nature communications

Article

https://doi.org/10.1038/s41467-026-71095-7

# Unveiling the biodiversity of large DNA viruses in intertidal mudflats via metagenomics

Received: 8 September 2025

Accepted: 6 March 2026

Published online: 24 March 2026

Check for updates

Mengzhi Ji[1], Yan Li[1], Mengqi Wang[1], Xia Liu[1], Xiaofan Gong[1] & Qichao Tu[1,2,3]

Large DNA viruses (LDVs) are unique members of the Earth's virosphere, remarkable for their extra-large genome sizes and broad metabolic potential. However, our knowledge of this viral group remains very limited, particularly in complex dynamic habitats. In this study, 237 metagenome-assembled LDV genomes are comprehensively recovered from intertidal mudflats using multiple sampling and sequencing strategies totaling 5.3 TB data. A phylogenetically distinct subgroup within *Imitervirales* is identified, showing broad associations with multiple eukaryotic lineages. Certain LDV populations can persist locally and exhibit significant genomic variations potentially driven by dynamic intertides. Ecological patterns are observed at both community and genetic levels, with giant viruses showing steeper community turnover but weaker nucleotide diversity variations than large phages. Moreover, LDVs exhibit similar macroecological patterns to their potential hosts, which substantially shape LDV community assembly. The intertidal LDVs encode diverse functional genes, most of which remain uncharacterized, with a 27.32% improvement for unknown phage genes using a protein language model. Although giant viruses and large phages share comparable functional gene composition, they exhibit distinct preferences for specific metabolic pathways, especially those associated with carbon and nitrogen cycling. This study broadens our understanding of the biodiversity and ecology of LDVs in the understudied intertidal ecosystems.

Viruses are ubiquitous in the Earth's biosphere, representing the most diverse and abundant biological entities[1]. In natural ecosystems, viruses are pivotal regulators of food web dynamics and nutrient cycling[2,3]. With the widespread application of meta-omics approaches in viral ecology, the diversity and functional potential of viruses have been actively investigated over the past decade, including their roles in shaping host community structure and driving various biogeochemical processes[4]. Despite ongoing technical and bioinformatic challenges, viromic and metagenomic analyses have emerged as indispensable methods for exploring viral biodiversity, ecological patterns, and their biotic and abiotic drivers[1,3,5]. More recently, enhanced binning and genome extension tools aimed to improve viral genome quality have been developed[6–8], providing important avenues for the study of large DNA viruses (LDVs).

LDVs are double-stranded DNA viruses characterized by remarkably large genomes. LDVs belonging to the phylum *Nucleocytoviricota*, also known as giant viruses (GVs) and having genome sizes up to 2.5 Mb, represent a major group infecting eukaryotic hosts[9,10].

[1]Institute of Marine Science and Technology, Shandong University, Qingdao, China. [2]Qingdao Key Laboratory of Ocean Carbon Sequestration and Negative Emission Technology, Shandong University, Qingdao, China. [3]Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, China. e-mail: tuqichao@sdu.edu.cn

According to a unified phylogenomic framework[11], known GVs are classified into six orders, including *Asfuvirales*, *Algalvirales*, *Chitovirales*, *Imitervirales*, *Pandoravirales*, and *Pimascovirales*. The members of *Asfuvirales*, *Chitovirales*, and *Pimascovirales* primarily infect protists and metazoan, while those belonging to *Algalvirales*, *Imitervirales*, and *Pandoravirales* mainly infect algae and heterotrophic protists[12]. Recent progress in recovering and analyzing GVs from shotgun metagenomes has substantially enhanced our knowledge of this unique life form across various habitats[13–15]. In particular, some GVs have been found to reprogram host cells by encoding abundant metabolic genes, such as those involved in the tricarboxylic acid cycle and glycolysis[15–17]. In contrast to GVs, another group of LDVs comprises jumbo phages infecting a broad range of bacterial lineages, typically defined as members of the phylum *Uroviricota* with genome sizes over 200 kb[18,19]. Jumbo phages have been extensively studied for their complex functional potential, such as the encoding of auxiliary metabolic genes (AMGs) and sophisticated anti-CRISPR defense systems[20,21]. Notably, most of the abovementioned studies have focused on LDVs from soil and marine environments, while the general characteristics of LDVs in complex dynamic habitats such as land-ocean interfaces warrant further investigation.

The intertidal zones are located at the interface of terrestrial and marine ecosystems, and represent one of the most ecologically complex habitats on Earth[22]. These coastal ecotones are shaped by a combination of geological, hydrological, biological, and physicochemical factors, playing vital roles in global biogeochemical cycles[22]. Tidal oscillations drive rapid shifts between aerobic and anoxic/suboxic states in intertidal mudflats, imposing substantial stress on organisms therein and contributing to their unique biodiversity[23]. Consequently, intertidal mudflats harbor intricate food webs composed of diverse prokaryotes, microeukaryotes, and multicellular organisms[23,24]. Among these, autotrophic protists act as primary producers forming the base of food webs, while heterotrophic protists act as key predators by consuming prokaryotes and phytoplankton and thus transfer energy and nutrients to higher trophic levels[25]. LDVs are known to infect many of these organisms, especially protists, and thus exert top-down controlling effects on host population dynamics and regulate ecosystem functioning and biogeochemical cycles[9,13,19]. Yet, the biodiversity of LDVs and their interactions with hosts remain poorly resolved in such complex ecosystems, limiting further exploration of their ecological roles.

In this study, we strived to recover LDV genomes from mudflat intertidal sediments via a large-scale data mining effort, aiming to address the following questions associated with LDV biodiversity, ecological patterns, and functional potential: (i) How are intertidal LDVs distributed across different scales and dimensions? (ii) How do different ecological processes and putative hosts shape intertidal LDV communities? (iii) How diverse are the metabolic capabilities encoded by intertidal LDVs? A total of 237 intertidal LDVs are recovered, encompassing a broad phylogenetic spectrum. The results reveal significant biogeographic patterns of LDVs and highlight their diverse metabolic capabilities involved in various elemental biogeochemical cycling processes. The findings offer valuable insights into the biodiversity, ecological drivers, and functional potential of intertidal LDVs within an integrated ecological framework.

## Results and discussion

### Metagenomic recovery of LDVs in intertidal mudflats

A total of 199 intertidal sediment metagenomes were recruited to recover LDVs, encompassing multiple sampling and sequencing strategies and approximately 5.3 TB of metagenomic data (Fig. 1a and Supplementary Data 1). By employing a comprehensive viral recovery pipeline (Fig. 1b), we reconstructed 171 viral metagenome-assembled genomes (vMAGs) whose genome sizes exceeded 200 kb and 66 vMAGs whose genome sizes ranged from 100 to 200 kb, as judged by the presence of multiple LDV markers[16]. Of these, 149 vMAGs were determined to be *Nucleocytoviricota*, while 88 vMAGs were *Uroviricota* (Fig. 1b). To minimize contamination, contigs containing substantial host genes were removed from vMAGs and those with ViralRecall scores < 0 were excluded from *Nucleocytoviricota* vMAGs[16,26]. Nevertheless, the lack of highly conserved markers in phages makes it challenging to eliminate contaminated contigs, potentially leading to overestimated genome sizes[18]. Given such limitations, *Uroviricota* vMAGs were considered as large phages, defined as phages with genome sizes significantly larger than typical phages[20].

Of the recovered 237 vMAGs, 15 were complete (circular with terminal repeats) and 53 were near-complete (>90% completeness) (Fig. 1b). This is a significant improvement in recovering high-quality LDV genomes, especially comparing to our previous study encompassing 96 shotgun metagenomes, in which only fragments were generated[27]. The genome sizes of *Nucleocytoviricota* vMAGs ranged from 103 kb to 645 kb with an average size of 234 kb, whereas those of *Uroviricota* vMAGs ranged from 200 kb to 452 kb with an average size of 241 kb (Fig. 1c). For most vMAGs, the more complete genomes were found with more viral hallmarks and lower host contamination (Fig. 1c).

At the species level, these 237 LDVs were clustered into 118 representative vMAGs at 95% average nucleotide identity (ANI) (Supplementary Data 2). To gain insights into the genomic diversity of the recovered *Nucleocytoviricota* vMAGs (i.e., GVs), their core marker genes were identified and analyzed[11,16]. All 71 GVs had at least one marker gene and exhibited high ViralRecall scores (Figure. S1). According to a previous criteria[16], 51 GVs containing ≥ 4 marker genes were identified as high-confidence (Supplementary Data 2). The most well-recognized markers[28], including major capsid protein (MCP) and DNA polymerase family B (PolB), were respectively detected in 49 and 40 GVs (Figure. S1).

Of the 47 *Uroviricota* vMAGs (i.e., large phages), 16 were of high-confidence by comprising 1-5 contigs[18]. These large phages encoded an average of 14 tRNA genes, with the highest achieving 64 in a complete vMAG designated as "ZJ_NB_2021_05_15_bin_21" (Figure. S2). Such high tRNA gene copies in vMAGs may facilitate the translational efficiency of large phage genes, potentially reducing their reliance on hosts[29]. Strikingly, a complete large phage with a genome size of 228.8 kb was consistently recovered from all datasets in Qingdao spanning various time points and depths (Figure. S3a). Both ANI and average amino acid identity (AAI) analyses showed that the variants of this large phage from different datasets shared more than 99.9% identity, demonstrating remarkable genome stability (Figure. S3a, b). The coverage and single nucleotide variation (SNV) of this large phage exhibited synchronized fluctuations over time and depth (Figure. S3c, d). Genome annotation suggested that this large phage harbored multiple lysis modules and lacked lysogenic elements, with a large number of genes encoding functions related to DNA metabolism and transcription regulation (Figure. S4). Comparative genomic analyses supported its stability, though exceptions were observed (Figure. S4). For example, a potential genomic inversion was identified in samples collected at a depth of 100 cm (Figure. S4). Overall, these findings revealed both the genomic conservation and variability of this large phage, reflecting its potential living strategies in the dynamic intertides.

### Phylogenetic diversity of intertidal LDVs

Using a highly conserved set of marker genes[11], phylogenetic trees were constructed for the recovered intertidal GVs and multiple reference genomes to infer their taxonomic and evolutionary diversity (Fig. 2 and Figure. S5). In total, 64 intertidal GVs (excluding 7 GVs with insufficient markers) and 251 reference GVs were clustered into five clearly distinguished phylogenetic clades affiliated with different taxonomic orders, including *Asfuvirales*, *Pimascovirales*, *Pandoravirales*, *Algavirales*, and *Imitervirales* (Fig. 2 and Supplementary Data 3).

https://doi.org/10.1038/s41467-026-71095-7

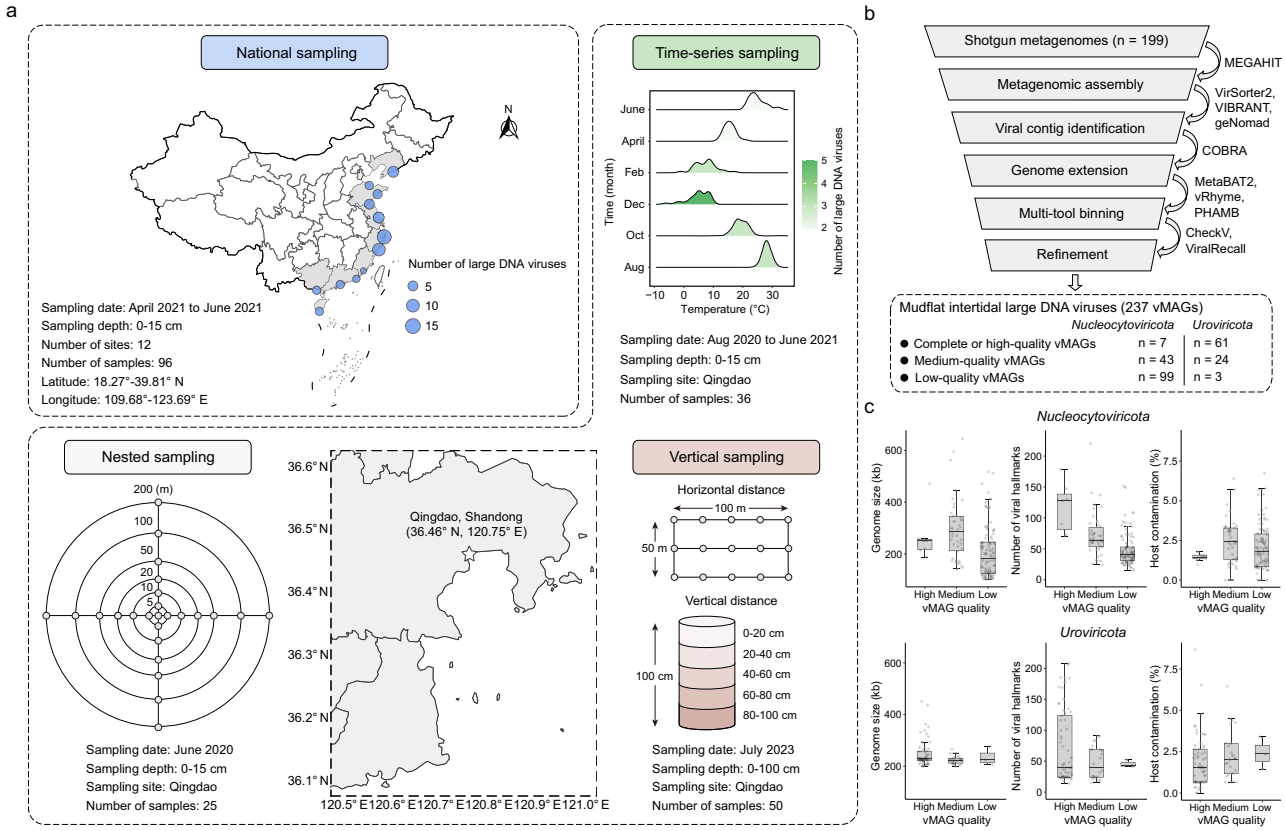


**Fig. 1 | An overview of large DNA viruses (LDVs) recovered from mudflat intertidal sediments. a** Different sampling schemes of intertidal sediments used in this study, including national sampling, time-series sampling, nested local sampling, and vertical sampling. The geographic locations of samples at the national scale are marked with blue circles, with the size of each circle representing the number of recovered LDVs. The time-series, nested, and vertical sampling schemes were conducted in Qingdao, indicated by the white star on the regional map. The ridgeline plot displays the temperature distribution across different sampling months, with the color gradient representing the number of recovered LDVs at each time point. A total of 199 sediment samples were collected through these schemes, with detailed information provided in Supplementary Data 1. The maps are generated using the "geom_sf" function in ggplot2. The geojson files of base map were obtained from the public source https://datav.aliyun.com/portal/school/atlas/area_selector. **b** The pipeline for recovering LDV genomes from metagenomic data used in this study. A total of 237 viral metagenome-assembled genomes (vMAGs) were obtained without dereplication. The quality of these LDV genomes was assessed by CheckV, and only those with terminal repeats were identified as complete. **c** Genomic features of 237 LDVs affiliated with *Nucleocytoviricota* and *Uroviricota*. The centered lines in the boxes indicate the median values. The bounds of the box represent the interquartile range, with the lower and upper bounds respectively corresponding to the first and third quartiles. The whiskers denote the lowest and highest values within 1.5 times the interquartile range. Source data are provided as a Source Data file.

The majority of intertidal GVs fell within *Imitervirales* (29 GVs), consistent with observations in marine ecosystems[12,13]. However, unlike marine environments where *Algavirales* dominate, they were relatively scarce in intertidal mudflats (4 GVs). *Pimascovirales*, typically abundant in soil ecosystems[14], were relatively enriched (15 GVs). These findings underscored the uniqueness of intertidal habitats. At the family level, intertidal GVs were clustered into 14 out of 32 subclades defined in a recently established phylogenomic framework[11]. Among these, 22 GVs were assigned to six subclades with recognized family names, including *Mesomimiviridae* (IM_01, 9 GVs), *Mimiviridae* (IM_16, 5 GVs), *Asfarviridae* (AF_01, 2 GVs), *Pithoviridae* (PM_07, 2 GVs), *Prasinoviridae* (AG_01, 2 GVs), and *Pandoraviridae* (PV_04, 2 GVs). The remaining 29 GVs were clustered into eight subclades represented by non-redundant identifiers, with the most abundant being PM_01 (10 GVs) and PV_02 (8 GVs).

In addition to phylogenetic analysis, a machine learning-based approach (TIGTOG)[30] and pairwise AAI comparisons were employed to validate the taxonomic assignment of these GVs (Figures. S6 and S7). TIGTOG also enabled the classification of 7 GVs lacking sufficient markers, of which six were assigned to *Imitervirales* and one to *Pimascovirales* (Figure. S6). Based on the AAI comparisons, several high-quality intertidal GVs exhibited strong similarity (e.g., >90%) to previously isolated complete genomes (Figure. S7). Further synteny analyses revealed a relatively conserved genomic architecture between them[31], with differences primarily arising from localized gene gain or loss (Figure. S8). These variable regions were mainly enriched in genes involved in auxiliary metabolism and host takeover, whereas those related to replication, transcription, and structure remained stable. This pattern may reflect the genomic plasticity of metabolic genes in GVs, potentially driven by environmental conditions and virus-host interactions[9,32].

Within *Imitervirales*, we identified a well-supported subgroup comprising 8 intertidal GVs together with a reference singleton genome (Fig. 2 and Figure. S5), which was evolutionarily adjacent to the IM_02 and IM_18 subclades, yet clearly distinct from other known subclades. Pairwise AAI analysis revealed observable sequence divergence between this subgroup and other reference genomes, while showing high intra-clade AAI values ranging from 42% to 99% (Figure. S9a, b). These results suggested that this subgroup may represent an unrecognized family-level lineage within *Imitervirales*, broadening our understanding of intertidal GV diversity and complementing the current phylogenetic framework[11].

Unlike GVs, large phages lack a concatenated set of marker genes, and their evolutionary relationships are typically inferred using MCP or

https://doi.org/10.1038/s41467-026-71095-7

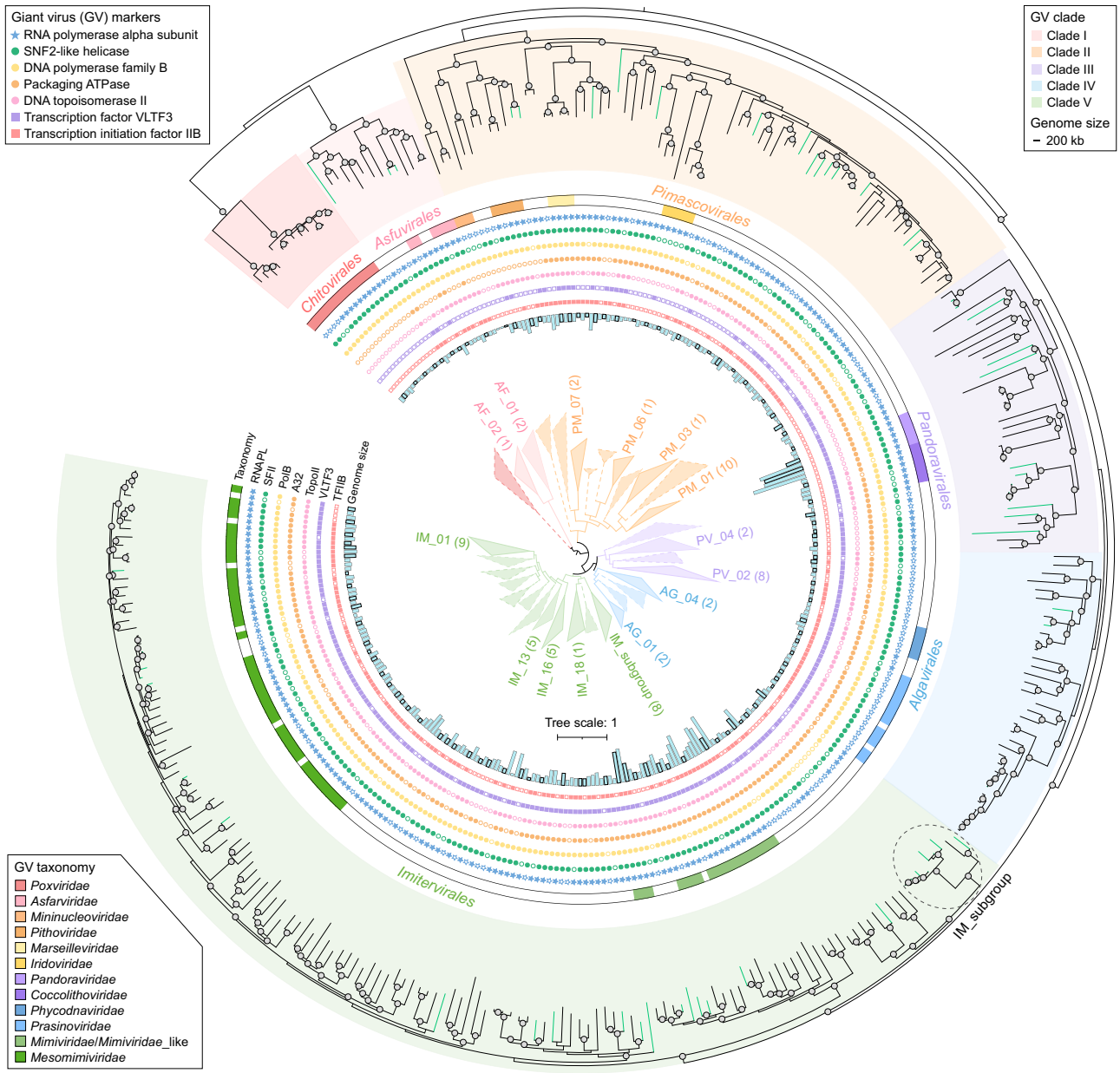


**Fig. 2 | Phylogenetic relationships of 64 intertidal giant viruses (GVs) recovered in this study and 251 reference genomes.** The phylogenetic tree was constructed from concatenated alignments of 7 highly conserved marker genes using the 'LG + I + F + G4' model in IQ-TREE with 1000 bootstrap replicates. The tree is visualized in two forms: the outer tree shows detailed branching relationships, while the inner tree displays family-level collapsed clades. In the outer tree, branches corresponding to GVs recovered in this study are colored green. The filled gray circles on the branches indicate bootstrap values ≥ 90%. The colored backgrounds represent order-level clades. A family-level subgroup comprising 8 intertidal GVs and a reference singleton genome with high bootstrap support is enclosed by dashed lines. In the inner tree, the non-redundant identifiers of collapsed clades containing intertidal GVs are shown, and the number of intertidal GVs within each clade is provided in brackets. All collapsed clades shown are supported by bootstrap values greater than 90%. For better visualization, branches not assigned to family-level taxonomy in the inner tree were removed. Identifiers with corresponding established family names are indicated as follows: AF_01, *Asfarviridae*; PM_07, *Pithoviridae*; PV_04, *Pandoraviridae*; AG_01, *Prasinoviridae*; IM_01, *Mesomimiviridae*; IM_16, *Mimiviridae*. Detailed descriptions of GV markers were: DNA-directed RNA polymerase alpha subunit (RNAPL), DEAD/SNF2-like helicase (SFII), DNA polymerase family B (PolB), Packaging ATPase (A32), DNA topoisomerase II (TopoII), Poxvirus Late Transcription Factor VLTF3 (VLTF3), and Transcription initiation factor IIB (TFIIB). Source data are provided as a Source Data file.

large terminase subunit (TerL) proteins[19,33]. Here, for LDVs belonging to *Uroviricota*, a phylogenetic tree was constructed for 43 intertidal large phages (excluding 4 lacking terminase) and 167 reference genomes using TerL[19] to infer their evolutionary diversity (Fig. 3 and Supplementary Data 4). A total of five clades were identified across the reference genomes and intertidal large phages, indicating their close affinities and shared biological origins (Fig. 3). The majority of large phages were clustered into the Mahaphage clade (clade V), which represented the largest group broadly distributed across almost all examined habitats[19]. Furthermore, multiple intertidal large phages were clustered with the reference genomes of Biggiephage, Judaphage, Jabbarphage, and Whopperphage, thereby expanding the phylogenetic and genomic diversity of these groups. Notably, phylogenetic signals produced by a single marker gene may be less resolved and robust than those derived from concatenated alignments[34]. Thus, the phylogenetic analysis of large phages is mainly used to identify clades

https://doi.org/10.1038/s41467-026-71095-7

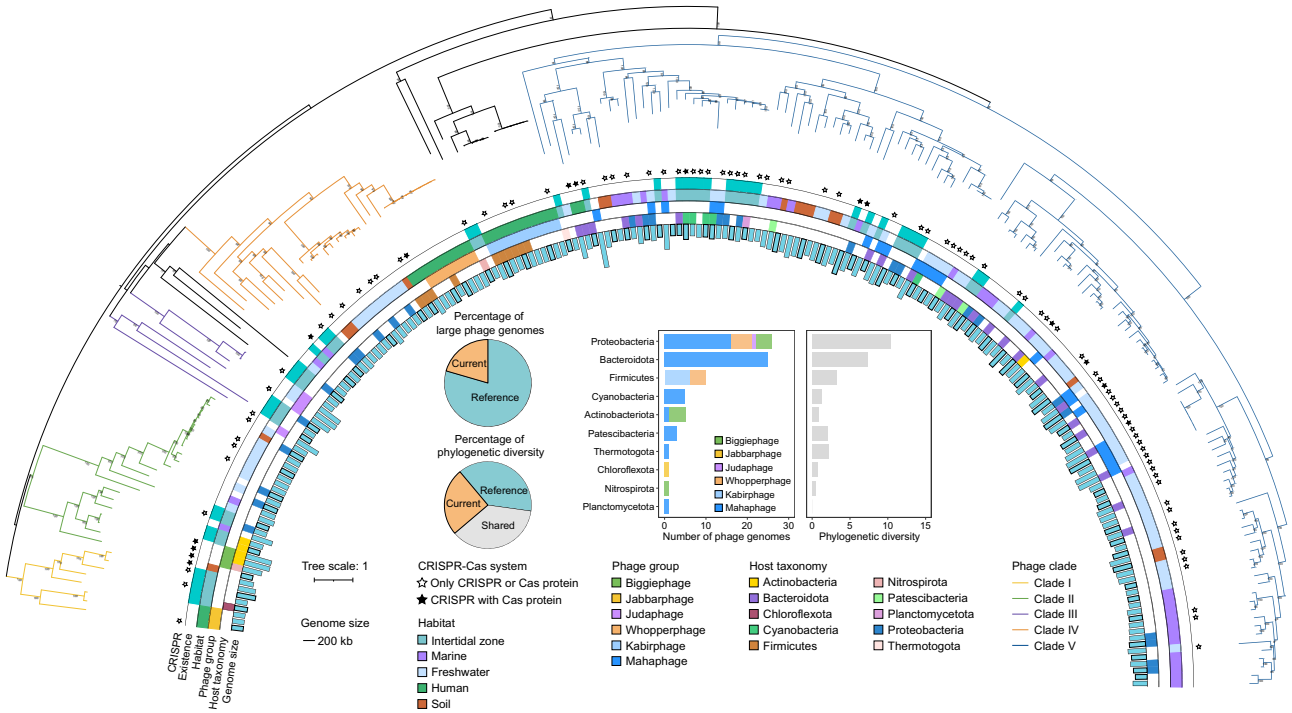


**Fig. 3 | Phylogenic relationships of 43 intertidal large phages recovered in this study and 167 reference genomes.** The tree was constructed based on the large terminase subunit (TerL) using the "Q.pfam + F + I + R8" model in IQ-TREE with 1000 bootstrap replicates. Bootstrap values are indicated on the branches. Tree branches are colored according to the clades with strong bootstrap support (≥70%). The inner to outer layers show the genome sizes, host taxonomy, previously defined phage groups, habitat types, large phages recovered from this study, and the presence of CRISPR-Cas systems. The phylogenetic diversity (PD) was calculated as the sum of branch lengths represented by each large phage. Source data are provided as a Source Data file.

with high bootstrap support, rather than to infer their taxonomic diversity[33,35,36]. Our results placed the intertidal large phages within several major clades characterized in prior work[19], highlighting their broad phylogenetic representation.

## Macroecological patterns and genetic variations of intertidal LDVs

Recent studies have shown that marine LDVs are heterogeneously distributed across diverse sampling sites, time points, and depths[12,13,18,37]. Accordingly, the multi-dimensional sampling scheme employed in this study allowed us to comprehensively explore the macroecological patterns followed by intertidal LDVs. To begin with, we investigated the macrodiversity of intertidal LDVs across various sampling scales. At the national scale, the richness of GVs and large phages generally peaked at mid-latitude and bottomed at low/high-latitude, with some exceptions (Figure. S10a). At the local scales, fluctuations in richness were observed for GVs and large phages, showing patterned seasonal and depth distributions (Figure. S10a). Principal coordinate analyses revealed that both GV and large phage communities were primarily clustered by sampling sites, followed by time points and depths, as suggested by permutational multivariate analyses of variance (Fig. 4a). Additionally, LDVs at both national and local scales were found to obey significant distance-decay relationship (DDR) (Fig. 4b), a fundamental macroecological pattern describing decreased community similarity with increasing geographic distance[38]. Comparatively, GVs exhibited steeper DDR patterns than large phages (Fig. 4b), suggesting stronger spatial turnover in community composition[39]. Beyond spatial patterns, GVs and large phages also exhibited shifts in community similarity along temporal and vertical distances (Fig. 4c), reflecting their community turnover across time and depth.

The microdiversity holds the potential to capture the underlying evolutionary trajectories and ecological mechanisms[40,41]. Through comparative analyses of the nucleotide diversity of GVs and large phages, deep insights into their genetic variations were gained (Supplementary Data 5). Our results demonstrated that large phages exhibited higher nucleotide diversity than GVs, either at the national or local scales (Fig. 4d and Figure. S10b). In addition, greater variations in nucleotide diversity were observed for large phages than for GVs across all sampling scales (Fig. 4d and Figure. S10b). Although double-stranded DNA viruses are generally more stable than other viral types, their mutation rates and genetic diversity can vary with host types and genome sizes[42–44]. Eukaryotic hosts typically replicate and reproduce more slowly and have lower mutation rates than prokaryotes, potentially contributing to the relatively stable nucleotide diversity of GVs[45,46]. Notably, the complex genomic features of GVs and large phages make them less dependent on host machinery for replication and transcription[9,36], suggesting that host mutations and replication rates may not entirely account for the differences in their nucleotide diversity. Alternatively, their infection strategies and evolutionary dynamics may also account for such differences. Unlike phages, GVs invade their host cells through phagocytosis rather than DNA injection[46], leading to slower replication rates and fewer accumulated SNVs. Moreover, GVs harbor a set of conserved marker genes that are absent in large phages[26,33,36], indicating stronger evolutionary constraints on their genomes. Taken together, multiple factors may explain the differences in genetic diversity between GVs and large phages.

## Associations between intertidal LDVs and their potential hosts

To explore the potential associations between GVs and eukaryotes, we first characterized the eukaryotic communities across intertidal mudflats (Figure. S11). In total, 41 eukaryotic lineages belonging to four major groups were detected, including algae ($n = 14$), animals ($n = 12$), protozoa ($n = 9$), and fungi ($n = 6$). Among these, animals (~35.9%) and fungi (~27.9%) exhibited higher relative abundances, followed by algae

https://doi.org/10.1038/s41467-026-71095-7

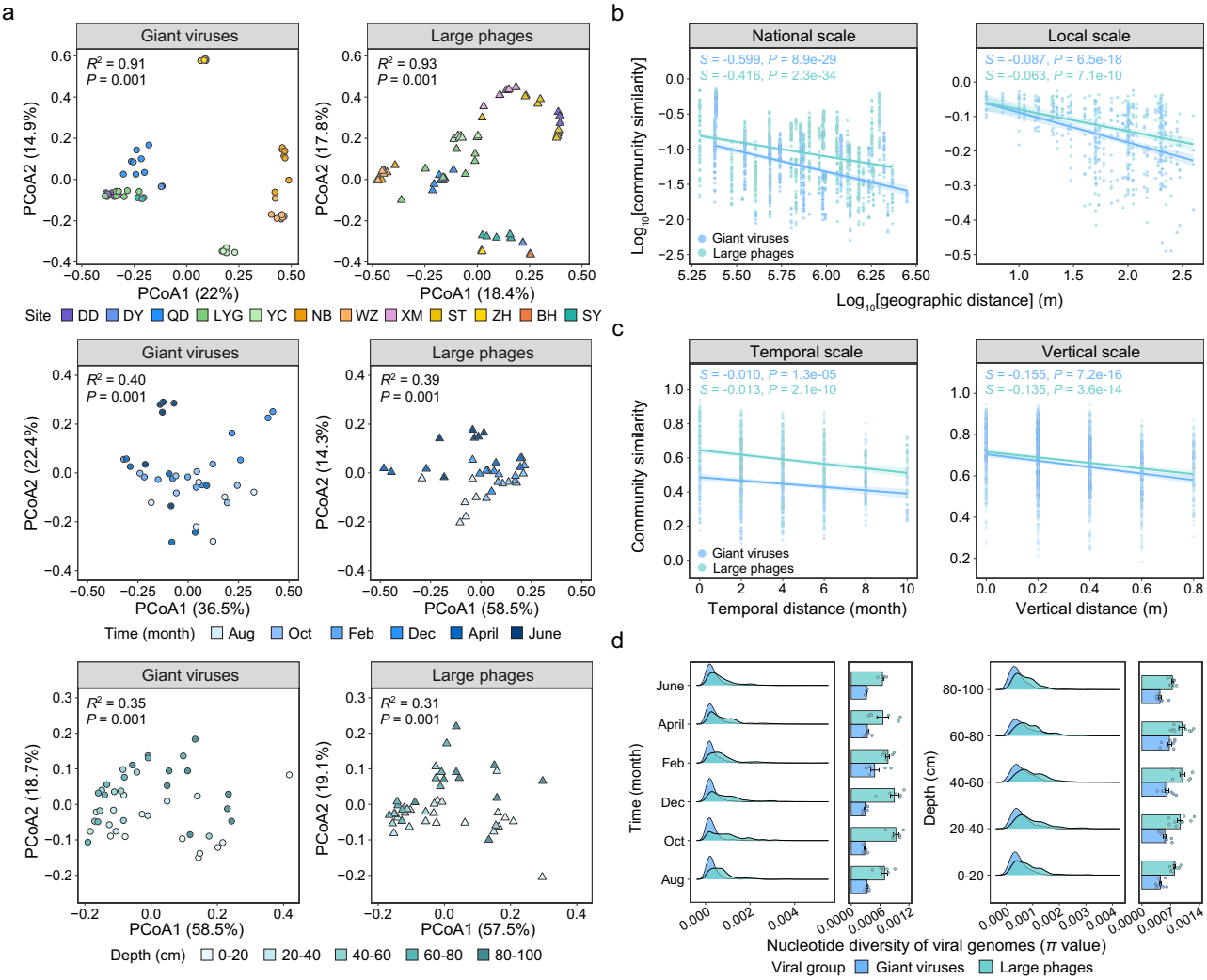


**Fig. 4 | Community- and genetic-level diversity of intertidal giant viruses (GVs) and large phages. a** Principal coordinate analysis (PCoA) of GV and large phage communities based on Bray-Curtis dissimilarity. Detailed descriptions of sampling sites were: Dandong (DD), Dongying (DY), Qingdao (QD), Lianyungang (LYG), Yancheng (YC), Ningbo (NB), Wenzhou (WZ), Xiamen (XM), Shantou (ST), Zhuhai (ZH), Beihai (BH), and Sanya (SY). The significance of differences between sampling sites, time points, or depths was assessed using permutational multivariate analysis of variance with 999 permutations. The statistical tests used were two-tailed. **b** Distance-decay relationships (DDRs) for GVs and large phages at the national and local scales. The relationships between geographic distance (log-transformed) and community similarity (log-transformed) were analyzed. The slope estimates from linear regression models are shown. Statistical significance of each regression model was evaluated using a two-sided *F* test. **c** Shifts in community similarity along temporal and vertical distances for GVs and large phages. The slope estimates from linear regression models are shown. Statistical significance of each regression model was evaluated using a two-sided *F* test. **d** Nucleotide diversity ($\pi$ value) of GVs and large phages across different sampling time points and depths. The ridgeline plot shows the distribution density of nucleotide diversity for GV and large phage genomes. The bar plot shows the nucleotide diversity for GVs and large phages in each group. Points denote nucleotide diversity for individual samples, while bars indicate the mean nucleotide diversity of each group. $n = 6$ biological replicates for each time point, and $n = 9$ biological replicates for each depth. All error bars are shown as mean ± SEM. Source data are provided as a Source Data file.

(~20.2%) and protozoa (~12.5%). Within each group, animals were mainly represented by *Arthropoda*, fungi by *Ascomycota*, algae by *Chlorophyta*, and protozoa by *Discoba* (Figure. S11). These results align with recent reports[23] and highlight the complex biodiversity of intertidal mudflats, particularly the diverse animals and fungi. Next, we investigated the distribution of major eukaryotic lineages across various sampling scales (Figure. S12). The eukaryotic communities showed clear spatial variation at the national scale, while maintaining a relatively stable composition at the local scale. Although temporal and vertical fluctuations were observable, they were much less pronounced.

Given the spatially structured distributions of both GVs and eukaryotes, we assessed their associations through partial Mantel tests controlling for geographic distance. As a result, 31 strong associations (Mantel's $r > 0.53$) were identified (Fig. 5a and Supplementary Data 6), 13 of which (~41.9%) had been reported as known symbiotic relationships[9,13,47]. Subsequently, co-occurrence network analysis based on Spearman's correlations was performed for GV-eukaryotic species pairs, which revealed 79 strong pairwise linkages (Supplementary Data 7). Based on these linkages, normalized abundances of intertidal GVs and their potential eukaryotic hosts displayed significant positive correlations at both the community and lineage levels (Fig. 5b and Figure. S13). In addition, GVs and their potential hosts exhibited similar macroecological patterns, although the turnover rate of GVs was slightly higher (Figure. S14a, b). Within the GV-eukaryotic network, fungi (26) and protozoa (25) accounted for the largest number of linkages, followed by algae (21) and animals (7) (Fig. 5c). As known hosts for multiple GV lineages[9,35], *Amoebozoa* and *Chlorophyta* exhibited the broadest connections with GVs across different families. Several fungal lineages, including *Ascomycota* and *Mucoromycota*, also showed

https://doi.org/10.1038/s41467-026-71095-7

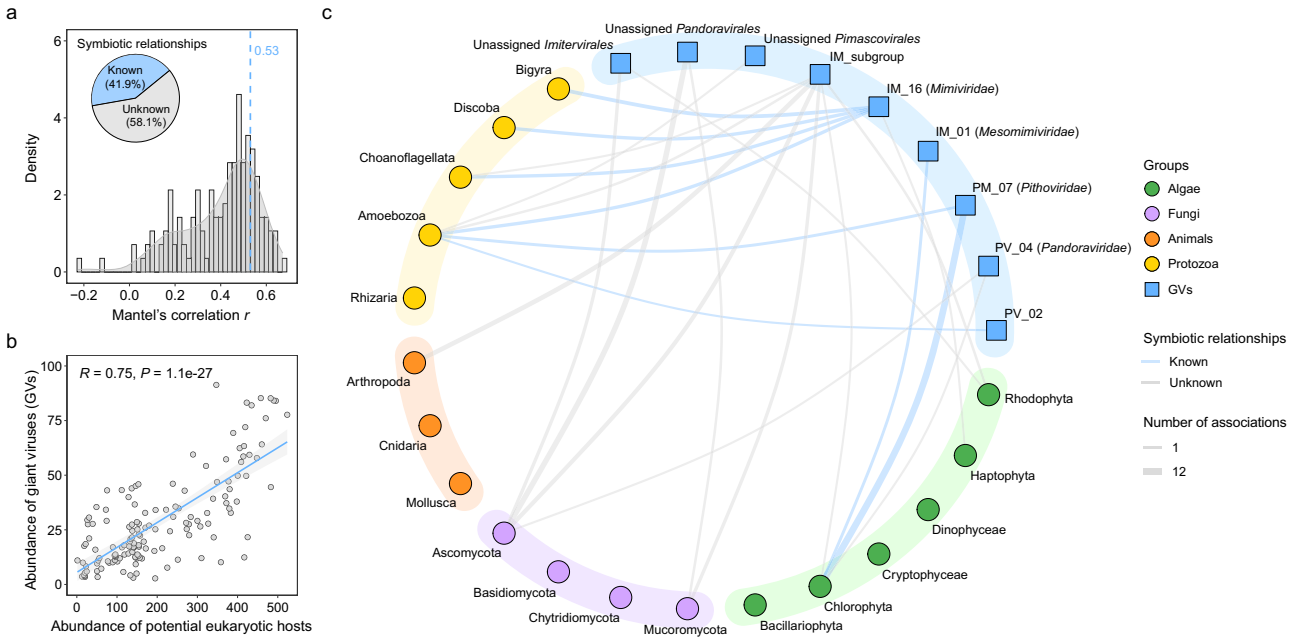


**Fig. 5 | Associations between giant viruses (GVs) and eukaryotic communities.** **a** Histograms show the frequency of partial Mantel correlation coefficients (*r*) between normalized abundances of GVs and eukaryotic lineages. Correlation coefficients $r > 0.53$ are drawn in blue dashed line. The pie chart shows the proportion of known and unknown GV-eukaryotic lineage associations. **b** The relationships between the normalized abundances of GVs and their potential eukaryotic hosts across intertidal samples. The shaded gray region reflects 95% confidence intervals of the fitted regression line. The linear regression represents Pearson correlation coefficient relationship. Statistical significance of the linear regression was evaluated using a two-sided *F* test. **c** Co-occurrence networks of GVs and eukaryotes. Squares represent GV lineages and circles represent eukaryotic lineages present in intertidal sediments. Edges represent the number of associations between GV and eukaryotic species. Known associations are shown in blue, whereas unknown associations are shown in gray. Source data are provided as a Source Data file.

strong associations with intertidal GVs despite previously not recognized as native hosts[9]. Strikingly, we found that 7 out of 8 GVs within the IM_subgroup were closely related with *Arthropoda*, indicating potential interactions between *Imitervirales* and animal-associated lineages[15,48].

Large phages usually carry CRISPR spacers and Cas proteins targeting various host taxa (Fig. 3 and Figure. S15), thereby shaping their evolutionary trajectories with bacterial hosts[19,35]. Through CRISPR spacer matching and the integrated iPHoP pipeline, we identified 24 phage-host pairs. The normalized abundances of large phages and their hosts showed a significant positive association ($R = 0.45$, $P = 1e\text{-}09$) (Figure. S16). The recovered intertidal large phages were predicted to infect nine bacterial phyla (Figure. S17). Of these, seven were linked to *Proteobacteria*, followed by *Bacteroidota* ($n = 5$) and *Cyanobacteria* ($n = 4$). This was generally consistent with findings in other habitats (Fig. 3), suggesting a close relationship between dominantly distributed bacterial phyla and large phages. *Firmicutes* represented an exception[19,35], as most intertidal large phages within the Whopperphage clade failed to establish phage-host links with them.

## Ecological drivers and community assembly of intertidal LDVs

To uncover the mechanisms that form and maintain LDV diversity, we further analyzed the ecological drivers and community assembly processes governing intertidal LDV communities using multiple statistical approaches[27,49]. First, we evaluated the associations between environmental variables and LDV compositional variation through partial Mantel tests. Factors including mean annual temperature (MAT), salinity, pH, nitrite nitrogen ($NO_2^-$-N), total phosphorus (TP), total organic carbon (TOC), and moisture, were significantly associated with the compositional variations of GVs and large phages (Fig. 6a). Second, null model analysis was used to quantify the relative importance of deterministic and stochastic processes in shaping LDV communities[50]. Stochastic processes respectively contributed 75% and 81% relative importance in shaping the compositional variations of GVs and large phages (Fig. 6b). This was also confirmed by the $RC_{bray}$ metric-based approach[51] that the compositions of intertidal LDVs were predominantly structured by stochastic processes (Fig. 6c). Finally, we quantified the relative contributions of individual ecological processes using the phylogenetic-bin-based null model analysis[52]. At the whole community level, GVs and large phages were mainly structured by drift (59% and 75%), followed by dispersal limitation (19.3% and 10.4%) and heterogeneous selection (13.2% and 7.8%) (Fig. 6d). For individual lineages, the relative importance of different ecological processes varied substantially. Of these, *Asfuvirales* exhibited a striking predominance of dispersal limitation (93.8%) in community assembly, largely driven by its members being restricted to only a few sites.

Compared to the ecological drivers of overall viral communities in intertidal mudflats[27], for which deterministic processes played relatively important roles, LDV communities were primarily shaped by stochastic processes, as revealed by null model and $RC_{bray}$ analyses of the compositional variations. Although geographic distance and multiple environmental variables significantly correlated with LDV communities, their contribution to the compositional variations was relatively limited. The results showed that ecological drift was the major process shaping the compositional variations of both GVs and large phages. Due to the large genome size and virion volume, LDVs typically exhibit smaller burst sizes and slower decay rates than other viruses[44,46], which may result in smaller population sizes and greater environmental stability. Consequently, the life history of LDVs appears to follow an ecological strategy analogous to K-selection, rather than the *r*-selection commonly observed in viruses[53]. The limited population sizes and distinct ecological strategies explain the major role of drift in LDV community assembly[49]. Importantly, the pairwise stochasticity indices for GVs (Mantel's $r = 0.416$, $P = 0.001$) and large phages (Mantel's $r = 0.212$, $P = 0.001$) exhibited significant correlations with the community composition of their potential hosts, indicating that host

https://doi.org/10.1038/s41467-026-71095-7

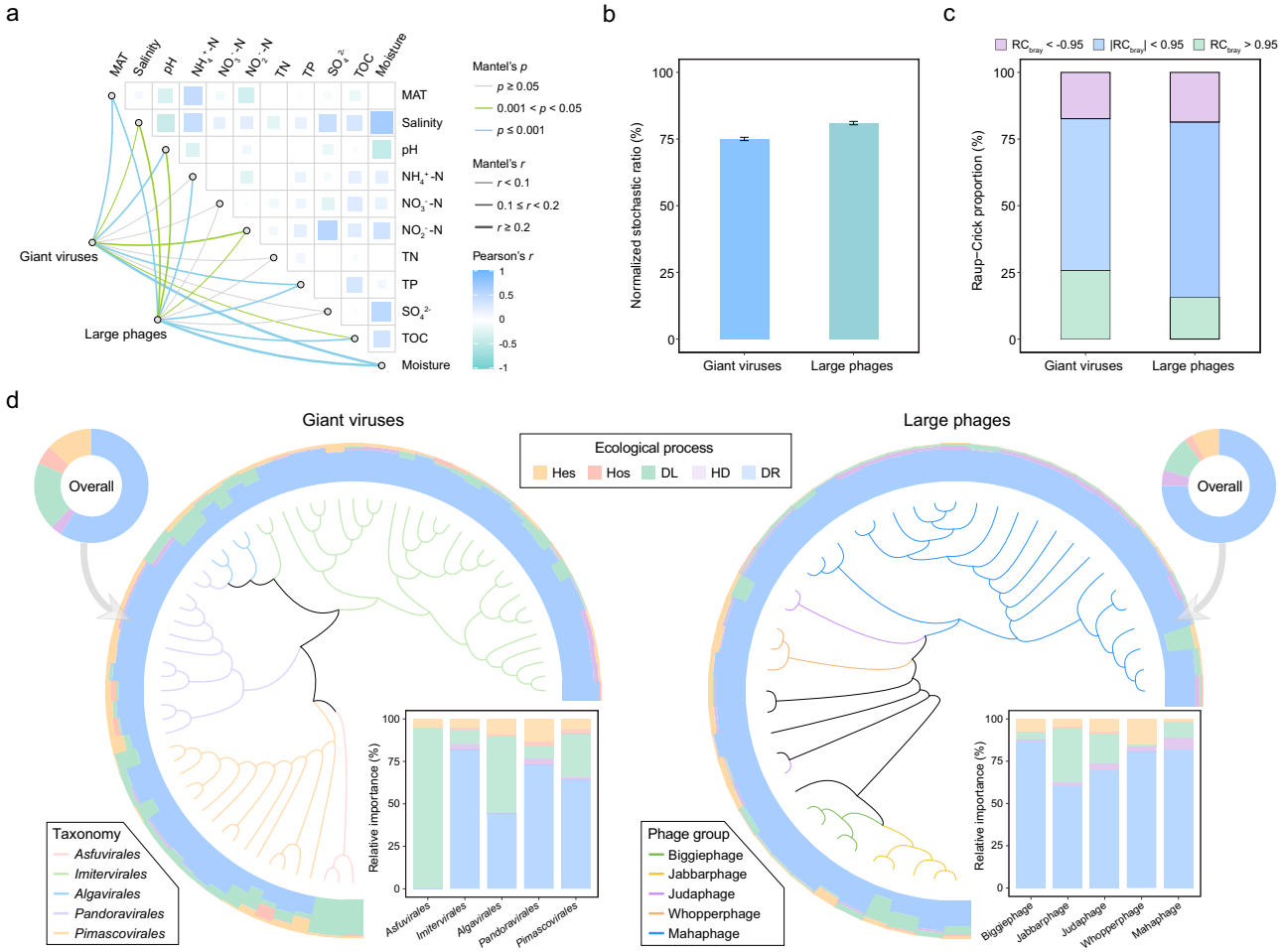


**Fig. 6 | Ecological drivers and community assembly of intertidal giant viruses (GVs) and large phages. a** Partial Mantel tests showing the relationships between environmental factors and community compositional variations of GVs and large phages. The edge color and width represent Mantel's *r* and *p* values, respectively. The color gradient in heatmap represents the Pearson correlation coefficients among different environmental factors. All statistical tests were two-tailed. Detailed Mantel's *r* and *p* values are provided in the Source Data. **b** Stochastic ratio analyses of GVs and large phages. The normalized stochastic ratios were calculated based on 1000 null models, using a threshold of 50% to distinguish between deterministic and stochastic processes. For GVs, the mean ± SEM of stochastic ratio values calculated from 13,861 pairwise comparisons among 167 intertidal samples were plotted. For large phages, the mean ± SEM of stochastic ratio values calculated from 17,766 pairwise comparisons among 189 intertidal samples were plotted for large phages. **c** The Raup-Crick proportion of GVs and large phages showing the contributions of different processes to community assembly, including deterministic processes ($|RC_{bray}| > 0.95$) and stochastic processes ($|RC_{bray}| \leq 0.95$). **d** Relative importance of different ecological processes, including heterogeneous selection (HeS), homogeneous selection (HoS), dispersal limitation (DL), homogenizing dispersal (HD), and drift (DR), in shaping the overall communities and individual genomes for GVs and large phages. The phylogenetic tree of GVs was constructed using a concatenated alignment of 7 highly conserved marker genes, while the phylogenetic tree of large phages was constructed based on large terminase subunit (TerL). Tree branches of GVs are colored according to the taxonomy, whereas tree branches of large phages are colored according to their groups. Source data are provided as a Source Data file.

communities substantially modulated LDV community assembly. Further analysis revealed that the stochasticity of LDV communities decreased significantly with increasing host abundance (Figure. S18). These findings are consistent with our earlier inference that higher host abundance can affect LDV population sizes, thereby reducing the strength of stochastic processes in LDV community assembly. Together with the similar macroecological patterns shared by LDVs and their hosts, our results underscore the pivotal role of virus-host interactions in shaping LDV community assembly.

## Metabolic potential and evolutionary selection of intertidal LDV genes

In total, 17,234 protein-coding genes were predicted for GVs, of which 4481 were assigned to known orthologous groups (Supplementary Data 8 and Figure. S19a). According to the COG annotation, 3000 GV orthologs were functionally known, with the class "replication, recombination and repair (L)" being the most abundant, potentially contributing to viral replication (Figure. S19b). For large phages, the relative proportions of dominant functional classes were generally similar to GVs (Figure. S19b). However, a higher proportion of large phage genes (58.09%) were characterized as "function unknown (S)" than those of GVs (33.05%). We then employed a prokaryotic viral protein language model[54] to improve the functional annotation of these unknown genes. As a result, 2954 unknown large-phage genes (27.32%) were assigned to 9 functional categories, expanding the overall annotated fraction of large phage genes by 21.41% (Figure. S20a). Of these, 55.48% were related to tail and packaging proteins, followed by those involved in DNA metabolism (18.99%) and auxiliary metabolism (8.09%).

To further explore the metabolic potential of intertidal LDVs, we identified 577 GV genes and 306 large-phage genes associated with various metabolic pathways (Supplementary Data 9). The dominant genes carried by GVs and large phages were involved in purine and pyrimidine metabolism, followed by those related to amino acid

https://doi.org/10.1038/s41467-026-71095-7

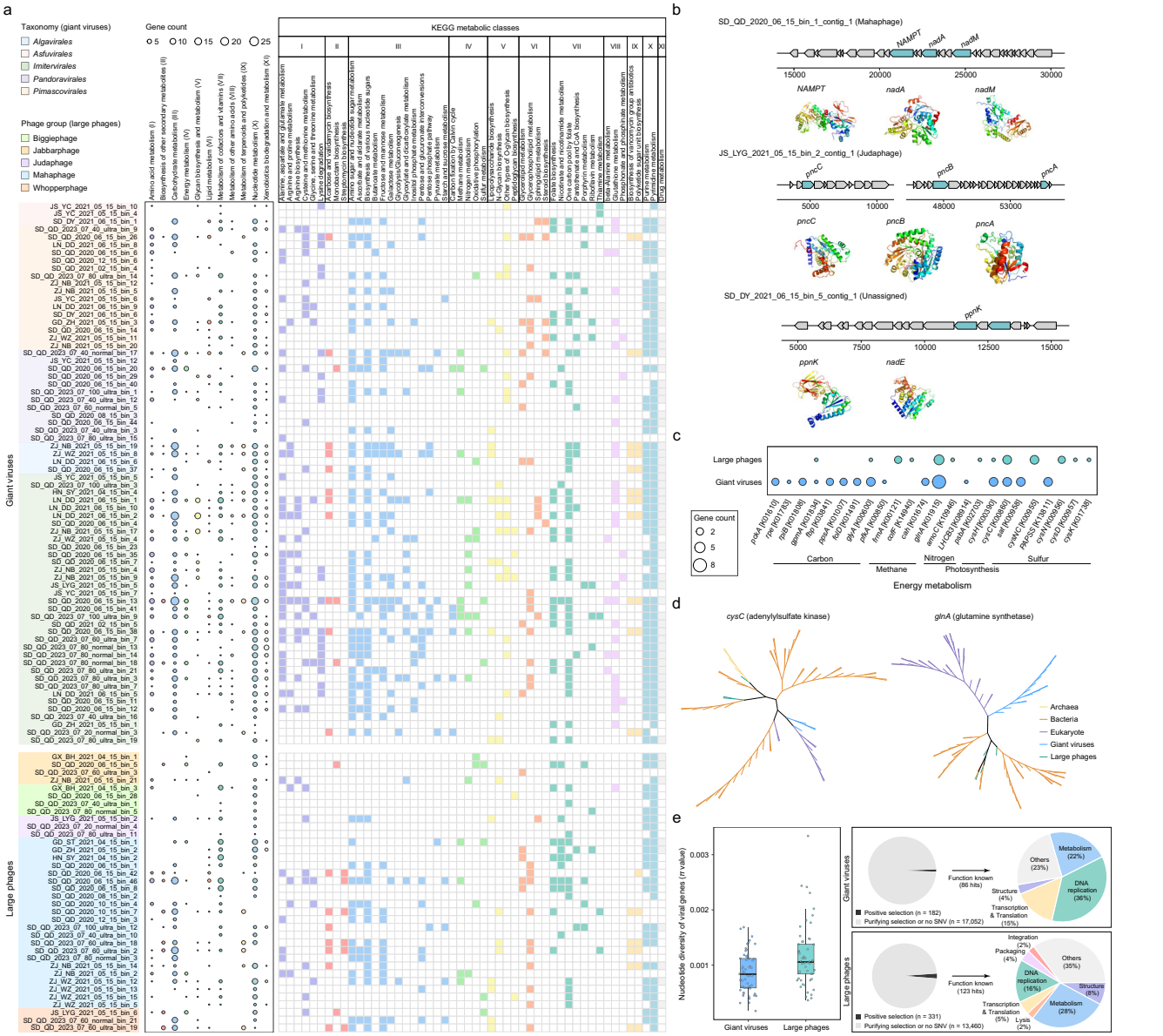


**Fig. 7 | Metabolic potential, phylogenetic relationships, and selection pressures of functional genes encoded by giant viruses (GVs) and large phages.** **a** The KEGG metabolic classes of functional genes encoded by each GV and large phage genome are shown. The colored backgrounds represent different GV taxonomy and large phage groups. The circle sizes in the bubble plot correspond to the number of genes encoded by each genome that are involved in specific metabolic classes. The presence or absence of each metabolic pathway was determined based on functional assignment against KEGG database. The pathways found with genes are shown as filled squares, while those without are shown as blank squares. **b** Genomic structures of representative large phages encoding key genes involved in $NAD^+$ synthesis. Cyan arrows indicate genes related to $NAD^+$ metabolism, while gray arrows indicate other functional genes. The protein structures of these genes were predicted using AlphaFold3. **c** Distribution of auxiliary metabolic genes (AMGs) related to energy metabolism in GVs and large phages. Bubble sizes represent the total number of AMGs encoded by GVs and large phages. **d** Phylogenetic trees of adenylylsulfate kinase (*cysC*) and glutamine synthetase (*glnA*) encoded by GVs (blue) and large phages (cyan), together with homologues from bacteria (orange), archaea (yellow), and eukaryotes (purple). **e** Nucleotide diversity ($\pi$ value) and selection pressures of functional genes encoded by GVs and large phages. The points in boxplot represent the average $\pi$ values of genes within each sample. The centered lines of the boxes indicate the median value. The bounds of the boxes represent the interquartile range, with the lower bound corresponding to the first quartile and the upper bound to the third quartile. The whiskers denote the lowest and highest values within 1.5 times the range of the first and third quartiles. Only samples ($n = 57$) in which single nucleotide variations of functional genes encoded by both GVs and large phages were simultaneously detected were included for the calculation and comparison. The pie charts on the left show the proportion of viral genes under positive selection (i.e., $pN/pS > 1$), while those on the right indicate the functional classes of positively selected genes. Detailed annotation for these genes is provided in Supplementary Data 10. Source data are provided as a Source Data file.

metabolism and carbohydrate metabolism (Fig. 7a). Although GVs and large phages had similar overall functional gene composition, they exhibited distinct preferences toward specific metabolic pathways. Of the 50 selected metabolic pathways, 39 were more prevalent in GVs than in large phages, suggesting broader metabolic versatility of intertidal GVs (Figure. S20b). For instance, genes involved in O-glycan biosynthesis (commonly carried by eukaryotes) were detected in over 20% of GVs but were absent in large phages (Fig. 7a and Figure. S20b). In contrast, we also found that more than 20% of the large phages carried genes involved in $NAD^+$ synthesis that were absent in GVs (Fig. 7b and Supplementary Data 9). Of these, five large phages possessed all genes necessary for the salvage pathway (i.e., *NAMPT* and

*nadD*/*M*), showing their capability of independent $NAD^+$ synthesis and thus expanding the known environmental distribution of NAD-large phages[33]. The presence of $NAD^+$ synthesis systems and rich tRNAs in these large phages supports their unique life strategies[33].

Reprogramming host metabolism by AMGs has been reported as a widespread strategy in both GVs and large phages[14,15,20]. Here, we found a variety of AMGs within intertidal GVs (30 AMGs) and large phages (19 AMGs), potentially contributing to different biogeochemical cycles (Fig. 7c). Of these, AMGs involved in sulfur and nitrogen metabolism were commonly detected in GVs and large phages, especially those encoding adenylylsulfate kinase (*cysC*) and glutamine synthetase (*glnA*). Phylogenetic analyses revealed that *cysC* and *glnA* from GVs and large phages clustered with eukaryotic and prokaryotic homologues, respectively (Fig. 7d). Similar to previous studies[16], AMGs involved in carbon fixation, glycolysis, and gluconeogenesis were detected in GVs affiliated with diverse lineages, though in relatively low abundance (Fig. 7a, c). However, AMGs involved in light-driven proton pump (e.g., rhodopsins)[15,48], which are frequently found in GVs in aquatic environments, were not detected in this study. Such results suggested that some functional traits encoded by GVs might be habitat specific, consistent with recent observations in soil studies[14].

Through detecting the density of SNVs, we also assessed the nucleotide diversity ($\pi$ value) and selection pressure ($pN/pS$) of GV and large-phage genes (Fig. 7e). Similar to genome-level analysis, large-phage genes exhibited greater nucleotide diversity than GV genes across intertidal samples (paired Student's $t$ test, df = 56, $\underline{t}$ = 4.04, $P$ = 0.0002, 95% CI = [0.00015, 0.00044]). In addition, we found that 182 GV (1.1% of total 17,234) and 331 large-phage genes (2.4% of total 13,791) were under positive selection (i.e., $pN/pS > 1$) (Supplementary Data 10). Of these, 86 GV and 123 large-phage genes were functionally annotated, with the majority (62% for GVs and 52% for large phages) associated with DNA replication, virion structure, and metabolism. Among the positively selected metabolic genes in both GVs and large phages, many encode glycosyl hydrolases and transferases (Supplementary Data 10), which are likely involved in host recognition and the reprogramming of host carbohydrate metabolism[4,48,55]. Several positively selected genes related to transcription and translation were also identified (Fig. 7e), including GV genes encoding RNA polymerase and large-phage genes encoding tRNA synthetase. Together, these findings reflect the strong selection pressure acting on LDV genes during their interactions with hosts.

In conclusion, the LDVs recovered in this study remarkably expand their taxonomic and phylogenetic diversity in mudflat intertidal zones. The distribution of LDV communities exhibits high spatial heterogeneity and shows strong fluctuations over time and across depth, with generally consistent patterns also observed for their potential hosts. Ecological drift is identified as the dominant process shaping LDV community compositions, and its strength is closely associated with host abundance. Furthermore, LDVs encode diverse metabolic capabilities that potentially reduce their dependence on hosts or enhance host metabolism. Comparative analyses of GVs and large phages revealed distinct characteristics between them. Our work extends the current knowledge of the diversity, potential hosts, auxiliary metabolism, and ecologcial mechanisms of LDVs in complex intertidal mudflats.

## Methods

### Intertidal sample collection and metagenomic sequencing

In this study, intertidal mudflats were sampled in four field experiment schemes (Fig. 1a): (i) National sampling: a total of 96 sediment samples (0–15 cm, eight replicates for each site) were collected from twelve coastal sites across China between April 2021 and June 2021; (ii) Nested local sampling: 25 sediment samples (0–15 cm) were collected from Qingdao (36.46° N, 120.75° E) in June 2020; (iii) Bimonthly time-series sampling: 36 sediments (0–15 cm, six replicates for each time point) were collected in Qingdao between August 2020 and June 2021; (iv) Vertical sampling: 50 sediments (0–100 cm, 20 cm as a column, ten replicates for each depth) were collected from Qingdao in July 2023.

For each sample, five intertidal sediment cores were collected within a 1 $m^2$ area and homogenized. The mixed samples were stored on ice and immediately transported to the laboratory. Total DNA of each sample was then extracted from 0.5 g of freeze-dried sediment using the FastDNA SPIN Kit for Soil (MP Biochemicals, USA). The quality of extracted DNA was assessed by measuring the 260/280 and 260/230 ratios with a Nanodrop ONE spectrophotometer (Thermo Fisher Scientific, MA, USA). The DNA samples meeting the required quality standards were sequenced by the Illumina NovaSeq platform with a PE150 strategy, conducted by NovoGene Co., Ltd. (Tianjin, China). Two types of sequencing depth were employed, including the regular 20 GB per sample and an ultra-deep 200 GB per sample. As a result, 199 shotgun metagenomes comprising approximately 5.3 TB of data were generated, including 194 with sequencing depths exceeding 20 GB and five with ultra-deep sequencing depths exceeding 200 GB. Among these, 124 metagenomes were derived from NCBI projects PRJNA957716 and PRJNA1029225, which were analyzed in our previous studies focusing on bulk viromes (primarily phages) at the contig level[5,27,41]. The remaining metagenomes were generated and analyzed for the first time in this study, and have been deposited under project numbers PRJNA1099403, PRJNA1100757, and PRJNA1099773. Detailed information of metagenomic data is provided in Supplementary Data 1.

### Recovery and refinement of LDV genomes

To reconstruct LDV genomes, raw reads of each shotgun metagenome were first trimmed and filtered with a minimum quality score of 20 and a minimum read length of 36[56,57]. Then, clean reads were input into MEGAHIT v1.2.9[57] to perform metagenomic assembly. VIBRANT v1.2.1[58], VirSorter v2.2.3[59], and geNomad v1.7.4[60] were used to identify viral fragments from assembled contigs with lengths ≥ 3 kb. Subsequently, viral contigs were screened on the basis of the following criteria: (i) identified by VirSorter2 and geNomad with scores ≥ 0.7 and containing viral hallmark genes; (ii) identified by VirSorter2, geNomad, or VIBRANT using default parameters and filtered by CheckV v1.0.1[61], with contigs marked as "not-determined" being excluded. VirSorter2, geNomad, and CheckV were also employed to filter out proviruses and complete genomes before genome extension and binning. Furthermore, viral contigs exhibiting significant host contamination (i.e., proviral extraction was unavailable) were excluded. The retained viral contigs were utilized as inputs for COBRA v1.2.3[6] to extend their lengths. MetaBAT v2.12.1[62], vRhyme v1.1.0[7], and PHAMB v1.0.1[8] were used to generate vMAGs with multi-coverage binning strategies for each assembly dataset (viral contig length ≥ 5 kb)[63].

Taxonomic assignment for LDVs was performed using geNomad to classify vMAGs as either *Nucleocytoviricota* or *Uroviricota*. To reduce cellular contamination, vMAGs classified within the phylum *Nucleocytoviricota* were searched against GVOG database in ViralRecall v2.1, and contigs with ViralRecall scores < 0 were removed[16,26]. Subsequently, *Nucleocytoviricota* vMAGs were also analyzed with TIGTOG v1.0[30] using a probability threshold of 0.5 to assign taxonomy at the order and family levels. Finally, all recovered LDV genomes were clustered and de-replicated at 95% ANI using dRep v3.4.0[64]. Representative genomes were manually selected based on their completeness and contamination as assessed by CheckV.

### Phylogenetic analyses of GVs and large phages

For GVs, the core genes were detected using ncldv_markersearch v1.1, including DEAD/SNF2-like helicase (SFII), DNA-directed RNA polymerase beta subunit (RNAPS), DNA-directed RNA polymerase alpha subunit (RNAPL), DNA topoisomerase II (TopoII), MCP, Packaging ATPase (A32), PolB, Poxvirus Late Transcription Factor VLTF3 (VLTF3), and Transcription initiation factor IIB (TFIIB)[16]. To infer the phylogeny

of intertidal GVs, 1372 reference genomes were retrieved from a recently established phylogenomic framework[11,16]. Seven highly conserved marker genes in these GVs, including SFII, RNAPL, PolB, A32, VLTF3, TFIIB, and TopoII, were identified and concatenated with ncldv_markersearch[11,16], aligned with Muscle v5.1[65], and trimmed with trimAl v1.4 using a gap threshold of 0.5[66]. GVs lacking sufficient markers were excluded owing to their incomplete genomes. A preliminary tree of 64 intertidal GVs and 1372 reference GVs was constructed using FastTree v2.1.11[67] to assess their phylogenetic placement. Subsequently, 251 reference GVs were selected to represent all major clades, with priority given to those closely related to intertidal GVs. A final phylogenetic tree including these references and intertidal GVs was inferred using IQ-TREE v2.2.5, applying the "LG + I + F + G4" model and 1000 bootstrap replicates[11,68].

For large phages, reference genomes were retrieved from IMG/VR v4[69] and recent studies[19]. TerL proteins from both intertidal and reference large phages were identified based on the geNomad and eggNOG-mapper annotations[60,70]. To reduce redundancy, TerL proteins from reference genomes were clustered at 95% identity using CD-HIT v4.8.1[71]. These de-replicated reference proteins were then run against TerL proteins from intertidal large phages using DIAMOND v2.0.14 with an e-value threshold of 1e-5, and only the hits were retained[72]. All TerL protein sequences were aligned with MAFFT v7.490 (parameters: --maxiterate 1000 --localpair)[73] and trimmed with trimAl v1.4 using a gap threshold of 0.1[66,74]. Finally, IQ-TREE v2.2.5[68] was used to infer their phylogenetic relationships with the best-fit model (Q.pfam + F + I + R8) and 1000 bootstrap replicates.

### Abundance profiling

To quantify the relative abundance of LDVs, clean reads of each sample were first mapped to 118 representative vMAGs using BWA-MEM v0.7.17[75]. Then, the coverage of each LDV genome was calculated by CoverM v0.7.0 with "trimmed_mean" mode (parameters: --trim-min 10 --trim-max 90 --min-read-percent-identity 95 --min-read-align-percent 90)[27]. Among these, the LDV genomes must have reads covering ≥ 20% of their length (parameters: --min-covered-fraction 20), and the coverage was reported as zero if this condition was not met[12,18]. Finally, the coverage of LDVs was normalized based on the read number of each sample[27]. To minimize the potential bias introduced by samples with substantially different sequencing depth, which could not be fully corrected by normalization, five ultra-deep metagenomes were excluded from downstream analyses.

### Correlation analyses between GVs and eukaryotic lineages

EukRep v0.6.6[76] and Tiara v1.0.3[77] were initially employed to identify eukaryotic fragments from assembled contigs with lengths ≥ 5 kb. Subsequently, all eukaryotic contigs were de-replicated at 95% identity using CD-HIT. MMseqs2 v14.7e284[78] was used to classify eukaryotic contigs and only the contigs assigned to eukaryotic species were retained. The coverage for each eukaryotic contig was calculated using CoverM with the same parameters as for LDVs. For each eukaryotic species, abundance was calculated as the average coverage across all contigs, weighting each contig by its length in base pairs, and normalized according to the read number of each sample[55].

To establish the relationship between GV and eukaryotic lineages, a partial Mantel test with 999 permutations was carried out to assess their correlations while controlling for the potential effects of geographic distance[13]. The community dissimilarity for GV and eukaryotic lineages was calculated using Bray-Curtis metric. Geographic distance among various sampling sites was measured using Euclidean metric. Mantel's $r > 0.53$ was used to indicate significant symbiotic relationships between GV and eukaryotic lineages[13]. Among these, *Asfuvirales* was excluded from the tests owing to the insufficient samples available for generating reliable community matrices. Subsequently, Spearman's correlation was used to evaluate the relationship of individual pairwise combinations of GVs and eukaryotes, with a cutoff of $P < 0.05$ (adjusted by applying a Bonferroni correction). GVs and eukaryotic species that were present in ≤10% of samples were excluded. Correlation coefficients of $\rho > 0.6$ and eukaryotic lineages consistent with partial Mantel test were included in co-occurrence network analyses, while a higher threshold ($\rho > 0.8$) was applied for previously unreported symbiotic relationships[9,13,79].

### Bacterial host prediction of large phages

To assign host information for large phages, bacterial MAGs were binned using metaWRAP v1.3.2 with 'binning' and 'bin_refinement' modules from each assembly dataset[80]. The CRISPR spacers within large phages were identified using CRISPRCasFinder v4.3.2[81]. Afterwards, the extracted spacer sequences were searched against bacterial MAGs with an identity ≥ 95% and a mismatch ≤ 1 to establish the phage-host interactions. Furthermore, iPHoP v1.3.2[82] was employed to predict the hosts of large phages, utilizing an expanded database comprising 6734 bacterial MAGs recovered from this study and 86,709 reference bacterial MAGs. For each computational method, only the best hits were retained.

### Functional gene identification and annotation

The proteins of LDVs were predicted using prodigal-gv[83]. These LDV proteins were subsequently categorized into different COG functional classes through eggNOG-mapper v2.1.12, employing an e-value threshold of 1e-5 and a bit score of 50[70]. Metabolic classes and pathways of LDV proteins were further annotated based on their assigned KEGG_ko numbers. For large phages, a protein language model was additionally applied to improve the annotation of their unknown proteins[54]. The Cas proteins in large phages were identified using CRISPRcasIdentifier v1.1.0[84], and their functional domains were manually verified by searching against UniProt and InterPro databases[85,86]. The tRNAs in large phages were identified using tRNAscan-SE v2.0.12[87]. The protein structures of metabolic genes were predicted using AlphaFold3[88]. To examine the phylogenetic relationships, protein sequences of *cysC* and *glnA* from GVs and large phages were queried against UniProt database to recruit homologous sequences, employing an e-value threshold of 1e-5. Multiple-sequence alignments were built using Muscle, and automatically trimmed using trimAl. Phylogenetic trees were constructed using IQ-TREE with the best-fit models ("LG + I + G4" for *cysC* and "LG + I + R4" for *glnA*) and 1000 bootstrap replicates.

### Diversity index calculation and ecological analyses

The alpha diversity (richness) and beta diversity (Bray-Curtis dissimilarity) indices were computed using the "vegdist" function of vegan package[89]. Principal coordinate analysis was performed based on Bray-Curtis dissimilarity, and the significance was tested using permutational multivariate analysis of variance with 999 permutations. The 'geoXY' function of SoDA package was used to calculate the geographic distance between various sampling sites[27]. The partial Mantel test with 999 permutations was used to assess the correlation between community distance and environmental factors while controlling for the potential effects of geographic distance. SNV detection and population nucleotide diversity calculation of LDV genomes were conducted using inStrain v1.9.0[90]. For subsequently microdiversity analysis, only LDV genomes with detectable abundance in each sample (i.e., ≥20% genomic breadth of coverage) were included. The MetaPop v1.0 was used to calculate the gene-level microdiversity, including the nucleotide diversity and selection pressure ($pN/pS$)[91].

The taxonomic stochastic ratios and $RC_{bray}$ metric for GVs and large phages were calculated using the NST package with 1000 null models generated by Bray–Curtis dissimilarity[50]. Subsequently, community assembly mechanisms of GVs and large phages were further inferred by the iCAMP package[52]. Within this framework, the relative

contributions of five ecological processes were quantified, including homogeneous selection (HoS), heterogeneous selection (HeS), dispersal limitation (DL), homogenizing dispersal (HD), and drift (DR). The normalized abundances of GVs and large phages across different samples and their phylogenetic trees were used as inputs, with a set of parameters: ds = 0.2, rand.time = 1000, sig.index = SES.RC, phylo.metric = $\beta$MPD.

### Reporting summary

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

## Data availability

The metagenomic data generated in this study have been deposited in NCBI Sequence Read Archive database under project numbers PRJNA957716, PRJNA1029225, PRJNA1099403, PRJNA1100757, and PRJNA1099773. Among these, PRJNA957716 and PRJNA1029225 were derived from our previous studies, while the remaining projects were newly generated in this study. The representative sequences of LDVs recovered from this work have been deposited in the ENA Sequence Read Archive database under project number PRJEB108004, and are also available in the Zenodo database [https://zenodo.org/records/18154325]. Source data are provided with this paper.

## Code availability

R code used for generating figures and performing data statistics in this study is provided with this paper and publicly available on GitHub [https://github.com/MengzhiJ/Unveiling-the-biodiversity-of-large-DNA-viruses-in-intertidal-mudflats-via-metagenomics].

## References


1. Chevallereau, A., Pons, B. J., van Houte, S. & Westra, E. R. Interactions between bacterial and phage communities in natural environments. *Nat. Rev. Microbiol.* **20**, 49–62 (2022).
2. Carreira, C. et al. Integrating viruses into soil food web biogeochemistry. *Nat. Microbiol.* **9**, 1918–1928 (2024).
3. 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. *Nat. Ecol. Evol.* **7**, 1809–1822 (2023).
4. Zimmerman, A. E. et al. Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat. Rev. Microbiol.* **18**, 21–34 (2020).
5. Ji, M. & Tu, Q. Towards comprehensive metagenomic recovery and data mining of complex environmental DNA viromes. *Environ. DNA* **7**, e70172 (2025).
6. Chen, L. & Banfield, J. F. COBRA improves the completeness and contiguity of viral genomes assembled from metagenomes. *Nat. Microbiol.* **9**, 737–750 (2024).
7. Kieft, K., Adams, A., Salamzade, R., Kalan, L. & Anantharaman, K. vRhyme enables binning of viral genomes from metagenomes. *Nucleic Acids Res.* **50**, e83–e83 (2022).
8. Johansen, J. et al. Genome binning of viral entities from bulk metagenomics data. *Nat. Commun.* **13**, 965 (2022).
9. Schulz, F., Abergel, C. & Woyke, T. Giant virus biology and diversity in the era of genome-resolved metagenomics. *Nat. Rev. Microbiol.* **20**, 721–736 (2022).
10. Bosmon, T., Abergel, C. & Claverie, J.-M. 20 years of research on giant viruses. *npj Viruses* **3**, 9 (2025).
11. Aylward, F. O., Moniruzzaman, M., Ha, A. D. & Koonin, E. V. A phylogenomic framework for charting the diversity and evolution of giant viruses. *PLoS Biol.* **19**, e3001430 (2021).
12. Ha, A. D., Moniruzzaman, M. & Aylward, F. O. Assessing the biogeography of marine giant viruses in four oceanic transects. *ISME Commun.* **3**, 43 (2023).
13. Endo, H. et al. Biogeography of marine giant viruses reveals their interplay with eukaryotes and ecological functions. *Nat. Ecol. Evol.* **4**, 1639–1649 (2020).
14. Liang, J.-L. et al. Unraveling the habitat preferences, ecological drivers, potential hosts, and auxiliary metabolism of soil giant viruses across China. *Microbiome* **12**, 136 (2024).
15. Schulz, F. et al. Giant virus diversity and host interactions through global metagenomics. *Nature* **578**, 432–436 (2020).
16. Moniruzzaman, M., Martinez-Gutierrez, C. A., Weinheimer, A. R. & Aylward, F. O. Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. *Nat. Commun.* **11**, 1710 (2020).
17. Minch, B. & Moniruzzaman, M. Expansion of the genomic and functional diversity of global ocean giant viruses. *npj Viruses* **3**, 32 (2025).
18. Weinheimer, A. R. & Aylward, F. O. Infection strategy and biogeography distinguish cosmopolitan groups of marine jumbo bacteriophages. *ISME J.* **16**, 1657–1667 (2022).
19. Al-Shayeb, B. et al. Clades of huge phages from across Earth's ecosystems. *Nature* **578**, 425–431 (2020).
20. Chen, L.-X. et al. Large freshwater phages with the potential to augment aerobic methane oxidation. *Nat. Microbiol.* **5**, 1504–1515 (2020).
21. Mendoza, S. D. et al. A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases. *Nature* **577**, 244–248 (2020).
22. Murray, N. J. et al. The global distribution and trajectory of tidal flats. *Nature* **565**, 222–225 (2019).
23. He, Q. et al. Coastal wetland resilience through local, regional and global conservation. *Nat. Rev. Biodivers.* **1**, 50–67 (2025).
24. Crump, B. C. & Bowen, J. L. The microbial ecology of estuarine ecosystems. *Annu. Rev. Mar. Sci.* **16**, 335–360 (2024).
25. Caron, D. A. et al. Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nat. Rev. Microbiol.* **15**, 6–20 (2017).
26. Aylward, F. O. & Moniruzzaman, M. ViralRecall—a flexible command-line tool for the detection of giant virus signatures in 'Omic Data. *Viruses* **13**, 150 (2021).
27. Ji, M. et al. Biodiversity of mudflat intertidal viromes along the Chinese coasts. *Nat. Commun.* **15**, 8611 (2024).
28. Perini, L. et al. Giant viral signatures on the Greenland ice sheet. *Microbiome* **12**, 91 (2024).
29. van den Berg, D. F. & Brouns, S. J. Phage tRNAs: decoding the enigma. *Trends Microbiol.* **33**, 1121–1131(2025).
30. Ha, A. D. & Aylward, F. O. Automated classification of giant virus genomes using a random forest model built on trademark protein families. *npj Viruses* **2**, 9 (2024).
31. Gilchrist, C. L. & Chooi, Y.-H. Clinker & clustermap. js: automatic generation of gene cluster comparison figures. *Bioinformatics* **37**, 2473–2475 (2021).
32. Koonin, E. V. & Yutin, N. Evolution of the large nucleocytoplasmic DNA viruses of eukaryotes and convergent origins of viral gigantism. *Adv. Virus Res.* **103**, 167–202 (2019).
33. Li, C. et al. Gastrointestinal jumbo phages possess independent synthesis and utilization systems of NAD+. *Microbiome* **12**, 268 (2024).
34. Low, S. J., Džunková, M., Chaumeil, P.-A., Parks, D. H. & Hugenholtz, P. Evaluation of a concatenated protein phylogeny for classification of tailed double-stranded DNA viruses belonging to the order Caudovirales. *Nat. Microbiol.* **4**, 1306–1315 (2019).
35. Liao, H. et al. Evolutionary diversification and succession of soil huge phages in glacier foreland. *Microbiome* **13**, 18 (2025).
36. Yuan, Y. & Gao, M. Jumbo bacteriophages: an overview. *Front. Microbiol.* **8**, 403 (2017).
37. Zhang, L., Meng, L., Fang, Y., Ogata, H. & Okazaki, Y. Spatiotemporal dynamics of giant viruses within a deep freshwater lake reveal a distinct dark-water community. *ISME J.* **18**, wrae182 (2024).

38. Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C. & Martiny, J. B. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat. Rev. Microbiol.* **10**, 497–506 (2012).
39. Anderson, M. J. et al. Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecol. Lett.* **14**, 19–28 (2011).
40. Gregory, A. C. et al. Marine DNA viral macro-and microdiversity from pole to pole. *Cell* **177**, 1109–1123. e1114 (2019).
41. Ji, M. et al. Temporal turnover of viral biodiversity and functional potential in intertidal wetlands. *npj Biofilms Microbiomes* **10**, 48 (2024).
42. Duffy, S., Shackelton, L. A. & Holmes, E. C. Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* **9**, 267–276 (2008).
43. Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M. & Belshaw, R. Viral mutation rates. *J. Virol.* **84**, 9733–9748 (2010).
44. Edwards, K. F., Steward, G. F. & Schvarcz, C. R. Making sense of virus size and the tradeoffs shaping viral fitness. *Ecol. Lett.* **24**, 363–373 (2021).
45. Lynch, M. et al. Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.* **17**, 704–714 (2016).
46. Chaudhari, H. V., Inamdar, M. M. & Kondabagil, K. Scaling relation between genome length and particle size of viruses provides insights into viral life history. *iScience* **24**, 102452 (2021).
47. Vieira, H. H. et al. Isolation of a widespread giant virus implicated in cryptophyte bloom collapse. *ISME J.* **18**, wrae029 (2024).
48. Moniruzzaman, M. et al. Virologs, viral mimicry, and virocell metabolism: the expanding scale of cellular functions encoded in the complex genomes of giant viruses. *FEMS Microbiol. Rev.* **47**, fuad053 (2023).
49. Zhou, J. & Ning, D. Stochastic community assembly: Does it matter in microbial ecology? *Microbiol. Mol. Biol. Rev.* **81**, 00002–00017 (2017).
50. Ning, D., Deng, Y., Tiedje, J. M. & Zhou, J. A general framework for quantitatively assessing ecological stochasticity. *Proc. Natl. Acad. Sci.* **116**, 16892–16898 (2019).
51. Stegen, J. C. et al. Quantifying community assembly processes and identifying features that impose them. *ISME J.* **7**, 2069–2079 (2013).
52. Ning, D. et al. A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. *Nat. Commun.* **11**, 4717 (2020).
53. Devoto, A. E. et al. Megaphages infect Prevotella and variants are widespread in gut microbiomes. *Nat. Microbiol.* **4**, 693–700 (2019).
54. Flamholz, Z. N., Biller, S. J. & Kelly, L. Large language models improve annotation of prokaryotic viral proteins. *Nat. Microbiol.* **9**, 537–549 (2024).
55. Emerson, J. B. et al. Host-linked soil viral ecology along a permafrost thaw gradient. *Nat. Microbiol.* **3**, 870–880 (2018).
56. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
57. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
58. Kieft, K., Zhou, Z. & Anantharaman, K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**, 1–23 (2020).
59. Guo, J. et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 1–13 (2021).
60. Camargo, A. P. et al. Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* **42**, 1303–1312 (2024).
61. Nayfach, S. et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2021).
62. Kang, D. D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
63. Wang, H. et al. Complementary insights into gut viral genomes: a comparative benchmark of short-and long-read metagenomes using diverse assemblers and binners. *Microbiome* **12**, 260 (2024).
64. 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).
65. Edgar, R. C. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat. Commun.* **13**, 6968 (2022).
66. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
67. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
68. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
69. Camargo, A. P. et al. IMG/VR v4: an expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Res.* **51**, D733–D743 (2023).
70. 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. Eevol.* **38**, 5825–5829 (2021).
71. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
72. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
73. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
74. Nayfach, S. et al. Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. *Nat. Microbiol.* **6**, 960–970 (2021).
75. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
76. West, P. T., Probst, A. J., Grigoriev, I. V., Thomas, B. C. & Banfield, J. F. Genome-reconstruction for eukaryotes from complex natural microbial communities. *Genome Res.* **28**, 569–580 (2018).
77. Karlicki, M., Antonowicz, S. & Karnkowska, A. Tiara: deep learning-based classification system for eukaryotic sequences. *Bioinformatics* **38**, 344–350 (2022).
78. Mirdita, M., Steinegger, M., Breitwieser, F., Söding, J. & Levy Karin, E. Fast and sensitive taxonomic assignment to metagenomic contigs. *Bioinformatics* **37**, 3029–3031 (2021).
79. Sun, T.-W. et al. Host range and coding potential of eukaryotic giant viruses. *Viruses* **12**, 1337 (2020).
80. Uritskiy, G. V., DiRuggiero, J. & Taylor, J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 1–13 (2018).
81. Couvin, D. et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* **46**, W246–W251 (2018).
82. Roux, S. et al. iPHoP: An integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLoS Biol.* **21**, e3002083 (2023).
83. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 1–11 (2010).
84. Padilha, V. A., Alkhnbashi, O. S., Shah, S. A., de Carvalho, A. C. & Backofen, R. CRISPRcasIdentifier: machine learning for accurate

identification and classification of CRISPR-Cas systems. *GigaScience* **9**, giaa062 (2020).
85. Consortium, U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
86. Paysan-Lafosse, T. et al. InterPro in 2022. *Nucleic Acids Res.* **51**, D418–D427 (2023).
87. 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).
88. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
89. Oksanen, J. et al. Package 'vegan'. *Community Ecology Package version* **2**, 1–295 (2013).
90. Olm, M. R. et al. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nat. Biotechnol.* **39**, 727–736 (2021).
91. Gregory, A. C. et al. MetaPop: a pipeline for macro-and microdiversity analyses and visualization of microbial and viral metagenome-derived populations. *Microbiome* **10**, 49 (2022).

## Acknowledgements

This study contributes to the scientific plan of Ocean Negative Carbon Emission (ONCE) Program. The supported funding includes the National Natural Science Foundation of China [32371598, 31971446, 92051110 (Q.T.)], National Key Research and Development Program of China [2020YFA0607600 (Q.T.)], the Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (SML2024SP022, SML2024SP002), the Taishan Young Scholarship of Shandong Province, the Natural Science Foundations of Shandong Province (ZR2020YQ21, 2020ZLYS04), and the Distinguished Young Scholarship of Shandong University. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## Author contributions

The research was conceived by Q.T. Sample collection and physicochemical characterization were carried out by M.J., Y.L., M.W., X.L., and X.G. Data analysis and integration were done by M.J. and guided by Q.T. The manuscript and figures were prepared by M.J. and Q.T. All authors have read and approved the submitted version of manuscript.

## 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/s41467-026-71095-7.

**Correspondence** and requests for materials should be addressed to Qichao Tu.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.