evolution (13, 16). While evidence suggests that some habitat and host diversity may exist among known phage satellites (6, 12, 15–17), they have been clearly demonstrated to date mainly in several gram-positive genera (17) and a few gram-negative bacterial orders (Enterobacterales, Pasteurellales, Vibrionales), most of which are pathogens (6, 8, 10, 16). The extant diversity of phage satellites—including their cellular host ranges, helper phages, ecological interactions, and habitats—remains largely unexplored. The global abundance and ecological impacts of bacterial viruses in marine plankton (18), together with a hint that phage satellites might be found in planktonic virion populations (19), prompted us to further investigate the putative existence,

Significance

Phage satellites are mobile genetic elements that parasitize viruses, exerting profound biological and ecological impacts. Phage satellites are known to infect several gram-positive genera and a few gram-negative bacterial species, most associated with the human microbiome. Direct inspection of "wild" virus particles, however, revealed that marine phage satellites are widely distributed and abundant in the global oceans. Their genetic diversity, gene repertoires, and host ranges appear much greater than has been previously reported. Genetic analyses now provide clues about the parasitic life cycles, helper bacteriophage interactions, and reproductive strategies of these newly recognized marine phage satellites. Their properties, diversity, and environmental distributions suggest they may exert substantial influence on microbial ecology and evolution in the sea.

properties, and abundance of marine phage satellites found within wild virion particles collected in surface waters of the North Pacific Subtropical Gyre (NPSG).

Many Wild Marine Phage Particles Contain Phage Satellite Concatemeric DNAs Instead of Phage Genomes

Previous work had shown that single-molecule genome sequencing of naturally occurring virus particles efficiently recovers complete phage genomes in single-DNA reads (19). In that same study, several apparently full-length, nonphage, concatemeric DNA sequences from viral fractions were also encountered (19), prompting the current study here. To enhance the purity of our virion preparations and further refine workflows and analyses, we used density gradient fractionation to separate tailed phages from other types of marine particles, such as membrane vesicles, tailless phages, and cellular DNA contamination (Fig. 1A) (20). Following single-molecule nanopore sequencing of virion-encapsidated high–molecular weight DNA, we found that a considerable fraction (~0.6%) of virion particles collected from a depth of 25-m in the NPSG contained DNA encoding concatemeric repeats (Fig. 1A, Dataset S2, and SI Appendix, Fig. S1). The length distribution of virion-encapsidated concatemeric DNA exhibited maxima near 35 and 60 kbp, coincident with the genome sizes of co-occurring bacteriophages, with the longest concatemers exceeding 100 kbp (Fig. 1B and Dataset S2). We also observed DNA concatemers in the lower-density membrane vesicle–enriched fraction, albeit at lower frequencies and with much shorter lengths, mostly <20 kbp (Fig. 1B).

Polishing and dereplication of the repeated sequences embedded within the concatemers revealed a diverse set of phage-mobilized genetic elements referred to here as virion-encapsidated integrative mobile elements (VEIMEs). Within the single seawater sample analyzed here, we identified 2,000 unique concatemer-derived monomeric VEIMEs in the tailed phage–enriched fraction and 63 in the membrane vesicle–enriched fraction (Fig. 1 A and B and Datasets S2 and S3). About 20% of the VEIMEs were also found integrated into the genomes of a variety of common bacterioplankton groups, including Pelagibacter and other Alphaproteobacteria, Prochlorococcus, Verrucomicrobia, Flavobacteria, and additional bacterial taxa, demonstrating that these mobile elements were widespread among diverse co-occurring marine planktonic bacteria (Fig. 1 C and D, Dataset S3, and SI Appendix, Fig. S3).

Marine Phage Satellites Have Diverse, Unique, and Distinctive Gene Suites That Differentiate Them from Previously Described Phage Satellites and Mobile Elements

We next asked whether the marine VEIMEs might share some features with previously characterized human microbiome–associated phage satellites (i.e., P4-like satellite phage, PICIs, and PLEs). Using a bipartite graph of genomes and gene annotations of VEIMEs along with known phage satellites, we found that marine phage satellites partitioned into 12 different modules (Fig. 2). The majority of VEIMEs (98%) were in modules composed solely of marine VEIMEs, all of which showed significant
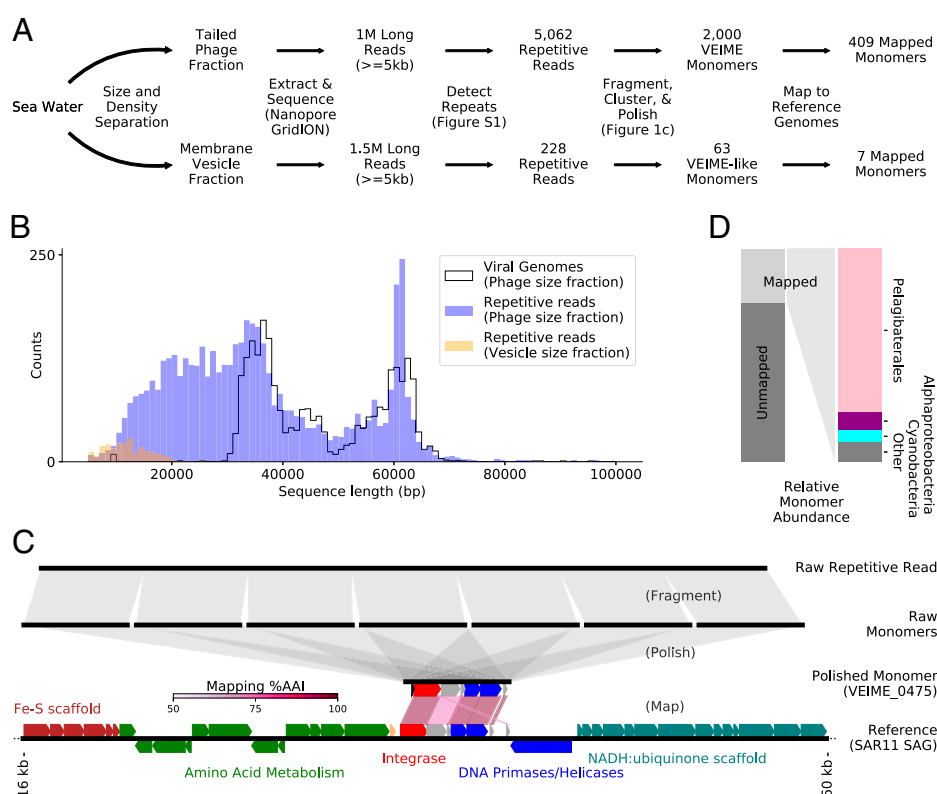

Fig. 1. Concatemeric DNAs within phage particles facilitate the discovery of phage-parasitizing MGEs. A shows the workflow for preliminary identification of VEIMEs in fractionated samples. The histogram of concatemeric read lengths (B) shows peaks that correspond to common phage genome lengths, represented here by a histogram of the lengths of reads with DTRs. C, Lower illustrates the fragmenting and self-polish steps of the workflow and shows a polished VEIME monomer aligned with a reference genome. In the initial step, raw concatemeric reads were identified in the virion sequencing pool (Dataset S2) followed by extraction of individual monomeric repeat sequences and subsequent monomer sequence polishing (Dataset S3). D, Left shows the relative abundance of VEIMEs that mapped to genomes compared with those that did not. D, Right displays the relative abundance of mapped VEIMEs and their frequency of occurrence in different bacterial host taxa. NADH, nicotinamide adenine dinucleotide; AAI, amino acid identity; SAG, single-cell amplified genome.
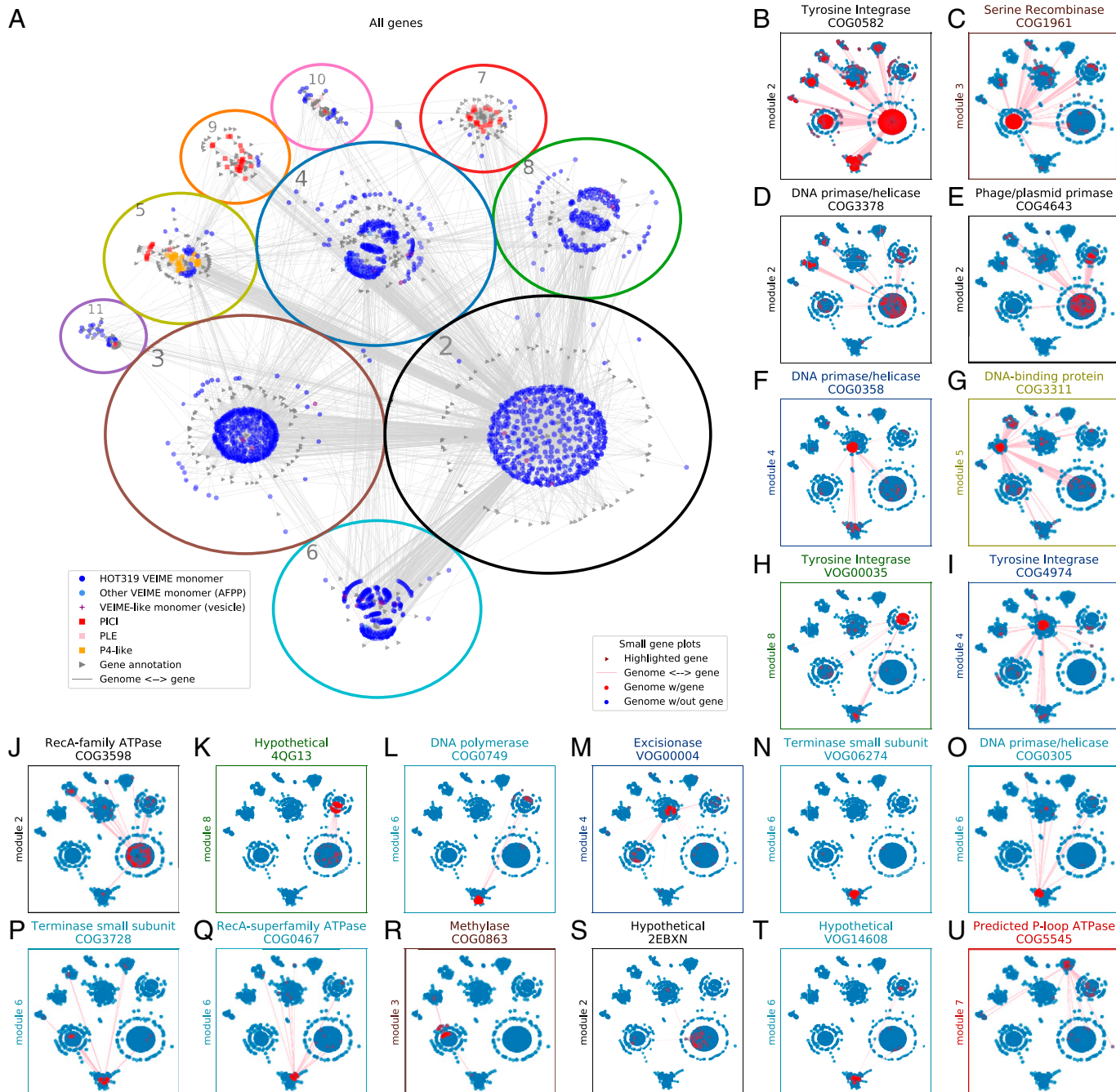
Fig. 2. Interrelationships among VEIMEs and reference phage satellites revealed by bipartite graph analyses. Bipartite graphs were constructed by connecting egg-NOG and VOG gene annotations (triangles) to the satellite genomes (other shapes) in which they were found. The displayed graph layout, used in all subplots, exaggerates module separation for clarity. (*A*) Module boundaries are proportional in area to the number of MGEs in each module. The shape and color of nodes indicate their type. All gene annotations are gray triangles. Published PICIs, P4-like phage, and PLEs are red, orange, and pink squares, respectively. VEIMEs are either dark blue (tailed phage fraction), or light blue (assembly-free putative PICIs from ref. 21), circles. (*B–U*) Each small plot highlights 1 of the top 20 most common gene annotations, arranged in order of decreasing abundance starting from the most frequently occurring COG0582 Integrase in *B*, to the least frequent COG5545 ATPase in *U*. The specific gene referenced is shown as a dark red triangle. MGEs are shown as circles, and they are colored red if they contain the gene and blue if not.

intramodule gene content variation (e.g., modules 1 to 4, 6, 8, and 12) (Fig. 2; Datasets S3, S5, and S6; and *SI Appendix*, Fig. S2). The remaining modules included both human microbiome–associated reference phage satellites as well as VEIMEs (Fig. 2 and Datasets S3–S5). Marine phage satellites in these modules, however, also encoded different gene suites than those found in previously characterized phage satellites (Dataset S6 and *SI Appendix*, Fig. S2). Different VEIME modules were also often associated with specific bacterial host taxa. For example, VEIMEs in modules 2, 6, and 8 were found mostly in *Pelagibacter* and other Alphaproteobacteria host genomes, while module 3 VEIMEs were associated with Cyanobacteria and Verrucomicrobia (Datasets S3 and S6 and *SI Appendix*, Figs. S2 and S3).

Analysis of VEIME gene repertoires indicated that most VEIMEs shared some common features that reflect their postulated integrative, replicative, and parasitic life cycles (Fig. 2 and Datasets S5 and S6). Most conspicuously, >80% of all VEIMEs encoded integrases: a tyrosine integrase, a serine recombinase, or both (Figs. 2 and 3; Datasets S3, S5, and S6; and *SI Appendix*, Figs. S2 and S4–S6). Notably, the only reference phage satellites in our analyses that encoded a serine recombinase were the three reference PLEs associated with module 10 (Fig. 2) (7, 10). Other
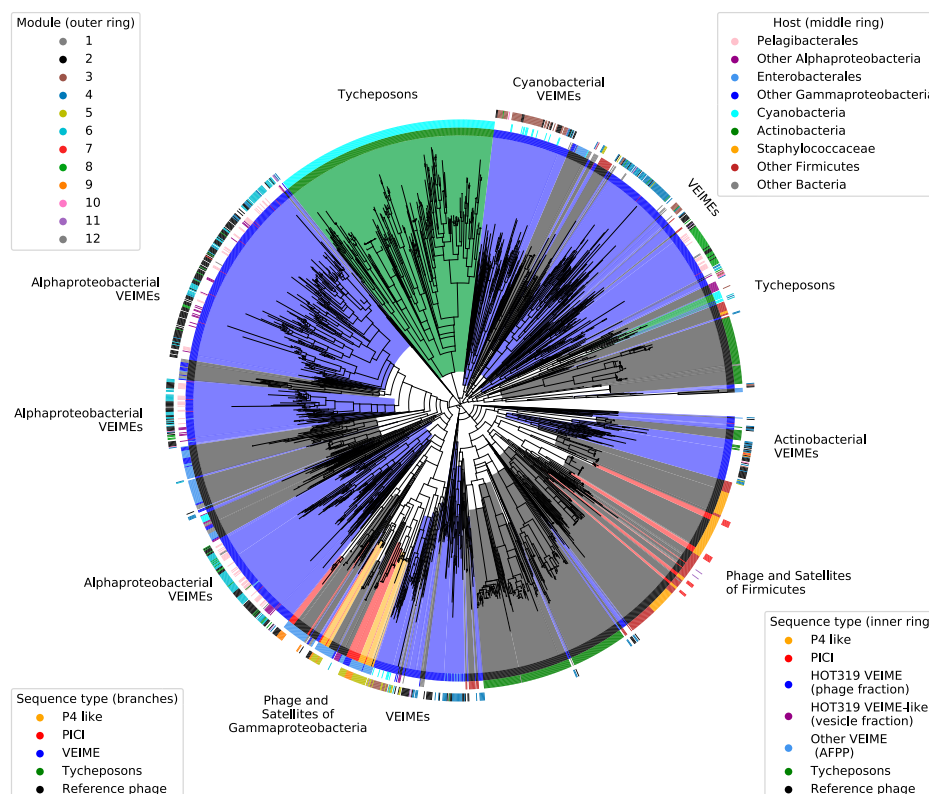
Fig. 3. Relationships of tyrosine integrases among VEIMEs, reference phage satellites, and bacteriophages. A phylogenetic tree of tyrosine recombinases found in environmentally derived and reference phage satellites, bacteriophages, and tycheposons. Branches are colored by sequence type, with environmental satellites broken down further by sample in the first colored ring. The second ring indicates the taxonomy of known hosts, with gaps for any unknown host associations. The outer ring indicates the module assignment from Fig. 2. AFPP, assembly-free putative phage-inducible chromosomal island.

mobile element–like genes shared across multiple VEIME modules included some involved in phage satellite replication, capsid assembly, and DNA packaging, including DNA primase/ helicases, the DNA-binding prophage transcriptional regulator AlpA, terminase small subunit homologs, motor-like ATPases, and excisionases (Fig. 2, Datasets S5 and S6, and *SI Appendix,* Figs. S2, S4, and S7).

The functional gene diversity and contents of marine phage satellites suggested that different groups may use a variety of mechanisms to reproduce and hijack helper phage replication cycles. All the marine VEIMEs we identified were initially detected as concatemeric repeat sequences in virus particles, which is consistent with a plasmid-like rolling circle DNA replication mechanism. Although the marine VEIMEs were found encapsidated in virion particles, they did not encode any major capsid structural genes (Datasets S2 and S3), which many known phage satellites encode and use to remodel and parasitize helper phage capsids. Together with the correspondence of viral genome and VEIME concatemer size ranges (Fig. 1*B*), these data suggest that capsid remodeling (and concomitant capsid size reduction) is not a common strategy among these marine VEIMEs, as it is in most previously described gram-positive and gram-negative phage satellite systems (6, 10, 17, 21).

While 91% of the module 3 VEIMEs encoded a serine recombinase, tyrosine integrases predominated in all the other modules (Figs. 2 and 3, Datasets S5 and S6, and *SI Appendix,* Figs. S2, S5, and S6). The VEIME-only, tyrosine integrase-dominated modules could be further differentiated by their module-specific gene content (Fig. 2, Datasets S5 and S6, and *SI Appendix,* Figs. S2, S4, and S5). For example, although VEIMEs in the *Pelagibacter*/Alphaproteobacteria-associated groups (modules 2, 6, and 8) shared some core replicative

genes, each of these modules also contained unique genes that were rare or absent among the others (Datasets S5 and S6 and *SI Appendix,* Figs. S2, S4, and S5). Module 2 VEIMEs encoded HNH endonucleases (22) and periplasmic serine proteases (23), whereas some module 6 VEIMEs contained small subunit terminases (*SI Appendix,* Fig. S7) and DNA polymerase genes (Dataset S6 and *SI Appendix,* Fig. S2). In contrast, module 8 VEIMEs encoded RecD-like helicases, hydrolases, and phage protein D14 genes (Fig. 2 and Datasets S5 and S6).

Some of these shared gene annotations are consistent with proposed phage-hijacking mechanisms found in previously described phage satellites. For example, phage satellite encoded small subunit terminases are thought to redirect the helper phage assembly process toward preferential packaging of phage satellite DNA into helper phage capsids (10). Notably, the VEIME-encoded terminase small subunit sequences shared the greatest identity with homologs from co-occurring bacteriophages from the same habitat (*SI Appendix,* Fig. S7). Additionally, HNH endonucleases are also common components of phage DNA packaging machinery (22), and some serine proteases are known to be involved in procapsid maturation (23). A variety of VEIME phage regulatory genes, restriction–modification genes, and some abortive infection–like genes, provided further evidence for the potential involvement of marine phage satellites in phage interference.

Although gene content analyses suggest that VEIMEs and known phage satellites may share some similarities in their life cycle strategies, some VEIMEs apparently use different replication strategies. Unique genes in module 3 included homologs of plasmid replication initiation and partitioning genes, hinting that some VEIMEs may employ a plasmid-like replication strategy (*SI Appendix,* Figs. S2 and S8). Module 4 VEIMEs contained the

highest proportion of excisionases, methyltransferases, YwqK-like antitoxins, and RecA-like recombinase genes (Fig. 2, Datasets S5 and S6, and *SI Appendix*, Figs. S2 and S4), further expanding the list of potential mechanisms for phage interference employed by marine phage satellites. The fact that some VEIMEs did not encode any genes known to be involved in phage satellite replication, further implies that different mechanisms of phage interference and satellite DNA packaging may be employed by marine phage satellites. Notably, integrase genes were not identified in ~20% of the VEIMEs, indicating that some VEIMEs might exist predominantly as plasmids, take advantage of host or helper phage–encoded integrases, or use as yet unidentified replication and phage interference strategies. We also found VEIME-like concatemeric elements in the lower-density membrane vesicle–enriched fraction of our seawater sample (Fig. 1*A*), suggesting that some of the mobile element concatemers might also be mobilized via extracellular vesicles (i.e., in the absence of any capsid) or potentially may be mobilized by hijacking tailless helper phages.

To further assess patterns of diversity among marine phage satellites, we next examined specific subgroups of VEIMEs that shared most of their gene content and organization, including both annotated and unannotated protein coding genes with sequence homology (*SI Appendix*, Fig. S8). Overall, the wide variety of functional gene suites observed among VEIMEs (Datasets S5 and S6 and *SI Appendix*, Fig. S8) suggests that considerable variability exists in specific parasitic molecular interactions between marine phage parasites and their helper phages. Notably, one VEIME subgroup of six VEIMEs shared nine genes among most members, including plasmid-like ParA-like and RepA-like genes, suggestive of a plasmid-like replication cycle (*SI Appendix*, Fig. S8). Other VEIME subgroups shared both a serine recombinase and a tyrosine integrase among all members (six subgroups representing a total of 48 polished

VEIMEs) (*SI Appendix*, Fig. S8). Many VEIME-encoded genes, similar to those found in the phages they parasitize, had no convincing homologs in existing databases.

Integrase Gene Phylogenies Reveal That Marine Phage Satellites Are Distinct from Previously Characterized Phage-Parasitizing Mobile Elements

VEIME gene phylogenies provided additional insight into the relationships among VEIMEs, reference phage satellites and other MGEs (Fig. 3 and *SI Appendix*, Figs. S4–S7). The tyrosine integrases found in the majority of VEIMEs formed several distinct clusters different from those of known phage satellites and bacteriophages (Fig. 3 and Figs. S4 and S5). The serine recombinases found in module 3 also exhibited considerable diversity, with one VEIME integrase clade specifically associated with known PLE serine recombinases (*SI Appendix*, Fig. S6). A few VEIME-encoded serine recombinases appeared most closely related to a new group of transposon-like MGEs (Tycheposons) recently reported in the cyanobacterium *Prochlorococcus* (24). Thus, VEIME gene content and integrase phylogenies both support the view that marine phage satellites are part of a large group composed of many diverse subtypes, most of which are distinct from previously characterized MGEs.

Marine Phage Satellites Are Ubiquitous and Abundant in the Ocean

To assess the spatiotemporal distribution of VEIMEs, we surveyed their relative abundances in metagenomic data collected at our NPSG sampling site and other global oceanic sampling stations (Fig. 4 and *SI Appendix*, Fig. S9) (25). The VEIMEs reported
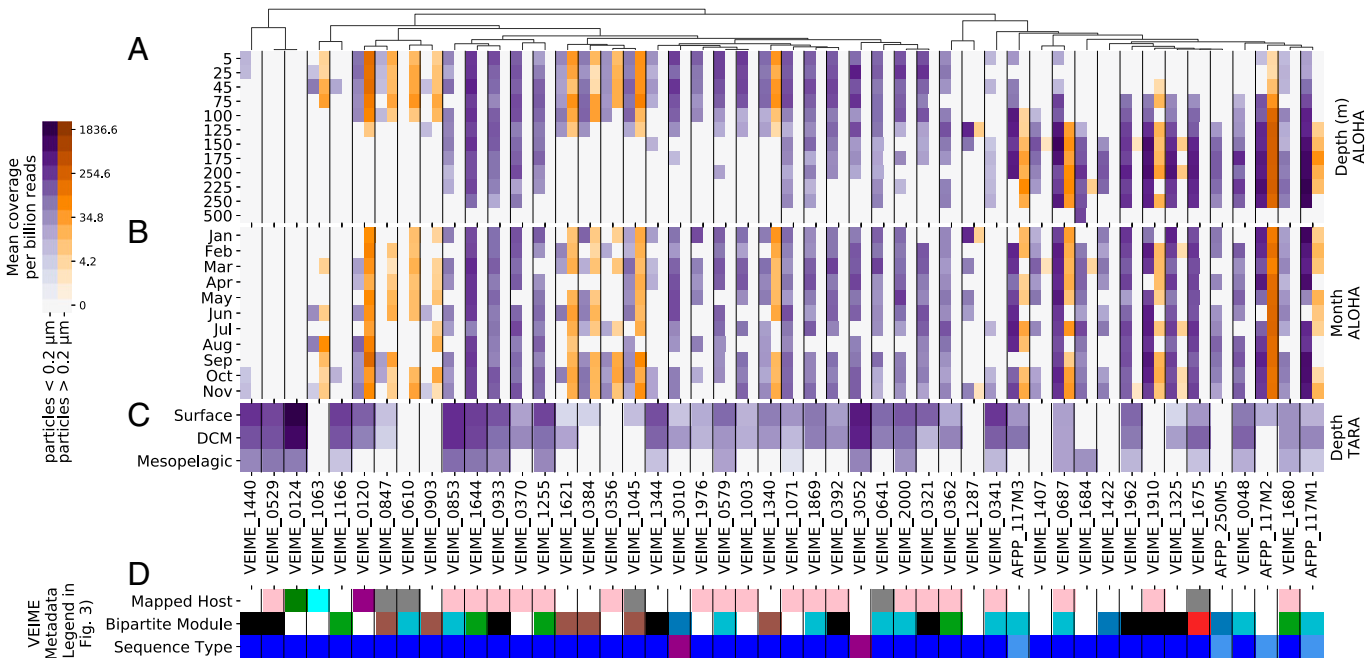

Fig. 4. VEIME DNA abundances in marine picoplankton and virioplankton size fractions. We compared VEIMEs with assembled contigs from published short-read metagenomes and used the normalized local metagenomic short-read coverage of the contigs in the aligned regions as a proxy for the environmental abundance of aligned VEIMEs. In *A*, VEIME abundances in bacterioplankton (orange) and picoplankton (purple) metagenomes from Station ALOHA are averaged by depth and displayed side by side for each monomer. In *B*, the same data are aggregated by calendar month. In *C*, VEIME abundance in Tara Oceans viral metagenomes is shown. VEIME placement along the *x* axis follows a hierarchical clustering (shown in *A*) by the correlation of abundances across all samples. In *D*, metadata on hosts, modules, and sequence types are shown using the same annotation color scheme as in Fig. 3. AFPP, assembly-free putative phage-inducible chromosomal island. DCM, deep chlorophyll maximum.

here, originally derived from a single 25-m deep NPSG seawater sample, were present primarily in shallower water virioplankton size fractions, with some more common in surface waters, others more common at greater depths, and still others with relatively uniform depth distributions (Fig. 4*A*). Temporally, the VEIMEs we identified were present year-round in the NPSG, with no clear seasonal pattern (Fig. 4*B*). VEIME homologs were also evident in virus-enriched samples in other oceanic regions (25) (Fig. 4*C* and *SI Appendix*, Fig. S9), primarily in shallow surface waters or around the deep chlorophyll maximum layer (Fig. 4*C*). Additionally, these marine phage satellites were found within the genomes of bacteria originating from diverse global ocean locales (Fig. 4D and *SI Appendix*, Fig. S3), further reflecting their ubiquity and abundance in marine plankton. Taken together, these results indicate that VEIMEs are widespread and persistent features of global ocean planktonic microbiomes worldwide.

The search for predominant cellular hosts of marine phage satellites provided additional details concerning their biology and ecology. Unlike previously reported phage satellites that have been found nearly exclusively in copiotrophic bacteria, the marine phage satellites we report here occurred in some of the most oligotrophic bacterial species known. Specifically, the most prevalent cellular hosts we found for surface water marine VEIMEs were associated with either *Pelagibacter* or the cyanobacterium *Prochlorococcus*, among the most abundant bacterial groups present in open ocean surface waters of the NPSG (26, 27). Recently, the presence of temperate phages in oligotrophic *Pelagibacter* species has been validated in the laboratory (28, 29), consistent with our observations of phage satellites within the genomes of this ubiquitous bacterioplankton group. Similarly, cyanophages of *Prochlorococcus* often encode integrases, which allows them to integrate into host genomes (30). Among VEIME-containing bacterial genomes, however (*SI Appendix,* Fig. S3), only a small percentage provided clear evidence of coexisting prophage within the same genome. It is unclear at present whether this may be just an artifact of incomplete genome assemblies from environmental samples. Oligotrophic shallow waters of the NPSG do however contain a greater proportion of lytic to temperate phages, when compared with deeper waters (27). It seems possible, therefore, that some marine VEIMEs, in analogy to the life cycle of known PLEs (10), may parasitize invading lytic phages instead of temperate phages.

The potential impacts of phage satellites on marine ecosystems may be substantial. At our sampling site in surface waters of the NPSG, viruses occur at concentrations of $\sim 1 \times 10^7$/mL (31). Since VEIMEs represented $\sim 0.6\%$ of the total virion population, we estimate that virion-encapsidated phage satellite particles occur at $\sim 60,000$ VEIMEs per milliliter in NPSG surface waters, about one-tenth the concentration of co-occurring bacterioplankton cells. Applying similar estimates using total global tropical and subtropical marine bacteriophage numbers ($\sim 5.34 \times 10^{28}$ from ref. 18), at any given time there may be as many as 3.2×10^{26} marine phage satellites packaged within virion particles in the world's oceans. Other DNA elements masquerading as phages, including generalized transducing particles and gene transfer agents (32, 33), may similarly influence virus-like particle (VLP) density estimates, but their numbers are currently difficult to constrain. Notably, the estimated 3.2×10^{26} phage satellites worldwide may represent an underestimate, since NPSG surface waters have fewer temperate phages than deeper waters (27, 34, 35), and deep water bacterioplankton have larger genomes, that may facilitate mobile element integration and propagation. The ratio of phage satellite to bacteriophage in viron particles we report here, therefore, may be greater at subsurface depths.

Conclusions

Phage satellites are known to have profound biological and ecological impacts, including the provision of phage defense and immunity (5, 9, 11, 12), the diminishment of viral productivity, and the reduction of virus-induced mortality. To date, phage satellite hosts have been demonstrated in several gram-positive genera (PICIs) (5, 17) and three gram-negative bacterial orders (Enterobacterales, Pasteurellales, Vibrionales; P4 like satellites, PICIs, PLEs) (2, 3, 6, 8, 10, 16). The data and analyses reported here considerably expand known phage satellite environmental distributions, genetic diversity, host ranges, and abundances. While most phage satellites reported to date have been found in animal-associated pathogenic bacteria, our results show that phage satellites are likely as widespread as are bacteriophages in many diverse habitats.

In the sea, aquatic viruses lyse as much as 20 to 40% of available prey cells per day (9, 10), contributing an estimated 145 gigatons to annual global carbon flux in tropical and subtropical seas (18, 34). Phage satellites are predicted to diminish (population-averaged) helper phage burst sizes and provide phage immunity to their host cells. Along with their considerable global abundance, these data indicate that marine phage satellites have the potential to influence the ocean carbon cycle; the biology of their bacterial hosts (and associated bacteriophages); and microbial dynamics, gene transfer, and genome evolution in the ocean.

Marine phage-parasitizing elements also may impact current quantitative estimates of bacteriophage numbers, virus production, and viral mortality in situ. For example, estimates of VLP numbers are typically based on epifluorescence microscopic (or flow cytometric) counts of virus-sized particles collected in situ. Similarly, electron microscopic counts of phage-like particles within infected cells have been used to estimate virus burst sizes and productivity estimates from counts of intracellular VLPs. Additionally, lysogenic phage production can be estimated by adding exogenous mitomycin C to induce lytic cycles, followed by epifluoresence microscopic or flow cytometric counts of induced phage particles (18, 34, 35). All the above methods would tend to overinflate bacteriophage abundances and production when phage satellites are present, since fluorescent or electron microscopic counts cannot discriminate bacteriophage from their hijackers, and would mistakenly count phage satellites as bacteriophages (36). Additionally, phage satellite to bacteriophage ratios may vary among different cellular hosts, among different helper phages, or in different environmental contexts (37). Considering all of the above, the recognition of marine phage satellites and their abundances may have both practical and empirical consequences as well theoretical implications for models of viral mortality.

Virtually all phage satellites described so far, including the marine phage satellites reported here, share a reliance on intracellularly co-occurring helper phage to propagate via transduction. Most known phage satellites have genome sizes of ~ 8 to 18 kbp, about one-third the size of their cognate helper phage genomes, facilitating this process. Mechanistically, many previously characterized phage satellites have been shown to modify and reduce the size of helper phage capsids, which allows for packaging of their smaller single genomes into modified capsids while simultaneously inhibiting the packaging of larger helper phage genomes. Recently, it was suggested that capsid size reduction might be a universally conserved feature of phage interference among all currently described phage satellite families (10). The marine phage satellites reported here, however, were packaged in virion

particles as concatemeric sequences, whose total lengths coincided with those of co-occurring bacteriophage genomes (Dataset S2). Individual marine phage satellite monomers had sequence lengths that ranged from <5 kbp to a maximum of 16.8 kbp, with repeat copy numbers ranging from ~8 to 17 phage satellite copies per virion-packaged concatemer (Dataset S3). Although their genome sizes are generally smaller than those of other known phage satellites, the marine phage satellites appear to be packaged within native full-sized helper phage capsids, without necessarily requiring any capsid size redirection or modification. Indeed, flexibility in packaged concatemeric copy number may represent a simple mechanism that marine phage satellites might use to adapt to and exploit other helper phages having different genome sizes and different capsid DNA content capacities.

Current data and theory suggest that both lytic and temperate bacteriophages are active in marine plankton, with lysogeny prevailing in regions of low nutrient concentrations and high virus to host cell ratios or alternatively, in habitats having greater bacterial production rates (37). A relevant feature here commonly shared among P4 like and PICIs (but not PLEs) is their strict reliance on temperate (vs. lytic) helper phages. In contrast, PLEs rely on strictly lytic helper phages (10). Some marine phage satellites may parasitize lytic phages as well. In addition, given that the ratio of temperate to lytic phages tends to increase with water column depth and bacterioplankton in deep waters tend to be copiotrophic and have larger genomes, greater proportions of phage satellites may exist in deeper ocean waters. Applying similar logic, we predict that virion particles of particle-attached bacteria may harbor greater proportions of phage satellites compared with planktonic free-living counterparts.

Accumulating evidence now suggests that MGEs frequently drive host-cell resistance to bacteriophage (38). In *Vibrio* species, for example, a variety of phage defense mechanisms appear to be associated with MGEs. Along with phage-receptor mutations, these can include integrative and conjugative elements, CRISPR-Cas, abortive–infective (Abi) systems, and restriction–modification systems that can occupy large percentages of the flexible genome (39–41). Antiphage systems are also frequently colocated with MGEs, consistent with the proposal that phage predation (and defenses against it) accelerates lateral gene transfer and thereby, influences bacterial evolution (40–43). Known phage satellites, like PICIs and PLEs, have been recently recognized as key elements of antiphage defense systems (12, 13, 44, 45). Some of the marine phage satellites reported here encoded known phage defense and interference genes (including small subunit terminases, restriction–modification systems, phage transcriptional regulators, and Abi-like genes), suggesting that they too may contribute to phage defense within their planktonic cellular hosts. As is the case with other MGEs (13, 39, 41, 43), these previously unrecognized marine phage satellites likely also impact the evolutionary rates and genetic diversity of the host cells and bacteriophage partners that they parasitize. Consistent with previous observations, while some general functional features appear well conserved among all phage satellites, diverse phage satellite lineages, including marine phage satellites, appear to have arisen independently via lateral transduction, recombination, and convergent evolution.

Many questions remain concerning the impact of phage-parasitizing mobile elements on microbial populations in the sea and elsewhere. It will be useful for example, to more completely identify the cognate helper phages and cellular hosts of the newly recognized marine phage satellites. In addition, specific helper phage genes (e.g., large subunit terminase sequences, major capsid structural genes, and tail genes)

presumably are involved in complementing marine phage satellite life cycles. Yet, specifics on how different marine phage satellites interact and interfere with their helper phages mechanistically (for example, via small subunit terminase interactions), remain to be determined. Although we were able to identify the cellular hosts of some marine phage satellites (~20% of the total), much more data and analyses will be required to disentangle the myriad of three-way interactions between phage satellites, helper phages, and their cellular hosts. Likewise, more detailed information on the biology of phage satellite–helper phage interactions will be required. Do marine phage satellites parasitize only temperate phages, or do some parasitize lytic phages, like PLEs? Can closely related phage satellites hijack different helper bacteriophages? Do some marine phage satellites have broad helper phage and/or cellular host ranges? Do some exist primarily in plasmid form, instead of integrating into cellular host genomes when helper phages are absent? Combined phage satellite–focused cultivation-independent surveys and quantitative analyses, along with cultivation-enabled model system studies, have potential to advance understanding of the biology and impacts of phage satellites in the sea. The considerable phage satellite diversity, abundance and host distributions reported here in just a single seawater sample, suggest that diverse and abundant phage satellites will likley be found in virtually any microbial habitat that also harbors bacteriophage.

Materials and Methods

Concentration of Virus- and Vesicle-Enriched Samples from Seawater. Seawater was collected from a depth of 25 m and prefiltered to minimize cellular contamination. The resulting virus- and membrane vesicle–enriched filtrate was concentrated via tangential flow filtration (TFF) before DNA extraction as outlined below. The 25 m seawater sample was collected on January 31st and February 1st, 2020 on Hawaii Ocean Time-series cruise 319 (HOT319) at Station ALOHA (22°45′ N, 158° W; https://hahana.soest.hawaii.edu/hot/).

A total of 440 L of seawater was collected using a Niskin bottle rosette attached to a conductivity–temperature–depth package. The seawater was prefiltered by peristatic pumping through a 0.1 μM Supor cartridge filter (Acropak 500; Pall). The resulting particle-enriched filtrate was concentrated by TFF over a 30-kDa filter (Biomax 30-kDa membrane, catalog no. P3B030D01; Millipore). Subsequently, the retentate was reduced to a volume of ~100 mL and stored at 4 °C. Next, ~10 L of <30 K_d permeate reserved from the TFF was added to the recirculation vessel, and the system was run for an additional 30 min to release virus particles trapped in the filter. After 30 min, the retentate volume was reduced to ~100 mL, and this flushing retentate was added to the ~100 mL of the initial retentate, resulting in a final retentate volume of ~200 mL. The virus-containing retentates were stored at 4 °C until further processing.

Small Particle Fractionation and Purification. Particles were isolated and purified from the TFF concentrate by ultracentrifugation at 32,000 rpm (~126,000 × g) for 2 h at 4 °C in a Beckman-Coulter SW32Ti rotor. The pelleted material was resuspended in residual seawater and separated across an iodixanol density gradient (Optiprep; Sigma-Aldrich). The gradient was formed as follows; successive 0.5-mL layers of iodixanol (45, 40, 35, 30, 25, 20, 15, and 10%; all in a 3.5% [wt/vol] NaCl, 3.75 mM TAPS, pH 8, 5 mM CaCl$_2$ buffer background) were placed in a 4-mL UltraClear ultracentrifuge tube (Beckman-Coulter) with the particle sample as the top layer. The gradient was spun at 45,000 rpm (~200,000 × g) for 6 h at 4 °C in an SW60Ti rotor (Beckman-Coulter). Successive 0.4-mL fractions were collected by pipetting, and their densities were measured. Fractions between 1.14 and 1.19 g/mL were pooled to form the "vesicle-enriched" sample (containing both extracellular vesicles and potentially, some nontailed viruses), and fractions >1.2 g/mL were combined to form the "tailed phage" sample (20, 46, 47). Particles in the tailed phage fraction were washed twice by diluting with filter-sterilized buffer (3.5% [wt/vol] NaCl, 3.75 mM TAPS, pH 8, 5 mM CaCl$_2$) followed by ultracentrifugal pelleting (32,000 rpm, 2 h, 4 °C, SW60Ti rotor). Particles from the vesicle-enriched sample

were washed as above, but in 1× PBS buffer. To remove any potential free DNA associated with the outside of the vesicle-enriched fraction, the sample was incubated with 2 U of TURBO DNase (Invitrogen) at 37 °C for 30 min. After a second round of TURBO DNase treatment as above, the enzyme was heat inactivated at 75 °C for 15 min.

DNA Purification from Tailed Phage and Vesicle-Enriched Fractions. Lysis and DNA purification were performed in 2-mL screw-cap vials using the Qiagen Genomic-tip 20/G protocol and buffers from the Qiagen Genomic DNA Buffer kit following the manufacturer's recommendations (Qiagen). First, an RNase A solution (200 μg/mL) was prepared by adding 20 μL of 10 mg/mL RNase A to 1 mL of Buffer B1 (50 mM Tris·HCl, pH 8, 50 mM EDTA, pH 8, 0.5% Tween-20, 0.5% Triton X-100). Next, 1 mL of the Buffer B1 lysis buffer containing RNase A was added to 35 μL of viral concentrate. Then, 45 μL of a Proteinase K solution (20 mg/mL) prepared in sterile water was added, followed by the addition of 350 μL of Buffer B2 lysis buffer (3 M guanidine HCl, 20% Tween-20). The samples were then incubated at 50 °C for 120 min.

The resulting lysate was then loaded onto a Qiagen Genomic-tip 20/G column, and purification was performed following the manufacturer's recommendations (Qiagen). The Genomic-tip 20/G column was equilibrated with 1 mL of Buffer QBT equilibration buffer (750 mM NaCl, 50 mM MOPS, pH 7.0, 15% isopropanol, 0.15% Triton X-100) by gravity flow. Sample lysate was then mixed by inverting several times and carefully pipetted sequentially onto an equilibrated Genomic-tip 20/G column, allowing the sample to enter the resin by gravity flow. Next, 1 mL Buffer B1 was combined with 350 μL Buffer B2; the sample lysate tube was gently rinsed with 1 mL of this solution, and the rinse solution was applied to the Genomic-tip 20/G column. The Genomic-tip 20/G column was washed by gravity flow by applying 1 mL of Buffer QC wash buffer (1.0 M NaCl, 50 mM MOPS, pH 7, 15% isopropanol) three times in succession. Finally, the high–molecular weight genomic DNA was eluted from the column by two successive applications of 1 mL Buffer QF elution buffer (1.25 M NaCl, 50 mM Tris·HCl, pH 8.5, 15% isopropanol) prewarmed to 50 °C, resulting in purified DNA preparations in a 2-mL final volume.

The column-purified DNA was concentrated by isopropanol precipitation as follows. The DNA eluant was split into two 2-mL conical screw-cap tubes, with 1 mL per tube. The DNA was precipitated by adding 0.7 mL of room temperature isopropanol per each 1 mL of DNA solution, followed by mixing by gentle inversion. After 2 h at room temperature, the DNA was pelleted by centrifugation at 10,000 × g for 30 min at 4 °C. The supernatant was removed, and the DNA precipitate was washed by the gentle addition of 1 mL of cold 70% ethanol and incubation for 60 s, followed by centrifugation at 10,000 × g for 15 min at 4 °C. The supernatant was removed, and the DNA pellet was air dried for 5 min. The purified DNA pellets were resuspended in a final volume of 26 μL of 1× TE buffer (10 mM Tris·HCl, pH 8.0, 1 mM EDTA, pH 8.0) and allowed to dissolve for a minimum of 10 min at room temperature before final storage at 4 °C. Final DNA quantity and quality were assessed initially by spectrophotometry and quantified via Quant-iT Picogreen dsDNA fluorometric assay (catalog no. P7589; Invitrogen). The virus-enriched sample collected from 440 L of 0.22-μm prefiltered seawater collected at a depth of 25 m yielded a total of 5.1 μg of purified high–molecular weight DNA. The total yield for the corresponding vesicle fraction was 8.2 μg of purified high–molecular weight DNA.

Oxford Nanopore Sequencing Methodology. Virus- and vesicle-enriched DNAs were processed using the Nanopore Ligation Sequencing Kit (LSK-109; Oxford Nanopore Technologies, Ltd.) following the manufacturer's instructions for the processing of high–molecular weight DNA. A total of two virus-enriched libraries were prepared using 2 and 1.5 μg of DNA each for the sequencing runs. All libraries were sequenced on a GridION X5 using FLO-MIN106 (R 9.4.1) flow cells (Oxford Nanopore Technologies, Ltd.). Read base calls were generated from the signal traces using Guppy version 3.0. The sequencing yield for the virus-enriched fraction totaled 31 Gbp, generating reads with an N50 of 37.67 kb (i.e., half of all bases are in reads at least 37,670 bases long). The sequencing yield for the corresponding membrane vesicle–enriched fraction totaled 46 Gbp, with an N50 of 4.6 kbp.

Concatemer Detection. Repetitive long reads were identified as follows. In the first phase, each read over 5 kb was compared with itself using minimap2 (48) and lastal (49). Only forward strand self-hits of at least 500 bp (*SI Appendix,*

Fig. S1A) were retained from either method. In the second phase, if a read had three or more self-hits (a full-length central hit and at least one hit offset in each direction), a repeat size was calculated from the hit positions using two different methods.

The "clust" method first calculates the offset of each hit as the average of the difference between that hit's start positions and the difference between its end positions (*SI Appendix,* Fig. S1B). Next, the clust method groups offsets using agglomerative clustering to allow for fragmented hits. Finally, the median distance between neighboring offset groups is taken as the repeat size. The "fft" method starts with hit offsets, smooths their locations with kernel density estimation (*SI Appendix,* Fig. S1C), uses fast Fourier transform to find the dominant frequency, and takes the corresponding wavelength as the repeat size.

For all four combinations of the two methods and two search tools, the calculated repeat size was compared with the actual hits. If the hits covered at least 50% of the expected single-offset self-hit (the red line in *SI Appendix,* Fig. S1A) and the repeat size was less than 60% of the total read length, then the read was marked as a concatemer by that method. If a read was marked as concatemeric by any method, then it was considered a concatemer using the median repeat size of all methods that flagged it (Dataset S2). Code for finding concatemers and calculating repeat size is available at https://github.com/jmeppley/concatemer_finding.

VEIME Generation. Polished VEIME sequences (Dataset S3) were generated from concatemeric reads as follows. Each read was broken up into nonoverlapping repeat-sized fragments. All fragments were compared with each other with lastal (49), retaining hits of at least 80% of the fragment length. Fragments from different reads were merged into one pool for polishing if the majority of the fragments from each read were highly similar (>80% sequence similarity over >80% of their lengths). The consensus sequence of each pool of fragments was determined using three passes of racon (50) with a final pass of medaka (https://github.com/nanoporetech/medaka), following the assembly-free viral genome (AFVG) polishing methods as previously described (19). The consensus sequences were further polished using two passes of racon with Illumina short reads derived from the same sample as previously described (19). Briefly, genomic DNA from each sample was sheared to an average size of 350 bp using a Covaris M220 Focused ultrasonicator (Covaris) with Micro AFA fiber tubes (no. 520166; Covaris). Libraries were sequenced using a 150-bp paired-end NextSeq High Output V2 reagent kit (FC-404-2004; Illumina). Finally, to improve gene predictions, erroneous frameshifts were corrected where possible using proovframe (51), and genes were retrieved from the Genome Taxonomy Database (GTDB) version 95 (52).

Bipartite Network Modules. Genomic sequences from the VEIMEs–including the VEIME-like monomers derived from the membrane vesicle fraction–were pooled with previously published phage satellites (*Reference Sequences*), assembled into a bipartite network, and partitioned into modules (Datasets S3, S5, and S6) as follows. Genes predicted using prodigal (53) were annotated independently with both EggNOG (54) using eggNOG-mapper v2 (55) and VOGdb using hmmsearch (56). Details of the tools used, including versions and parameters, are provided in Dataset S1.

A bipartite network (Fig. 2) was defined with satellites as one set of nodes and gene families as the other, with an edge between a satellite and gene family if that satellite includes a gene annotated with that gene family. The bipartite network comprised two connected components. The smaller component, containing three VEIMEs, was dubbed module 1, and the larger was partitioned into modules 2 through 12 using the constant Potts model (57) as implemented by the leidenalg Python package (Dataset S1). For most analyses, modules 1 and 12, each containing fewer than five sequences, were ignored.

Smaller groups of VEIMEs with significant shared gene content were found with a brute force search of the gene-sharing network (*SI Appendix,* Fig. S8). Twelve groups of 5 to 12 VEIMEs were found to share four to five gene annotations. No groups were found sharing more than five annotations. Unannotated genes were clustered by using the bit scores from a lastal (49) all vs. all homology search as the distance matrix for clustering with mcl (58) (Dataset S1).

Reference Sequences. Polished VEIMEs were searched for in three databases of marine microbial genomes: GTDB (59), GORG (60), and Mar. Micro. DB (61) using minimap2 (Dataset S1). Only hits covering at least 80% of the VEIME

sequence were kept (*SI Appendix*, Fig. S3). The taxonomy of the best hit was assumed to be the host taxonomy for the VEIME (Dataset S3).

To provide context for the concatemer-derived mobile element sequences, previously published phage satellite sequences and viral genomes were downloaded from the National Center for Biotechnology Information (NCBI). Phage satellite references include the PICIs listed in refs. 2, 6, 44, and 62. Reference phage genomes were chosen for the tailed phage with complete genomes and known bacterial hosts. In all, 90 satellites and 1,700 viral genomes were downloaded (Dataset S4).

Additionally, 16 monomers, taken from previously published concatemeric assembly-free putative PICIs (22), were included in the analyses (Dataset S3). Similar to the VEIMEs, these were generated from concatemeric Nanopore reads sequenced from 0.2-μm-filtered, TFF-concentrated seawater collected at Station ALOHA.

Finally, long reads with direct terminal repeats (DTRs) were gathered from the tailed phage fraction and self-polished into AFVGs using the DTR–phage pipeline (22). These were used to provide a distribution of phage genome sizes in this environment (Fig. 1*B*) and a set of environmental terminase small subunit genes for comparison with VEIMEs (*SI Appendix*, Fig. S7).

Environmental Metagenome Abundances. The historical presence of VEIME sequences in Station ALOHA waters was assessed as follows. Raw metagenomic reads were collected from 456 metagenomic samples collected from Station ALOHA between 2014 and 2018 at depths ranging between 5 and 500 m (Dataset S7) and assembled into contigs as described previously (27). VEIME sequences were mapped against contigs using minimap2, and hits covering at least 80% of the VEIME were kept. Raw reads were mapped against contigs using Burrows-Wheeler Aligner's BWA-MEM algorithm (63) and used to calculate base by base coverage of contigs at VEIME hit locations. The mean contig coverage across the hit location was normalized to the number of sequenced reads and used as a proxy for environmental VEIME abundance.

Additionally, the global distribution of VEIMEs was assessed using using Tara Oceans' Global Ocean Virome GOV 2.0 (GOV 2.0) reads and assemblies (25). Assembled GOV contigs were downloaded from iVirus and matched by sampling station and depth to raw reads obtained from NCBI (Dataset S7). VEIME abundance in these samples was calculated as for the Station ALOHA metagenomes.

Phylogenetic Trees. Phylogenetic trees were inferred for tyrosine integrases using the VOG00035 hmm (Fig. 3 and *SI Appendix*, Fig. S4), for serine recombinases using VOG00893 (*SI Appendix*, Fig. S6), and for terminase small subunits using VOG06274 (*SI Appendix*, Fig. S7). For each selected gene annotation, genes from VEIMEs and reference sequences were aligned individually using hmmalign (56) and combined into a single multiple sequence alignment (MSA) for each gene. MSAs were trimmed of columns present in fewer than half of the genomes, and genomes containing fewer than 5% of the amino acids were removed from the alignments. Trees were inferred from MSAs using IQ-TREE (64) with the partitioned LG + GAMMA model (65).

Author affiliations: ^aDaniel K. Inouye Center for Microbial Oceanography: Research and Education, University of Hawaii, Honolulu, HI 96822; and ^bDepartment of Biological Sciences, Wellesley College, Wellesley, MA 02481

1. R. Calendar, Ed., *The Bacteriophages* (Oxford University Press, New York, NY, ed. 2, 2006).
2. E. W. Six, C. A. Klug, Bacteriophage P4: A satellite virus depending on a helper such as prophage P2. *Virology* **51**, 327–344 (1973).
3. B. H. Lindqvist, G. Dehò, R. Calendar, Mechanisms of genome propagation and helper exploitation by satellite phage P4. *Microbiol. Rev.* **57**, 683–702 (1993).
4. J. Chen, R. P. Novick, Phage-mediated intergeneric transfer of toxin genes. *Science* **323**, 139–141 (2009).
5. G. Ram *et al.*, Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16300–16305 (2012).
6. A. Fillol-Salom *et al.*, Phage-inducible chromosomal islands are ubiquitous within the bacterial universe. *ISME J.* **12**, 2114–2128 (2018).
7. K. D. Seed *et al.*, Evidence of a dominant lineage of *Vibrio cholerae*-specific lytic bacteriophages shed by cholera patients over a 10-year period in Dhaka, Bangladesh. *MBio* **2**, e00334-10 (2011).
8. Z. K. Barth, Z. Netter, A. Angermeyer, P. Bhardwaj, K. D. Seed, A family of viral satellites manipulates invading virus gene expression and can affect cholera toxin mobilization. *mSystems* **5**, e00358-20 (2020).
9. K. N. LeGault, Z. K. Barth, P. DePaola, K. D. Seed, A phage parasite deploys a nicking nuclease effector to inhibit viral host replication. *Nucleic Acids Res.* **50**, 8401–8417 (2022).
10. R. Ibarra-Chavez, M. F. Hansen, R. Pinilla-Redondo, K. D. Seed, U. Trivedi, Phage satellites and their emerging applications in biotechnology. *FEMS Microbiol. Rev.* **45**, fuab031 (2021).
11. J. R. Penadés, G. E. Christie, The phage-inducible chromosomal islands: A family of highly evolved molecular parasites. *Annu. Rev. Virol.* **2**, 181–201 (2015).
12. F. Rousset *et al.*, Phages and their satellites encode hotspots of antiviral systems. *Cell Host Microbe* **30**, 740–753.e5 (2022).
13. R. Ibarra-Chávez, A. Brady, J. Chen, J. R. Penadés, A. F. Haag, Phage-inducible chromosomal islands promote genetic variability by blocking phage reproduction and protecting transductants from phage lysis. *PLoS Genet.* **18**, e1010146 (2022).
14. J. Chen, G. Ram, J. R. Penadés, S. Brown, R. P. Novick, Pathogenicity island-directed transfer of unlinked chromosomal virulence genes. *Mol. Cell* **57**, 138–149 (2015).
15. R. Rezaei Javan, E. Ramos-Sevillano, A. Akter, J. Brown, A. B. Brueggemann, Prophages and satellite prophages are widespread in *Streptococcus* and may play a role in pneumococcal pathogenesis. *Nat. Commun.* **10**, 4852 (2019).
16. J. A. Moura de Sousa, E. P. C. Rocha, To catch a hijacker: Abundance, evolution and genetic diversity of P4-like bacteriophage satellites. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **377**, 20200475 (2022).
17. R. Martínez-Rubio *et al.*, Phage-inducible islands in the Gram-positive cocci. *ISME J.* **11**, 1029–1042 (2017).
18. E. Lara *et al.*, Unveiling the role and life strategies of viruses from the surface to the dark ocean. *Sci. Adv.* **3**, e1602565 (2017).
19. J. Beaulaurier *et al.*, Assembly-free single-molecule sequencing recovers complete virus genomes from natural microbial communities. *Genome Res.* **30**, 437–446 (2020).
20. S. J. Biller *et al.*, Bacterial vesicles in marine ecosystems. *Science* **343**, 183–186 (2014).
21. H. Murialdo, M. Feiss, Enteric chromosomal islands: DNA packaging specificity and role of lambda-like helper phage terminase. *Viruses* **14**, 818 (2022).
22. S. Kala *et al.*, HNH proteins are a widespread component of phage DNA packaging machines. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6022–6027 (2014).
23. H. Cheng, N. Shen, J. Pei, N. V. Grishin, Double-stranded DNA bacteriophage prohead protease is homologous to herpesvirus protease. *Protein Sci.* **13**, 2260–2269 (2004).
24. T. Hackl *et al.*, Novel integrative elements and genomic plasticity in ocean ecosystems. bioRxiv [Preprint] (2020). 10.1101/2020.12.28.424599 (Accessed 1 May 2021).
25. A. C. Gregory *et al.*; Tara Oceans Coordinators, Marine DNA viral macro- and microdiversity from pole to pole. *Cell* **177**, 1109–1123.e14 (2019).
26. D. R. Mende *et al.*, Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nat. Microbiol.* **2**, 1367–1373 (2017).
27. E. Luo, J. M. Eppley, A. E. Romano, D. R. Mende, E. F. DeLong, Double-stranded DNA virioplankton dynamics and reproductive strategies in the oligotrophic open ocean water column. *ISME J.* **14**, 1304–1315 (2020).
28. R. M. Morris, K. R. Cain, K. L. Hvorecny, J. M. Kollman, Lysogenic host-virus interactions in SAR11 marine bacteria. *Nat. Microbiol.* **5**, 1011–1015 (2020).
29. S. Du *et al.*, Genomic diversity, life strategies and ecology of marine HTVC010P-type pelagiphages. *Microb. Genom.* **7**, 000596 (2021).
30. D. Shitrit *et al.*, Genetic engineering of marine cyanophages reveals integration but not lysogeny in T7-like cyanophages. *ISME J.* **16**, 488–499 (2022).
31. J. R. Brum, Concentration, production and turnover of viruses and dissolved DNA pools at Stn ALOHA, North Pacific Subtropical Gyre. *Aquat. Microb. Ecol.* **41**, 103–113 (2005).
32. A. S. Lang, A. B. Westbye, J. T. Beatty, The distribution, evolution, and roles of gene transfer agents in prokaryotic genetic exchange. *Annu. Rev. Virol.* **4**, 87–104 (2017).
33. E. S. Esterman, Y. I. Wolf, R. Kogay, E. V. Koonin, O. Zhaxybayeva, Evolution of DNA packaging in gene transfer agents. *Virus Evol.* **7**, veab015 (2021).
34. R. Zhang, M. G. Weinbauer, P. Peduzzi, Aquatic viruses and climate change. *Curr. Issues Mol. Biol.* **41**, 357–380 (2021).
35. M. G. Weinbauer, Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* **28**, 127–181 (2004).
36. P. Forterre, N. Soler, M. Krupovic, E. Marguet, H. W. Ackermann, Fake virus particles generated by fluorescence microscopy. *Trends Microbiol.* **21**, 1–5 (2013).
37. B. Knowles *et al.*, Lytic to temperate switching of viral communities. *Nature* **531**, 466–470 (2016).
38. S. Doron *et al.*, Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018).
39. A. Angermeyer *et al.*, Evolutionary sweeps of subviral parasites and their phage host bring unique parasite variants and disappearance of a phage CRISPR-Cas system. *mBio* **13**, e0308821 (2022).
40. K. N. LeGault *et al.*, Temporal shifts in antibiotic resistance elements govern phage-pathogen conflicts. *Science* **373**, eabg2166 (2021).
41. F. A. Hussain *et al.*, Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages. *Science* **374**, 488–492 (2021).

42. S. Benler, E. V. Koonin, Recruitment of mobile genetic elements for diverse cellular functions in prokaryotes. *Front. Mol. Biosci.* **9**, 821197 (2022).

43. E. V. Koonin, K. S. Makarova, Y. I. Wolf, M. Krupovic, Evolutionary entanglement of mobile genetic elements and host defence systems: Guns for hire. *Nat. Rev. Genet.* **21**, 119–131 (2020).

44. B. J. O'Hara, Z. K. Barth, A. C. McKitterick, K. D. Seed, A highly specific phage defense system is a conserved feature of the *Vibrio cholerae* mobilome. *PLoS Genet.* **13**, e1006838 (2017).

45. A. Fillol-Salom, L. Miguel-Romero, A. Marina, J. Chen, J. R. Penadés, Beyond the CRISPR-Cas safeguard: PICI-encoded innate immune systems protect bacteria from bacteriophage predation. *Curr. Opin. Microbiol.* **56**, 52–58 (2020).

46. S. G. John *et al.*, A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ. Microbiol. Rep.* **3**, 195–202 (2011).

47. K. M. Kauffman *et al.*, Viruses of the Nahant Collection, characterization of 251 marine Vibrionaceae viruses. *Sci. Data* **5**, 180114 (2018).

48. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

49. S. M. Kiełbasa, R. Wan, K. Sato, P. Horton, M. C. Frith, Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).

50. R. Vaser, I. Sović, N. Nagarajan, M. Šikić, Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).

51. T. Hackl *et al.*, proovframe: Frameshift-correction for long-read (meta)genomics. bioRxiv [Preprint] (2021). https://doi.org/10.1101/2021.08.23.457338 (Accessed 26 August 2021).

52. D. H. Parks *et al.*, GTDB: An ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* **50**, D785–D794 (2022).

53. D. Hyatt *et al.*, Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

54. J. Huerta-Cepas *et al.*, eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47** (D1), D309–D314 (2019).

55. C. P. Cantalapiedra, A. Hernández-Plaza, I. Letunic, P. Bork, J. Huerta-Cepas, eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).

56. S. R. Eddy, Accelerated profile HMM searches. *PLOS Comput. Biol.* **7**, e1002195 (2011).

57. V. A. Traag, P. Van Dooren, Y. Nesterov, Narrow scope for resolution-limit-free community detection. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **84**, 016114 (2011).

58. A. J. Enright, S. Van Dongen, C. A. Ouzounis, An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).

59. C. Rinke *et al.*, A standardized archaeal taxonomy for the Genome Taxonomy Database. *Nat. Microbiol.* **6**, 946–959 (2021).

60. M. G. Pachiadaki *et al.*, Charting the complexity of the marine microbiome through single-cell genomics. *Cell* **179**, 1623–1635.e11 (2019).

61. J. W. Becker, S. L. Hogle, K. Rosendo, S. W. Chisholm, Co-culture and biogeography of *Prochlorococcus* and SAR11. *ISME J.* **13**, 1506–1519 (2019).

62. L.-M. Bobay, M. Touchon, E. P. C. Rocha, Pervasive domestication of defective prophages by bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 12127–12132 (2014).

63. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [Preprint] (2013). https://arxiv.org/abs/1303.3997 (Accessed 6 August 2020).

64. B. Q. Minh *et al.*, IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

65. O. Chernomor, A. von Haeseler, B. Q. Minh, Terrace aware data structure for phylogenomic inference from supermatrices. *Syst. Biol.* **65**, 997–1008 (2016).

66. J. M. Eppley, HOT cruise 319 virus and vesicles that were concentrated, purified and sequenced from Station ALOHA seawater. NCBI BioProject. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA855972. Deposited 19 July 2022.