

Community-level diversity of viral counter-defensomes

Angelina Beavogui^{1†}, Lucas da Silva e Silva^{1†},
Izabel de G. M. Gomes¹, Luke Doladille¹, Nicolas Wiart²,
Patrick Wincker¹, Pedro H. Oliveira^{1*}

¹Génomique Métabolique, Genoscope, Institut François Jacob, Commissariat à l’Energie Atomique (CEA), CNRS, Université Evry, Université Paris-Saclay, 2 Rue Gaston Crémieux, Evry, 91057, France.

²Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine (CNRGH), Evry, 91057, France.

*Corresponding author(s). E-mail(s): pcoutool@genoscope.cns.fr;

[†]These authors contributed equally to this work.

Abstract

Bacteria and phages have co-evolved tit-for-tat strategies to combat one another for survival. To overcome phage infection, bacteria have developed various defense systems (the defensome), and in response, phages have evolved a counter-defensome to inhibit or evade such defensive strategies. While significant progress has been achieved on mechanistic, structural and functional aspects of the defensome, the study of the phage counter-defensome is relatively new, with very limited information available on the diversity and distribution of these systems across complex environmental phageomes. Here we present a large-scale analysis of the counter-defensome of 50,595 DNA phage population genomes reconstructed from soil, marine, and human gut environments. We observed substantial variation in the frequency and composition of the counter-defensome across phage families, which correlated with host range, habitat, and geographic context. A substantial fraction of the counter-defensome was clustered in islands, with gene expression restricted to a limited subset of families. Moreover, both phage counter-defense and bacterial-type defense families were detected across genomes of nucleocytoplasmic large DNA viruses, suggesting cross-clade horizontal gene transfer. Our results reveal the diverse counter-defense strategies across environmentally distinct viral communities and provide a foundation for uncovering novel mechanisms that shape conflicts and alliances in host–virus immunity networks.

Keywords: phage–bacteria arms race; horizontal gene transfer; counter-defense islands; immune transcriptional landscape; nucleocytoplasmic large DNA viruses

Introduction

Viruses are the most abundant, pervasive, and genetically diverse biological entities on Earth. Among them, bacteriophages (phages) are profoundly influential in shaping microbial ecosystem dynamics, biogeochemical nutrient cycling, and gene flow through horizontal gene transfer (HGT). With an estimated global population size of 10^{31} particles¹, phage abundance relative to their hosts, often quantified as a virus-to-prokaryote ratio (VPR), spans more than six orders of magnitude across environments as disparate as the human gut, the global ocean, and soils^{2,3}. Most known prokaryotic viruses are tailed double-stranded (ds) DNA viruses pertaining to the class *Caudoviricetes* (realm *Duplodnaviria*), with additional modest contributions from *Varidnaviria* and archaeal viruses of *Adnaviria*^{4,5}. They typically possess genomes ranging from tens to hundred of kilobases, with characteristically mosaic architectures, diverse replication strategies, and variable host ranges. Beyond prokaryotes, certain dsDNA viruses can also infect unicellular eukaryotes. An interesting example is provided by virophages (class *Maveriviricetes*), which parasitize giant viruses of the order *Imitervirales* (phylum *Nucleocytoviricota*) also known as nucleocytoplasmic large DNA viruses (NCLDV)s⁶. Virophages depend on the viral factory provided by the co-infecting NCLDV for their own replication, eventually impairing the latter's fitness and benefiting the host cell⁷. Comparatively less studied than phages, virophages and NCLDV)s appear to comprise only a small fraction of the virosphere (often less than 1% of the viral reads in environmental metagenomes)^{8–10}, although NCLDV)s can reach 10^3 – 10^5 genomes per milliliter in aquatic ecosystems¹¹. At evolutionary time scales, dsDNA viruses have the capacity to mediate gene transfer across deep evolutionary lineages, as previously reported for NCLDV)s and eukaryotes¹², and potentially between cellular domains through indirect or mobile genetic element (MGE)-mediated mechanisms^{6,12,13}.

This pervasive viral pressure has driven bacteria to evolve a sophisticated repertoire of defense systems, collectively termed the defensome^{14,15}. Classical examples include restriction–modification (R–M) systems¹⁶ and CRISPR/Cas adaptive immunity¹⁷, which detect and degrade invading nucleic acids. Beyond these well-characterized mechanisms, the past decade has revealed a vast arsenal of previously unknown anti-MGE defense systems, many of which have begun to be mechanistically characterized^{18,19}. Phages, in turn, have responded by evolving a growing array of counter-defense strategies capable of targeting bacterial defensomes at multiple levels, with more than 150 systems and 27 families described to date²⁰. These include direct inhibition of defense proteins (*e.g.*, anti-CRISPR (Acr)^{21,22}, anti-RecBCD²³, anti-restriction enzymes^{24,25}), interference with toxin–antitoxin (anti-T–A) systems^{26,27}, sequestration of immune signaling molecules (*e.g.*, anti-CBASS^{28,29}, anti-Thoeris³⁰), and chemical modification of host defense components^{31,32}. Apart from counter-defenses, phage genomes, particularly in the lysogenic state, display pronounced genomic mosaicism and often carry anti-MGE genes implicated in inter-viral competition, capable of increasing host resistance against infection from competing phages^{33,34}. Such layer of accessory ‘moron’ genes comprising defense and counter-defense is often acquired and lost via homologous recombination and lateral transduction³⁵. Arms races are also present among eukaryotic viruses. For example, giant viruses have evolved defense systems to counter virophage parasitism, mirroring bacterial anti-MGE strategies. The

best-characterized example is MIMIVIRE, a CRISPR-like adaptive immune system identified in mimiviruses, which incorporates virophage-derived sequences to mediate sequence-specific resistance³⁶. Emerging evidence suggests that giant viruses encode a broader repertoire of virophage defense mechanisms, including DNA modification and restriction nucleases³⁷, underscoring evolutionary parallels between prokaryotic and viral immunity.

Recent advances in high-throughput sequencing and computational metagenomics have begun to illuminate largely hidden dimensions of viral biology. Enhanced metavirome extraction, combined with culture-independent, genome-resolved approaches, now enables systematic exploration of the vast viral ‘dark matter’ across diverse ecosystems^{38–42}, providing opportunities to study its abundance and diversity across environments. These advances reveal pronounced temporal and spatial variation in VPRs, suggesting that fluctuating viral predation exerts a major influence on the composition and evolution of immune repertoires across biomes. Here, we systematically characterized the counter-defensome of dsDNA phage communities across soil, marine, and human gut ecosystems. We quantified its abundance, diversity, and genomic organization, examined patterns of co-localization within counter-defense islands, and assessed associations with host taxa and biogeography. We further mapped the landscape of bacterial anti-MGE defense genes carried by phages, and the overall expression repertoire of phage immune modules. Finally, we extended our analysis to virophages and NCLDV, revealing a continuum of conserved strategies across viral lineages. Together, our findings contribute to a better understanding of the complexity of viral immune networks and highlight their central role in shaping microbial community structure and evolutionary dynamics across the biosphere.

Results

Abundance and distribution of counter-defense elements in phageomes

We first examined the distribution of phage-encoded counter-defense elements across a comprehensive collection of 50,595 high-quality viral operational taxonomic units (vOTUs, $\geq 90\%$ completeness and $\leq 5\%$ contamination or redundancy, see Methods) recovered from soil, marine, and human gut environments (Fig. 1a, Supplementary Data 1–2, Supplementary Fig. 1a). In parallel, we extended our analysis to identify and quantify phage-encoded anti-MGE systems of bacterial origin, which are typically co-opted for inter-MGE conflicts⁴³. Throughout this study, we considered a counter-defense or anti-MGE system to be complete when it retains an experimentally validated genetic architecture sufficient to confer immune activity. By contrast, individual counter-defense and anti-MGE genes are interpreted as remnants of these systems, arising either through HGT or through the progressive decay of once-intact loci. The composition of our vOTU collection is summarized in the Sankey diagrams in Fig. 1b, which connect phage morphology, family- and subfamily-level taxonomy, and sample provenance across the three environments. As expected given their global prevalence, head-tailed phages of the class *Caudoviricetes* predominate in all datasets, with the *Siphoviridae*, *Podoviridae*, and *Autographiviridae* families accounting for most recovered vOTUs. Across specific environments, soil vOTUs are dominated by *Siphoviridae* and uncategorized lineages, marine vOTUs are more evenly distributed among multiple families across ocean regions, while human-gut viromes exhibit the highest overall family richness, with *Siphoviridae* as the single most represented group across geographically diverse cohorts.

In our vOTU dataset, we identified a total of 15,080 counter-defense systems and 17,756 counter-defense genes spanning 18 counter-defense families (Fig. 2a, Supplementary Data 3). The relative distribution of counter-defense systems varied markedly across environments, with phages from the human gut exhibiting both higher diversity and greater system density per genome compared to soil and marine counterparts. Acr, anti-R–M, and anti-T–A systems emerged among the most abundant counter-defense strategies across all environments and were consistently detected in both virulent and temperate phages (Fig. 2a). This pattern mirrors the widespread prevalence of corresponding anti-MGE defense families in bacterial genomes¹⁵, suggesting a tight co-evolutionary coupling between bacterial immunity and phage countermeasures.

Also consistent with trends previously reported for bacterial anti-MGE defenses¹⁵, we observed a negative correlation between counter-defense density and medium-to-large sized phage genomes, particularly in human gut biomes (Supplementary Fig. 1b). This relationship is compatible with the notion that smaller phage genomes (typically associated with temperate lifestyles) engage more frequently in horizontal genetic exchange⁴⁴, and favor the acquisition and maintenance of mechanisms to evade host defenses. The number of counter-defense systems per vOTU followed a geometric-like distribution with most phages (75.1%) showing a limited arsenal of counter defense systems, and a global mean value of 0.3 (Fig. 2b). The latter is approximately ten-fold lower than the average number of defense systems reported per bacterial genome¹⁵ across similar environments, likely reflecting both biological asymmetries in host–virus

interactions and the current bias in our knowledge of defensomes relative to counter-defensomes. Still, a subset of human gut-associated *Myoviridae* phages displayed markedly expanded repertoires with up to 11 systems (Fig. 2b). Such outliers consist essentially in anti-R-M and likely reflect adaptation to hosts with particularly R-M-rich genomes, including species such as *Helicobacter pylori*. Stratification by phage family and subfamily revealed finer-scale patterns of counter-defensome enrichment. For example, members of *Schitoviridae* and *Podoviridae* appear to act as preferential reservoirs of specialized inhibitors, including anti-Thoeris and anti-Dnd systems, while others such as *Intestiviridae* are relatively depleted (Fig. 2c, Supplementary Data 4). Across environments, *Siphoviridae*- and *Podoviridae*-like phages encoded the broadest repertoire of counter-defense systems, including Acr, anti-R-M, anti-T-A and anti-SOS response modules (Figs. 2c,d, Supplementary Fig. 1c). *Proteobacteria* represented the predominant host phylum for soil and marine vOTUs, supporting a wide diversity of counter-defense strategies, whereas human-gut vOTUs were primarily associated with *Firmicutes* and *Bacteroidota* hosts and enriched in Acr, anti-SOS, and anti-Thoeris systems. Soil and marine environments showed a higher prevalence of counter-defense systems carried by virulent phages, whereas the human gut biome showed a higher prevalence of these systems carried temperate phages (Supplementary Fig. 1d). This pattern likely reflects: i) the lysogenic lifestyle of temperate phages in the human gut; and ii) the comparatively short and less frequent phage-bacteria contact in soil and marine environments, which favors virulent phages that carry counter-defense elements enabling them to overcome host immunity within a single infection attempt. While not directly implicated in neutralizing host defense systems, we also examined the presence of anti-MGE defenses in phages, which are thought to mediate inter-MGE conflicts and thereby favor their retention within host cells. We found 2,874 anti-MGE systems and 60,145 anti-MGE genes pertaining to a total of 191 families (Supplementary Data 3). These genes proved to be unevenly distributed across viral genomes and were globally dominated by R-M, Cas and All-UG families (Supplementary Fig. 2a). Particularly for human gut biomes, smaller genomes tended to be denser (per kb) in anti-MGE defense genes (Supplementary Fig. 2b), and most phages (~69%) harbored at most one anti-MGE defense gene (Supplementary Fig. 2c). At the phage-family level, the dominant families in each ecosystem carry the bulk of the anti-MGE load, with prominence in *Kyanoviridae*, *Autographoviridae* and *Siphoviridae* (Supplementary Fig. 2d). Finally, and similarly to counter-defense, soil and marine environments showed a higher frequency of anti-MGE genes in virulent phages, whereas the human gut biome showed a slightly higher prevalence in temperate phages (Supplementary Fig. 2e).

Overall, these results show that counter-defense and anti-MGE systems are widespread but unevenly distributed across phageomes, with clear stratification by environment, genome size and phage taxonomy. Their enrichment in specific habitats, particularly the human gut, smaller-sized genomes, and their concentration in distinct phage lineages indicate that the acquisition and maintenance of these loci are tightly linked to local host ecology and lifestyle.

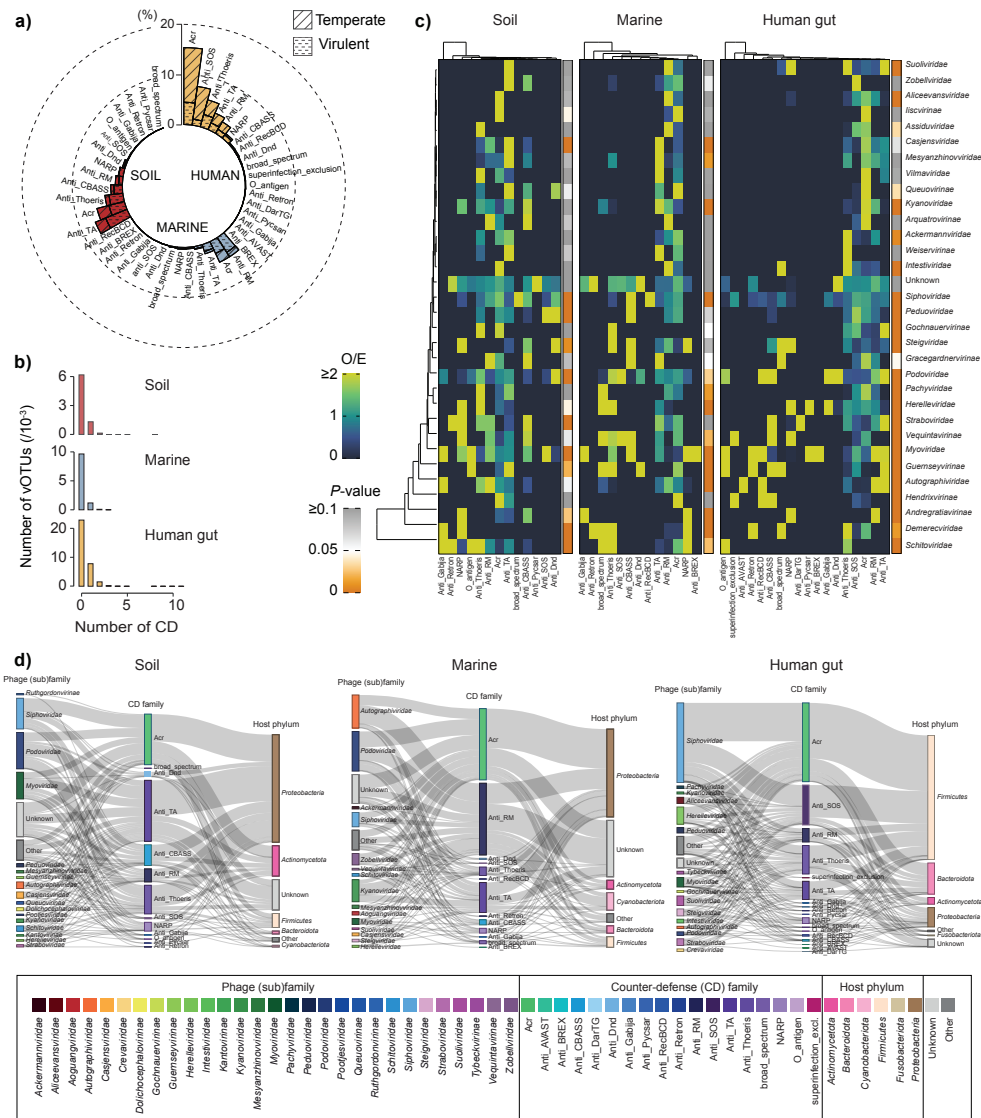


Fig. 2 Landscape of counter-defense (CD) systems across viral populations. **a)** Proportion of soil, marine, and human gut vOTUs encoding each CD family. **b)** Frequency distribution of CD system counts (per vOTU) across environments. **c)** Heatmap showing observed-to-expected (O/E) ratios for distinct CD families. Only families detected across all three environments (soil, marine, and human gut) are displayed. Expected values were calculated by multiplying the total number of systems within each counter-defense family by the overall proportion of all counter-defense-associated systems assigned to each viral (sub)family across the full dataset. Statistical significance was assessed using a two-sided Chi-square test. **d)** Sankey diagrams tracing the connections among phage family or subfamily, CD family, and host phylum across environments.

Co-localization of phage-encoded immune modules

Given the widespread distribution of counter-defense and anti-MGE systems/genes across phage genomes, we next quantified the extent of their co-occurrence, co-localization, and assessed whether they cluster into genomic islands (here termed counter-defense islands), in analogy to bacterial defense islands (Methods). Restricting the analysis to the five most abundant counter-defense and anti-MGE families, we found that genes pertaining to Dpd, Anti-Thoeris, and Anti-SOS more frequently co-occurred with other counter-defense gene families within the same phage genome (Fig. 3a). Among all pairwise combinations, Dpd/All_UG, Acr/Anti-SOS, Dpd/Anti-Thoeris, and Anti-SOS/Anti-Thoeris exhibited the highest observed-expected ratios, whereas Anti-TA/Acr and Anti-SOS/All_UG showed the lowest ones. Consistent with these findings, several immune gene families were found to co-localize (Fig. 3b), with counter-defense and anti-MGE modules frequently interspersed within the same genomic loci. In one representative example, a phage from an unknown family recovered from the human gut (uvig_530261) harbored a Type II R-M system flanked by a P1 *dar* Anti-R-M system of Type I, a *cas12f* (type V-F) effector gene, and the Anti-Thoeris *tad2* gene. This organization expands the repertoire of genomic regions that combine immune evasion and immune interference functions, extending previous observations of multi-layered counter-defense architectures²⁰. At a global scale, and irrespective of environmental context, we observed preferential co-localization between R-M / SspBCDE and to a lesser extent MADS / Acr (Fig. 3c, Supplementary Data 5). While the former suggests interplay between methylation and phosphorothioation, the latter parallels the MADS-CRISPR/Cas association observed in bacteria⁴⁵. In contrast, other combinations exhibited stronger environment-dependent patterns. For example, Cas / anti-Thoeris, anti-T-A / anti-Thoeris, anti-T-A / anti-CBASS co-localization was more prevalent in soil and/or human gut viral samples, whereas Gabija / Zorya co-localization was enriched in marine samples (Fig. 3c). Similar co-localization trends were observed if only counter-defense genes were considered in the analysis (Supplementary Fig. 3, Supplementary Data 5).

We next quantified the extent to which these co-localized genes are organized within counter-defense islands (Methods). Counter-defense and anti-MGE genes co-localized within 10 CDS of one another in 3.7% of phage genomes, with 26.7% co-localizing as immediately adjacent blocks. In total, we identified 99 counter-defense islands in 93 vOTUs from our dataset (Supplementary Data 6, Methods). Although more than 70% of the genes within counter-defense islands could not be assigned a clear immune-related function, approximately 20 major defense-associated families were identified, with R-M, Cas, Acr and NARP being the most abundant (Fig. 3d). These islands displayed similar median sizes across environments (approximately 10 genes), although human-associated samples contained the largest islands of up to 40 genes (Fig. 3e). Counter-defense island density likewise showed comparable median values across environments, but differed in its distribution, with soil and marine samples exhibiting the narrowest ranges and human-associated samples the greatest variability with as much as 3 DIs per vOTU (Fig. 3f). These patterns are consistent with earlier observations of an expanded defensome and increased island content in human gut MAGs relative to those from other biomes¹⁵.

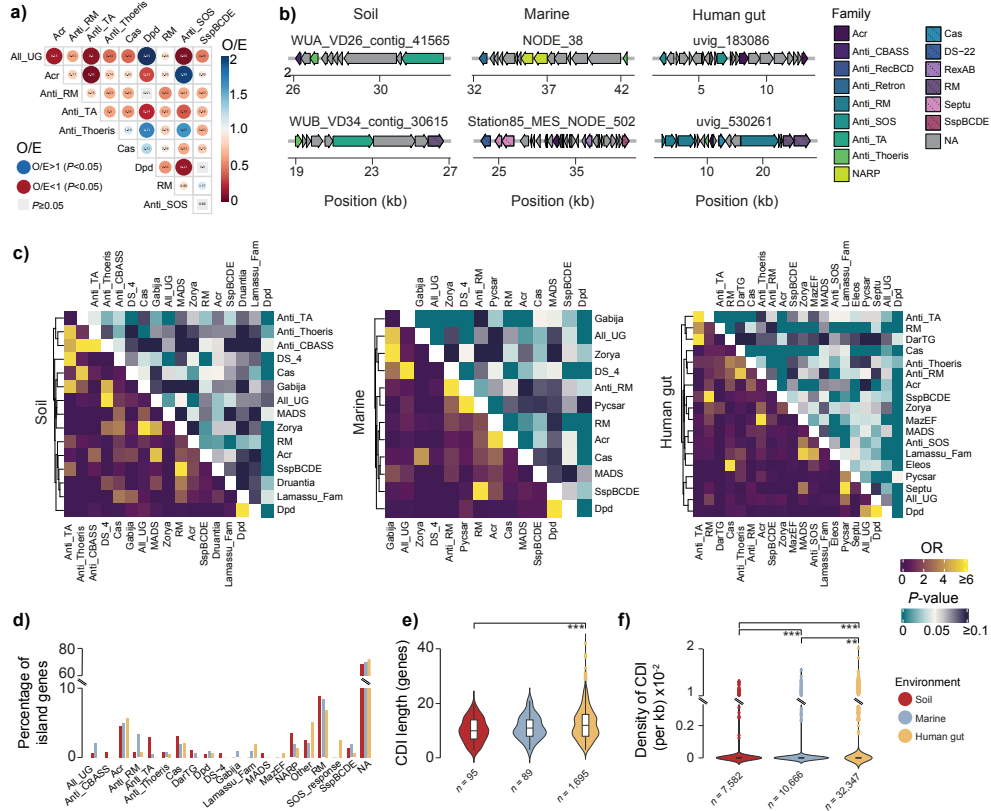


Fig. 3 Co-occurrence and co-localization of counter-defense and anti-MGE immune modules across soil, marine, and human gut vOTUs. **a)** Pairwise co-occurrence of presence/absence patterns for the ten most abundant counter-defense and anti-MGE families across 36,239 phage genomes. For each family pair, presence/absence was tested for independence using Pearson's chi-square test, and the ratio of observed to expected (O/E) joint presence was computed, where the expected joint presence was calculated as the product of each family's marginal presence frequency across the dataset. Circle color indicates the direction of the deviation from the null expectation of independence (red, O/E < 1; blue, O/E > 1), and both size and color intensity scale with the magnitude of that deviation. The O/E value is given within each symbol. Gray squares indicate non-significant correlations ($P \geq 0.01$, Chi-square test). **b)** Representative examples of co-localizing immune modules across environments. **c)** Odds ratio (OR) of co-localization (bottom heatmaps) and associated two-sided Fisher's exact test P values (upper heatmaps) among anti-MGE defense and counter-defense gene families, restricted to families with a co-localization frequency $\geq 1\%$ across soil, marine, and human gut vOTUs. **d)** Relative abundance (%) of gene families within counter-defense islands across environments. **e)** Counter-defense island (CDI) length distribution (in genes) across environments. **f)** Distribution of CDI densities (per kb, per vOTU). Statistical comparisons were performed using two-sided Wilcoxon rank-sum tests (** $P < 10^{-2}$, *** $P < 10^{-3}$).

205 In summary, co-localization patterns point to potential unrecognized epistatic relationships among specific subsets of counter-defense and anti-MGE genes in phage genomes. Approximately 8% of phage immune-related modules are concentrated within counter-defense islands, which display environmental variation and contain a substantial proportion of non-defense genes. Similar to bacterial defense islands, these genomic hubs are likely to shape phage–host antagonistic interactions.

210 **Biogeography shapes phage counter-defense repertoires and transcriptome**

Environmental conditions shape phage diversity, and phages in turn shape the microbial composition of their environments, driving continuous co-evolution. To explore this interplay, we examined how the phage counter-defense varies with biogeography. Counter-defense systems exhibited marked heterogeneity in both density and family composition across finer ecological sub-habitats in soil and marine environments and across geographic cohorts in the human gut (Fig. 4a, Supplementary Data 7). Soil vOTUs from desert environments displayed particularly high counter-defense densities and anti-SOS modules, whereas volcanic soils were comparatively depleted and dominated by Acr modules. Marine phages maintained lower overall densities with a homogeneous breakdown of counter-defense families across oceanic regions. In contrast, human-gut phages sustained consistently higher densities across countries, with broadly similar frequency spectra interrupted by clear country-specific differences (for example, virtual absence of Acr systems in Italian samples contrasted with their marked over-representation in Tanzanian cohorts). Counter-defense modules, including anti-SOS elements, were more abundant in human gut phages than in marine or soil phages, in line with the higher density and diversity of bacterial defenses, as well as frequent SOS induction in the gut due to genotoxic stressors and high prophage induction rates^{46–48}. The results were qualitatively similar when performed with counter-defense genes (Supplementary Fig. 4a, Supplementary Data 7) or anti-MGE genes (Supplementary Fig. 4b, Supplementary Data 7), with the exception that overall densities were globally more elevated in the latter.

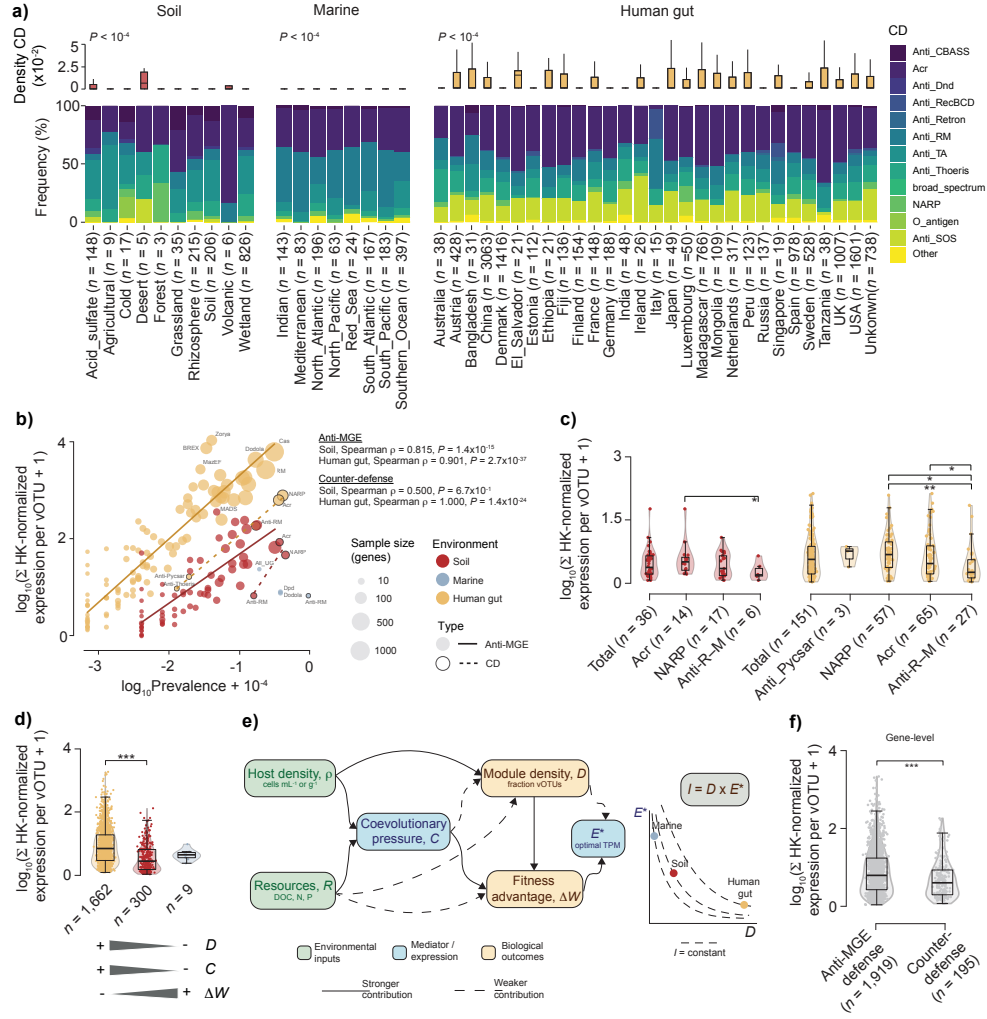


Fig. 4 Interplay between phage biogeography, counter-defense repertoires, and their transcriptome. **a)** Density and frequency of counter-defense (CD) systems across distinct ecological (soil, marine) and geographical (human gut) contexts. Boxplots depict the interquartile range (25th - 75th percentiles), with the median indicated by a central black line. Whiskers extend to 1.5x the interquartile range, and outliers were omitted for improved visualization. Statistical significance was assessed using a Kruskal-Wallis test. The number of vOTUs included in each category is provided in parentheses. **b)** Relationship between gene prevalence and sum of HK-normalized expression per vOTU across anti-MGE defense (solid trend lines) and counter-defense (dashed trend lines) genes in soil (red), human gut (yellow), and marine (blue) samples. Each point represents one immune family sized by the number of contributing genes. Spearman correlation coefficients and P values are shown for each immune class and environment. Marine was excluded from correlation testing owing to small sample size. **c)** Summed housekeeping (HK)-normalized expression per vOTU of counter-defense genes carried by soil (red) and human gut (yellow) vOTUs, shown for all counter-defense-carrying vOTUs pooled (Total) and split by family. n corresponds to the number of vOTUs per category. **d)** Same expression metric as in (b, c), pooled across all counter-defense + anti-MGE genes and compared across environments. Beneath the plot, schematic gradients summarize how immune module density (D), coevolutionary pressure (C), and per-module fitness advantage (ΔW) are inferred to vary.

across environments. **e)** Schematic model linking environmental inputs to realized immune investment. Host density (ρ) and resource availability (R) jointly shape co-evolutionary pressure (C), which in turn shapes the fitness advantage (ΔW) conferred by carrying a given module. Host density additionally shapes immune module density (D , fraction of vOTUs carrying a given module) directly, independent of co-evolutionary pressure, with C exerting a minor, secondary influence on D over longer timescales. D feeds directly into ΔW (as a module becomes more prevalent, its marginal fitness benefit erodes through negative frequency-dependence). ΔW is the primary driver of each module's optimal expression level (E^*). D contributes a smaller, secondary effect, in line with a redundancy/free-riding dynamic. Dashed arrows denote weaker, secondary contributions alongside the primary (solid) pathways above. Total realized immune investment is modeled as $I = D \times E^*$. The graph on the right represents optimal expression (E^*) against module density (D) for the three environments. Realized total investment (I) is highest in human gut and comparably low in soil and marine. Human gut and marine sit at opposite, constrained extremes of the D - E^* trade-off. In the gut, D saturation (via HGT) reduces the fitness advantage of the module, pulling E^* down even as D itself stays high. In marine, a genome-streamlining selection filter caps D directly, even though per-carrier E^* remains comparatively high. Soil is unconstrained by either extreme, but its overall inputs (ρ , C) are modest, so it remains at low D and low-to-moderate E^* . **f)** Summed HK-normalized expression per vOTU of immune genes split by anti-MGE or counter-defense function. Statistical comparisons were performed using two-sided Wilcoxon rank-sum tests (* $P < 0.05$, ** $P < 10^{-2}$, *** $P < 10^{-3}$)

Having analyzed the impact of biogeography, we next asked which counter-defense and anti-MGE families were transcriptionally active in viral populations, and at what magnitude. To this end, we mapped metatranscriptomic reads from soil, human gut, and marine viromes onto their cognate vOTUs, and related summed per-vOTU transcript abundance for each family to that system's prevalence within the corresponding metagenomic population. This analysis revealed that counter-defense and anti-MGE immune modules are transcribed highly unevenly, with transcriptional level tightly coupled to how common a given system is at the population level. Summed per-vOTU expression correlated positively with system prevalence in both soil and human gut metatranscriptomes (Fig. 4b, Supplementary Data 8). Marine metatranscriptomes, by contrast, were comparatively depleted in immune modules, and only a handful of systems passed detection thresholds, precluding a statistically robust correlation. Genes from widely distributed families (Cas, R-M, NARP, Acr, Anti-R-M) mapped at the high-prevalence, high-expression end of this relationship in both environments, while narrowly distributed genes were both rare and weakly transcribed. When comparing the most prevalent counter-defense families in each environment, anti-R-M genes were significantly more weakly expressed than Acr and NARP in both soil and human gut vOTUs (Fig. 4c, Supplementary Data 9). This might be explained by the fact that R-M cleaves unmodified phage DNA shortly after cell entry, and anti-R-M genes are correspondingly restricted to a brief, immediate burst at the point of genome entry, a narrow window that a cross-sectional sample is less likely to catch. Instead, CRISPR interference and NAD^+ depletion unfold over longer and/or later windows of infection^{49–51}, so Acr and NARP may remain transcriptionally active across a substantially larger share of the infection cycle, making them proportionally easier to detect in cross-sectional data. At the community level, total per-vOTU immune-related expression differed sharply between environments, with human gut vOTUs showing significantly higher expression than soil vOTUs, and marine vOTU data being less abundant but falling within the same low range as soil (Fig. 4d, Supplementary Data 10). This trend correlates with lower host density, co-evolutionary pressure, and module density in

marine environments relative to the human gut. Yet this same rarity is expected to confer, via negative frequency-dependence, a higher fitness advantage on the few marine phages that carry a given module. These observations can be illustrated with a simple resource–density model (Fig. 4e). In densely populated ecosystems like the human gut, elevated viral-host encounter rates typically drive extensive neutral HGT, leading to a ubiquitous prevalence of immune modules across vOTUs. However, this genetic saturation is expected to diminish the marginal fitness advantage of the trait during viral-viral and viral-host competition, as a module carried by nearly every genome ceases to confer a competitive edge. To offset the metabolic burden of transcription, selection will likely drive optimal expression to basal levels. Conversely, in less populated marine environments, module prevalence is expected to decrease under stringent selection conditions and reduced contact times, but this same decrease will likely maximize the competitive payoff for the few phages possessing these systems, driving localized expression to its theoretical peak. In soil, by contrast, neither constraint dominates, and module prevalence and per-carrier expression both settle at intermediate values. This release from either extreme, however, does not seem to translate into the highest total investment, because soil’s underlying host density and co-evolutionary pressure are themselves comparatively modest, and its realized immune investment remains comparable to that of marine communities. Realized investment is instead highest in the human gut, where the scale of host density-driven module acquisition outweighs the accompanying reduction in per-carrier expression.

Similar trends were obtained when performing the analysis at the system level (Supplementary Figs. 5a,b, Supplementary Data 11) and when restricting it to anti-MGE modules alone (Supplementary Fig. 5c,d, Supplementary Data 12). Globally, anti-MGE genes showed significantly higher expression than counter-defense genes (Fig. 4f). As noted above, this asymmetry may reflect a genuine biological difference in transcriptional investment between the two module classes, or the current annotation bias between them. Regardless, system-level resolution revealed that phage-encoded CRISPR/Cas systems dominated anti-MGE expression in the human gut, whereas R–M systems were the most actively expressed in soil (Supplementary Fig. 5d).

Collectively, these patterns demonstrate that phage counter-defense repertoires are finely tuned to local ecological and geographic contexts, shaped by host defenses, environmental stressors, and evolutionary pressures. They also support a model in which transcriptional investment, and the specific systems recruited to deliver it, scale primarily with host density and the resulting intensity of host-phage arms races.

Immune-related modules are absent in virophages but present in NCLDV

The widespread occurrence of anti-MGE defenses in bacteria and counter-defenses in bacteriophages, together with evidence of HGT involving bacteria, phages, virophages, and NCLDVs, suggests that immune-related genes may also be present in virophages and NCLDVs. To explore this hypothesis, we systematically surveyed the presence and distribution of bacterial anti-MGE defense genes and phage counter-defense genes in virophage and NCLDV genomes (Supplementary Fig. 6a, Supplementary Data 13).

We detected no significant homology between virophage genes and any characterized immune modules from prokaryotes or phages. Given the highly compact nature

of virophage genomes and their obligate dependence on the replication machinery of
NCLDV, virophages are likely to rely on minimal, structurally simple counter-defense
strategies. Furthermore, the selective pressures imposed by giant virus-encoded defense
310 systems such as MIMIVIRE, are expected to drive accelerated sequence evolution,
recombination, or other genomic innovations, that enable virophages to evade target-
ing while maintaining parasitic compatibility with their hosts. On the other hand,
across 5,284 NCLDV genomes, we identified only three instances of MIMIVIRE, and
315 a limited number of phage counter-defense systems and genes (126 and 776 respec-
tively) (Fig. 5a, Supplementary Data 14). In contrast, a diverse repertoire of bacterial
anti-MGE defense genes was observed, unevenly distributed across the dataset (Sup-
plementary Data 14). Some families, notably CRISPR/Cas and R-M, were present in
a substantial fraction of genomes ($> 39\%$), whereas others were confined to a limited
number of viral lineages. To assess whether these genes occur as solitary elements or as
320 components of larger systems, we examined their genomic organization. Solitary genes
were the most abundant, particularly DNA MTases, while multi-gene configurations
were comparatively rare. An exception was the presence of digenic arrangements, con-
sistent with the presence of R-M modules (Fig. 5b). Among MTases, Type II genes
dominated, whereas Type I were less frequently observed (Fig. 5c). We also identi-
325 fied loci in which multiple defense genes co-localize, although these were less frequent
than islands reported in bacterial or phage genomes (Supplementary Fig. 6b, Supple-
mentary Data 15). When comparing NCLDV-encoded defense-gene abundance across
biomes, marine was the most NCLDV-rich environment overall, and also the one in
which these viruses harbor the highest density of defense genes (Fig. 5d). Within
330 marine samples specifically, no significant differences were observed in defense-gene
family composition across different oceans (Supplementary Fig. 6c). When splitting
per defense and NCLDV families, the most abundant viral families (*Mimiviridae* and
Phycodnaviridae) exhibited a statistically significant enrichment of multiple defense
genes, whereas a small subset of families, such as Dodola, Lamassu, and PsyrTA,
335 were enriched across most viral families. Concomitantly, the least abundant viral fam-
ilies, were largely devoid of most known defense genes, with particular emphasis to
Asfarviridae, in which only Dnd was detected (Fig. 5e).
To further characterize the genomic context of immune-related genes in NCLDVs,
we analyzed protein domain composition within a ± 4 coding sequence window
340 around known defense gene anchors. This analysis revealed significant enrichment of
several domain clans, many of which have established or putative roles in antivi-
ral defense (Fig. 5f-g, Supplementary Data 16). Examples include: *i*) Cas4-like
domains, a hallmark component of the MIMIVIRE system; *ii*) members of the PD-
(D/E)XK nuclease superfamily^{52,53}; *iii*) RNase H and RNase H-like domains⁵⁴;
345 *iv*) OB-fold kinases⁵⁵; *v*) RING finger domains, which function as putative E3
ubiquitin ligases⁵⁶; *vi*) ribonucleotide reductases (RNRs)⁵⁷; and *vii*) GT-A fold gly-
cosyltransferases⁵⁸. Observed-to-expected enrichment analysis revealed non-random
co-occurrence between defense-family homologs and specific accessory domain clans
(Fig. 5h). As these families are assigned by homology to defense systems characterized
350 in bacteria and archaea, whether the associated domains reflect genuine functional
coordination, shared horizontal acquisition, or broader features of NCLDV genome
organization remains to be determined.

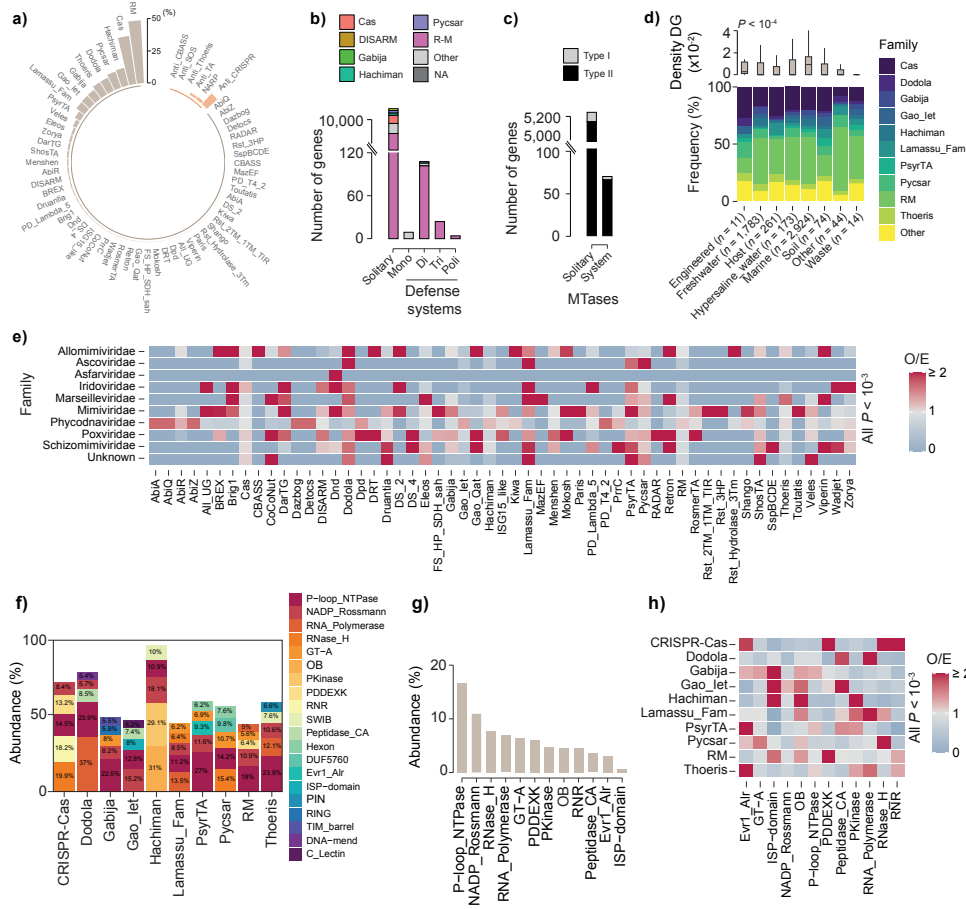


Fig. 5 Bacterial anti-MGE defense genes are abundant in NCLDVs. **a)** Fraction of NCLDVs encoding each identified anti-MGE defense family (beige) compared to counter-defense families (light orange). **b)** Distribution of defense genes (DG) according to genomic organization (solitary or arranged within mono-, di-, tri-, or polygenic defense systems). **c)** Classification of DNA MTase genes as either solitary or belonging to complete R-M systems. **d)** Density and frequency of anti-MGE genes in NCLDVs recovered from distinct environments. Boxplots depict the interquartile range (25th - 75th percentiles), with the median indicated by a central black line. Whiskers extend to 1.5x the interquartile range, and outliers were omitted for improved visualization. Statistical significance was assessed using a Kruskal-Wallis test. The number of NCLDV genomes included in each category is provided in parentheses. **e)** Heatmap of observed-to-expected (O/E) ratios for distinct defense families across major NCLDV families. Expected values were calculated by multiplying the total number of genes within each defense family by the overall proportion of all defense-associated genes assigned to each viral family across the full dataset. **f)** Most abundant protein domain clans found within a ± 4 CDS window surrounding the ten most abundant NCLDV defense gene families (top 5 clans per family shown). **g)** Rank-ordered abundance of these domain clans within the same ± 4 CDS genomic radius. **h)** Heatmap of observed-to-expected (O/E) ratios for domain clans across major defense gene families. Expected values were calculated by multiplying the total number of domains within each clan by the overall proportion of all domains assigned to each defense family across the full dataset.

Overall, these results reveal a clear contrast between virophages and NCLDV in
the presence of immune-related genes. While no immune modules were detected in
355 virophages, NCLDVs encode a diverse set of genes homologous to bacterial anti-MGE
defense systems, unevenly distributed across viral lineages and predominantly present
as solitary loci, with occasional multi-gene configurations. These loci show non-random
genomic association with putative defense-associated accessory domains, suggesting
that NCLDVs have evolved a mosaic defensive landscape either reflecting functional
360 innovation, differential retention, or recurrent horizontal acquisition.

Discussion

In this study, we present a comprehensive analysis of the abundance and diversity of phage-encoded counter-defensomes across three representative biomes: soil, marine environments, and the human gut. This work complements previous analyses of bacterial MAG defensomes¹⁵, enabling a parallel, biome-resolved comparison of defense and counter-defense strategies across both hosts and their viruses. Consistent with prior studies on Refseq genomes²⁰, we found predominance of anti-R-M and Acr proteins, presumably because their targets are also the most abundant in bacterial populations. Although counter-defense genes were detected across a broad diversity of vOTUs, their distribution was skewed toward phages predicted to infect *Gammaproteobacteria*. This apparent enrichment likely reflects historical discovery bias, as early characterization of counter-defense systems relied heavily on model Gammaproteobacterial hosts and their associated phages. Furthermore, detection biases may contribute to these patterns, as taxa overrepresented in reference databases such as RefSeq are not necessarily the most abundant in natural environments, whereas ecologically dominant but under-sequenced organisms may be systematically underrepresented in current annotations of counter-defense genes. Challenges remain in the detection of distantly related counter-defense proteins, largely due to their small size and extensive sequence divergence. Recent advances in protein structure prediction and comparative structural analyses are helping to overcome these limitations, enabling a shift toward structure-guided discovery of novel counter-defense systems. This approach has already proven effective in identifying new Acr proteins in bacteriophages⁵⁹. In any case, the diversity of counter defense systems we detected is consistent with recent simulations predicting the number of immune modules sustained in coexisting ecosystems^{60,61}.

Counter-defense modules were found to be globally more abundant in human gut phages than in marine or soil phages (Fig. 2a). This enrichment likely stems from the elevated density and diversity of bacterial defense systems in the gut¹⁵, which impose stronger and more continuous selective pressure on phages despite the characteristically low VPR and predominance of temperate lifestyles. Constant exposure to a high density and diversity of defense systems is expected to drive rapid phage adaptation through bursts of acquisition of compatible counter-defense systems or the expansion of existing ones⁶², with similar adaptive dynamics likely to occur in other defense-rich environments, including the urethra, vagina, oral cavity, and skin⁶³. We also observed an enrichment of anti-SOS response modules in gut-derived phages. This pattern is consistent with the higher frequency of SOS response induction in the human gut microbiome than in marine or soil habitats, likely driven by persistent genotoxic stressors such as host-derived reactive oxygen and nitrogen species, microbial secondary metabolites (*e.g.*, colibactin), bile salts, and frequent prophage induction linked to high lysogeny rates^{46–48}.

Similarly to the bacterial defensome¹⁵, the phage counter-defensome also displayed marked biogeographic variation. In particular, desert environments exhibited a higher density of counter-defense genes, whereas volcanic environments were characterized by comparatively lower diversity. Drought conditions were previously shown to nearly double negative prokaryote–phage interactions, coinciding with an increased abundance and diversity of bacterial antiphage defense systems⁶⁴. In the particular case

of microbial communities in hyper arid desert soils, they experience episodic hydration alongside persistent exposure to desiccation, osmotic, and UV-induced oxidative stresses. These environments are further characterized by a high prevalence of lysogeny and low extracellular viral abundances⁶⁵. Such conditions may explain the selection
410 for vOTUs enriched in anti-SOS systems (Fig. 4a), potentially as part of an adaptive strategy to suppress host stress-induced prophage induction. In the case of volcanic soils, SO₂ exposure creates a strongly filtered habitat with the most affected deposits being acidic, carbon- and nitrogen-poor, and dominated by a small set of pioneer autotrophs. In these rapidly assembling bacterial communities of newly formed vol-
415 canic soils, phage-encoded counter-defense systems are expected to exhibit limited diversity, reflecting the transient and taxonomically dynamic host landscape⁶⁶. Instead of maintaining a wide repertoire of highly specialized modules tailored to individual hosts, phages in these environments are likely to rely on broadly effective strategies, which helps explain the observed predominance of anti-R-M and Acr mechanisms
420 (Fig. 4). As microbial communities stabilize during ecological succession, the increasing diversity and persistence of host interactions are expected to drive the expansion and diversification of the phage counter-defensome.

We found extensive co-localization between the phage counter-defensome and the anti-MGE defensome, as well as general evidence for the existence of counter-defense
425 islands. This pattern supports the view that MGEs are major drivers of Red Queen dynamics by promoting the rapid dissemination and co-accumulation of both defense and their counter-defense mechanisms within microbial populations. We also provided, the first snapshot of the metatranscriptome landscape of immune modules for phage populations across diverse biomes. To date, no community metatranscrip-
430 tomic study has systematically quantified the expression of known anti-MGE defense and counter-defense genes across natural microbiomes. Instead, most ecological studies infer immune repertoires from gene presence in metagenomes, whereas expression analyses have largely been restricted to laboratory infection models^{67–69}. We caution, however, that assigning expression signals to viral or bacterial origins is not
435 unequivocal. Since read assignment relies on competitive pseudoalignment, transcripts originating from conserved or recently horizontally transferred loci may align equally well to homologs encoded in both viral and bacterial genomes and are therefore assigned probabilistically according to their relative transcript abundance. Yet, since bacterial transcripts typically dominate community metatranscriptomes, ambiguous
440 reads are expected to be preferentially assigned to the bacterial compartment, making our estimates of phage defense and counter-defense expression conservative.

The diversification of immune systems is often facilitated by their modular organization, which allows frequent recruitment and exchange of defensive modules. In recent years, our understanding of the anti-MGE defensome has expanded dramatically,
445 largely driven by the ‘guilty by association’ principle, *i.e.*, the observation that multiple immunity-related modules frequently cluster within specific genomic loci in bacteria⁷⁰. In this study, we extended this concept to NCLDV, querying domains preferentially co-localized with known immune modules. While the functional significance of these co-localizations remains uncertain, their enrichment in defense-associated domains is
450 consistent with a role in coordinated antiviral activity. Our observations in NCLDV

align with recent findings using machine-learning frameworks, which identified PD-(D/E)XK nucleases, P-loop NTPases, and PIN domains among the most prevalent features in newly predicted defense modules in bacteria⁵⁹. Also, our identification of numerous Type II MTases helps explain the methylome diversity of giant viruses³⁷.
455 Moreover, the predominance of solitary enzymes suggests that at least some may play functional roles in gene expression regulation. It remains to be determined to what extent bacterial MTases carried by NCLDV genomes are transferred to eukaryotic hosts, how they evolve within NCLDV genomes, and whether, similarly to bacteria, viruses with comparable methylation repertoires promote genetic exchange^{71,72}. Hence, our study
460 provides a biome-resolved view of phage anti-MGE and counter-defense repertoires, extended to dsDNA viruses. These insights refine our understanding of host–virus interactions across ecosystems and establish a framework for the discovery of novel immunity-related modules across the virosphere.

Methods

465 Data

In this study, we analyzed a large dataset of 50,595 publicly available phage vOTUs from soil^{73–76}, marine^{38,77–79}, and human gut environments⁴⁰. These vOTUs were filtered using PhaBOX⁸⁰ v2.1.12 (default parameters) to retain only viral contigs, further screened for genome quality using CheckV⁸¹ v1.0.3 (end_to_end mode, completeness $\geq 90\%$, contamination $\leq 5\%$), and dereplicated using dRep⁸² v3.6.2 (options: -pa 1, -l 18000) to remove redundant sequences and retain only high-quality contigs. vOTU annotation was performed with Prokka⁸³ v1.14.5 (default parameters). In parallel, we analyzed dedicated datasets of 1,682 virophage⁸⁴ and 5,284 NCLDV genomes^{85,86}. Contigs were dereplicated using dRep (-pa 1), with an additional minimum length threshold of 10 kb applied to virophage sequences. For the detection and quantification of immune gene expression, 19 publicly available metatranscriptomic datasets were used, spanning marine, soil (^{87–89}), and human gut environments (^{90–92}) (Supplementary Table 8). These studies also provided corresponding metagenomic data, enabling: i) the recovery of vOTUs from CheckV-filtered contigs ≥ 1 kb, and ii) the recovery of bacterial metagenome-assembled genomes (MAGs) quality-assessed with CheckM v1.2.3 (completeness $\geq 90\%$, contamination $\leq 10\%$). Both viral and bacterial genome catalogs were dereplicated with dRep at a 95% average nucleotide identity (ANI) threshold. Gene prediction was performed with Prodigal⁹³ v2.6.3.1.

Taxonomic assignment and viral morphology

485 Taxonomic assignment of phage vOTUs was performed using a combination of PhaBOX (default parameters), Kaiju⁹⁴ v1.7.3 (default parameters), and vConTACT3⁹⁵ v3.1.6 (options: -nucleotide -db-version 230). vOTUs sharing $\geq 20\%$ of their homologous proteins were considered to belong to the same viral family. Unclassified sequences were further assigned at the family level using VPF-class⁹⁶ (membership ratio > 0.5 , confidence score > 0.5), retaining only the highest-confidence prediction per genome. Host taxonomy was predicted using PhaBOX and iPhoP⁹⁷ v1.4.1 (default parameters). Viral morphology was inferred from the functional annotation of structural proteins by comparing predicted proteins against the RefSeq viral database using MMseqs2⁹⁸ and the non-redundant database using DIAMOND⁹⁹ v2.0.15, with assignments based on the detection of characteristic marker proteins (major capsid protein, portal protein, and tail proteins for head-tailed viruses, filamentous capsid proteins for rod-shaped viruses, and double jelly-roll capsid proteins for icosahedral viruses). For vOTUs that remained unclassified, morphological information was further refined by propagation within gene-sharing clusters defined by vConTACT3. Although *Archaea* are also infected by viruses, there is ongoing debate about whether these viruses should be classified as phages. In this study, archaea-infecting viruses were considered phages, despite the fact that most viruses considered in this study typically infect bacteria. For virophage and NCLDVs, taxonomic assignments were verified using PhaBOX and Kaiju, retaining only sequences classified as *Lavidaviridae* for virophages and 505 *Nucleocytoviricota* for giant viruses.

Identification of anti-MGE defense and counter-defense genes / systems and islands

Anti-MGE defense genes and systems were identified in vOTUs using DefenseFinder¹⁴ v2.0.0 (option: `-preserve-raw`). Counter-defense genes and systems were identified by combining DefenseFinder anti-defense mode (options: `-antidefensefinder-only -preserve-raw`) with the dbAPIS framework¹⁰⁰. For dbAPIS-based detection, predicted proteins were screened using both profile-based searches with HMMER v3.4 (<http://hmmer.janelia.org/>) (hmmsearch) and sequence similarity searches with DIAMOND⁹⁹ v2.0.15 (blastp), and annotations were integrated using the dbAPIS parsing workflow. Hits were filtered to retain only alignments with a minimum coverage of 80%, and for each gene only the best-supported assignment was kept. Finally, overlapping annotations between defense and counter-defense catalogs were removed at both the gene and system levels. MIMIVIRE systems were identified through a BLASTp search between the protein sequences R349 (YP_003986852.1), R350 (YP_003986853.1), and R354 (YP_003986857.1) from *Acanthamoeba polyphaga* mimivirus (NC_014649.1) and the full NCLDV proteome dataset. Hits were filtered at a minimum of 80% identity and 80% query coverage. To assess co-localization between anti-MGE / counter-defense families, we computed odds ratios and associated Fisher’s exact test *P* values. For this purpose, counter-defense genes were ordered along each contig and co-localization was evaluated using a bidirectional sliding window in which each annotated gene was paired with up to the next five annotated genes on the same contig. Because genes belonging to the same predicted complete system are inherently clustered, gene pairs assigned to the same system were deliberately excluded to avoid inflating co-localization frequencies. Counter-defense islands were defined as clusters of annotated genes separated by ten CDS or less, and only clusters containing at least three counter-defense / anti-MGE genes were retained. For islands containing only counter-defense genes, the same identification and filtering strategy was applied.

Metatranscriptomic analyses of immune genes

Raw paired-end RNA-seq reads were quality- and adapter-trimmed with fastp¹⁰¹ v0.24.0 (options: `-q 25, -e 25, -u 20, -length_required 50, -detect_adapter_for_pe`). Prokaryotic and eukaryotic non-protein-coding rRNA sequences were then removed from the trimmed paired reads with SortMeRNA¹⁰² v4.3.4 (options: `-paired_in, -out2, -num_alignments 1`). Read pairing was restored after filtering with repair.sh from BBTools¹⁰³ v39.19, and pseudoaligned against a combined Kallisto¹⁰⁴ v0.52.0 index built from the pooled CDS catalogs of the dereplicated bacterial MAGs and viral vOTUs using kallisto quant (options: `paired-end mode, -pseudobam, -fr-stranded, -rf-stranded, or unstranded`). Pseudoalignment BAM files were coordinate-sorted and indexed with SAMtools¹⁰⁵ v1.21, and per-base read depth was computed with samtools depth -aa. For each CDS, mapped reads and effective length were taken from Kallisto’s output, and reads per kilobase (RPK) were calculated as $RPK_{\text{gene}} = \frac{\text{mapped reads}_{\text{gene}}}{\text{effective length}_{\text{gene}} [\text{kb}]}$. Breadth of coverage was computed from the samtools depth output as the proportion of each gene’s length covered by ≥ 1 read. A gene was considered reliably supported and retained for downstream analysis if it had ≥ 5 mapped reads and $\geq 50\%$ breadth of coverage. Each immune gene’s RPK was normalized to a community-internal housekeeping-gene baseline composed of bacterial *gyrA* and *rpoB* as $RPK_{\text{HK}} = \frac{RPK_{\text{gyrA}} + RPK_{\text{rpoB}}}{2}$. Every supported gene’s housekeeping (HK)-normalized expression was then calculated as $\text{HK-normalized expression}_{\text{gene}} = \frac{RPK_{\text{gene}}}{RPK_{\text{HK}}}$. For

each vOTU, HK-normalized expression values of all supported defense- or counter-defense-annotated genes were summed to obtain the Σ HK-normalized expression per
555 vOTU.

Gene neighborhood analysis

To explore the genomic context of defense genes, we performed a gene neighborhood analysis on giant virus contigs. For each defense gene predicted by DefenseFinder, neighboring genes within a sliding window of four coding sequences upstream and
560 downstream on the same contig were retrieved. Defense genes themselves were subsequently excluded from this neighborhood dataset to focus on putative novel defense-associated genes. The resulting protein sequences were functionally annotated using InterProScan¹⁰⁶ v5.75.106.0, and only domains assigned to Pfam entries were retained (e-value $\leq 10^{-5}$). When available, Pfam clans were used to group related
565 domains into higher-level functional categories.

Statistical and graphical analyses of data

Statistical and graphical analysis of the results was conducted using R v4.3.3. Sankey diagrams were built with the R package networkD3.

Code availability

570 Wrapper scripts supporting all key analyses of this work are publicly available at
<https://github.com/oliveira-lab/Counter-Defensome>.

Author contributions

P.H.O. supervised the project. A.B., L.S.S., I.G.M.G., L.D. and P.H.O. designed the
computational methods. A.B., L.S.S., I.G.M.G., and L.D. performed most of the com-
575 putational analyses and developed most of the scripts that support the analyses. A.B.,
L.S.S., I.G.M.G., L.D., N.W., P.W., and P.H.O. analyzed the data. A.B. and P.H.O.
wrote the manuscript with additional information inputs from other co-authors.

Declaration of interest

The authors declare no competing interests.

580 Acknowledgments

We thank Eduardo P.C. Rocha (Institut Pasteur, Paris) for his critical reading of
the manuscript. This work was supported by the Genoscope, the Commissariat à
l'Énergie Atomique et aux Énergies Alternatives (CEA), France Génomique (ANR-
10-INBS-09-08), and the Fondation pour la Recherche Médicale (grant number
585 FDT202504020397) to A.B.

References

- [1] Mushegian, A. R. Are there 10^{31} virus particles on Earth, or more, or fewer? *Journal of Bacteriology* **202**, e00052–20 (2020). URL <https://journals.asm.org/doi/10.1128/JB.00052-20>.
- 590 [2] Sausset, R., Petit, M. A., Gaboriau-Routhiau, V. & De Paepe, M. New insights into intestinal phages. *Mucosal Immunology* **13**, 205–215 (2020). URL <https://linkinghub.elsevier.com/retrieve/pii/S1933021922002872>.
- [3] Parikka, K. J., Le Romancer, M., Wauters, N. & Jacquet, S. Deciphering the virus-to-prokaryote ratio (VPR): insights into virus-host relationships in a variety of ecosystems. *Biological Reviews* **92**, 1081–1100 (2017). URL <https://onlinelibrary.wiley.com/doi/10.1111/brv.12271>.
- 595 [4] Koonin, E. V. *et al.* Global organization and proposed megataxonomy of the virus world. *Microbiology and Molecular Biology Reviews* **84**, e00061–19 (2020). URL <https://journals.asm.org/doi/10.1128/MMBR.00061-19>.
- 600 [5] Krupovic, M. *et al.* *Adnaviria*: a new realm for archaeal filamentous viruses with linear A-form double-stranded DNA genomes. *Journal of Virology* **95**, e00673–21 (2021). URL <https://journals.asm.org/doi/10.1128/JVI.00673-21>.
- [6] Roux, S. *et al.* Ecogenomics of virophages and their giant virus hosts assessed through time series metagenomics. *Nature Communications* **8**, 858 (2017). URL <https://www.nature.com/articles/s41467-017-01086-2>.
- 605 [7] Mougari, S., Sahmi-Bounsiair, D., Levasseur, A., Colson, P. & La Scola, B. Virophages of giant viruses: An update at eleven. *Viruses* **11** (2019). URL <https://www.mdpi.com/1999-4915/11/8/733>.
- [8] Wu, Z. *et al.* Unveiling the unknown viral world in groundwater. *Nature Communications* **15**, 6788 (2024). URL <https://www.nature.com/articles/s41467-024-51230-y>.
- 610 [9] Hingamp, P. *et al.* Exploring nucleo-cytoplasmic large DNA viruses in tara oceans microbial metagenomes. *The ISME Journal* **7**, 1678–1695 (2013). URL <https://academic.oup.com/ismej/article/7/9/1678/7590279>.
- 615 [10] Cheng, R. *et al.* Virus diversity and interactions with hosts in deep-sea hydrothermal vents. *Microbiome* **10** (2022). URL <http://dx.doi.org/10.1186/s40168-022-01441-6>.
- [11] Villain, A., Gallot-Lavallée, L., Blanc, G. & Maumus, F. Giant viruses at the core of microscopic wars with global impacts. *Current Opinion in Virology* **17**, 130–137 (2016). URL <https://www.sciencedirect.com/science/article/pii/S1879625716300372>.
- 620

- [12] Irwin, N. A. T., Pittis, A. A., Richards, T. A. & Keeling, P. J. Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. *Nature Microbiology* **7**, 327–336 (2022). URL <https://www.nature.com/articles/s41564-021-01026-3>.
625
- [13] Husnik, F. & McCutcheon, J. P. Functional horizontal gene transfer from bacteria to eukaryotes. *Nature Reviews Microbiology* **16**, 67–79 (2018). URL <https://doi.org/10.1038/nrmicro.2017.137>.
- [14] Tesson, F. *et al.* Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nature Communications* **13**, 2561 (2022). URL <https://www.nature.com/articles/s41467-022-30269-9>.
630
- [15] Beavogui, A. *et al.* The defensome of complex bacterial communities. *Nature Communications* **15**, 2146 (2024). URL <https://www.nature.com/articles/s41467-024-46489-0>.
- [16] Kobayashi, I. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Research* **29**, 3742–3756 (2001). URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/29.18.3742>.
635
- [17] Koonin, E. V. & Makarova, K. S. Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20180087 (2019). URL <https://royalsocietypublishing.org/doi/10.1098/rstb.2018.0087>.
640
- [18] Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **577**, 327–336 (2020). URL <https://www.nature.com/articles/s41586-019-1894-8>.
645
- [19] Georjon, H. & Bernheim, A. The highly diverse antiphage defence systems of bacteria. *Nature Reviews Microbiology* **21**, 686–700 (2023). URL <https://www.nature.com/articles/s41579-023-00934-x>.
- [20] Tesson, F. *et al.* Exploring the diversity of anti-defense systems across prokaryotes, phages and mobile genetic elements. *Nucleic Acids Research* **53**, gkae1171 (2025). URL <https://academic.oup.com/nar/article/doi/10.1093/nar/gkae1171/7919512>.
650
- [21] Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432 (2013). URL <https://doi.org/10.1038/nature11723>.
655
- [22] Pawluk, A., Davidson, A. R. & Maxwell, K. L. Anti-CRISPR: discovery, mechanism and function. *Nature Reviews Microbiology* **16**, 12–17 (2018). URL <https://www.nature.com/articles/nrmicro.2017.120>.

- 660 [23] Zhang, Z. *et al.* Kiwa is a membrane-embedded defense supercomplex activated at phage attachment sites. *Cell* **188**, 5862–5877.e23 (2025). URL <https://linkinghub.elsevier.com/retrieve/pii/S0092867425007913>.
- 665 [24] Bandyopadhyay, P. K., Studier, F., Hamilton, D. L. & Yuan, R. Inhibition of the type I restriction-modification enzymes EcoB and EcoK by the gene 0.3 protein of bacteriophage T7. *Journal of Molecular Biology* **182**, 567–578 (1985). URL <https://linkinghub.elsevier.com/retrieve/pii/0022283685902426>.
- [25] Belogurov, A. A., Yussifov, T. N., Kotova, V. U. & Zavilgelsky, G. B. The novel gene(s) ARD of plasmid pKM101: alleviation of EcoK restriction. *Molecular and General Genetics MGG* **198**, 509–513 (1985). URL <http://link.springer.com/10.1007/BF00332948>.
- 670 [26] Wan, H. *et al.* Structural insights into the inhibition mechanism of bacterial toxin LsoA by bacteriophage antitoxin Dmd. *Molecular Microbiology* **101**, 757–769 (2016). URL <https://onlinelibrary.wiley.com/doi/10.1111/mmi.13420>.
- 675 [27] LeRoux, M. *et al.* The DarTG toxin-antitoxin system provides phage defence by ADP-ribosylating viral DNA. *Nature Microbiology* **7**, 1028–1040 (2022). URL <https://www.nature.com/articles/s41564-022-01153-5>.
- [28] Cao, X. *et al.* Phage anti-CBASS protein simultaneously sequesters cyclic trinucleotides and dinucleotides. *Molecular Cell* **84**, 375–385.e7 (2024). URL <https://linkinghub.elsevier.com/retrieve/pii/S1097276523009711>.
- 680 [29] Hobbs, S. J., Nomburg, J., Doudna, J. A. & Kranzusch, P. J. Animal and bacterial viruses share conserved mechanisms of immune evasion. *Cell* **187**, 5530–5539.e8 (2024). URL [https://www.cell.com/cell/fulltext/S0092-8674\(24\)00889-4](https://www.cell.com/cell/fulltext/S0092-8674(24)00889-4).
- 685 [30] Yirmiya, E. *et al.* Phages overcome bacterial immunity via diverse anti-defence proteins. *Nature* **625**, 352–359 (2024). URL <https://www.nature.com/articles/s41586-023-06869-w>.
- [31] Zeng, Z. *et al.* Base-modified nucleotides mediate immune signaling in bacteria. *Science* **388**, eads6055 (2025). URL <https://www.science.org/doi/10.1126/science.ads6055>.
- 690 [32] Zang, Z. *et al.* Chemical inhibition of a bacterial immune system. *Cell Host & Microbe* **34**, 263–277.e11 (2026). URL [https://www.cell.com/cell-host-microbe/fulltext/S1931-3128\(26\)00003-X](https://www.cell.com/cell-host-microbe/fulltext/S1931-3128(26)00003-X).
- [33] Bondy-Denomy, J. *et al.* Prophages mediate defense against phage infection through diverse mechanisms. *The ISME Journal* **10**, 2854–2866 (2016). URL <https://academic.oup.com/ismej/article/10/12/2854/7538083>.

- 695 [34] Rousset, F. *et al.* Phages and their satellites encode hotspots of antiviral systems. *Cell Host & Microbe* **30**, 740–753.e5 (2022). URL <https://linkinghub.elsevier.com/retrieve/pii/S1931312822001044>.
- [35] Kuang, X. *et al.* Bacteriophages mobilize bacterial defense systems via lateral transduction. *Science Advances* **12**, eadx5749 (2026). URL <https://www.science.org/doi/10.1126/sciadv.adx5749>.
700
- [36] Levasseur, A. *et al.* MIMIVIRE is a defence system in mimivirus that confers resistance to virophage. *Nature* **531**, 249–252 (2016). URL <https://www.nature.com/articles/nature17146>.
- [37] Jeudy, S. *et al.* The DNA methylation landscape of giant viruses. *Nature Communications* **11**, 2657 (2020). URL <https://www.nature.com/articles/s41467-020-16414-2>.
705
- [38] Mizuno, C. M., Rodriguez-Valera, F., Kimes, N. E. & Ghai, R. Expanding the marine virosphere using metagenomics. *PLoS Genetics* **9**, e1003987 (2013). URL <https://dx.plos.org/10.1371/journal.pgen.1003987>.
- 710 [39] Coutinho, F. H. *et al.* Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nature Communications* **8**, 15955 (2017). URL <https://www.nature.com/articles/ncomms15955>.
- [40] Camarillo-Guerrero, L. F., Almeida, A., Rangel-Pineros, G., Finn, R. D. & Lawley, T. D. Massive expansion of human gut bacteriophage diversity. *Cell* **184**, 1098–1109.e9 (2021). URL <https://linkinghub.elsevier.com/retrieve/pii/S0092867421000726>.
715
- [41] Graham, E. B. *et al.* A global atlas of soil viruses reveals unexplored biodiversity and potential biogeochemical impacts. *Nature Microbiology* **9**, 1873–1883 (2024). URL <https://www.nature.com/articles/s41564-024-01686-x>.
- 720 [42] Wirbel, J. *et al.* Long-read metagenomics reveals phage dynamics in the human gut microbiome. *Nature* **649**, 982–990 (2026). URL <https://www.nature.com/articles/s41586-025-09786-2>.
- [43] Getz, L. J. & Maxwell, K. L. Diverse antiphage defenses are widespread among prophages and mobile genetic elements. *Annual Reviews of Virology* **11**, 343–362 (2024). URL <https://www.annualreviews.org/content/journals/10.1146/annurev-virology-100422-125123>.
725
- [44] Moura de Sousa, J. A., Pfeifer, E., Touchon, M. & Rocha, E. P. C. Causes and consequences of bacteriophage diversification via genetic exchanges across lifestyles and bacterial taxa. *Molecular Biology and Evolution* **38**, 2497–2512 (2021). URL <https://academic.oup.com/mbe/article/38/6/2497/6133237>.
730

- [45] Maestri, A. *et al.* The bacterial defense system MADS interacts with CRISPR-Cas to limit phage infection and escape. *Cell Host Microbe* **32**, 1412–1426.e11 (2024). URL <https://www.sciencedirect.com/science/article/pii/S1931312824002646>.
- 735 [46] Samuels, A. N., Roggiani, M., Zhu, J., Goulian, M. & Kohli, R. M. The SOS response mediates sustained colonization of the mammalian gut. *Infection and Immunity* **87**, 10.1128/iai.00711–18 (2019). URL <https://journals.asm.org/doi/abs/10.1128/iai.00711-18>.
- 740 [47] Azulay, G. *et al.* A dual-function phage regulator controls the response of cohabiting phage elements via regulation of the bacterial SOS response. *Cell Reports* **39**, 110723 (2022). URL <https://www.sciencedirect.com/science/article/pii/S2211124722004843>.
- 745 [48] Jancheva, M. *et al.* A phage-selective trigger hints at an SOS-independent mechanism of prophage induction by oxidative stress. *Chem. Sci.* **17**, 1151–1157 (2026). URL <https://pubs.rsc.org/sc/article/17/2/1151/910863/>.
- [49] Vander Zee, A., Natarajan, A. & Maxwell, K. L. Timing is everything: regulation of bacterial defences and phage counter-defences. *Current Opinion in Microbiology* **91**, 102750 (2026). URL <https://www.sciencedirect.com/science/article/pii/S1369527426000445>.
- 750 [50] Maguin, P., Varble, A., Modell, J. W. & Marraffini, L. A. Cleavage of viral DNA by restriction endonucleases stimulates the type II CRISPR-Cas immune response. *Mol. Cell* **82**, 907–919.e7 (2022). URL [https://www.cell.com/molecular-cell/fulltext/S1097-2765\(22\)00055-7](https://www.cell.com/molecular-cell/fulltext/S1097-2765(22)00055-7).
- 755 [51] Workman, R. E. *et al.* Anti-CRISPR proteins trigger a burst of CRISPR-Cas9 expression that enhances phage defense. *Cell Reports* **43**, 113849 (2024). URL <https://www.sciencedirect.com/science/article/pii/S2211124724001773>.
- [52] Zhang, H. *et al.* The two-component nuclease-active KELShedu system confers broad antiphage activity via abortive infection. *Science Advances* **11**, eadv4747 (2025). URL <https://www.science.org/doi/10.1126/sciadv.adv4747>.
- 760 [53] Zhou, R. *et al.* Structural and mechanistic insights into the dual-nuclease defense protein Upx as an anti-phage system. *Nature Communications* **17**, 3692 (2026). URL <https://www.nature.com/articles/s41467-026-70435-x>.
- 765 [54] Palka, C., Fishman, C. B., Bhattarai-Kline, S., Myers, S. A. & Shipman, S. Retron reverse transcriptase termination and phage defense are dependent on host RNase H1. *Nucleic Acids Research* **50**, 3490–3504 (2022). URL <https://academic.oup.com/nar/article/50/6/3490/6549382>.
- [55] Puteikiene, R. *et al.* Viral SSB-bound ssDNA activates the bacterial anti-phage defense system DARN (2026). URL <http://biorxiv.org/lookup/doi/10.64898/>

2026.02.22.705581.

- 770 [56] Chambers, L. R. *et al.* A eukaryotic-like ubiquitination system in bacterial antiviral defence. *Nature* **631**, 843–849 (2024). URL <https://www.nature.com/articles/s41586-024-07730-4>.
- [57] Song, X.-Y. *et al.* Bacterial reverse transcriptase synthesizes long poly(A)-rich cDNA for antiphage defense. *Science* **388**, eads4639 (2025). URL <https://www.science.org/doi/10.1126/science.ads4639>.
775
- [58] Pyle, J. D., Lund, S. R., O’Toole, K. H. & Saleh, L. Virus-encoded glycosyl-transferases hypermodify DNA with diverse glycans. *Cell Reports* **43**, 114631 (2024). URL <https://linkinghub.elsevier.com/retrieve/pii/S2211124724009811>.
- [59] DeWeirdt, P. C., Mahoney, E. M. & Laub, M. T. DefensePredictor: A machine learning model to discover prokaryotic immune systems. *Science* **392**, eadv7924 (2026). URL <https://www.science.org/doi/10.1126/science.adv7924>.
780
- [60] Kimchi, O., Meir, Y. & Wingreen, N. S. Bacterial defense and phage counter-defense lead to coexistence in a modeled ecosystem. *Proceedings of the National Academy of Sciences* **121**, e2414229121 (2024). URL <https://www.pnas.org/doi/abs/10.1073/pnas.2414229121>.
785
- [61] Ispolatov, Y., Lekontseva, A. & Severinov, K. The number of immune defence and counter-defence systems sustained in the arms race between prokaryotes and viruses. *Scientific Reports* **15**, 28083 (2025). URL <https://www.nature.com/articles/s41598-025-11335-w>.
- 790 [62] Srikant, S., Guegler, C. K. & Laub, M. T. The evolution of a counter-defense mechanism in a virus constrains its host range. *eLife* **11**, e79549 (2022). URL <https://doi.org/10.7554/eLife.79549>.
- [63] Meaden, S., Westra, E. R. & Fineran, P. C. Phage defence-system abundances vary across environments and increase with viral density. *Philosophical Transactions of the Royal Society B: Biological Sciences* **380**, 20240069 (2025). URL <https://doi.org/10.1098/rstb.2024.0069>.
795
- [64] Liu, C. *et al.* Response of soil phage communities and prokaryote–phage interactions to long-term drought. *Environmental Science & Technology* **59**, 3054–3066 (2025). URL <https://pubs.acs.org/doi/10.1021/acs.est.4c08448>.
- 800 [65] Zablocki, O., Adriaenssens, E. M. & Cowan, D. Diversity and ecology of viruses in hyperarid desert soils. *Applied and Environmental Microbiology* **82**, 770–777 (2016). URL <https://journals.asm.org/doi/10.1128/AEM.02651-15>.
- [66] Song, H. *et al.* Metagenomic analysis reveals rapid development of soil biota on fresh volcanic ash. *Scientific Reports* **10**, 21419 (2020). URL <https://www.nature.com/articles/s41598-020-78413-z>.
805

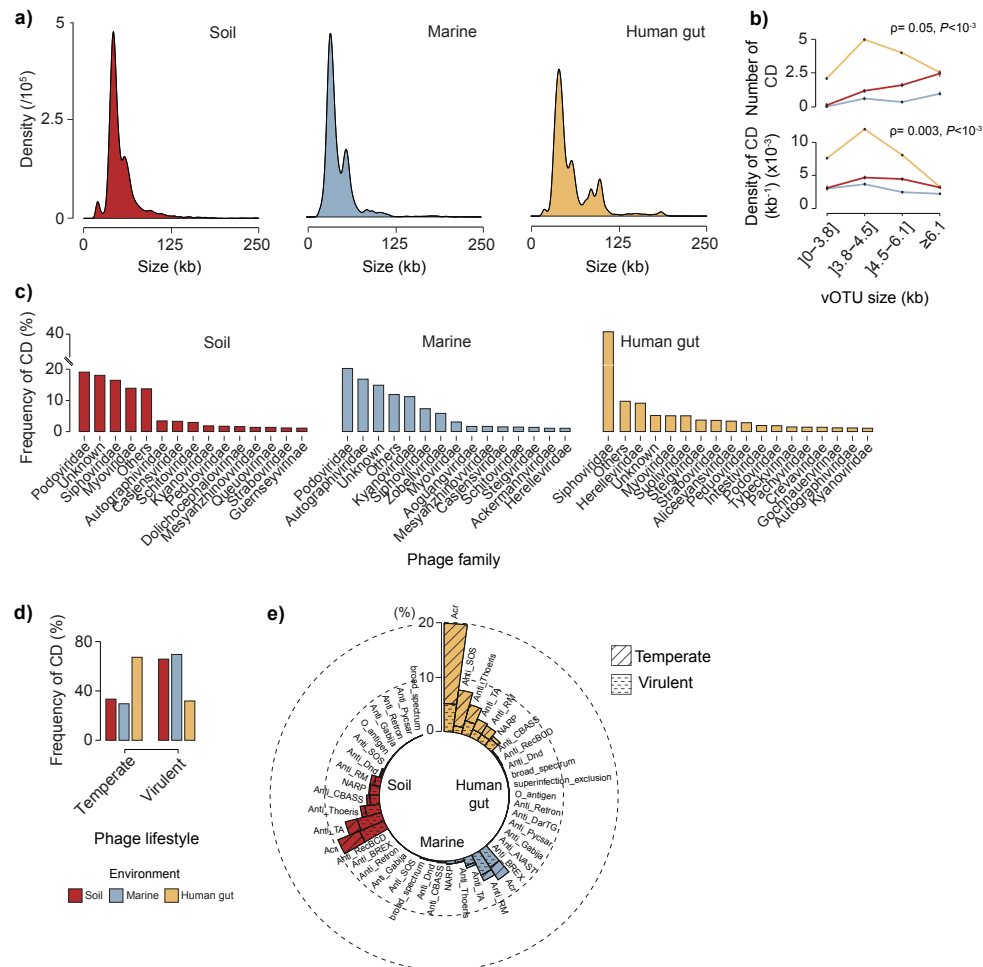
- [67] Yao, J. *et al.* Phages-bacteria interactions underlying the dynamics of polyhydroxyalkanoate-producing mixed microbial cultures via meta-omics study. *mSystems* **10**, e00200–25 (2025). URL <https://journals.asm.org/doi/abs/10.1128/msystems.00200-25>.
- 810 [68] Wang, G. *et al.* Temporal transcriptomics uncover dynamic interactions between pathogenic *Escherichia coli* and phage vB_Eco_K1B4. *Frontiers in Microbiology* **16**, 1668341 (2025). URL <https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2025.1668341>.
- 815 [69] Ghiotto, G. *et al.* Single nucleotide variants drive evolutionary phage-host arms race in anaerobic carbon dioxide-converting microbiome. *Nature Communications* **17**, 6927 (2026). URL <http://dx.doi.org/10.1038/s41467-026-73084-2>.
- [70] Liu, Y. *et al.* Leveraging defense system modularity to discover anti-phage systems. *bioRxiv* (2025). URL <https://www.biorxiv.org/content/early/2025/12/08/2025.12.08.692795>.
- 820 [71] Oliveira, P. H., Touchon, M. & Rocha, E. P. C. Regulation of genetic flux between bacteria by restriction–modification systems. *Proceedings of the National Academy of Sciences* **113**, 5658–5663 (2016). URL <https://pnas.org/doi/full/10.1073/pnas.1603257113>.
- 825 [72] Kavagutti, V. S., Beavogui, A., Wiart, N., Wincker, P. & Oliveira, P. H. Defensomes, counter-defensomes, and the remodeling of microbial communities. *PNAS Nexus* **5**, pgag073 (2026). URL <https://academic.oup.com/pnasnexus/article/doi/10.1093/pnasnexus/pgag073/8525634>.
- [73] Braga, L. P. P. & Schumacher, R. I. Awakening the dormant virome in the rhizosphere. *Molecular Ecology* **32**, 2985–2999 (2023). URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.16893>.
- 830 [74] Santos-Medellín, C., Blazewicz, S. J., Pett-Ridge, J., Firestone, M. K. & Emerson, J. B. Viral but not bacterial community successional patterns reflect extreme turnover shortly after rewetting dry soils. *Nature Ecology & Evolution* **7**, 1809–1822 (2023). URL <https://www.nature.com/articles/s41559-023-02207-5>.
- 835 [75] Bi, L., He, J.-Z. & Hu, H.-W. Total metagenomes outperform viromes in recovering viral diversity from sulfuric soils. *ISME Communications* **4**, ycae017 (2024). URL <https://doi.org/10.1093/ismeco/ycae017>.
- [76] Ma, B. *et al.* Biogeographic patterns and drivers of soil viromes. *Nature Ecology & Evolution* **8**, 717–728 (2024). URL <https://www.nature.com/articles/s41559-024-02347-2>.
- 840 [77] Brum, J. R. *et al.* Patterns and ecological drivers of ocean viral communities. *Science* **348**, 1261498 (2015). URL <https://www.science.org/doi/10.1126/science.1261498>.

- 845 [78] Gregory, A. C. *et al.* Marine DNA viral macro- and microdiversity from pole to pole. *Cell* **177**, 1109–1123.e14 (2019). URL <https://www.sciencedirect.com/science/article/pii/S0092867419303411>.
- [79] Michniewski, S. *et al.* A new family of "megaphages" abundant in the marine environment. *ISME Communications* **1**, 58 (2021). URL <https://www.nature.com/articles/s43705-021-00064-6>.
- 850 [80] Shang, J., Peng, C., Liao, H., Tang, X. & Sun, Y. Phabox: a web server for identifying and characterizing phage contigs in metagenomic data. *Bioinformatics Advances* **3**, vbad101 (2023). URL <https://doi.org/10.1093/bioadv/vbad101>.
- [81] Nayfach, S. *et al.* CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2021). URL <https://doi.org/10.1038/s41587-020-00774-7>.
- 855 [82] Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017). URL <https://doi.org/10.1038/ismej.2017.126>.
- 860 [83] Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014). URL <https://doi.org/10.1093/bioinformatics/btu153>.
- [84] Paez-Espino, D. *et al.* Diversity, evolution, and classification of virophages uncovered through global metagenomics. *Microbiome* **7**, 157 (2019). URL <https://doi.org/10.1186/s40168-019-0768-5>.
- 865 [85] Aylward, F. O., Moniruzzaman, M., Ha, A. D. & Koonin, E. V. A phylogenomic framework for charting the diversity and evolution of giant viruses. *PLOS Biology* **19**, 1–18 (2021). URL <https://doi.org/10.1371/journal.pbio.3001430>.
- [86] Gaïa, M. *et al.* Mirusviruses link herpesviruses to giant viruses. *Nature* **616**, 783–789 (2023). URL <https://doi.org/10.1038/s41586-023-05962-4>.
- 870 [87] Chowdhury, P. R. *et al.* The transcriptional response of soil bacteria to long-term warming and short-term seasonal fluctuations in a terrestrial forest. *Frontiers in Microbiology* **12** (2021). URL <https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2021.666558>.
- 875 [88] Wilson, R. M. *et al.* Stability of peatland carbon to rising temperatures. *Nature Communications* **7**, 13723 (2016). URL <https://doi.org/10.1038/ncomms13723>.
- [89] Sorensen, P. O. *et al.* Multi-omics reveals nitrogen dynamics associated with soil microbial blooms during snowmelt. *Nature Microbiology* **11**, 359–374 (2026). URL <https://doi.org/10.1038/s41564-025-02213-2>.

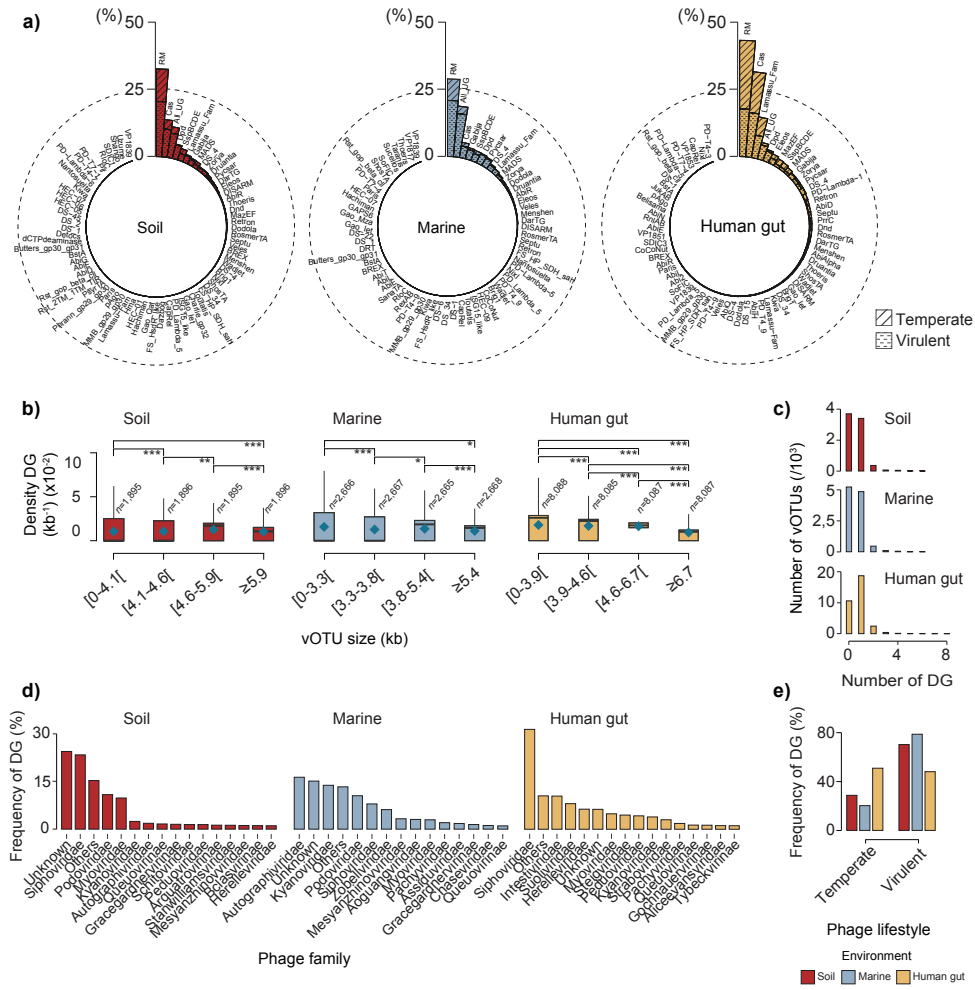
- 880 [90] Peng, Y. *et al.* A metagenome-assembled genome inventory for children reveals early-life gut bacteriome and virome dynamics. *Cell Host & Microbe* **32**, 2212–2230.e8 (2024). URL <https://doi.org/10.1016/j.chom.2024.10.017>.
- [91] Heintz-Buschart, A. *et al.* Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. *Nature Microbiology* **2** (2016). URL <https://doi.org/10.1038/nmicrobiol.2016.180>.
- 885 [92] Asnicar, F. *et al.* Studying vertical microbiome transmission from mothers to infants by strain-level metagenomic profiling. *mSystems* **2**, e00164–16 (2017). URL <https://doi.org/10.1128/msystems.00164-16>.
- [93] Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11** (2010). URL <http://dx.doi.org/10.1186/1471-2105-11-119>.
890
- [94] Menzel, P., Ng, K. L. & Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications* **7**, 11257 (2016). URL <https://doi.org/10.1038/ncomms11257>.
- [95] Bolduc, B. *et al.* Machine learning enables scalable and systematic hierarchical virus taxonomy. *Nature Biotechnology* (2025). URL <https://doi.org/10.1038/s41587-025-02946-9>.
895
- [96] Pons, J. C. *et al.* Vpf-class: taxonomic assignment and host prediction of uncultivated viruses based on viral protein families. *Bioinformatics* **37**, 1805–1813 (2021). URL <https://doi.org/10.1093/bioinformatics/btab026>.
- 900 [97] Roux, S. *et al.* iPHoP: An integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLOS Biology* **21**, 1–26 (2023). URL <https://doi.org/10.1371/journal.pbio.3002083>.
- [98] Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology* **35**, 1026–1028
905 (2017). URL <https://www.nature.com/articles/nbt.3988>.
- [99] Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods* **18**, 366–368 (2021). URL <https://www.nature.com/articles/s41592-021-01101-x>.
- [100] Yan, Y., Zheng, J., Zhang, X. & Yin, Y. dbAPIS: a database of anti-prokaryotic immune system genes. *Nucleic Acids Research* **52**, D419–D425 (2024). URL <https://academic.oup.com/nar/article/52/D1/D419/7331021>.
910
- [101] Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018). URL <https://doi.org/10.1093/bioinformatics/bty560>.

- 915 [102] Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering
of ribosomal rnas in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217
(2012). URL <https://doi.org/10.1093/bioinformatics/bts611>.
- [103] Bushnell, B. BBMap: A fast, accurate, splice-aware aligner. *Conference: 9th
Annual Genomics of Energy Environment Meeting, Walnut Creek, CA, March
17-20* (2014). URL <https://sourceforge.net/projects/bbmap/>.
920
- [104] Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic
rna-seq quantification. *Nature Biotechnology* **34**, 525–527 (2016). URL <https://doi.org/10.1038/nbt.3519>.
- [105] Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma-
tics* **25**, 2078–2079 (2009). URL [https://academic.oup.com/bioinformatics/
article/25/16/2078/204688](https://academic.oup.com/bioinformatics/article/25/16/2078/204688).
925
- [106] Jones, P. *et al.* InterProScan 5: genome-scale protein function classifica-
tion. *Bioinformatics* **30**, 1236–1240 (2014). URL [https://doi.org/10.1093/
bioinformatics/btu031](https://doi.org/10.1093/bioinformatics/btu031).

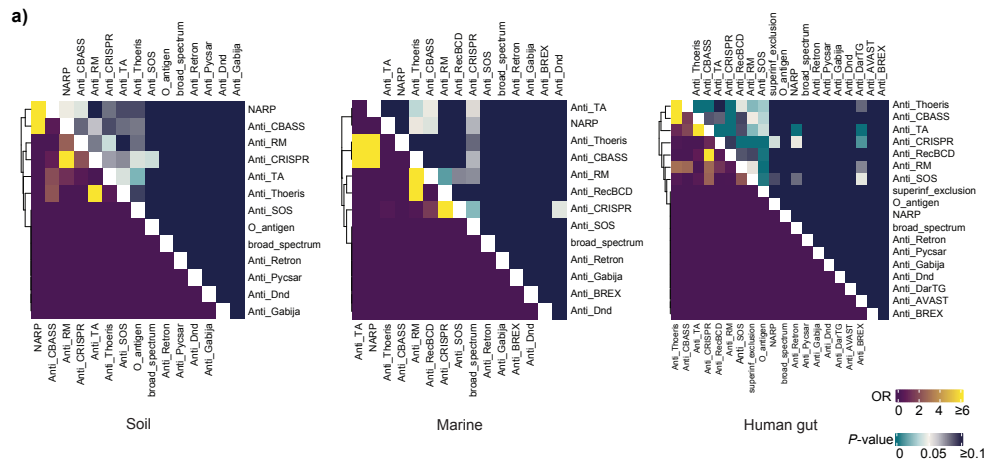
930 **Supplementary Information**



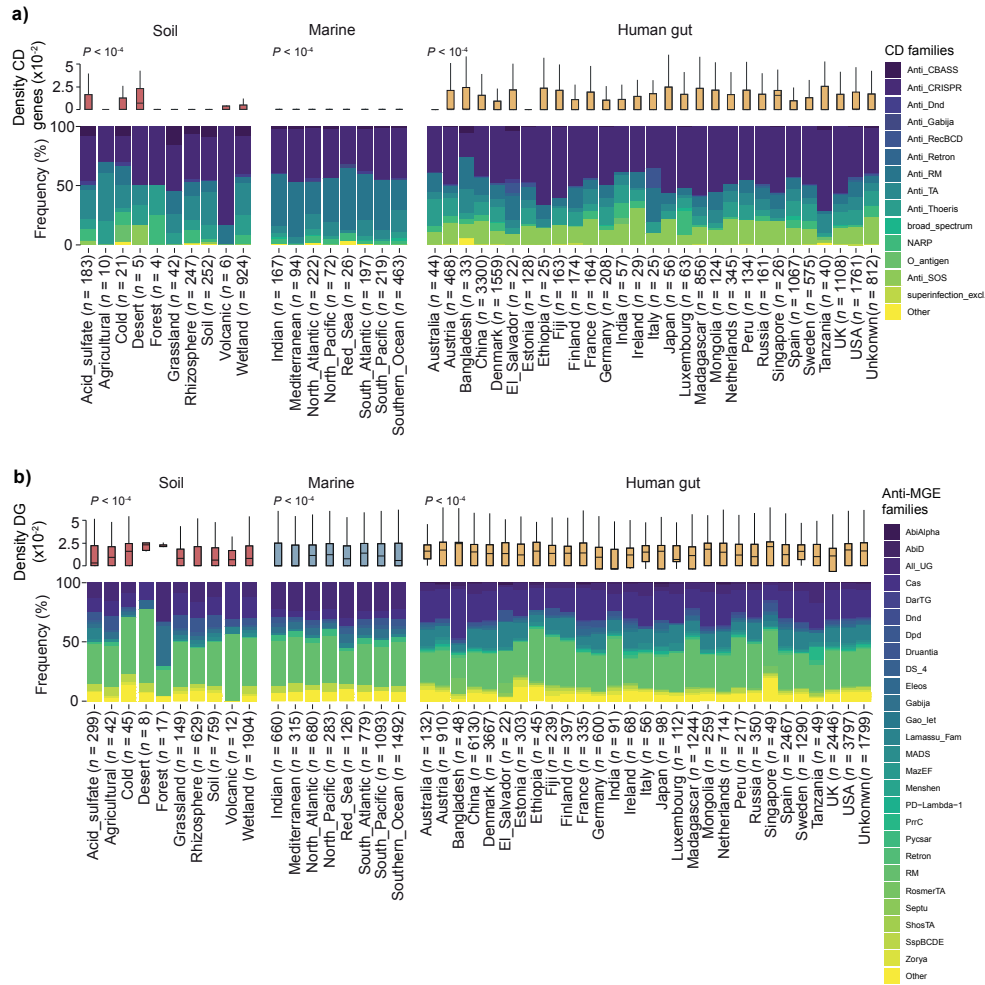
Suppl. Fig. 1 Phage size distribution and ecological partitioning of viral counter-defenses. **a)** Size distribution patterns of the phage dataset. **b)** Changes in the number and density of counter-defense (CD) systems per vOTU (per genome and per kb) as a function of vOTU genome size (kb) across biomes. Error bars indicate standard deviations of the mean. Correlations were assessed by a two-sided Spearman's rank test. **c)** Distribution of CD systems across phage families and environments, shown as percentages. **d)** Distribution of CD systems across phage lifestyles and environments, expressed as percentages. **e)** Proportion of soil, marine, and human gut vOTUs encoding each CD gene family.



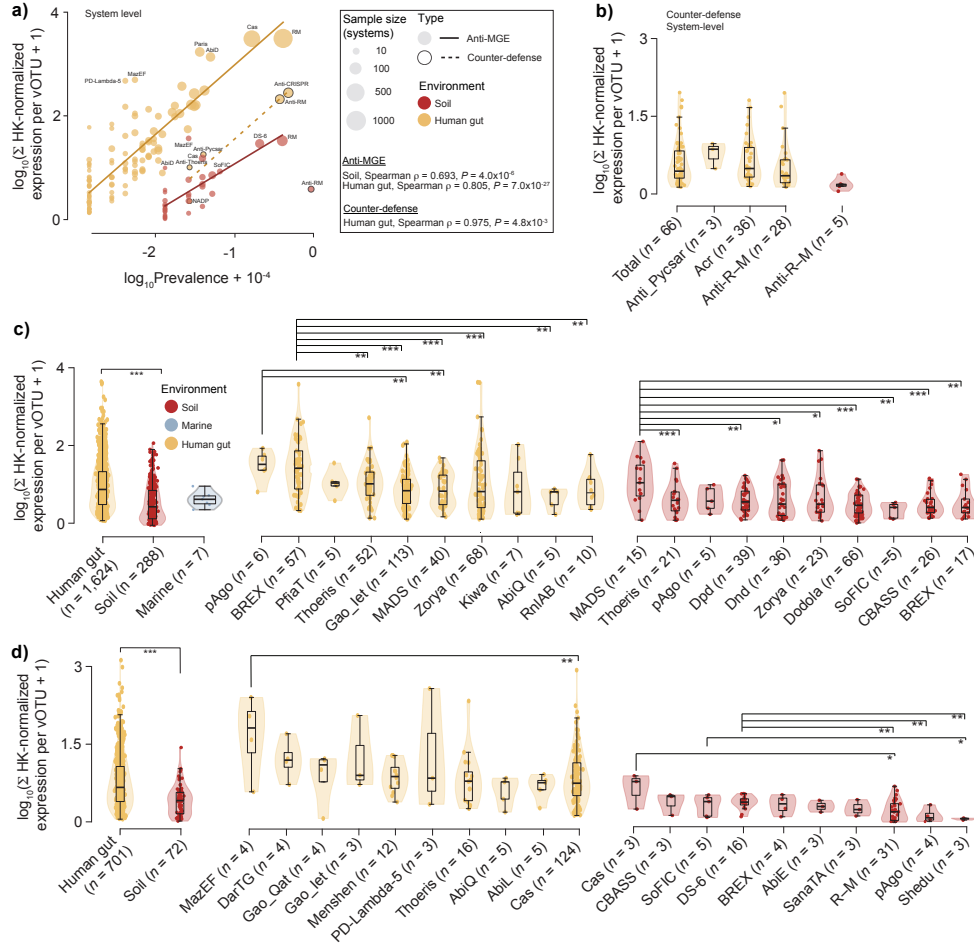
Suppl. Fig. 2 Distribution of anti-MGE defense genes (DG) across viral genomes and environments. **a)** Percentage of soil, marine and human gut vOTUs encoding each family of anti-MGE DG. Data are shown for both temperate and virulent phages. **b)** Relationship between DG density (per kb) and vOTU genome size (kb) across environments. Statistical comparisons were performed using two-sided Wilcoxon rank-sum tests (* $P < 0.05$, ** $P < 10^{-2}$, *** $P < 10^{-3}$). **c)** Distribution of DG counts (per vOTU) across environments. **d)** Proportion of DG across phage families within each environment. **e)** Distribution of DG across phage lifestyles and environments, expressed as percentages.



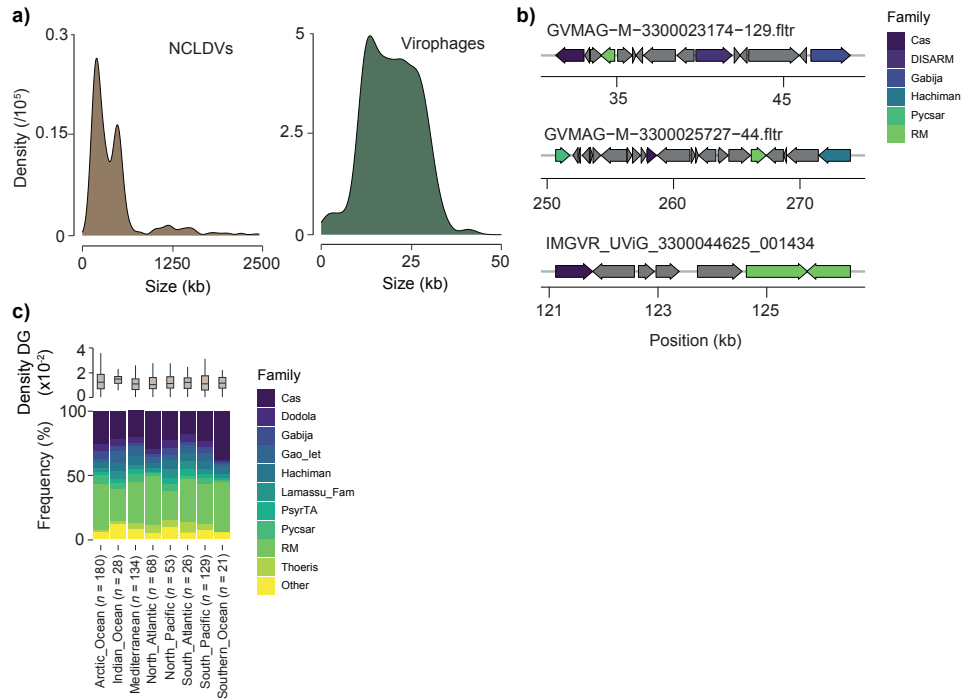
Suppl. Fig. 3 Counter-defense gene colocalization across soil, marine, and human gut vOTUs. **a)** Odds ratio (OR) of co-localization (bottom heatmaps) and associated two-sided Fisher's exact test P values (upper heatmaps) among counter-defense gene families, restricted to families with a co-localization frequency $\geq 1\%$ across soil, marine, and human gut vOTUs.



Suppl. Fig. 4 Interplay between phage biogeography, counter-defense (CD) gene abundance, and anti-MGE gene abundance. **a)** Density and frequency of CD genes across distinct ecological (soil, marine) and geographical (human gut) contexts. **b)** Density and frequency of anti-MGE genes across the same biomes. Boxplots depict the interquartile range (25th - 75th percentiles), with the median indicated by a central black line. Whiskers extend to 1.5x the interquartile range, and outliers were omitted for improved visualization. Statistical significance was assessed using a Kruskal-Wallis test. The number of vOTUs included in each category is provided in parentheses.



Suppl. Fig. 5 Interplay between counter-defense repertoires and their transcriptome. **a)** Relationship between immune system prevalence and sum of housekeeping (HK)-normalized expression per vOTU across complete anti-MGE (solid trend lines) and counter-defense (dashed trend lines) system families in soil (red) and human gut (yellow) samples. Each point represents one immune family sized by the number of contributing systems. Spearman correlation coefficients and P values are shown for each immune class and environment. Marine was devoid of complete systems passing filters. **b)** Summed HK-normalized expression per vOTU of counter-defense systems carried by soil (red), and human gut (yellow) vOTUs. **c)** Summed HK-normalized expression per vOTU of anti-MGE genes carried by soil (red), marine (blue), and human gut (yellow) vOTUs (left), or split by the top 10 most-expressed families for soil and human gut vOTUs. Not enough statistical power was obtained to represent marine samples when split by family. **d)** Summed HK-normalized expression per vOTU of anti-MGE systems carried by soil (red), and human gut (yellow) vOTUs (left), or split by the top 10 most-expressed families for soil and human gut vOTUs. Statistical comparisons were performed using two-sided Wilcoxon rank-sum tests (* $P < 0.05$, ** $P < 10^{-2}$, *** $P < 10^{-3}$).



Suppl. Fig. 6 Clustering and biogeographical distribution of immune modules in NCLDVs. **a)** Size distribution patterns of the virophage and NCLDV dataset. **b)** Representative examples of immune module clustering in NCLDVs. **c)** Density and frequency of anti-MGE and counter-defense gene families in NCLDVs recovered from distinct marine regions. Boxplots depict the interquartile range (25th - 75th percentiles), with the median indicated by a central black line. Whiskers extend to 1.5x the interquartile range, and outliers were omitted for improved visualization. Statistical significance was assessed using a Kruskal-Wallis test. The number of NCLDV genomes included in each category is provided in parentheses.