nature communications

Article

https://doi.org/10.1038/s41467-025-57276-w

Distinct horizontal gene transfer potential of extracellular vesicles versus viral-like particles in marine habitats

Received: 27 June 2024

Accepted: 13 February 2025

Published online: 03 March 2025

Check for updates

Steven J. Biller[1] ✉, M. Gray Ryan[1], Jasmine Li[1], Andrew Burger[2], John M. Eppley[2], Thomas Hackl[3] & Edward F. DeLong[2]

Horizontal gene transfer (HGT) is enabled in part through the movement of DNA within two broad groups of small (<0.2 μm), diffusible nanoparticles: extracellular vesicles (EVs) and virus-like particles (VLPs; including viruses, gene transfer agents, and phage satellites). The information enclosed within these structures represents a substantial portion of the HGT potential available in planktonic ecosystems, but whether some genes might be preferentially transported through one type of nanoparticle versus another is unknown. Here we use long-read sequencing to compare the genetic content of EVs and VLPs from the oligotrophic North Pacific. Fractionated EV-enriched and VLP-enriched subpopulations contain diverse DNA from the surrounding microbial community, but differ in their capacity and encoded functions. The sequences carried by both particle types are enriched in mobile genetic elements (MGEs) as compared with other cellular chromosomal regions, and we highlight how this property enables novel MGE discovery. Examining the *Pelagibacter* mobilome reveals >7200 distinct chromosomal fragments and MGEs, many differentially partitioned between EVs and VLPs. Together these results suggest that distinctions in nanoparticle contents contribute to the mode and trajectory of microbial HGT networks and evolutionary dynamics in natural habitats.

The structure, function, and evolution of the marine microbiome are profoundly influenced by horizontal gene transfer (HGT). HGT is a primary driver of microbial genome and pangenome evolution, new gene acquisition, and functional innovation[1–4]. The process of HGT is ubiquitous and promiscuous, enabling gene transfer between both closely and distantly related organisms[1,5,6]. Gene transfers between microbes ultimately impact metabolic and functional repertoires and interorganismal interactions[2,7–10], and are particularly important for adaptive niche specialization of bacteria and archaea[11–13]. A successful HGT event requires that DNA first physically moves between cells, that the fragment is stably replicated within the recipient, and then that the information is maintained following cell division. Microbial evolutionary dynamics are thus influenced by variables influencing every step of this process beginning with factors affecting the HGT potential, or the total DNA available to be exchanged, in an environment.

The canonical view of microbial HGT has historically considered three primary mechanisms that move genetic information between cells: conjugation (movement of DNA via direct contact enabled by specialized pili), natural transformation (cell-directed uptake of free DNA from the environment), and transduction (transfer of DNA via viruses). In aquatic environments, HGT via conjugation is primarily used within surface-associated microbiomes on particles or biofilms

[1]Department of Biological Sciences, Wellesley College, Wellesley, MA, USA. [2]Department of Oceanography, Daniel K. Inouye Center for Microbial Oceanography: Research and Education (C-MORE), University of Hawai'i at Manoa, Honolulu, HI, USA. [3]Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, the Netherlands. ✉e-mail: sbiller@wellesley.edu

where cells can maintain the direct and sustained contact required[14,15]. Free DNA, required for natural transformation, is found at relatively high concentrations in the oceans—sometimes equaling the amount of DNA found in the co-occurring cellular fraction per liter[16]. Despite this, natural transformation appears to be far from a universal process in the marine microbiome, as it depends on genetically encoded systems common in surface-associated communities[14] but rarer among free-living cells[17–19]. Viruses, by contrast, are likely key mediators of marine HGT[20]. Phage infecting marine cyanobacteria mispackage chromosomal DNA at a substantial frequency, suggesting that viral transduction could be a common occurrence[21,22]. Other virus-like particles (VLPs), such as gene transfer agents or phage satellites, are similarly abundant in the oceans and can influence microbial communities[23,24]. However, viral host ranges are generally, though not always, considered to be taxonomically limited[25–29], raising questions as to whether viral transduction alone is sufficient to explain the breadth and frequency of HGT events observed in the marine microbiome.

Among the more recently recognized HGT modalities is DNA transfer via extracellular vesicles (EVs)—small (~20–200 nm diameter), spherical, membrane-bound structures released from most, if not all, cells[30]. EVs can be continually released from intact bacteria when sections of outer membranes bleb away from the surface, or following cell lysis events when small membrane fragments re-anneal[31–33]. These structures provide a vehicle for transporting and delivering diverse biological molecules—including nucleic acids, proteins, and metabolites—between cells, thereby contributing to a range of biological processes[34,35]. Bacterial EVs can move DNA into bacteria, archaea, and eukaryotes, indicating that they function as versatile vehicles for HGT[36,37]. DNA is heterogeneously distributed within EV populations[38], though how this material becomes packaged in EVs remains unclear. Nucleic acids likely enter EVs via multiple mechanisms, including the formation of vesicles containing both inner and outer membranes in Gram-negative cells[39]; capture of DNA during re-annealing of lytic membrane fragments[33]; directed packaging[40]; or other unknown mechanisms that may exist for moving DNA across membranes[34,41].

EVs are ubiquitous in coastal and open ocean environments, reaching concentrations of at least ~10^5–10^6 per mL[42]. VLPs are also widespread in the oceans where they can be found at concentrations exceeding ~10^7 per mL, often more than ~10X that of planktonic cells[20]. Together, these two groups of nanoparticles enclose a substantial fraction of the total dissolved extracellular DNA within the oceans enclosed within discrete, protected entities and distinct from free DNA[16,43]. Marine EVs and VLPs can contain both chromosomal regions that integrate into a recipient chromosome through homologous recombination as well as mobile genetic elements (MGEs) such as plasmids, transposons, integrative and conjugative elements (ICEs), phage-inducible chromosomal islands (PICIs), phage satellites, and tycheposons[24,43,44]. These elements can be maintained either as semi-autonomous entities within the cell, or else utilize integrase/recombinase enzymes to integrate into the cellular genome. The many differences between EVs and VLPs, including their physical properties, mechanisms of DNA delivery to recipient cells, and host ranges, indicate that they may play distinct roles in mediating marine HGT. For instance, while biases likely exist in vesicle interaction networks[45,46], abundant examples of EV-mediated interactions among bacteria and across domains[37] suggest that EVs may be able to deliver DNA to a broader range of cells than viruses[47]. Thus, distinctions in the genetic content contained within these two types of vehicles could represent a factor influencing HGT patterns.

In this work, we examine how the nanoparticle-associated portion of the marine dissolved information pool—i.e. the raw material available for HGT—is partitioned between EVs and VLPs. We use long-read sequencing of DNA contained within fractionated nanoparticle populations from the surface ocean to determine the origins and capacities of different particle types. We then compare their genetic cargo, including MGEs, and characterize the HGT potential of an abundant marine heterotroph. Together, our results show that EVs and VLPs contain distinct genetic potential, likely affecting the mode and trajectory of microbial HGT dynamics in the ocean microbiome.

Results

Marine particles differ in their DNA-carrying capacity

To compare the HGT potential of marine EVs and VLPs, we examined naturally occurring nanoparticles (<0.2 μm) from a 440 L seawater sample collected at 25 m depth in the oligotrophic North Pacific Subtropical Gyre (Fig. 1a). As EVs and VLPs are heterogeneous structures with overlapping size ranges (generally between 50 and 250 nm), we used density gradient ultracentrifugation to separate the nanoparticles into two fractions. This fractionation process is imperfect, but density gradients can partition most EVs from tailed phage and some tailless phages[42,43,48]; hence we refer to the resulting fractions as EV-enriched and VLP-enriched. As expected, the VLP-enriched fraction contained larger particles than the EV-enriched fraction (Table 1; Supplementary Fig. 1a), consistent with expected morphological and compositional differences[42].

To directly examine the genetic content of nanoparticle populations, we isolated and sequenced DNA from the two fractions using both short and long-read sequencing. As our purification procedures removed free DNA found outside of particles (Methods), we infer that the obtained DNA was located within or otherwise tightly associated with the nanoparticles. We observed a significant difference in the overall carrying capacity of the two-particle fractions as measured by read length (two-sided Wilcoxon test $p < 2.2e\text{-}16$; Table 1; Fig. 1b). DNA reads from the EV-enriched fraction ranged between 100 s of bp to 100 s of kb (N50 = ~3 kb; maximum 183 kb). The VLP-enriched particles contained markedly longer DNA fragments (N50 of ~37 kb; maximum 233 kb), exhibiting distinct peaks at lengths corresponding to known phage genome sizes[24] (Fig. 1b). These DNA fragment lengths represent a lower bound measurement of particle capacity in the surface ocean and are broadly consistent with previous reports of EV content in ocean samples[49] and cultured isolates[40,50]. Thus, both marine EVs and VLPs can transport genetic material of sufficient length to mediate HGT of individual genes, complete operons, or more. We hypothesize that the observed differences in DNA lengths reflect mechanistic differences in how EVs and VLPs package DNA. For instance, while viruses typically package long strands of DNA selectively and actively into capsids via ATP-requiring proteins, EVs appear to enclose DNA using relatively passive and nonspecific mechanisms that may limit their genetic capacity[34].

Cellular sources of DNA within marine nanoparticles

Sequences of microbial and viral origin were present in both nanoparticle fractions (Fig. 1c). Taxonomically classifiable reads indicated contributions of at least 75 different bacterial and archaeal phyla to the particle-associated DNA pool, consistent with previous findings on marine EV contents[42,43,49]. While not the focus of our analysis, reads with significant homology to eukaryotes such as protozoa, fungi, and coccolithophores were also present. Each taxa's relative representation within the EV-enriched and VLP-enriched fractions was positively correlated with cellular abundance (Pearson's correlation = 0.93 and 0.95; df = 26, $t = 13.41$ and 16.44; $p = 3.44 \times 10^{-13}$ and 2.94×10^{-15}, respectively; Fig. 1d), suggesting that most marine microbes contribute proportionally to extracellular DNA pools. Across all taxa, ~14.5% of the cellular (>0.2 μm) metagenome was found mobilized within marine particles. The true total representation of cellular genomic potential within EVs and VLPs is undoubtedly higher, as our sequencing depth did not saturate particle-associated DNA diversity (Supplementary Fig. 1b). Read length distributions in the EV-enriched fraction were broadly similar among taxa (Supplementary Fig. 2a), with the longest reads reaching 40–60 kb. Manual examination confirmed

https://doi.org/10.1038/s41467-025-57276-w

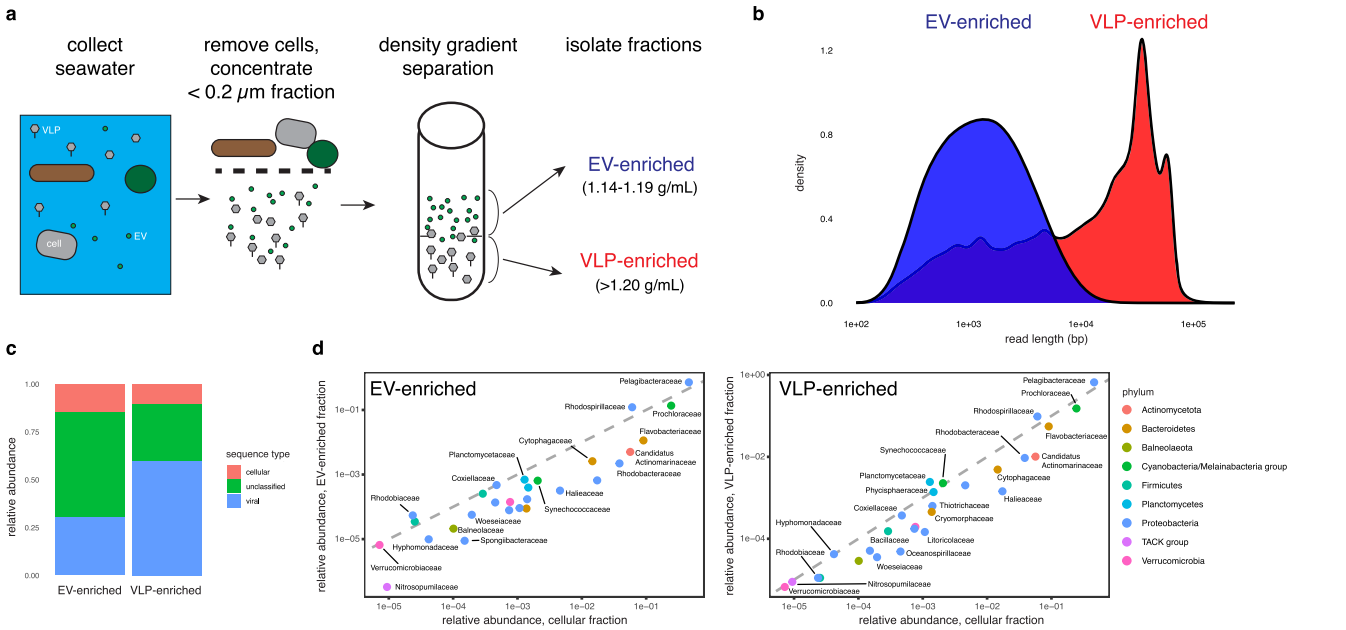

Fig. 1 | Composition of DNA within subfractions of marine particles. a Schematic overview of sample collection and particle fractionation methodology. **b** Size distribution of DNA fragments contained within either the EV-enriched (blue) or VLP-enriched (red) fractions, as determined by nanopore sequence length. **c** Relative abundance of cellular, viral, and unclassified reads in the EV-enriched and VLP-enriched particle fractions. Colors indicate sequence type. **d** Relative abundance of sequences taxonomically classifiable to the family level in the EV-enriched and VLP-enriched particle fractions, as compared with the cellular community from the same water sample. Colors indicate phylum; selected families are labeled. Source data are provided as a Source Data file.

Table 1 | Particle fraction and sequencing summary statistics

	EV-enriched fraction	VLP-enriched fraction
Mode particle size (nm)	83	95
Total sequences (>1 kb)	11,574,853	1,097,084
Total yield (Gbp)	33.86	25.51
Sequence N50 (bp)	3138	36,976
Predicted protein-coding genes	38,489,055	27,565,952

that many of these long EV-enriched fraction reads were likely cellular in origin, such as fragments that matched *Prochlorococcus* and *Pelagibacter* reference genomes at >87% nucleotide identity across ~10–25 kb (Supplementary Fig. 2b, c).

Sequences with similarity to known viral genomes comprised 30% of the EV-enriched data and 60% of the VLP-enriched reads (Fig. 1c). Of taxonomically classifiable viral reads, >92% were members of the *Caudoviricetes*, and others showed significant homology to previously described viral genomes from the sampling site[51] (Supplementary Table 1). We propose that the presence of viral DNA within the EV-enriched fraction arose from a combination of sources. First, some reads undoubtedly represent bona fide viral particles, including lipid-enveloped viruses, which could not be physically separated from EVs. However, the lack of distinct peaks in the distribution of DNA lengths from the EV-enriched fraction, which were present in the VLP-enriched fraction (Fig. 1b, Supplementary Fig. 3a), suggests that this was not the primary factor. Secondly, some of the viral DNA could have been truly enclosed within EVs. Such particles could result from viral genomic DNA becoming enveloped in membrane fragments during cell lysis[33], or via phage binding to and delivering their DNA into planktonic EVs[42,52], potentially representing an alternate route for viral transmission[52,53]. Third, a portion of these sequences appears to reflect the random incorporation of prophage-containing cellular genome fragments into vesicles or misclassifications of mobile elements as viruses. While much remains to be determined, our data imply an intricate and complex relationship between viruses and EVs in the oceans.

Extensive diversity of genetic cargo within marine particles

EV and VLP-associated read compositions were similar at a broad taxonomic level, but their specific DNA cargo was quite distinct: only 0.008% of sequences from one fraction dataset matched a read from the other fraction at >80% nucleotide identity over >80% of the read length. The differences in genetic content were further reflected in their encoded proteins. Looking across both fractions, only 10% of all protein family level (20% identity) clusters, and just 3% of unique (90% identity) proteins, were found in both EVs and VLPs (Fig. 2a). The fact that >40% of family-level protein clusters in our data contained only one sequence further highlights the vast diversity within the marine mobilome.

As expected, reads encoding annotatable phage structural proteins were more abundant in the VLP-enriched fraction than the EV-enriched data (Fig. 2b). Together, the particles contained non-singleton protein family clusters encompassing 4552 KEGG orthologs and all COG categories (Supplementary Fig. 3b). This functional genetic diversity was differentially distributed between particle fractions (Fig. 2c). Approximately 16% of annotatable KEGG orthologs were found exclusively in either the EV-enriched or VLP-enriched fraction (Supplementary Data 1), encompassing a wide variety of functional proteins. EV-associated reads contained a higher relative abundance of KEGG BRITE functions, including secretion system genes, kinases, two-component systems, putative toxin-related genes, and ion channels. Conversely, VLP-enriched particles were relatively enriched with ribosomal and translation-related proteins, transporters, amino acid metabolism, lipid biosynthesis, and motility genes. These categories are commonly identified as auxiliary metabolic genes in viral genomes, where their expression during infection can improve viral reproduction efficiency[54], and the genetic components required to produce functional virus propagules may account for some of the observed functional differences between fractions. Among the most abundant EV-exclusive KEGG orthologs were putative archaeal and eukaryotic

https://doi.org/10.1038/s41467-025-57276-w

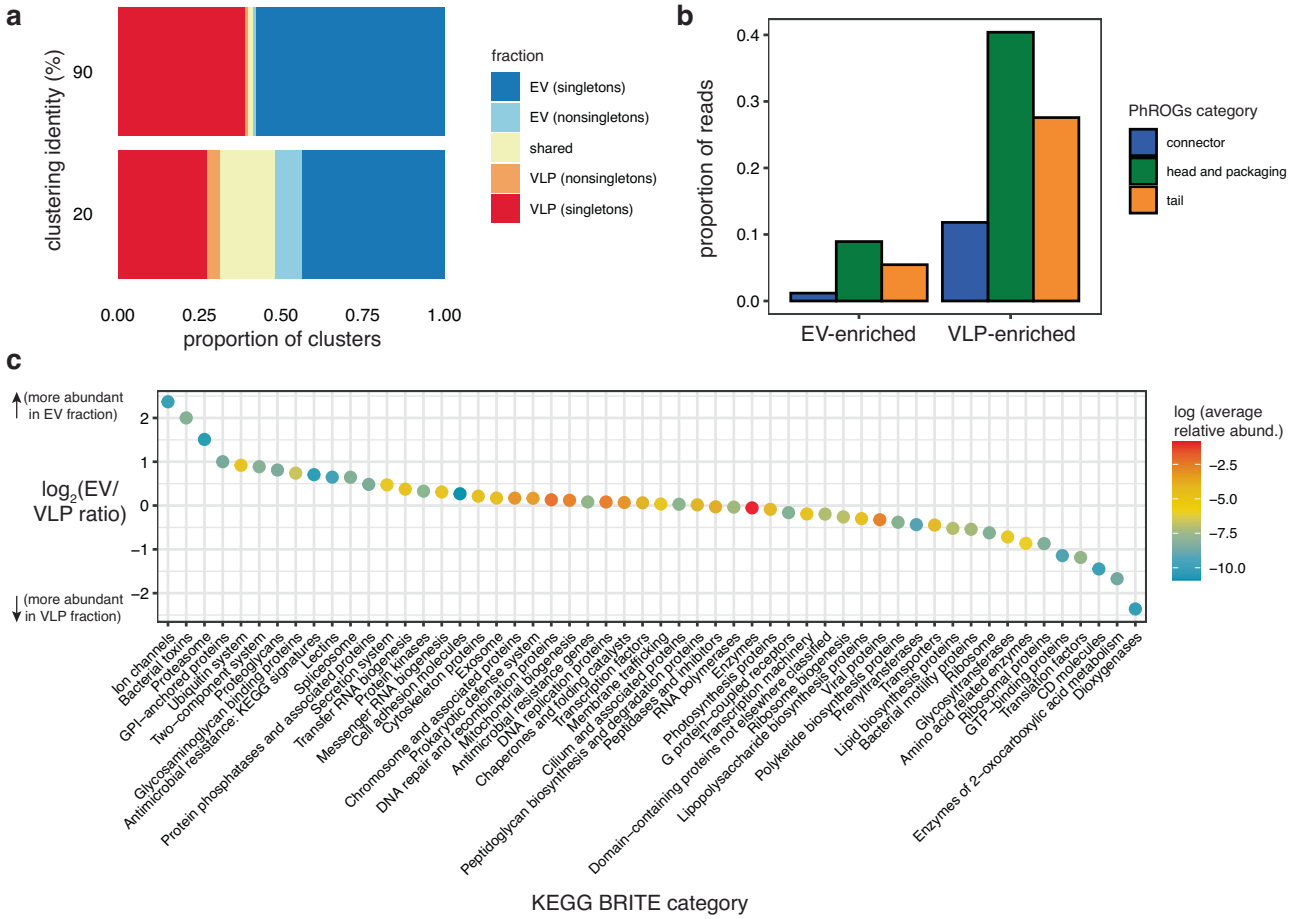

Fig. 2 | Differences in the protein-coding gene content encoded within EVs and VLPs. a Distribution of unique protein-coding genes between the EV-enriched vs VLP-enriched fraction samples. Predicted protein sequences were clustered at either the level of individual proteins (90% identity) or protein families (20% identity). Colors indicate particle fraction, and singleton protein sequence clusters (i.e., clusters with only one member) are noted. **b** Relative abundance of reads containing putative phage structural genes in the particle fractions. Colors indicate category definitions from the PhROGs database. **c** Relative abundance of KEGG-annotatable proteins in both particle fractions, grouped by KEGG BRITE C category. Positive values are more abundant in the EV-enriched fraction; negative values are more abundant in the VLP-enriched fraction. Values indicate the mean ±SD of 10 random resamplings to compare equivalent numbers of proteins from both fraction datasets (all error bars are within the points); colors indicate the average relative abundance of proteins in each category. BRITE categories with fewer than 100 total identified proteins are not shown. Source data are provided as a Source Data file.

cellular proteins, suggesting that EVs may be important vehicles for HGT by these groups in the surface ocean. Together, our findings indicate that functional genetic potential is not equally distributed between EVs and VLPs.

MGEs are prevalent within marine particles

To explore the origin and local genomic context of mobile DNA in detail, we aligned the long reads to a reference dataset of diverse marine microbial genomes. Alignment coverages of particle-associated reads across individual reference genomes were highly skewed in some cases (Fig. 3a; Supplementary Fig. 4a), with some regions recruiting at >10–100-fold excess compared to the surrounding areas. Overall reference coverage skew did not differ significantly between the two fractions, but differences in relative read recruitment patterns on individual references were consistent with differential compositions of the EV and VLP pools (Supplementary Fig. 4b). Manual examination revealed that the highest-recruiting regions represented both potential viruses as well as non-viral MGEs. The skew in reference genome coverage was also not significantly correlated with the number of MGE hallmark genes in each genome, suggesting that only a subset of all potential mobile elements in the community was present within particles at a detectable level (Supplementary Fig. 4c). To better characterize this MGE content we next looked for putative mobile element hallmark genes, including viral hallmarks, in all particle-associated reads. As expected, 81% of sequences from the VLP-enriched fraction contained at least one putative MGE-associated hallmark gene, and at least 74% of those appeared viral. 50% of reads from the EV-enriched fraction also contained MGE signature genes, indicating that MGEs represent a substantial portion of marine EV content as well. Overall, reference genome regions associated with putative mobile elements had significantly higher read coverage than non-MGE-associated regions across all taxa (two-sided Wilcoxon test, $p < 2.2e\text{-}16$; Fig. 3b). These data indicate that marine nanoparticle populations contain two tiers of content: a relatively low-abundance set of particles containing putatively nonspecific genomic DNA fragments and a comparatively higher frequency of particles carrying MGEs.

Nanoparticle-associated MGEs originated from 75 different phyla of marine Archaea and Bacteria (Fig. 3c). The largest taxonomically classifiable MGE sources were abundant marine microbes belonging to the Pseudomonadota (numerically dominated by *Pelagibacter*), Bacteroidota, Cyanobacteriota (primarily *Prochlorococcus*), and Actinomycetota. We also note that 34% of all putative MGE-associated long reads could not be recruited to a reference genome, whether viral or cellular, reinforcing a large amount of remaining diversity left to be understood.

https://doi.org/10.1038/s41467-025-57276-w

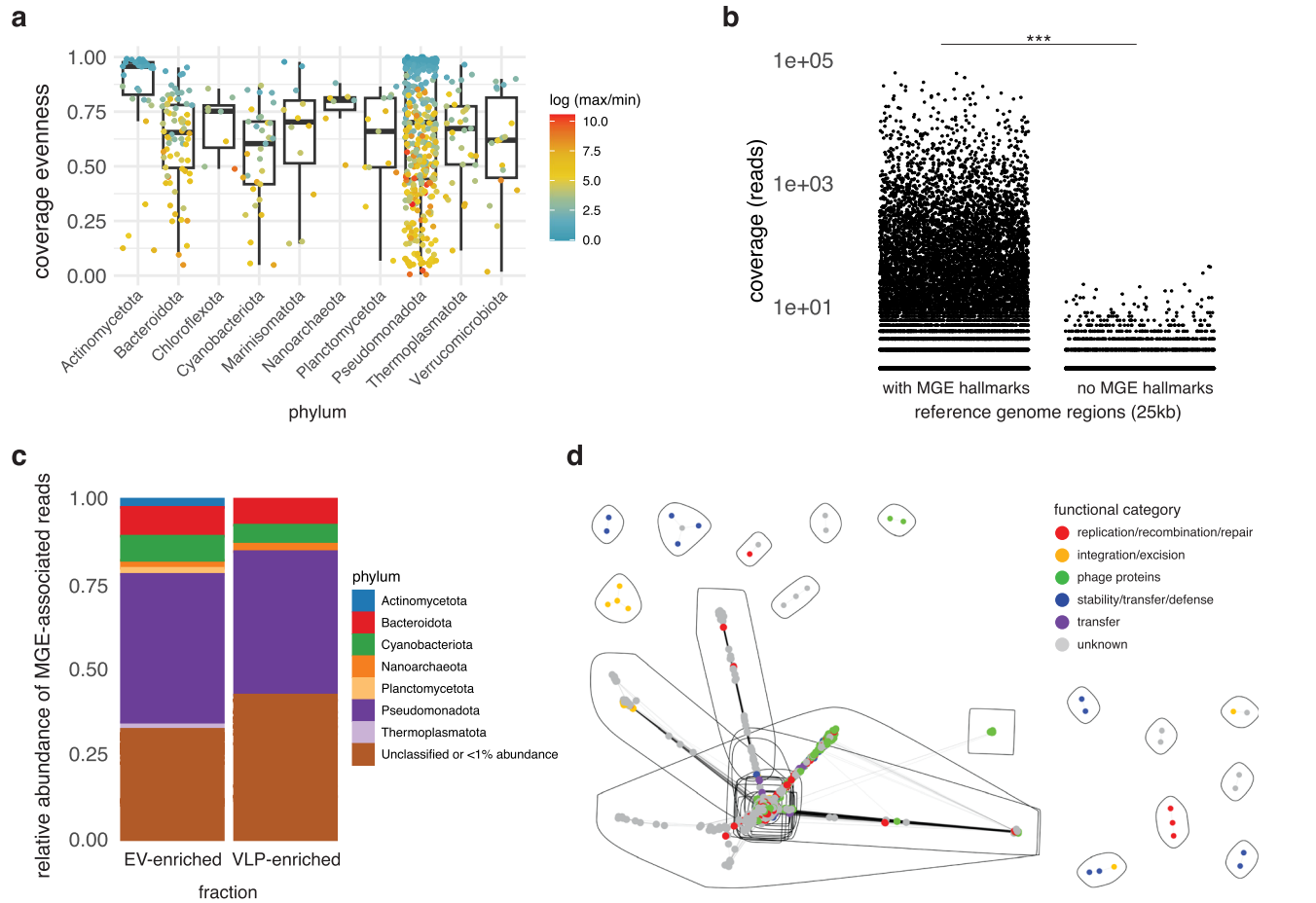

Fig. 3 | High frequencies of mobile elements in marine particles. a Variability in particle-associated DNA read recruitment, as measured by Pielou's evenness index, across individual reference genomes (larger values = more even recruitment). Recruitment evenness was determined using the combined EV-enriched and VLP-enriched fraction data, and the ten most well-represented phyla are shown. Boxplot values: center line, median; box limits, first and third quartiles; whiskers, 1.5× interquartile range. Colors reflect the ratio of maximum to minimum coverage observed for each individual reference. **b** DNA read recruitment levels across individual reference genomes. Points indicate coverage of individual 25 kb genome regions, separated by whether aligning reads contained mobile genetic element (MGE) hallmark genes (***: two-sided Wilcoxon test, $p < 2.2e{-}16$). **c** Taxonomic origin of putative mobile elements carried in particle fractions; colors indicate phylum-level annotation. **d** Gene-sharing network of mobile genetic element hallmark genes found within the same particle-associated DNA fragments. Lines indicate boundaries of inferred mobile element clusters; colors indicate the predicted functional category of MGE hallmark genes. Source data are provided as a Source Data file.

To broadly characterize the MGE diversity in our data we clustered reads from both particle fractions based on MGE hallmark gene co-occurrence, identifying 31 major groups of elements (Fig. 3d, Supplementary Fig. 5a). MGE clusters encompassing the greatest number of reads contained viral-like sequences (including true viruses and viral satellites) from both the EVs and VLP-enriched fractions (Supplementary Fig. 5b). We confirmed the presence of putative viruses and prophages, including examples of abundant viral-like sequences in marine Cyanobacteria, Alphaproteobacteria, and Acidimicrobiia (e.g., Supplementary Fig. 6, Clusters 1–5). Other clusters represented non-viral mobile elements such as tycheposon-like elements[44], concatemeric phage satellites[24], and prophage-like sequences. Beyond these, additional clusters delineated sequences encoding Cas9-like enzymes (including some with associated integrases), small serine recombinases, and restriction-modification systems (Supplementary Fig. 6). As mobile elements comprise a diverse spectrum of genetic entities that propagate using a wide range of mechanisms and which are difficult to delineate[55], the true number of different MGEs within these clusters remains elusive.

We next asked whether particle-enclosed MGEs have the potential to move genes that confer ecologically relevant traits. In total, 85% of all KEGG-annotatable metabolic enzymes were found on reads also containing MGE signature genes, including both viral and non-viral elements. Similarly, 98% of genes classified as potential auxiliary metabolic genes by the PhROGs database were associated with potential mobile elements. For instance, integrase-associated *Prochlorococcus* mobile elements were observed on the same read with genetic cargo such as metal-binding genes, sugar metabolism enzymes, and central carbon enzymes like aldolase (Supplementary Fig. 7a, b). We also noted a few potential cargo-carrying plasmids in the data encoding genes involved in cobalamin biosynthesis, sugar transporters, and a Type IV secretion system (Supplementary Fig. 7c, d). These results are consistent with previous short-read data implicating marine EVs as vehicles for mobile elements[24,43,44] and provide a basis for examining the specific genetic context in which functional genes travel between cells.

Extensive diversity of mobilized DNA from *Pelagibacter*

Pelagibacter is one of the most abundant and ubiquitous bacterial groups in the ocean, with an extensive, globally distributed pangenome associated with ecologically relevant functional differentiation[56]. This diversity is shaped by several evolutionary processes, including homologous recombination of replacement genomic islands, pelagiphage activity, and integrase-mediated mobile elements[44,57,58]. To

https://doi.org/10.1038/s41467-025-57276-w

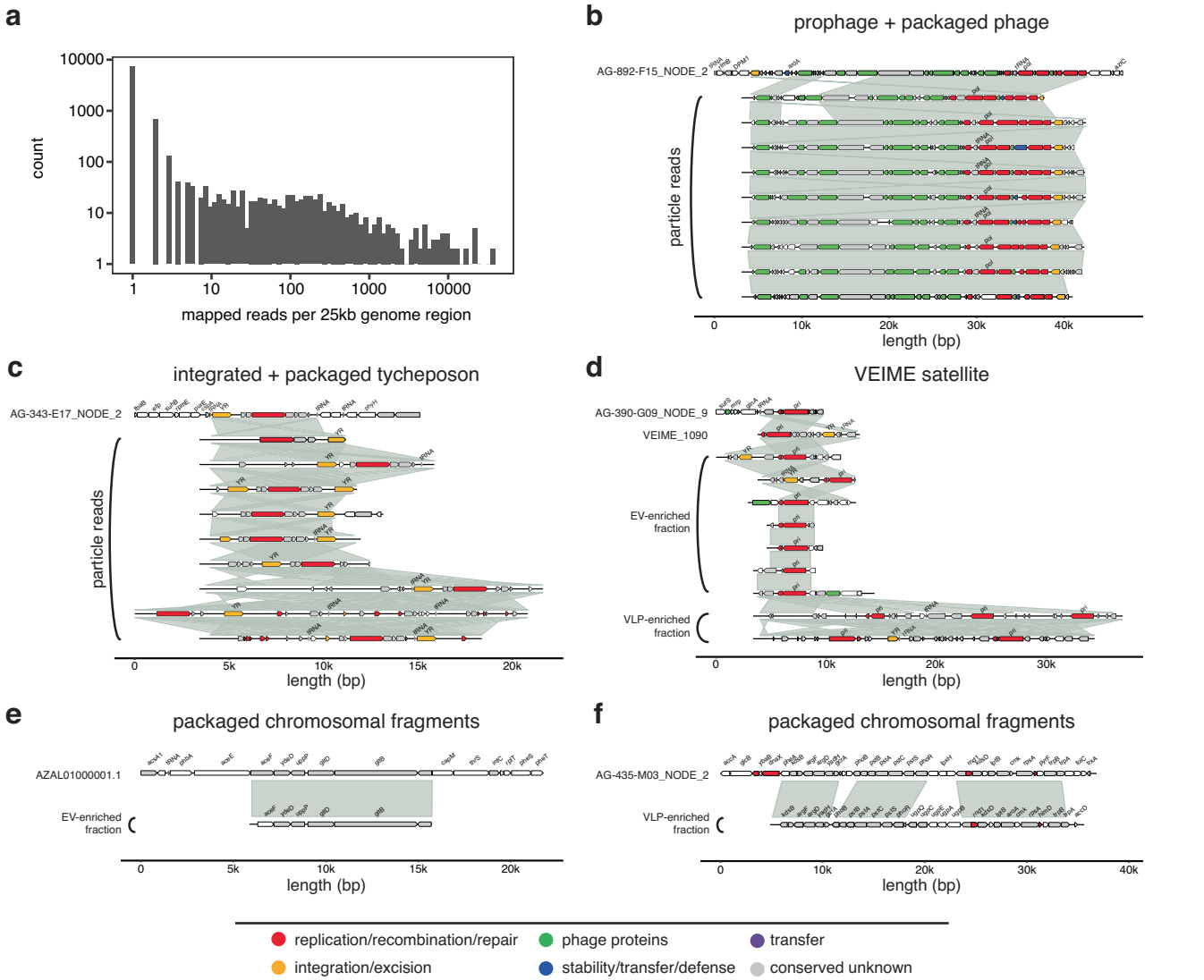

Fig. 4 | *Pelagibacter*-associated genetic information mobilized within marine particles. a Distribution of particle-associated read coverage among 25 kb regions of *Pelagibacter* reference genomes. **b**–**f** Selected examples of *Pelagibacter* mobile regions. The top line of each plot represents a *Pelagibacter* reference assembly; lines below show particle-associated reads with significant similarity to the reference genome. Gray shading indicates regions of adjacent genome plots with high nucleotide-level similarity; gene colors indicate functional categories of mobile element-associated proteins as indicated at bottom. Conserved unknown, unannotated gene found in >1 sequence; YR tyrosine recombinase, pri primase primase/polymerase or primase/helicase, pol polymerase.

better understand what EV and VLP analysis can teach us about the overall landscape of HGT potential in a single taxon, we next characterized the composition of the *Pelagibacter* mobilome.

The Pelagibacteraceae were the most abundant family in our sample (Fig. 1d), representing 5.5% and 7.2% of reads in EV- and VLP-enriched fractions, respectively. Particle-associated reads mapped to 7266 distinct chromosomal regions within 2775 different *Pelagibacter* reference genomes, allowing us to characterize diverse types of mobilized regions. Among the 25 best-represented *Pelagibacter* reference genomes, an average of 17% (±6.5%) of each genome (maximum 34.6%) was found in the mobile pool, suggesting that a notable proportion of total *Pelagibacter* coding capacity is found within marine particles at any given time. Individual *Pelagibacter* DNA molecules were found across lengths up to 64 kb in the EV-enriched fraction and 146 kb in the VLP-enriched fraction (Supplementary Fig. 8a). Specific *Pelagibacter* content was not equally distributed between EVs and VLPs; for instance, reads aligning to 9% of reference genome regions were found exclusively in either the EV-enriched or VLP-enriched particle fraction (Supplementary Fig. 8b), again consistent with preferential incorporation of some regions into specific nanoparticles.

Read recruitment to local *Pelagibacter* genome regions varied across five orders of magnitude (Fig. 4a), and coverage levels generally correlated with the type of element identified. As in the bulk particle analysis above, the highest-recruiting reference regions (here, corresponding to >1000 particle reads) typically contained *Pelagibacter* phage or prophage (Fig. 4b), as well as other viral-like regions ranging in length from ~10 to >60 kb. The high relative abundance of these sequences in our data suggests that such *Pelagibacter* prophages are commonly mobilized in the environment. Reference genome regions with moderate coverage levels (e.g., 11–1000 mapped reads) revealed a variety of other local genomic architectures indicative of both viral-like and non-viral MGEs. Alignments of particle-associated reads to the reference genomes served to define the boundaries of tyrosine recombinase-associated mobile DNA regions such as tycheposons (Fig. 4c), phage satellites (Fig. 4d), and other classes of integrase-associated gene regions (Supplementary Fig. 9a–d). Such regions were frequently located near tRNAs, suggesting potential integration sites,

https://doi.org/10.1038/s41467-025-57276-w

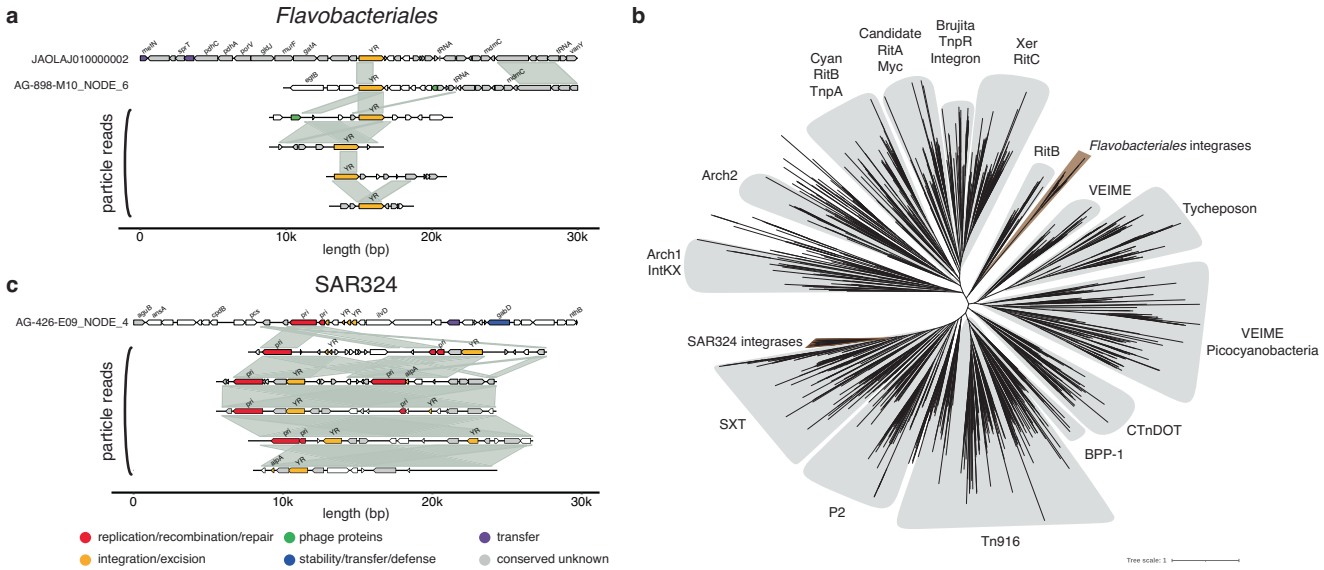

Fig. 5 | Identification of previously uncharacterized mobile elements within marine particles. a Example of a novel tyrosine integrase-associated element within marine *Flavobacteriales*. The top two lines represent reference genomes; lines below show particle-associated reads with significant similarity to these regions. Gray shading indicates regions of adjacent genome plots with high nucleotide-level similarity; gene colors indicate functional categories of mobile element-associated proteins. **b** Maximum likelihood phylogeny of the novel tyrosine recombinase sequences in context with previously defined families[24,44,59]. **c** Example of a novel tyrosine integrase-associated element within SAR324 genomes, plotted as in part a. Conserved unknown, unannotated gene found in >1 sequence; YR tyrosine recombinase, pri primase, primase/polymerase, or primase/helicase.

and typically contained other common MGE-associated genes such as primases and helicases (Fig. 4c, d, Supplementary Fig. S9a–d). In total, tyrosine recombinases were identified within or proximal to 585 of the mobilized *Pelagibacter* genome regions, including families typically associated with phages, ICEs/PICIs, tycheposons, and VEIME phage satellites (Supplementary Fig. 10)[24,44,59]. Together, these data suggest that recombinase-associated mobile elements represent a substantial component of *Pelagibacter* HGT potential in our sample. Finally, *Pelagibacter* reference genome regions with the lowest relative abundance (1–10 reads; Fig. 4a) in the particles were putatively nonspecific fragments of DNA with no clear MGE hallmark genes (e.g., Fig. 4e, f; Supplementary Fig. 2c). Such chromosomal-like sequences were found in both nanoparticle fractions (Fig. 4e–f), indicating the existence of *Pelagibacter* transducing EVs and VLPs in the surface ocean. Though each region in the low-abundance category was individually rare, these sequences represented a long tail of diversity collectively encoding ~93% of the total overall unique *Pelagibacter* genetic content in our dataset.

The distinct cargo capacities of EVs and VLPs may contribute to the observed partitioning of *Pelagibacter* regions between particle types. For instance, an element aligning to a *Pelagibacter* region encoding a potential MobA/MobL family protein and closely related to a VEIME phage satellite was found in distinct forms in EVs vs. VLPs (Fig. 4d). The element was found as a concatemer within the VLP fraction, as is common among marine VEIMEs found in viral capsids and which likely reflects their replication via a plasmid-like rolling circle replication mechanism[24]. The EV-enriched fraction, by contrast, contained only monomeric versions of this element (Fig. 4d). In another example, reads recruiting to a *Pelagibacter* genome were also differentially partitioned by size. Reads matching this contig from the EV-enriched fraction were typically ~4.6 kb, whereas the VLP-enriched fraction contained both these short sequences as well as a set of much longer (40–50 kb) sequences encoding putative viral genomes (Supplementary Fig. 9e). Though we cannot rule out that these shorter reads may derive from a sequencing artifact, the frequency and relatively consistent size of these elements suggest a possible biological origin and role for the shorter version. While the active and/or passive mechanisms mediating the differential packaging seen in either case remain unclear, these examples further highlight the distinct roles and impacts of elements that might move through EVs versus within viral capsids.

Particle-associated DNA enables identification of novel candidate mobile elements

Horizontally transferred genes are frequently concentrated within specific local regions of bacterial chromosomes, but many such regions lack known MGEs[60]. Thus, there exists a need for additional methods to identify novel elements. The enrichment of potential MGE-encoding reads, in their mobilized forms, within particle fractions suggested that relative read recruitment to reference genomes could represent a homology-independent metric for MGE discovery. To examine this possibility, we surveyed highly recruiting genome regions of less well-characterized marine taxa, which revealed a diverse range of potential phage/prophage and other MGEs. We then focused on regions lacking either an annotated integrase/recombinase or with low homology to known tyrosine recombinases.

One such region identified from the EV-enriched fraction encoded a novel mobile element within two marine *Flavobacteriales* isolated from distinct ocean regions (Fig. 5a). The most highly conserved protein in these reads was unannotated, but domain-level analysis suggested that it encoded a site-specific recombinase. Phylogenetic analysis showed this protein forms a novel branch within the tyrosine recombinases (Fig. 5b) most closely related to the RitB family found in bacterial MGEs. In both references the putative integrase is found proximal to a tRNA gene, suggesting a probable integration site. The particle-associated reads contain few other annotatable genes, though some short genes surrounding the integrase were conserved across examples. In addition, some of the longer recruited reads (up to 13 kb) with this enzyme have a potentially concatemeric structure, similar to VEIME satellites[24].

In another example, our approach identified a potential PICI-like element within a SAR324 genome (Fig. 5c). Particle-associated reads (88 in the EV-enriched fraction and 2 in the VLP-enriched data) aligned

to a surface ocean SAR324 genome region containing a putative primase, and encoded a novel branch of the Int_{SXT} family of tyrosine recombinases found within phages, PICIs, and ICEs[59] (Fig. 5b). Multiple reads also contained AlpA-like regulatory proteins associated with PICIs and phage-inducible chromosomal minimalist islands[61]. Though these individual reads exhibited some general overlap in hallmark gene content and relative order of the putative integrase and primase, there was clear variation in gene content including the presence of potential cargo genes. The reference genome region also contained a second partial primase enzyme and multiple short regions containing putative Arm DNA-binding domains found within tyrosine recombinases, perhaps reflecting a complex evolutionary history of imprecise integrations/excisions. Together, these examples highlight the potential for reference-guided recruitment analysis of the mobile DNA pool to provide a homology-independent method for identifying novel replicative mobile elements, serving as a complementary approach to existing MGE discovery methods.

Discussion

Here we show that EVs and VLPs represent genetically distinct vehicles for HGT in planktonic marine ecosystems. While not all EVs will contain DNA[38] and only a subset of VLPs are true transducing particles[22], the vast populations of these oceanic nanoparticles collectively encompass a diverse pool of genetic information available for microbial gene exchange. Since these discrete packets of DNA likely serve as major vehicles for moving novel genes to organisms that are not naturally competent for free DNA uptake, their composition is critical to deciphering the details of HGT networks in aquatic systems. HGT dynamics are predicted to be influenced by the heterogeneity and diversity of EVs and VLP content we report here, and need to be considered alongside differences in their physical and functional properties such as population sizes, host ranges, ability to protect DNA from environmental damage, and overall stability in the environment.

Though VLPs harbor longer fragments of DNA than EVs on average, this does not mean that they necessarily deliver more functional cargo to a recipient cell. Viruses have strict genetic requirements to propagate, and viral genomes are comprised of interdependent genes encoding proteins for structural components, replication, virus assembly, and other essential functions. Viral genomes, thus, typically have less space available per particle for auxiliary metabolic genes or other cargo than EVs, which lack such constraints. Viral transduction of mispackaged chromosomal segments or foreign mobile elements can evade these limits, potentially resulting in a particle with the same content as a vesicle but with different attachment and delivery efficiencies. Viruses are abundant entities in the oceans, but their formation requires an infection or prophage induction event to occur. EVs, by contrast, appear to be produced continually by essentially all microbes during reproduction cycles[42]. While evidence suggests that EVs may be able to deliver DNA across broader taxonomic ranges than VLPs[37], whether this is universally the case remains uncertain. Additionally, it remains unknown how variations in taxonomic, spatial, or temporal contexts might impact ultimate HGT outcomes. Biases in the incorporation of genomic regions into different particle types could represent a mechanism influencing the movement of genetic information between close versus distant relatives. Clearly, there are many tradeoffs and relevant variables to consider when comparing the potential dynamics of EV-mediated vs. VLP-mediated HGT in the oceans, particularly in different ecological contexts and across evolutionary timescales[47].

MGEs make up a highly enriched proportion of the nanoparticle-associated DNA gene pool, and our data suggest a general role for EVs and VLPs as vehicles for transporting such elements in planktonic settings. The role of EVs as carriers of MGEs may either present a selective constraint on the size of MGEs in the oligotrophic oceans, or serve to bias the nanoparticle type that can enable a given MGE to mobilize between cells. Consistent with previous data and models, we speculate that such size constraints or biases could select for the formation of flexible genomic islands by additive means, rather than wholesale replacement by large DNA fragments[44,62]. Our analysis suggests that individually sequencing different mobile DNA pools represents a fruitful approach for identifying active mobile elements in an environmental context, highlighting actively mobilized phage, prophage, and other MGEs in their respective vectors and replicative stages. Though our data are generally consistent with previous hypotheses that the mobility of bacterial chromosomal DNA is relatively limited in comparison to MGE-associated transfer[63], we show that the diversity of putatively non-MGE-containing DNA in marine particles is extensive and likely important over evolutionary timescales. Directed strategies to access and directly interrogate EVs and VLP gene contents and DNA fragment context should prove useful for comprehensively understanding HGT dynamics across the entire process, from gene acquisition to maintenance, replication, mobilization, and selection, in any microbial ecosystem.

Methods

Sample collection and particle purification

Sample collection, filtration, and concentration were described previously[24]. Sequences were derived from a 440 L seawater sample collected from 25 m depth at Station ALOHA (22°45′ N, 158° W) on Hawaii Ocean Time-series cruise 319 (HOT319), 31 January–1 February 2020. Samples were shipped, stored, and manipulated at 4 °C except as noted to help maintain the integrity of the high molecular weight DNA. Briefly, seawater was filtered through a 0.2 uM Supor cartridge filter (Acropak 500, Pall) and then concentrated by tangential flow filtration (TFF) over a 30 kDa filter. Particles from the resulting concentrate were then pelleted by ultracentrifugation at 32,000 rpm (~126,000 × *g*) for 2 hours at 4 °C in a Beckman-Coulter SW32Ti rotor. Particle fractions were then collected by fractionating the sample across an iodixanol (Optiprep; Sigma-Aldrich) gradient. For this, successive 0.5 mL layers of iodixanol (45%, 40%, 35%, 30%, 25%, 20%, 15%, and 10%; all in a 3.5% (w/v) 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 hours at 4 °C in a SW60Ti rotor (Beckman-Coulter). Successive 0.4 mL fractions were collected by pipetting, and their densities were measured.

Fractions between 1.14–1.19 g/mL were pooled to form the EV-enriched sample (containing both extracellular vesicles and potentially some non-tailed viruses), and fractions >1.2 g/mL were combined to form the VLP-enriched sample[42,48]. Particles in the tailed phage fraction were washed twice by diluting with filter-sterilized buffer (3.5% (w/v) NaCl, 3.75 mM TAPS pH 8, 5 mM $CaCl_2$) followed by pelleting by ultracentrifugation (32,000 rpm/~105,000 × *g*, 2 hrs, 4 °C, SW60Ti rotor). Free DNA in Optiprep gradients migrated to upper layers of ~1.1 g/mL that were not included in these samples, but which were in physical proximity to the EV-enriched fraction. While some amount of DNA may be naturally associated with the surface of vesicles[64], to remove any potential contaminating free DNA samples were treated with DNase. To do this, particles from the EV-enriched sample were washed as above, but in 1x PBS buffer, and the sample was incubated with 2U of TURBO DNase (Invitrogen) at 37 °C for 30 minutes. After a second round of TURBO DNase treatment as above, the enzyme was heat-inactivated at 75 °C for 15 minutes. Control DNase reactions with purified plasmid DNA showed that this procedure resulted in a $>10^6$-fold reduction in the concentration of free DNA as determined by qPCR. Particle diameters and concentrations were measured using a Nanosight LM10HS instrument (Malvern/NanoSight), equipped with a LM14 blue laser module and NTA software V3.1. Three technical replicate videos (60 s each) were collected from each sample using a

camera level setting of 11 and analyzed with a threshold setting of 1. In total, this process recovered ~2.3×10^{12} and 2.4×10^{12} particles between 50 and 250 nm diameter from the EV-enriched and VLP-enriched fractions, respectively.

DNA isolation and sequencing

DNA isolation and sequencing for both particle fractions were described previously[24]. Particle lysis and DNA purification were performed in 2 mL screwcap vials using the Qiagen Genomic-tip 20/G protocol and buffers from the Qiagen Genomic DNA Buffer kit following manufacturer's recommendations. The column purified DNA was concentrated by isopropanol precipitation, and resuspended in 1X TE buffer (10 mM Tris•HCl, pH 8.0; I mM EDTA pH 8.0) 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 (Invitrogen). The VLP-enriched sample yielded a total of 5.1 µg of high molecular weight DNA. The total yield for the corresponding EV-enriched fraction was 8.2 µg of high molecular weight DNA.

VLP-enriched and EV-enriched DNA pools were processed using the Nanopore Ligation Sequencing Kit (LSK-109, Oxford Nanopore Technologies, Ltd.) following manufacturer's instructions for the processing of high molecular weight DNA. A total of 2 VLP-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) flowcells (Oxford Nanopore Technologies, Ltd.). Read basecalls were generated from the signal traces using Guppy v3.0. The sequencing yield for the VLP-enriched fraction totaled 31 Gbp, generating reads with an N50 of 37.67 kbp. The sequencing yield for the corresponding membrane EV-enriched fraction totaled 46 Gbp, with an N50 of 4.6 kbp. To aid in error correction, we also generated short-read sequences from both fractions. To do this, 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 (Covaris #520166). Libraries were sequenced using a 150 bp paired-end NextSeq High Output V2 reagent kit (Illumina, FC-404-2004).

Short-read sequence assembly and analysis

Metagenomic assembly of the cellular fraction data was conducted with metaSPAdes V3.15.5[65] on paired-end Illumina sequencing data. Cellular metagenome coverage was determined by mapping both short-read and long-read sequencing data from the vesicle and VLP-enriched data using Bowtie2 V2.5.1[66] and minimap2 V2.26 (-x map-ont)[67], respectively. Contig coverage from combined sorted and merged.bam files for each fraction were determined using coverM V0.6.1 (https://github.com/wwood/CoverM) from the single-stranded nanopore reads. Overall coverage breadth was calculated as the number of reference bases to which ≥1 sample base was mapped divided by the total number of bases in the assembled reference metagenome, and per-base coverage of the EV- and VLP-enriched fractions were plotted using ggplot2.

Sampling saturation was examined by rarefaction. First, reads were randomly selected at varying depths in triplicate. The metagenomic contigs assembled from each particle-enriched fraction were divided into 25 kb regions, and for each randomly sampled subset of reads, the number of unique regions to which reads in that subset were mapped was tallied. The mean number of unique chunks represented and an associated 95% confidence interval were calculated for each read depth value and particle fraction.

Nanopore sequence analysis

Frameshifts introduced from Nanopore sequencing error were corrected to the extent possible using the proovframe algorithm V0.9.7[68] and Illumina short-read sequences from the identical DNA samples. The algorithm was only applied to reads from both particle fractions ≥1 kb, and reads shorter than this were not included in subsequent analyses. Alignment-free taxonomic classification of reads was carried out using the GORG-Classifier[69]. Detailed reference genome recruitment was carried out using minimap2 V2.26[67] (options -x map-ont) against a dataset of microbial genomes from the GORG-Tropics dataset[69], MARMICRODB[70], GTDB release 08-RS214[71], and other sources[72,73]. To reduce the impact of spurious short alignments on the data, the top alignment for each nanopore read was only kept if it had a match length of ≥ 1000 bp and aligned over ≥25% of the length of the read. Reference genome coverage evenness was calculated as Pielou's index[74] using functions in the vegan R package (V 2.6–4)[75]. Evenness statistics were based on number of particle read alignments per 25 kb genome region, calculated for all genomes with >10 total alignments across >3 different 25 kb regions.

To identify sequences as viral while minimizing the bycatch of other types of MGEs that are frequently also identified as viral, we used a combination of approaches. We counted reads as putatively viral if they met any one of the following criteria: (1) Were classified as either a provirus, medium-quality, or high-quality viruses by CheckV 1.0.1[76]; (2) Were classified as viral by geNomad[77] with a score ≥ 0.9 and >1 hallmark gene; (3) Had a significant minimap2 alignment (alignment match length ≥1000 bp over ≥25% of the read) to a viral sequence in either the INPHARED database of viral genomes (version 1 May 2023)[78] or a set of marine assembly-free viral genomes[51]. This approach is intended to be conservative but is certainly not foolproof, complicated by factors including genetic overlap between viral hallmark genes and those associated with other types of mobile elements; errors in alignment and classification arising from sequencing error rates and minimap2 alignment filters; and artifacts which could be due to incomplete sequencing of DNA fragments. Potential plasmid sequences and viral taxonomy of individual reads were based on geNomad classification. Statistical analysis and bulk processing of outputs from these algorithms was done in R V4.3.0 using Tidyverse tools[79] and plotted with ggplot2[80].

Complete protein sequences were inferred for all proovframe-corrected reads in both fractions using prodigal V2.6.3[81] (options -p meta -c). Protein clusters were generated using MMseqs2 (version 139e4502a9d7056b33698b685c8812d453180627)[82]. Family-level protein clusters were generated at 20% similarity (options –min-seq-id 0.2 -c 0.5 –cov-mode 2 –cluster-mode 0) and at the individual protein level at a 90% identity threshold (options –min-seq-id 0.9 -c 0.9 –cov-mode 1 –cluster-mode 2)[83]. KEGG annotations were obtained by annotating cluster representative sequences using both eggNOGmapper V2.1.4-2[84] (against the eggnog V5.0 dataset[85]) and KofamScan against the February 2023 release of the KOfam database[86]. KEGG metabolic enzymes were defined as those annotated to the KEGG BRITE category "B: Protein families: metabolism". Gene annotations against the PhROGs dataset V4[87] were carried out using the mmseqs2-based pipeline as indicated by the authors. Additional annotations for individual genes were carried out using a viral protein function prediction protein language model[88]. Protein domains within unannotated proteins were identified using InterPro[89].

Mobile element analysis

Identification of potentially MGE-associated proteins was carried out by first building sequence profiles of MGE hallmark genes compiled from mobileOG-db V1.6[90], Viral RefSeq, and genes from PICIs and *Prochlorococcus* tycheposons using MMseqs2. MGE hallmark genes in reference genomes were identified by comparing all prodigal-identified coding regions against this profile database using Mmseqs2 (options search -s 7). Hallmark genes in the long reads were identified similarly, and all reads containing 2 or more sequences with an annotatable hallmark gene were retained. Due to the size of the dataset, de novo clustering of reads based on hallmark gene co-

occurrence was carried out using the cluster_fast_greedy algorithm as implemented by igraph for R[91].

Tyrosine integrase phylogenies were constructed from a set of tyrosine recombinases extracted from the UniRef50 database (http://www.uniprot.org/uniref) using HMM models[59]; a set of integrases associated with *Prochlorococcus* tycheposons and cryptic elements[44]; and representative sequences of VEIME-associated integrases[24] based on 40% identity clusters as generated by MMSeqs2). Sequences were aligned with Mafft V7.520 (options –maxiterate 1000 –genafpair)[92], a maximum likelihood phylogeny was generated using FastTree v2.1.11 using default settings[93], and the tree was plotted using iTOL[94].

Plots of nanopore reads against each other and/or reference genome regions were made in R using gggenomes V0.9.7.9000[95]. For each plot, protein clusters were determined using the mmseqs2 easy-cluster pipeline with default options. tRNA annotations were determined using tRNAscan-SE V2.0.9[96]. Annotations were inferred by eggNOGmapper (options –sensmode very-sensitive). All-vs-all nucleotide-level alignments were determined by minimap2 (options -X -x map-ont).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All original sequence data are available from the NCBI Sequence Read Archive under Bioproject PRJNA855972. Proovframe-corrected nanopore reads and cellular fraction assembled metagenomes can be obtained from Zenodo record 11551382. Source data are provided with this paper.

References

1. Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304 (2000).
2. Welch, R. A. et al. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci.* **99**, 17020–17024 (2002).
3. Smillie, C. S. et al. Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480**, 241–244 (2011).
4. Glasner, M. E., Truong, D. P. & Morse, B. C. How enzyme promiscuity and horizontal gene transfer contribute to metabolic innovation. *FEBS J.* **287**, 1323–1342 (2020).
5. Karberg, K. A., Olsen, G. J. & Davis, J. J. Similarity of genes horizontally acquired by *Escherichia coli* and *Salmonella enterica* is evidence of a supraspecies pangenome. *Proc. Natl Acad. Sci. USA* **108**, 20154–20159 (2011).
6. Nagies, F. S. P., Brueckner, J., Tria, F. D. K. & Martin, W. F. A spectrum of verticality across genes. *PLoS Genet.* **16**, e1009200 (2020).
7. Lerat, E., Daubin, V., Ochman, H. & Moran, N. A. Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol.* **3**, e130 (2005).
8. Koonin, E. V. Horizontal gene transfer: essentiality and evolvability in prokaryotes, and roles in evolutionary transitions. *F1000Res* **5**, F1000 (2016).
9. Hehemann, J.-H. et al. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**, 908–912 (2010).
10. Gillings, M. R. Lateral gene transfer, bacterial genome evolution, and the Anthropocene. *Ann. N. Y. Acad. Sci.* **1389**, 20–36 (2017).
11. Coleman, M. L. et al. Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**, 1768–1770 (2006).
12. Frigaard, N.-U., Martinez, A., Mincer, T. J. & DeLong, E. F. Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature* **439**, 847–850 (2006).
13. Tettelin, H., Riley, D., Cattuto, C. & Medini, D. Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* **11**, 472–477 (2008).
14. Abe, K., Nomura, N. & Suzuki, S. Biofilms: hot spots of horizontal gene transfer (HGT) in aquatic environments, with a focus on a new HGT mechanism. *FEMS Microbiol. Ecol.* fiaa031 (2020).
15. Dubnau, D. & Blokesch, M. Mechanisms of DNA uptake by naturally competent bacteria. *Annu. Rev. Genet* **53**, 217–237 (2019).
16. Linney, M. D., Schvarcz, C. R., Steward, G. F., DeLong, E. F. & Karl, D. M. A method for characterizing dissolved DNA and its application to the North Pacific Subtropical Gyre. *Limnol. Oceanogr. Methods* **19**, 210–221 (2021).
17. Ganesh, S., Parris, D. J., Delong, E. F. & Stewart, F. J. Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. *ISME J.* **8**, 187–211 (2014).
18. Fontanez, K. M., Eppley, J. M., Samo, T. J., Karl, D. M. & DeLong, E. F. Microbial community structure and function on sinking particles in the North Pacific Subtropical Gyre. *Front. Microbiol.* **6**, 469 (2015).
19. Pelve, E. A., Fontanez, K. M. & Delong, E. F. Bacterial succession on sinking particles in the ocean's interior. *Front. Microbiol.* **8**, 605–615 (2017).
20. Suttle, C. A. Marine viruses — major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
21. Clokie, M. R. J., Millard, A. D., Wilson, W. H. & Mann, N. H. Encapsidation of host DNA by bacteriophages infecting marine *Synechococcus* strains. *FEMS Microbiol. Ecol.* **46**, 349–352 (2003).
22. Laurenceau, R., Raho, N., Forget, M., Arellano, A. A. & Chisholm, S. W. Frequency of mispackaging of *Prochlorococcus* DNA by cyanophage. *ISME J.* **15**, 129–140 (2021).
23. McDaniel, L. D. et al. High frequency of horizontal gene transfer in the oceans. *Science* **330**, 50–50 (2010).
24. Eppley, J. M., Biller, S. J., Luo, E., Burger, A. & DeLong, E. F. Marine viral particles reveal an expansive repertoire of phage-parasitizing mobile elements. *Proc. Natl. Acad. Sci.* **119**, e2212722119 (2022).
25. Kauffman, K. M. et al. Resolving the structure of phage–bacteria interactions in the context of natural diversity. *Nat. Commun.* **13**, 372 (2022).
26. Hwang, Y., Roux, S., Coclet, C., Krause, S. J. E. & Girguis, P. R. Viruses interact with hosts that span distantly related microbial domains in dense hydrothermal mats. *Nat. Microbiol.* **8**, 946–957 (2023).
27. Waterbury, J. B. & Valois, F. W. Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Appl. Environ. Microbiol.* **59**, 3393–3399 (1993).
28. Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. Cyanophages infecting the oceanic cyanobacterium. *Nature* **424**, 1047–1051 (2003).
29. Holmfeldt, K., Middelboe, M., Nybroe, O. & Riemann, L. Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their flavobacterium hosts. *Appl. Environ. Microbiol.* **73**, 6730–6739 (2007).
30. Deatherage, B. L. & Cookson, B. T. Membrane vesicle release in bacteria, eukaryotes, and archaea: a conserved yet underappreciated aspect of microbial life. *Infect. Immun.* **80**, 1948–1957 (2012).
31. Toyofuku, M., Nomura, N. & Eberl, L. Types and origins of bacterial membrane vesicles. *Nat. Rev. Microbiol.* **17**, 13–24 (2019).
32. Mandal, P. K., Ballerin, G., Nolan, L. M., Petty, N. K. & Whitchurch, C. B. Bacteriophage infection of *Escherichia coli* leads to the formation of membrane vesicles via both explosive cell lysis and membrane blebbing. *Microbiology* **167**, 001021 (2021).
33. Turnbull, L. et al. Explosive cell lysis as a mechanism for the biogenesis of bacterial membrane vesicles and biofilms. *Nat. Commun.* **7**, 11220–13 (2016).

34. Toyofuku, M., Schild, S., Kaparakis-Liaskos, M. & Eberl, L. Composition and functions of bacterial membrane vesicles. *Nat. Rev. Microbiol.* **21**, 415–430 (2023).
35. Sjöström, A. E., Sandblad, L., Uhlin, B. E. & Wai, S. N. Membrane vesicle-mediated release of bacterial RNA. *Sci. Rep.* **5**, 15329 (2015).
36. Yaron, S., Kolling, G. L., Simon, L. & Matthews, K. R. Vesicle-mediated transfer of virulence genes from *Escherichia coli* O157:H7 to other enteric bacteria. *Appl. Environ. Microbiol.* **66**, 4414–4420 (2000).
37. Bitto, N. J. et al. Bacterial membrane vesicles transport their DNA cargo into host cells. *Sci. Rep.* **7**, 7072 (2017).
38. Biller, S. J. et al. Membrane vesicles in sea water: heterogeneous DNA content and implications for viral abundance estimates. *ISME J.* **11**, 394–404 (2017).
39. Pérez-Cruz, C. & Delgado, L. López-Iglesias, C. & Mercade, E. Outer-inner membrane vesicles naturally secreted by gram-negative pathogenic bacteria. *PLoS One* **10**, e0116896 (2015).
40. Erdmann, S., Tschitschko, B., Zhong, L., Raftery, M. J. & Cavicchioli, R. A plasmid from an Antarctic haloarchaeon uses specialized membrane vesicles to disseminate and infect plasmid-free cells. *Nat. Microbiol.* **2**, 1446–1455 (2017).
41. Aktar, S. et al. Incorporation of plasmid DNA into bacterial membrane vesicles by peptidoglycan defects in *Escherichia coli*. *Front. Microbiol.* **12**, 747606 (2021).
42. Biller, S. J. et al. Bacterial vesicles in marine ecosystems. *Science* **343**, 183–186 (2014).
43. Lücking, D., Mercier, C., Alarcón-Schumacher, T. & Erdmann, S. Extracellular vesicles are the main contributor to the non-viral protected extracellular sequence space. *ISME Commun.* **3**, 112 (2023).
44. Hackl, T. et al. Novel integrative elements and genomic plasticity in ocean ecosystems. *Cell* **186**, 47–62 (2023).
45. Tashiro, Y. et al. Interaction of bacterial membrane vesicles with specific species and their potential for delivery to target cells. *Front. Microbiol.* **8**, 873–13 (2017).
46. Biller, S. J. et al. *Prochlorococcus* extracellular vesicles: molecular composition and adsorption to diverse microbes. *Environ. Microbiol.* **24**, 420–435 (2022).
47. Nazarian, P., Tran, F. & Boedicker, J. Q. Modeling multispecies gene flow dynamics reveals the unique roles of different horizontal gene transfer mechanisms. *Front. Microbiol.* **9**, 2978 (2018).
48. Kauffman, K. M. et al. A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. *Nature* **554**, 118–122 (2018).
49. Linney, M. D. et al. Microbial sources of exocellular DNA in the ocean. *Appl. Environ. Microbiol.* **88**, e02093–21 (2022).
50. Hagemann, S. et al. DNA-bearing membrane vesicles produced by *Ahrensia kielensis* and *Pseudoalteromonas marina*. *J. Basic Microbiol.* **54**, 1062–1072 (2013).
51. Beaulaurier, J. et al. Assembly-free single-molecule sequencing recovers complete virus genomes from natural microbial communities. *Genome Res.* **30**, 437–446 (2020).
52. Manning, A. J. & Kuehn, M. J. Contribution of bacterial outer membrane vesicles to innate bacterial defense. *BMC Microbiol.* **11**, 258 (2011).
53. Moulin, C., Crupi, M. J. F., Ilkow, C. S., Bell, J. C. & Boulton, S. Extracellular vesicles and viruses: two intertwined entities. *IJMS* **24**, 1036 (2023).
54. Hurwitz, B. L. & U'Ren, J. M. Viral metabolic reprogramming in marine ecosystems. *Curr. Opin. Microbiol.* **31**, 161–168 (2016).
55. Koonin, E. V., Dolja, V. V., Krupovic, M. & Kuhn, J. H. Viruses defined by the position of the virosphere within the replicator space. *Microbiol Mol. Biol. Rev.* **85**, e00193–20 (2021).
56. López-Pérez, M., Haro-Moreno, J. M., Coutinho, F. H., Martinez-Garcia, M. & Rodriguez-Valera, F. The evolutionary success of the marine bacterium SAR11 analyzed through a metagenomic perspective. *mSystems* **5**, e00605–e00620 (2020).
57. López-Pérez, M., Martin-Cuadrado, A.-B. & Rodriguez-Valera, F. Homologous recombination is involved in the diversity of replacement flexible genomic islands in aquatic prokaryotes. *Front. Genet.* **5**, 147 (2014).
58. Zhao, Y. et al. Pelagiphages in the *Podoviridae* family integrate into host genomes. *Environ. Microbiol.* **21**, 1989–2001 (2019).
59. Smyshlyaev, G., Bateman, A. & Barabas, O. Sequence analysis of tyrosine recombinases allows annotation of mobile genetic elements in prokaryotic genomes. *Mol. Syst. Biol.* **17**, e9880 (2021).
60. Oliveira, P. H., Touchon, M., Cury, J. & Rocha, E. P. C. The chromosomal organization of horizontal gene transfer in bacteria. *Nat. Commun.* **8**, 841 (2017).
61. Barcia-Cruz, R. et al. Phage-inducible chromosomal minimalist islands (PICMIs), a novel family of small marine satellites of virulent phages. *Nat. Commun.* **15**, 664 (2024).
62. Rodriguez-Valera, F., Martin-Cuadrado, A.-B. & López-Pérez, M. Flexible genomic islands as drivers of genome evolution. *Curr. Opin. Microbiol.* **31**, 154–160 (2016).
63. Humphrey, S. et al. Bacterial chromosomal mobility via lateral transduction exceeds that of classical mobile genetic elements. *Nat. Commun.* **12**, 6509 (2021).
64. Johnston, E. L. et al. Planktonic and biofilm-derived *Pseudomonas aeruginosa* outer membrane vesicles facilitate horizontal gene transfer of plasmid DNA. *Microbiol. Spectr.* e05179-22 (2023).
65. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P.A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
66. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
67. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
68. Hackl, T. et al. Proovframe: frameshift-correction for long-read (meta)genomics. *bioRxiv*, https://doi.org/10.1101/2021.08.23.457338 (2021).
69. Pachiadaki, M. G. et al. Charting the complexity of the marine microbiome through single-cell genomics. *Cell* **179**, 1623–1635 (2019).
70. Becker, J. W., Hogle, S. L., Rosendo, K. & Chisholm, S. W. Co-culture and biogeography of *Prochlorococcus* and SAR11. *ISME J.* **13**, 1506–1519 (2019).
71. Parks, D. H. et al. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat. Biotechnol.* **38**, 1079–1086 (2020).
72. Tully, B. J., Sachdeva, R., Graham, E. D. & Heidelberg, J. F. 290 metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology. *PeerJ* **5**, e3558 (2017).
73. Munson-McGee, J. H. et al. Decoupling of respiration rates and abundance in marine prokaryoplankton. *Nature* **612**, 764–770 (2022).
74. Pielou, E. C. The measurement of diversity in different types of biological collections. *J. Theor. Biol.* **13**, 131–144 (1966).
75. Oksanen, J. et al. Vegan: Community Ecology Package. https://github.com/vegandevs/vegan (2022).
76. Nayfach, S. et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2021).
77. Camargo, A. P. et al. Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* **42**, 1303–1312 (2023).
78. Cook, R. et al. INfrastructure for a PHAge REference database: identification of large-scale biases in the current collection of cultured phage genomes. *Phage* **2**, 214–223 (2021).
79. Wickham, H. et al. Welcome to the Tidyverse. *JOSS* **4**, 1686 (2019).

80. Wickham, H. *Ggplot2: elegant graphics for data analysis*. (Springer-Verlag New York, 2016).
81. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
82. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
83. Coelho, L. P. et al. Towards the biogeography of prokaryotic genes. *Nature* **601**, 252–256 (2022).
84. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
85. Huerta-Cepas, J. et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
86. Aramaki, T. et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
87. Terzian, P. et al. PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genom. Bioinform.* **3**, lqab067 (2021).
88. Flamholz, Z. N., Biller, S. J. & Kelly, L. Large language models improve annotation of prokaryotic viral proteins. *Nat. Microbiol.* **9**, 537–549 (2024).
89. Paysan-Lafosse, T. et al. InterPro in 2022. *Nucleic Acids Res.* **51**, D418–D427 (2023).
90. Brown, C. L. et al. mobileOG-db: a manually curated database of protein families mediating the life cycle of bacterial mobile genetic elements. *Appl. Environ. Microbiol.* **88**, e00991–22 (2022).
91. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal Complex Syst.* **1695**, 1–9 (2006).
92. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
93. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).
94. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
95. Hackl, T., Ankenbrand, M., Adrichem, B. van, Wilkins, D. & Haslinger, K. gggenomes: effective and versatile visualizations for comparative genomics. *arXiv*, https://doi.org/10.48550/arXiv.2411.13556 (2024).
96. Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096 (2021).

Acknowledgements

We thank Libusha Kelly and members of the Biller lab for providing valuable suggestions on the manuscript. We also thank Elaine Luo, the captain and crew of R/V *Kilo Moana*, and the HOT program for assistance in sample collection and expedition support. This work was supported by the NSF (OCE-2049004 and OCE-2304066 to S.J.B.), the Simons Foundation (Award 917971 to S.J.B.; Awards 329108 and 721223 to E.F.D.), and the Gordon and Betty Moore Foundation (Grant 3777 to E.F.D.).

Author contributions

S.J.B., T.H., and E.F.D. conceived and designed the study. S.J.B. and A.B. carried out the experiments. S.J.B., M.G.R., J.L., J.M.E., T.H. and E.F.D. analyzed the data. S.J.B. wrote the paper with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-57276-w.

Correspondence and requests for materials should be addressed to Steven J. Biller.

Peer review information *Nature Communications* thanks Yosuke Nishimura and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.