Continental-scale multi-omics reveals a distinct microbial-viral biome in sandy beach ecosystems structured by tidal zonation

Xiyang Dong[1†], Feng Cai[1†*], Yingchun Han[1†], Chuwen Zhang[1†], Hongshuai Qi[1], Shaohua Zhao[1], Liang Wang[1], Zhong Pan[1], Yanbin Chen[1], Zhengran Li[1], Zijian Lu[1,2], Xiaomei Guo[1], Yiting Ji[1], Jianhui Liu[1], Shengjie Li[3], Chujin Ruan[2], Lu Zhang[4]

[1] Third Institute of Oceanography, Ministry of Natural Resources, Xiamen 361005, China

[2] College of Ocean & Earth Sciences, Xiamen University, Xiamen 361102, China

[3] State Key Laboratory of Submarine Geoscience, and School of Oceanography, Shanghai Jiao Tong University, Shanghai 200240, China

[4] Department of Health and Environmental Sciences, School of Science, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China

† These authors contributed equally to this work.

* Correspondence to: Feng Cai (caifeng@tio.org.cn).

Abstract

Sandy beaches are dynamic coastal interfaces shaped by strong physical forcing and intense exchange between marine and terrestrial environments, yet their microbiomes remain poorly resolved at the genomic scale. Here we present a genome-resolved survey of microbial and viral communities across sandy beaches spanning a continental-scale latitudinal gradient along the Chinese coastline. By integrating cross-shore sampling, coastal geochemistry and large-scale multi-omics, we generated 978 metagenomes, 63 viromes and 72 metatranscriptomes, reconstructing 13,337 metagenome-assembled genomes and 38,255 viral populations. Sandy beach microbiomes exhibit exceptionally high genomic novelty, with more than 90% of species-level genomes representing previously undescribed taxa, suggesting that permeable coastal sediments constitute a distinct microbial and viral reservoir. Tidal zonation emerged as a dominant ecological driver structuring microbial diversity, metabolic strategies and virus-host interactions across cross-shore gradients. Genome-resolved analyses revealed systematic metabolic shifts from oxic heterotrophy in supratidal sediments toward increasingly chemolithotrophic and autotrophic pathways toward the low-intertidal and subtidal zone. Sandy beach microbiomes further encode broad potential for hydrocarbon and plastic transformation, together with diverse biosynthetic and antibiotic resistance repertoires that may mediate microbial chemical interactions. Together, these findings identify sandy beaches as a previously under-recognized microbial-viral biome shaped by tidal forcing, providing insight into microbiome evolution and coastal ecosystem resilience under increasing anthropogenic pressure.

Introduction

Sandy beaches occupy approximately one-third of the world's ice-free coastline and represent one of the most intensively used land-sea interfaces[1]. Despite their apparent geomorphic simplicity, sandy beaches are highly dynamic systems shaped by waves, tides, and sediment transport that continually reorganize pore structure and hydrological connectivity across spatial and temporal scales. Permeable sands can function as advective biogeochemical reactors, facilitating rapid exchange among seawater, groundwater and the atmosphere[2]. Tidal pumping and wave-driven circulation generate steep gradients in oxygen availability, redox potential and nutrient concentrations, enabling intense microbial processing of carbon, nitrogen and trace elements before solutes are returned to coastal waters[3-5]. Because tidal elevation determines the frequency and duration of inundation, desiccation and oxygen exposure, cross-shore zonation provides a predictable physical template that may structure microbial metabolism and biogeochemical transformations along the land-sea interface.

Despite the recognized importance of permeable sediments for coastal biogeochemistry, the microbial and viral communities inhabiting sandy beaches remain poorly resolved, particularly at the genomic scale[6-8]. Previous studies have shown that microbial assemblages vary across intertidal microhabitats, reflecting gradients in hydrodynamic forcing, redox conditions and resource availability[9-14]. However, most investigations rely on single-site surveys or marker-gene approaches, limiting understanding of whether consistent ecological patterns emerge across coastlines spanning different climatic regimes, morphodynamic settings and environmental pressures. This limitation reflects the intrinsic properties of permeable beach sands, where rapid wetting-drying cycles, salinity fluctuations and ultraviolet exposure coincide with low biomass, strong microscale heterogeneity and advective redox oscillations, jointly selecting for metabolically flexible microorganisms while complicating genome-resolved characterization[15]. Viruses influence microbial mortality, horizontal gene transfer and host metabolic potential in many ecosystems[16]. However, they remain especially understudied in permeable coastal sediments, leaving unresolved how virus-host interactions contribute to microbial adaptation and biogeochemical functioning across tidal gradients.

Human activity introduces an additional layer of environmental complexity. Sandy beaches are among the most heavily visited coastal ecosystems and receive diverse anthropogenic inputs, including nutrients, trace metals, antibiotics, hydrocarbons and microplastics[16, 17]. These compounds may alter microbial resource availability, impose selective pressures on resistance and detoxification mechanisms, and influence horizontal gene transfer[18, 19]. Because contaminant accumulation and transport depend strongly on hydrodynamic flushing, sediment permeability and residence time, anthropogenic impacts are likely modulated by the same physical processes that structure microbial habitats. Understanding how physical forcing, environmental chemistry and human pressures interact to shape sandy-beach microbiomes therefore represents an important challenge for predicting coastal ecosystem functioning and resilience.

Here we present a continental-scale genome-resolved investigation of sandy beach microbiomes and viromes across 18 beaches spanning tropical to temperate climatic zones, contrasting morphodynamic settings and varying levels of human use along the Chinese coastline. By integrating cross-shore sampling, coastal chemistry and large-scale multi-omics analyses spanning sediment and water samples, we generated 978 metagenomes, 63 viromes and 72 metatranscriptomes, enabling reconstruction of 13,337 metagenome-assembled genomes and 38,255 viral populations. Our analyses identify sandy beaches as a distinct microbial-viral biome characterized by high genomic novelty, strong tidal structuring and broad functional potential for organic matter turnover, redox flexibility and pollutant transformation. These results provide a genomic framework for understanding how physical forcing structures microbiomes in permeable coastal sediments and establish a reference resource for predicting ecosystem responses to anthropogenic pressure and climate-driven coastal change.

Results

Environmental gradients across continental-scale sandy beach ecosystems

To characterize environmental variation across sandy beach ecosystems, we conducted a coordinated survey along the Chinese coastline spanning a latitudinal gradient from 18°N to 40°N. In total, 18 sandy beaches were investigated, including 15 urban recreational beaches located in major coastal tourism cities and three natural beaches

serving as reference sites (**Fig. 1a–b and Tables S1–S2**). These sites span tropical, subtropical and temperate climatic zones along the South China Sea, East China Sea, Yellow Sea and Bohai Sea coasts, capturing broad environmental variation across the Chinese coastline.

The surveyed beaches displayed pronounced variation in geomorphology and hydrodynamic regimes (**Fig. 1c–f and Tables S3–S4**). Based on beachface slope and sediment characteristics, the beaches were classified into reflective, intermediate and dissipative morphodynamic types. Beachface slopes ranged from steep reflective systems (~1/8–1/10) to gently sloping dissipative beaches (<1/40) (**Fig. 1c**). Dry-beach widths varied from ~14 to 102 m, whereas intertidal widths ranged from ~15 m to more than 300 m, indicating substantial cross-site variation in beach morphology (**Fig. 1d**). Hydrodynamic forcing also varied among sites, with tidal ranges of approximately 1.0–4.5 m and significant wave heights of ~0.28–1.33 m (**Fig. 1e**). Surface sediments were dominated by sand fractions across all sites, typically exceeding 80% of total sediment composition, with minor contributions from gravel and mud (**Fig. 1f**).

Environmental conditions varied markedly across the coastline. Surface-water temperature ranged from 23.8 °C in northern temperate regions to 34.0 °C at southern tropical beaches, whereas pH remained consistently alkaline (7.67–8.51) (**Fig. 1g and Table S5**). Nutrient concentrations, including total nitrogen, total phosphorus and dissolved inorganic nitrogen and phosphorus species, also varied substantially among beaches. Metal concentrations (Fe, Cu, Zn, Pb and Cr) also showed strong spatial variability and clear partitioning between dissolved and sediment-associated fractions (**Fig. 1h and Tables S5-S6**). Metal concentrations were generally higher in sediments than in the overlying water column. Antibiotics were widely detected across beaches **(Table S7**). Concentrations in seawater typically occurred at ng L^{-1} levels, whereas sediment concentrations were generally observed at pg g^{-1} levels. Within individual beaches, antibiotic concentrations were often higher in mid-intertidal sediments than in upper beach zones, consistent with an influence of tidal exchange and hydrodynamic processes on contaminant accumulation in sandy beach sediments.

Tidal zonation structures prokaryotic and viral communities across sandy beaches

To characterize community organization across sandy beach ecosystems, we analyzed

978 metagenomes together with 63 viromes, enabling integrated comparisons of prokaryotic and viral communities across tidal zones and environmental compartments. Across all samples, prokaryotic communities spanned 81 phyla, including 74 bacterial and 7 archaeal phyla (**Fig. 2a and Table S8**). Community composition was consistently dominated by *Pseudomonadota*, *Bacteroidota*, *Actinomycetota* and *Bacillota*, which together accounted for approximately 80% of total relative abundance across sites. Clear cross-shore patterns were observed: *Pseudomonadota* increased toward intertidal sediments, whereas *Actinomycetota* and *Bacillota* were relatively enriched in supratidal sands. Archaeal lineages, primarily *Thermoproteota* and *Thermoplasmatota*, were more abundant in lower intertidal sediments, consistent with stronger redox gradients and organic matter accumulation in periodically inundated zones[20]. Viral communities were dominated by fewer major taxonomic groups but were strongly dominated by *Caudoviricetes*, with additional contributions from *Faserviricetes*, *Megaviricetes* and *Tectiliviricetes* (**Fig. 2b, Fig. S1 and Table S9**). Viral assemblages also exhibited zonation across tidal habitats. *Faserviricetes* were more abundant in supratidal sands, whereas *Megaviricetes* contributed more prominently in several intertidal samples, although *Caudoviricetes* remained dominant across most sites.

Across the coastline, tidal zonation emerged as the strongest driver of microbial community structure. Prokaryotic α-diversity increased progressively from supratidal to subtidal zones (Simpson index, $P < 2.2 \times 10^{-16}$; **Fig. 2c, Fig. S2 and Tables S10–S11**), consistent with reduced desiccation stress and enhanced hydrological exchange in frequently inundated sediments[11]. Correspondingly, β-diversity analyses revealed clear compositional differentiation among tidal zones (ANOSIM $R = 0.27$, $P = 0.001$; **Fig. 2d and Fig. S3a**), with the largest dissimilarities observed between supratidal and lower intertidal communities. Viral diversity showed a similar but not identical cross-shore pattern. Viral α-diversity peaked in mid- and low-intertidal sediments (**Fig. 2e, Fig. S4 and Tables S12–S13**), and viral community turnover closely paralleled prokaryotic compositional shifts across tidal zones (ANOSIM $R = 0.27$, $P = 0.001$; **Fig. 2f and Fig. S5a**). Habitat matrix further structured microbial communities. Prokaryotic α-diversity was higher in water samples than in sediments, whereas water samples showed a narrower range of Simpson diversity values ($P < 2.2 \times 10^{-16}$; **Fig. 2g and Fig. S3b**). In contrast, viral diversity was significantly higher in sediments than in water ($P = 1.5 \times 10^{-10}$; **Fig. 2h and Fig. S5b**), consistent with the possibility that porous

sediments act as viral retention zones that enhance virus-host interactions[21].

Geographic gradients imposed additional structure across the coastline. Although prokaryotic α-diversity did not differ significantly between northern and southern beaches (Simpson index, $P = 0.23$; **Fig. 2i**), community composition exhibited clear regional separation (ANOSIM $R = 0.27$, $P = 0.001$; **Fig. S3c**), suggesting distance-decay relationships and climatic filtering[9]. Viral communities showed stronger geographic differentiation, with higher α-diversity and greater compositional turnover observed in northern sites (**Fig. 2j and Fig. S5c**). In contrast, anthropogenic disturbance exerted comparatively modest effects. Prokaryotic communities did not show significant shifts in diversity or composition between natural and recreational beaches (**Fig. 2k and Fig. S3e–f**), suggesting limited detectable effects of beach usage on prokaryotic community diversity and composition in this dataset. Viral richness, however, was lower at natural reference sites ($P = 4.6 \times 10^{-9}$; **Fig. 2l and Fig. S5e–f**), indicating that beach usage may be associated with differences in local virus-host dynamics[13, 22]. Mantel tests revealed that both prokaryotic and viral community structures were significantly associated with environmental variables (**Fig. 2m–n and Tables S14–S15**). Geographic position, physicochemical parameters, nutrients and trace metals collectively contributed to microbial biogeographic patterns, with viral communities showing particularly strong correlations with nutrient and metal gradients[23].

Sandy beaches harbor extensive and largely distinct microbial genomic diversity

To characterize the genomic diversity of sandy beach microbiomes, we reconstructed microbial genomes from 978 metagenomic assemblies, generating a comprehensive genome catalogue spanning sandy beach sediments and adjacent water samples. After dereplication at 95% average nucleotide identity (ANI) and quality filtering according to MIMAG standards[24], a total of 13,337 medium- to high-quality metagenome-assembled genomes (MAGs) were retained to construct the Sandbeach Microbiome Genome Catalogue (SMGC) (**Fig. 3 and Table S16**). These genomes exhibited a mean completeness of 86.7% and mean contamination of 2.4%, including 5,607 near-complete genomes, 6,494 high-quality genomes and 1,236 medium-quality genomes (**Fig. S6a**). Genome sizes ranged from 0.25 to 14.8 Mb, reflecting substantial life-history diversity among sandy beach microorganisms (**Fig. S6b**). The smallest genomes

were primarily affiliated with ultrasmall bacterial lineages such as *Patescibacteriota*, whereas the largest genomes belonged to *Myxococcota*, a group known for complex predatory and social behaviors[25, 26].

Taxonomic classification using the Genome Taxonomy Database (GTDB r226) revealed extensive phylogenetic diversity spanning 64 bacterial and 11 archaeal phyla (**Table S17**). Dominant bacterial phyla included *Pseudomonadota*, *Actinomycetota*, *Bacteroidota*, and *Patescibacteriota* (**Fig. 3a and Fig. S7**). Archaeal diversity was also substantial, with MAGs assigned to *Thermoproteota*, *Nanoarchaeota* and several additional archaeal lineages (**Figs. 3b and Fig. S8**). Despite this broad taxonomic representation, most genomes corresponded to previously undescribed species. Across major phyla, more than 80% of MAGs could not be assigned to known species-level taxa in the GTDB database (**Fig. 3c and Fig. S9**). Several dominant phyla, including *Pseudomonadota*, *Patescibacteriota*, *Actinomycetota* and *Bacteroidota*, showed substantial phylogenetic gains, ranging from 2.69% to 23.98% (**Fig. 3d**). Genomic novelty was strongly associated with habitat type. The relative abundance of novel species-representative genomes was significantly higher in sand-associated communities (mean 88.9%) than in overlying seawater (30.8%; $P < 0.001$) (**Fig. 3e**), suggesting that the heterogeneous microhabitats within sandy sediments may provide ecological niches that favor microbial diversification. Across tidal zones, novel species were most abundant in mid-intertidal sediments, where periodic tidal flushing and resource renewal create dynamic environmental conditions (**Fig. S10a**). In contrast, northern and southern regions showed no significant differences in novelty (**Fig. S10b**), suggesting that microhabitat heterogeneity within beaches exerts a stronger influence on microbial diversification than broad geographic gradients[10].

To place sandy beach microbiomes in a broader ecological context, we compared the SMGC with genome catalogues from other major environments under a unified taxonomic framework (**Fig. 3f**). The SMGC exhibited an exceptionally high proportion of species-representative genomes representing previously uncharacterized taxa (91.7%), exceeding levels reported for freshwater[27], soil[28], and open-ocean microbiomes[29] and approaching those observed in extreme deep-sea environments such as hadal trenches[30]. Consistent with this pattern, cross-habitat comparisons revealed remarkably limited genomic overlap between sandy beach microbiomes and other

229 ecosystems. At a 95% ANI threshold, only 4.75% of SMGC genomes overlapped with
230 oceanic datasets, whereas overlaps with soil and freshwater microbiomes were
231 negligible (<0.25%; **Fig. S11a**). Most shared genomes originated from epipelagic
232 seawater communities (**Fig. S11b**), consistent with microbial exchange between coastal
233 waters and beach sediments driven by waves and tidal processes[31]. Together, these
234 results indicate that sandy beaches harbor extensive and largely distinct microbial
235 genomic diversity.

236 **Sandy beaches are a vast and previously unrecognized viral diversity hotspot**

237 To characterize viral diversity across sandy beach ecosystems, we reconstructed viral
238 genomes from 978 metagenomes together with 63 viromes (1,041 datasets in total),
239 generating a large-scale viral genome collection from sandy beach sediments and
240 adjacent coastal waters. In total, 207,056 viral contigs were identified and clustered into
241 38,255 viral operational taxonomic units (vOTUs) at 95% ANI, forming the Sandbeach
242 Virome Catalogue (SVC) (**Fig. 4a, Fig. S12 and Table S18**). Genome completeness
243 assessment indicated that the catalogue contains 6,627 complete, 6,683 high-quality,
244 and 24,920 medium-quality viral genomes (**Fig. S13**). The majority of viral genomes
245 ranged between 20 and 70 kb, consistent with genome sizes commonly reported for
246 double-stranded DNA viruses infecting prokaryotes[32].

247 Taxonomic annotation revealed that the vast majority of vOTUs belonged to the viral
248 realm *Duplodnaviria*, with the class *Caudoviricetes* overwhelmingly dominating the
249 viral assemblage (**Fig. 4a, Fig. S14 and Table S19**). Despite this dominance, taxonomic
250 resolution beyond the class level remained extremely limited. More than 95% of
251 annotated vOTUs could not be assigned to currently recognized viral families,
252 underscoring the extensive unexplored diversity of viruses inhabiting sandy beach
253 sediments. Metatranscriptomic analyses revealed transcriptional activity for a subset of
254 vOTUs in intertidal water samples. In total, 2,255 vOTUs exhibited detectable
255 transcriptional activity across the dataset (**Table S20**). Transcriptionally active viruses
256 were detected across both tidal zones, with higher transcriptional coverage observed in
257 water samples from the mid-intertidal zone (**Fig. 4b and Fig. S15**), suggesting
258 enhanced virus-host interactions in frequently inundated regions of the beach.

259 Structural annotation of viral proteins further highlighted extensive diversity in capsid

260 architectures (**Fig. 4c and Table S21**). A total of 367,320 major capsid proteins (MCPs)
261 were identified across the SVC, most of which displayed the canonical HK97-like fold
262 characteristic of tailed bacteriophages (n = 313,093). A smaller subset of MCPs
263 exhibited double jelly-roll (DJR) folds (n = 2,657), indicating structural diversity and
264 potential evolutionary links among distinct viral lineages. The remaining MCPs could
265 not be confidently assigned to known structural folds, suggesting the presence of
266 additional, yet uncharacterized capsid architectures[33]. To assess habitat specificity, the
267 SVC was compared with viral genome catalogues from multiple ecosystems, including
268 global ocean viromes[34], groundwater[35], soils[36], mudflat sediments[23] and oil reservoirs[37]
269 (**Fig. 4d**). Remarkably, only 26 vOTUs were shared across all environments, and
270 overlaps between sandy beach viruses and any individual ecosystem were extremely
271 limited. These patterns suggest that sandy beach viral communities represent a largely
272 distinct viral reservoir shaped by the unique physicochemical conditions and
273 hydrodynamic disturbances characteristic of permeable coastal sediments[38]. Viral
274 replication strategies were predicted for 6,540 vOTUs using deep-learning approaches[39]
275 (**Fig. S14c and Table S22**). Among these, lytic viruses (4,561 vOTUs) were more
276 prevalent than temperate viruses (1,979 vOTUs), consistent with the possibility that
277 rapid host turnover in dynamic sandy beach environments favors lytic infection
278 strategies[40].

279 Integration of viral genomes with microbial MAGs enabled reconstruction of virus-host
280 interaction networks. A total of 6,858 virus-host linkages were identified, connecting
281 6,655 vOTUs to 4,472 microbial genomes (**Fig. 4e and Table S23**). Predicted hosts
282 spanned 41 bacterial and 9 archaeal phyla, with most linkages associated with dominant
283 bacterial groups including *Pseudomonadota*, *Actinomycetota*, and *Bacteroidota*.
284 Notably, the distribution of virus-host associations did not scale directly with host
285 abundance across phyla (**Fig. 4f**), indicating that viral host range and infection patterns
286 are shaped by ecological interactions beyond simple host availability[41]. Furthermore,
287 based on CRISPR-informed iPHoP predictions[42], 14.6% of virus-host clusters
288 contained more than one viral or host genome, suggesting that multi-host associations
289 and shared host ranges occur in sandy beach viral networks[43].

290 **Tidal forcing drives metabolic stratification across sandy beach habitats**

291 To characterize the metabolic potential of sandy beach microbiomes, we constructed a

292 nonredundant gene catalogue comprising more than 4.08×10^8 protein-coding gene
293 clusters, a gene-catalogue size exceeding those reported for several previously
294 published oceans[44-46], soils[47], and glacier microbiome datasets[48] (**Fig. S16a**). Despite
295 annotation with KEGG, eggNOG, Pfam, and CAZy, approximately 40% of gene
296 clusters remained functionally uncharacterized (unannotated and of unknown function;
297 **Fig. S16**), indicating substantial unexplored metabolic potential[45]. Functional profiles
298 were dominated by heterotrophic carbon utilization and energy metabolism, with genes
299 related to carbohydrate and amino acid metabolism, energy production, and inorganic
300 ion transport accounting for >60% of annotated functions (**Fig. S16c**).

301 Genome-resolved analyses revealed a clear reorganization of microbial metabolic
302 strategies along the tidal gradient, with oxic heterotrophic processes dominating upper
303 beach sediments and chemolithotrophic and autotrophic pathways becoming
304 increasingly important toward the subtidal zone (**Fig. 5a and Table S24**). Genes
305 associated with complex organic matter degradation, including starch (*amyA*; mean 102
306 GPM), hemicellulose (*manA*; mean 99 GPM) and cellulose degradation (*bglB*; mean
307 106 GPM), were widely distributed and generally showed higher abundances in
308 supratidal and upper intertidal sediments (**Fig. S17a**). These substrates are further
309 processed through central carbon metabolism, as reflected by the widespread presence
310 of glycolysis (*pfkA*; mean 99 GPM) and tricarboxylic acid cycle genes (*gltA*; mean 417
311 GPM). Together with the high abundance of aerobic respiration genes (**Fig. S17b**),
312 particularly cytochrome c oxidase (*coxA*; mean 584 GPM), these patterns indicate that
313 oxic heterotrophic metabolism represents the primary energy-generating strategy in
314 upper beach sediments. These pathways were primarily associated with bacterial phyla
315 including *Pseudomonadota*, *Bacteroidota* and *Actinomycetota*, which together
316 contributed the majority of genes involved in organic carbon degradation and aerobic
317 respiration.

318 Toward the intertidal zone, microbial communities displayed increasing reliance on
319 redox-flexible metabolic pathways (**Fig. 5a**). Genes associated with alternative electron
320 acceptors, including denitrification (*napA*, *nirS*, *nirK*, *norB* and *nosZ*; mean 16-124
321 GPM), arsenite oxidation (*aoxA*; mean 13 GPM), arsenate reduction genes (*arsC*; mean
322 722 GPM) and formate oxidation (*fdhA*; mean 376 GPM), were widely detected and
323 tended to show relatively higher abundance in intertidal sediments (**Figs. S17c-d**).

These genes were mainly affiliated with *Pseudomonadota* and *Desulfobacterota* lineages, suggesting that these taxa contribute to respiration under fluctuating oxygen conditions (**Tables S5-S6**). In contrast, autotrophic and chemolithotrophic pathways became increasingly prominent toward the subtidal zone (**Fig. 5a**). RuBisCO genes (*rbcL*; mean 50 GPM), representing the Calvin–Benson–Bassham cycle, were significantly enriched in subtidal sediments (Kruskal–Wallis test, $P < 0.001$), while sulfur oxidation genes, including sulfide:quinone oxidoreductase (*sqr*; mean 155 GPM) and thiosulfate oxidation (*soxB*; mean 54 GPM), showed similar seaward enrichment (**Figs. S17f-g**). These pathways were predominantly associated with *Pseudomonadota* and *Campylobacterota*-related lineages, which include known chemolithoautotrophic microorganisms[49]. Phototrophic marker genes (*psaA* and *psbA*) were detectable but generally low in abundance, although relatively higher signals were observed in subtidal zones (**Fig. S17h**).

Phosphorus acquisition strategies were pervasive across all tidal zones (**Fig. S17e**). Alkaline phosphatase genes (*phoA*, *phoD* and *phoX*; mean 146 GPM), which mediate organic phosphorus mineralization, were substantially more abundant than acid phosphatases, while genes involved in mineral phosphate solubilization (*gcd*) and polyphosphate metabolism (*ppk* and *ppx*) were also widespread. These patterns indicate that microbial communities employ diverse strategies to access phosphorus under nutrient-limited conditions (**Table S6**), independent of tidal position[50].

Metatranscriptomic analyses further showed that these pathways were actively expressed in water samples from the mid- and low-intertidal zones, with broadly similar expression patterns between the two zones (**Fig. 5b–g and Table S25**). Transcripts associated with carbon fixation (*rbcL*) and sulfur oxidation (*sqr*) were consistently abundant in intertidal waters, supporting the ecological importance of chemolithotrophic and autotrophic processes in these environments[7]. These results indicate that tidal zonation structures not only metabolic potential but also in situ metabolic activity.

Beyond cellular microorganisms, sandy beach viruses may further shape these metabolic processes through auxiliary metabolic genes (AMGs). Among 38,255 identified vOTUs, 4,641 AMGs (143 gene families encoded by 2,556 vOTUs) were detected that overlap with dominant host metabolic pathways, including photosynthesis

356 (*psbA*), aerobic respiration (*coxA*, *coxB* and *atpA*), phosphorus metabolism (*phoD* and
357 *ppk*) and carbon metabolism genes (e.g., *icd*, *ppdK*, *xyn*, *pulA, csn*) (**Fig. 5h and Table**
358 **S26**). Several of these AMGs (n = 233, spanning 57 gene families; e.g., *psbA*, *atpA*,
359 *phoD*, *icd*, *ppdK*) also exhibited detectable transcriptional activity, with nearly 40% of
360 gene families being actively expressed, highlighting the widespread transcriptional
361 engagement of viral AMGs. This indicates that viral infection can influence actively
362 expressed pathways. The strong functional overlap between viral AMGs and host
363 metabolic pathways suggests that viruses may modulate and potentially amplify key
364 biogeochemical processes along the tidal gradient.

365 **Sandy beach microbiomes encode broad capacities for pollutant transformation**
366 **and microbial chemical interactions**

367 Hydrocarbon degradation genes were broadly represented, with 2,915 proteins
368 spanning 24 functional categories covering pathways for short- and medium-chain
369 alkanes, long-chain alkanes, monoaromatic hydrocarbons and polycyclic aromatic
370 hydrocarbons (**Fig. 6a and Table S27**). Most sequences (2,841 genes; 97.5%) were
371 associated with aerobic pathways, consistent with efficient oxygen renewal in
372 permeable sediments[51-53]. Key enzymes included AlkB, CYP153, AlmA, LadA,
373 MAHα/β and NdoB/NdoC, indicating capacity for oxidation of chemically diverse
374 petroleum-derived substrates. Among these, AlkB (mean: 31 GPM) and CYP153 (mean:
375 37 GPM) were particularly abundant and showed significant variation across tidal zones
376 (**Fig. 6b and Table S28**; $P < 0.001$). Moreover, these hydrocarbon-degradation genes
377 were actively expressed in water samples from the mid- and low-intertidal zones, with
378 CYP153 showing especially high expression (mean: 59 TPM; **Table S29**). Phylogenetic
379 reconstruction resolved nine AlkB clades, including a previously unrecognized lineage
380 lacking close reference homologues, and structural modelling showed strong similarity
381 to experimentally validated AlkB enzymes (PDB: 8F6T; TM-score = 0.98) with
382 conserved catalytic residues (**Figs. S18–S19**). Hydrocarbon degradation genes were
383 detected in 1,606 MAGs, including bacterial lineages from *Gammaproteobacteria*,
384 *Actinomycetota* and *Binatia*, as well as archaeal homologues affiliated with
385 *Thermoplasmatota* and *Halobacteriota* (**Table S27**), extending the phylogenetic
386 breadth of aerobic hydrocarbon oxidation[54-56]. In total, 38 MAGs encoded multiple
387 hydrocarbon degradation modules, frequently combining *alkB*, *CYP153*, *MAHα/β* and

388 *ndo* genes, with co-localized gene clusters suggesting coordinated oxidation of multiple
389 hydrocarbon classes **(Fig. 6c and Fig. S20**).

390 In parallel, the SMGC revealed an extensive repertoire of putative plastic-degrading
391 enzymes, indicating a broader xenobiotic transformation network in sandy beach
392 ecosystems (**Fig. S21 and Table S27**). We identified 11,929 polyethylene terephthalate
393 (PET) hydrolases, 10,528 polylactic acid (PLA) depolymerases, 9,332 serine
394 hydrolases, 5,583 polyhydroxyalkanoates (PHA) depolymerases, 4,765 esterases, and
395 45 cutinases, collectively targeting diverse polyester and aliphatic polymers[57]. These
396 genes were widely distributed across 10,949 bacterial MAGs spanning 64 phyla,
397 dominated by *Pseudomonadota*, *Actinomycetota* and *Bacteroidota*, with additional
398 archaeal homologues (n = 327) primarily affiliated with *Thermoplasmatota*, indicating
399 broad phylogenetic involvement in plastic transformation. Except for cutinases, all
400 enzyme classes showed significant variation across tidal zones (**Fig. 6d and Table S30**;
401 $P < 0.001$), with highest abundances generally observed in intertidal sediments (e.g.,
402 PET hydrolases mean 252 GPM), consistent with enhanced accumulation of
403 microplastics and organic substrates at the land-sea interface[58]. Furthermore, intertidal
404 water metatranscriptomes confirmed *in situ* expression of the plastic-degrading genes
405 (e.g., PET hydrolases, mean 84 TPM; **Table S31**). Structural prediction of
406 representative enzymes yielded TM-scores of 0.82–0.98 (**Fig. 6e**), revealing conserved
407 α/β-hydrolase folds and catalytic residues characteristic of functional polyester
408 hydrolases[59, 60]. Notably, plastic- and hydrocarbon-degrading genes frequently co-
409 occurred within the same genomes (**Table S27**), indicating multi-substrate metabolic
410 versatility and suggesting adaptation to chemically heterogeneous pollutant mixtures
411 shaped by tidal transport and retention[61, 62].

412 Beyond xenobiotic degradation, sandy beach microbiomes encoded extensive
413 biosynthetic capacities, shaping microbial chemical ecology. Across the SMGC, we
414 identified 58,956 biosynthetic gene clusters (BGCs), including 57,546 bacterial and
415 1,410 archaeal clusters, spanning 29 biosynthetic classes, including both canonical and
416 hybrid pathways (**Figs. S22–S23 and Table S32**). The dominant BGC categories
417 included RiPPs (14,842 clusters), NRPS (12,288), terpenes (11,845), and PKSs (8,873),
418 together accounting for more than 80% of bacterial biosynthetic diversity. Near-
419 complete genomes (completeness ≥90%, contamination ≤5%) and BGC-rich (≥15

420 BGCs) genomes (n = 111) also contained extracellular polysaccharide (exoPS)
421 biosynthesis clusters, of which 81% corresponded to uncharacterised Wzy-dependent
422 or ABC transporter-dependent pathways (**Table S33**), particularly in *Acidobacteriota*
423 and *Myxococcota*, potentially supporting adhesion, aggregation and biofilm formation
424 in dynamic sediments[63]. BGC distribution largely followed community composition
425 (**Fig. 6f**), with major contributions from *Pseudomonadota*, *Actinomycetota*,
426 *Acidobacteriota* and *Myxococcota*, the latter showing the highest biosynthetic density
427 (up to 14 BGCs per genome), consistent with strong metabolic specialization.

428 Functionally, many BGCs were associated with adaptation to environmental stress
429 characteristic of sandy beaches (**Fig. 6g and Table S34**). Osmoprotectant pathways
430 such as ectoine[64] (mean 40,059 GPM) and N-acetylglutaminylglutamine amide[65]
431 (NAGGN; mean 10,412 GPM) were highly abundant, consistent with fluctuating
432 salinity and recurrent wetting-drying cycles, while pigment-associated arylpolyene
433 clusters[66] (25,874 GPM) and redox cofactor pathways (mean 15,569 GPM) support
434 tolerance to oxidative and UV stress. Metal acquisition pathways, including
435 siderophores (mean 5,995 GPM) and metallophores (mean 3,341 GPM), indicate
436 adaptation to nutrient limitation in permeable sands, and polyunsaturated fatty acid
437 pathways (mean 211 GPM) suggest membrane-level responses to environmental
438 variability. Antimicrobial-associated BGCs were also prevalent, including lanthipeptide,
439 lassopeptide and thiopeptide pathways[67], with lassopeptides (mean 31,173 GPM)
440 representing the most abundant antimicrobial class, together with phosphonate
441 pathways (mean 13,784 GPM). Metatranscriptomic analyses further showed that stress-
442 adaptation BGCs exhibited higher expression than antimicrobial clusters (**Fig. 6g and**
443 **Table S35**; $P < 0.001$), with particularly high transcript abundance for ectoine (mean
444 129,151 TPM), NAGGN (mean 20,395 TPM) and arylpolyene (mean 19,351 TPM)
445 pathways, indicating strong selection imposed by salinity fluctuation, desiccation-
446 rewetting dynamics and oxidative stress.

447 Microbial chemical defense capacity was further reflected in the extensive antibiotic
448 resistome (**Fig. 6h**). The SMGC contained 29,067 antibiotic resistance genes (ARGs)
449 identified in 10,505 MAGs spanning 62 phyla, indicating widespread adaptation to
450 bioactive compounds derived from both microbial competition and anthropogenic
451 inputs (**Table S36**). Archaeal genomes encoded 304 ARGs, whereas bacterial genomes

452 carried 28,763 ARGs across 28 resistance classes, dominated by glycopeptide (39.1%),
453 multidrug (24.7%) and aminoglycoside (8.5%) resistance genes. ARGs frequently co-
454 occurred with BGCs within the same genomes (**Table S37**), particularly in
455 *Actinomycetota* (1,163 MAGs) and *Pseudomonadota* (3,137 MAGs), indicating tight
456 evolutionary coupling between secondary metabolite production and resistance
457 capacity[68]. Environmental measurements further detected 12 antibiotics in 276 sand
458 samples and 14 antibiotics in 54 water samples (**Table S7**), including azithromycin (up
459 to 101 pg g^{-1}) and sulfaquinoxaline (up to 60 pg g^{-1}) in sand, consistent with persistent
460 exposure to antimicrobial compounds.

461 **Discussion**

462 Sandy beaches are physically dynamic environments in which hydrodynamic forcing
463 strongly shapes sediment structure, oxygen availability and solute transport. Across a
464 continental-scale dataset spanning diverse climatic and morphodynamic settings, tidal
465 elevation consistently structured microbial diversity, metabolic organization and virus-
466 host interactions. This repeatable zonation pattern indicates that cross-shore position
467 provides a primary constraint on microbiome assembly in permeable sediments,
468 outweighing the effects of geographic variation and differences in beach usage detected
469 in this dataset[7, 15, 17]. Because inundation frequency controls oxygen exposure,
470 porewater residence time and resource renewal, tidal forcing generates predictable
471 environmental gradients that organize microbial community composition along the
472 land-sea interface. The metabolic profiles observed across tidal zones are consistent
473 with these physical constraints. Upper beach sediments were dominated by aerobic
474 heterotrophic pathways associated with rapid turnover of labile organic matter, whereas
475 sediments toward the subtidal zone showed increasing representation of
476 chemolithotrophic and autotrophic processes linked to fluctuating redox conditions.
477 Permeable sands differ from diffusion-dominated sediments in that advective exchange
478 rapidly redistributes electron donors and acceptors, creating transient oxic-anoxic
479 interfaces at millimetre scales[17, 69]. Such conditions favor metabolically flexible
480 microorganisms capable of switching between respiratory strategies, explaining the
481 coexistence of aerobic respiration, denitrification and sulfur oxidation pathways across
482 short spatial distances.

483 At the genomic scale, sandy beach sediments harbor exceptionally high diversity and
484 novelty, consistent with a distinct evolutionary reservoir rather than a simple admixture
485 of marine and terrestrial microorganisms. The limited overlap between the SMGC and
486 existing environmental genome collections indicates that permeable sediments support
487 specialized lineages adapted to dynamic microscale gradients, repeated disturbance and
488 pulsed resource supply. Elevated novelty in mid-intertidal sediments further supports
489 the idea that fluctuating ecotones promote diversification by generating heterogeneous
490 niches and episodic selection pressures[11, 14, 70]. Viral communities exhibit similarly high
491 levels of novelty and habitat specificity, with limited overlap with global virome
492 datasets. The strong correspondence between viral assemblages and host distributions,
493 together with extensive virus-host linkages and auxiliary metabolic genes, supports the
494 view that viruses are a major component of sandy beach ecosystems, potentially
495 influencing microbial turnover, metabolic evolution and biogeochemical cycling[71, 72].

496 In addition to core carbon and nutrient cycling, sandy beach microbiomes encode broad
497 potential for hydrocarbon and plastic transformation. The predominance of aerobic
498 hydrocarbon degradation genes is consistent with efficient oxygen renewal driven by
499 tidal pumping, while the widespread occurrence of candidate polyester hydrolases is
500 consistent with adaptation to recurrent exposure to synthetic polymers accumulating at
501 the land-sea interface[73-75]. The frequent co-occurrence of hydrocarbon and plastic
502 degradation genes within the same genomes is consistent with metabolic versatility
503 under chemically mixed conditions. These traits are consistent with the role of
504 permeable sediments as zones of intense solute exchange, where both natural organic
505 matter and anthropogenic compounds are repeatedly transported and retained.
506 Biosynthetic gene clusters and antibiotic resistance genes were also widely distributed,
507 reflecting the importance of both abiotic stress tolerance and microbial interactions in
508 sandy sediments. High abundances of osmoprotectant and oxidative stress-related
509 pathways indicate adaptation to fluctuating salinity, desiccation-rewetting cycles and
510 strong irradiance. The co-occurrence of biosynthetic and resistance genes within the
511 same genomes suggests a potential linkage between secondary metabolite production
512 and resistance capacity, consistent with competitive interactions in spatially structured
513 microbial communities. The detection of diverse antibiotic residues in sand and water
514 further indicates that microbial communities are recurrently exposed to bioactive
515 compounds derived from both natural processes and human activities[18, 38, 76].

Together, these results indicate that sandy beaches represent physically structured microbial ecosystems characterized by high genomic novelty, strong metabolic flexibility and broad functional potential for processing chemically diverse substrates. By integrating cross-shore sampling with genome-resolved multi-omics analyses, this study provides a reference framework for investigating microbiome dynamics in permeable coastal sediments and for assessing how physical forcing and environmental variability jointly shape microbial community function in permeable coastal sediments.

Materials and Methods

Study sites and sampling design

We conducted a coordinated coast-wide survey of sandy beaches along the Chinese coastline during August-September 2024, corresponding to the summer tourism season. The survey spanned a latitudinal gradient from 18° N to 40° N and covered tropical, subtropical and temperate climatic zones across the South China Sea, East China Sea, Yellow Sea, and Bohai Sea. In total, 18 sandy beaches were investigated, including 15 urbanized recreational beaches and three natural reference beaches distributed across temperate, subtropical and tropical regions, respectively (**Fig. 1 and Table S1**). Sampling sites were selected to ensure broad geographic coverage of China's coastline, representation of beaches experiencing high human activity, inclusion of natural reference beaches for baseline comparison and coverage of multiple beach morphodynamic types including reflective, intermediate and dissipative beaches. At each beach, depending on beach width / accessibility / tidal exposure, 7-10 cross-shore transects were established extending from the supratidal zone through the high-, mid- and low-intertidal zones, resulting in 174 standardized transects. Because the lower intertidal zone was accessible only during brief periods around extreme low tide, sampling in this zone was temporally constrained and yielded fewer samples.

Field sampling and sample processing

Sediment sampling was conducted during low tide along each transect. Surface sediments (0–5 cm depth) were collected using metal-free polypropylene scrapers after removing visible debris such as shell fragments and plant material. Samples for omics

545 analyses were transferred into sterile polypropylene tubes or pre-combusted aluminum
546 containers, flash-frozen in liquid nitrogen *in-situ* and stored at −80 °C until nucleic acid
547 extraction. Sediment samples for physicochemical analyses were transported under
548 cooled and dark conditions, air-dried and sieved in the laboratory prior to analysis.

549 Water samples were collected adjacent to each transect using pre-cleaned
550 polycarbonate or glass sampling bottles. For microbial metagenomic and
551 metatranscriptomic analyses, water samples were sequentially prefiltered through 200
552 μm and 20 μm meshes to remove large debris and plankton and subsequently filtered
553 through 0.22 μm polyethersulfone membranes to capture microbial cells. For viral
554 metagenomes and viral metatranscriptomes, water samples were prefiltered using the
555 same meshes and then filtered through 0.1 μm polyethersulfone membranes under low
556 vacuum pressure. Filters were immediately frozen in liquid nitrogen and stored at
557 −80 °C until nucleic acid extraction. For chemical analyses, water samples were filtered
558 through 0.45 μm membranes to obtain the dissolved fraction. Samples for metal
559 analysis were acidified to pH <2 using ultrapure HNO_3 (65–69%, trace-metal grade)
560 and stored at 4 °C in the dark. Nutrient samples were refrigerated and analyzed within
561 24 h, whereas antibiotic samples were frozen immediately after filtration.

562 **Morphodynamic characterization**

563 Cross-shore transects were surveyed using a STONEX S9II PRO RTK-GPS system
564 with a horizontal accuracy of ±8 mm and vertical accuracy of ±15 mm. Beach
565 morphology was quantified using beach slope, dry-beach width and intertidal width.
566 Based on these characteristics and surf-zone features, beaches were classified as
567 reflective, intermediate or dissipative systems. Hydrodynamic conditions including
568 tides and waves were characterized using field observations, nearby instrumental
569 records and numerical simulations. Wave parameters including significant wave height,
570 wave period and dominant wave direction were simulated using the TOMAWAC wave
571 model[77]. Bathymetry was derived from the GEBCO global dataset combined with
572 CMAP coastal bathymetric data. Sediment grain-size distributions were measured
573 using dry sieving with a POWTEQ SS2000 vibrating sieve shaker after oven drying
574 samples at 60 °C.

575 **Physicochemical analyses**

576 In situ water physicochemical parameters including temperature and pH were measured
577 using a YSI ProDSS multiparameter probe calibrated prior to deployment. Nutrient
578 samples were collected in 500-mL polyethylene (PE) bottles and frozen before
579 measurement. Dissolved inorganic nitrogen (nitrate, nitrite, and ammonium) and
580 phosphate concentrations were analyzed using a 723C spectrophotometer
581 (Shanghaijingke®, China) following standardized colorimetric methods. Ammonium
582 was measured using the Nessler reagent colorimetric method (HJ 535-2009). Total
583 nitrogen was determined after alkaline persulfate digestion by UV spectrophotometry
584 (HJ 636-2012). Total phosphorus was measured after digestion using the ammonium
585 molybdate spectrophotometric method (GB/T 11893-1989). Dissolved silicate was
586 analyzed using a standard molybdenum blue spectrophotometric method. Metal
587 concentrations were determined using inductively coupled plasma mass spectrometry
588 (ICP-MS; Agilent 7900). Water samples were filtered through 0.45 μm membranes and
589 acidified prior to analysis. Sediment samples were digested using mixed-acid digestion
590 (HNO_3-HCl-HF) in a microwave digestion system and subsequently analyzed by ICP-
591 MS. Antibiotics in water samples were extracted using solid-phase extraction with
592 Oasis HLB cartridges and quantified using liquid chromatography-tandem mass
593 spectrometry (UPLC-MS/MS; Thermo TSQ) operating in multiple reaction monitoring
594 (MRM) mode. Sediment antibiotics were extracted from freeze-dried sediments using
595 methanol-dichloromethane mixtures followed by SPE purification prior to UPLC-
596 MS/MS analysis.

597 **DNA extraction and metagenomic sequencing**

598 Metagenomic DNA was extracted from 978 sand and water samples. For water
599 metagenomes, the cellular fraction was collected on 0.22 μm filters prior to DNA
600 extraction. For an additional 63 water samples used for virome analysis, seawater was
601 first filtered through 0.22 μm membranes to remove cells, amended with 0.1% ferric
602 chloride for 30 min to induce viral flocculation, and then filtered to recover virus-
603 enriched material for DNA extraction. To optimize DNA recovery from samples with
604 different moisture conditions, moist samples (>20% moisture content) were extracted
605 using the E.Z.N.A. Soil DNA Kit (Omega Bio-Tek, Norcross, GA, USA), whereas drier
606 samples were extracted using the DNeasy PowerSoil Pro Kit (QIAGEN, Netherlands).

607 Samples yielding <500 ng DNA were re-extracted and pooled to obtain at least 1 μg
608 total DNA. DNA concentration and purity were evaluated spectrophotometrically.
609 Samples with $OD_{260/280}$ ratios of 1.8–2.2 and $OD_{260/230}$ values ≥2.0 were used for library
610 preparation. Sequencing libraries were prepared using a standard genomic DNA library
611 preparation procedure workflow and sequenced by Biozeron Biological Technology
612 Co., Ltd. (Shanghai, China) on the MGI-T7 platform in paired-end mode (2 × 150 bp).

613 **Community profiling, metagenomic assembly and genome reconstruction**

614 Raw reads were quality-filtered and adaptor-trimmed using Trimmomatic (v0.39)[78].
615 Clean reads were then profiled using sylph (v0.8.0)[79] with default parameters against
616 the GTDB r220-c200 and IMG/VR c200 reference databases to characterize the
617 composition of prokaryotic and viral communities, respectively; prokaryotic profiling
618 was performed on 978 samples, whereas viral profiling included these samples together
619 with an additional 63 virome samples. Clean reads were also assembled using
620 MEGAHIT (v1.1)[80] with the parameter "--min-contig-len 500". Contig coverage was
621 calculated using jgi_summarize_bam_contig_depths from MetaBAT2 (v2.12.1)[81] based
622 on sorted BAM files generated with BWA-MEM (v0.7.17)[82] and SAMtools (v1.21)[83].
623 Contigs were binned using MetaBAT2 (v2.12.1) and SemiBin2 (v 2.1.0)[84]. Bins
624 generated by the two methods were refined using the bin_refinement module in
625 MetaWRAP (v1.3.2; parameters: -c 50 -x 10)[85]. Species-level dereplication was
626 performed using dRep (v3.5.0; parameters: -comp 50 -con 10 -sa 0.95 --S_algorithm
627 skani)[86]. Genome quality was assessed using CheckM2 (v1.1.0)[87]. MAGs were retained
628 if they met the criterion of medium or higher quality of the MIMAG standard
629 (completeness ≥50%, contamination ≤10%, and quality score QS = completeness − 5 ×
630 contamination ≥50). In total, 13,337 MAGs were retained for downstream analyses and
631 designated as the SMGC.

632 Taxonomic annotation of MAGs was performed using GTDB-Tk (v2.4.1)[88] against
633 GTDB release 226. Phylogenetic diversity within each phylum was assessed using
634 GenomeTreeTk (v0.1.6; https://github.com/dparks1134/GenomeTreeTk). Relative
635 abundances of MAGs were calculated using CoverM (v0.7.0)[89] with reads-per-
636 kilobase-million (RPKM) normalization. Comparative analyses were performed
637 between the SMGC and published genome catalogues from other ecosystems, including
638 the Global Ocean Microbiome Catalogue (GOMC)[29], the Soil Microbiome Catalogue[28],

639 the Genome-Resolved Open Watersheds Database (GROWdb)[27], the Mariana Trench
640 Environment and Ecology Research dataset (MEER)[30] and the Global Cold Seep
641 Microbiome Catalogue[46].

642 **Viral genome identification and characterization**

643 Potential viral sequences were identified from assembled contigs across 978
644 metagenomes and 63 viromes using geNomad (v1.8.1; end-to-end) [90] and VirSorter2
645 (v2.2.4; default parameters)[91]. Viral completeness and contamination were evaluated
646 using CheckV (v1.0.3; database v1.5)[92], and contigs with estimated completeness ≥50%
647 were retained. Viral contigs were clustered into species-level vOTUs using a 95%
648 average nucleotide identity threshold and 85% minimum aligned fraction, following
649 CheckV recommendations. This yielded 38,255 vOTUs, collectively defined as the
650 SVC. Taxonomic assignments were generated using geNomad (v1.8.1)[90] against ICTV
651 release MSL39. Open reading frames were predicted using Prodigal (v2.6.3; -p meta)
652 and functionally annotated using DRAM-v (v1.3.5)[93]. Major capsid proteins were
653 identified by searching predicted viral proteins against the PDB100 database[94] using
654 Foldseek[95] in easy-search mode with an E-value cutoff of 1×10^{-5}. Auxiliary metabolic
655 genes were inferred from DRAM-v annotations, and functional filters were applied to
656 obtain high-confidence AMGs. Viral lifestyle was predicted using phabox (v2.1.13)
657 with the phatyp task[39]. Viral novelty and environmental distribution were assessed by
658 comparing SVC vOTUs with published virome catalogues from mudflat intertidal
659 sediments[23] (n = 20,102), soils[36] (n = 80,750), groundwater[35] (n = 280,420), the Global
660 Ocean Viromes 2.0[34] (GOV2, n = 488,133), and oil reservoirs[37] (n = 3,886)[34] using
661 BLASTn (v2.16.0; -perc_identity 90 -qcov_hsp_perc 90)[96].

662 **Virus-host association prediction**

663 Virus-host associations were predicted using iPHoP (v1.4.1)[42] with the iPHoP reference
664 database and the 13,337 MAGs recovered in this study. The pipeline integrated
665 sequence-based and composition-based approaches, including BLASTn searches
666 against host and CRISPR spacer databases and host prediction by RaFAH (v0.3)[97],
667 WIsH (v1.0)[98], VirHostMatcher-Net[99], and PHP[100]. Virus-host associations supported
668 by multiple independent lines of evidence were considered high-confidence, with
669 CRISPR spacer matches treated particularly strong evidence. In cases where multiple

670 candidate hosts were detected, the host with the highest confidence score was selected
671 as the final assignment.

672 **Gene prediction and functional annotation**

673 Protein-coding sequences were predicted from assembled contigs >1,000 bp using
674 Prodigal (v2.6.3; -meta)[101]. The resulting sequences were clustered hierarchically at 90%
675 amino acid identity and 90% coverage using MMseqs2 v13.45111 with the easy-linclust
676 module (-c 0.9 -min-seq-id 0.9 --cov-mode 1)[102]. This generated a non-redundant gene
677 catalogue comprising 407,984,259 representative clusters. Representative amino acid
678 sequences were functionally annotated using eggNOG-mapper (v2.1.12)[103, 104].
679 Functional assignments included eggNOG 5.0, Pfam 33.1, KEGG, EC, GO and CAZy.

680 Protein-coding sequences of genomes in SMGC were predicted with Prodigal (v2.6.3)
681 in single mode, and then annotated with eggNOG-mapper (v2.1.12) following the same
682 procedure as described above. Annotated KO numbers were used for inferring the
683 metabolic pathway encoded in each MAG. Hydrocarbon degradation genes were
684 identified using hmmsearch against 37 HMM profiles with noise cutoff from the CANT-
685 HYD database[105]. Plastic-degrading proteins were identified using hmmsearch against
686 the RemeDB database with default parameters[106]. Secondary metabolite BGCs were
687 identified using antiSMASH (v7.1)[107] with the parameters "-fullhmmer -cb-general -
688 cb-subclusters -cb-knownclusters -genefinding-tool prodigal-m -clusterhmmer -asf -
689 smcog-trees -pfam2go". Among SMGC genomes, 111 BGC-rich MAGs (≥15 BGCs;
690 completeness ≥90%, contamination ≤5%) were selected for exopolysaccharide BGC
691 prediction using epsSMASH (v1.0)[108] with --genefinding-tool prodigal. ARGs were
692 inferred using RGI v6.0.3 in strict mode with CARD (v3.2.4)[109] and independently
693 annotated using DeepARG v2 in long-sequence mode (--min-prob 0.8 --id 50 -e 1e-
694 10)[110]. DeepARG results were normalized to ARO terms using ArgNorm (v1.0.0)[111]. A
695 sequence was retained as an ARG if it was identified by either approach. Putative
696 dehalogenases annotated by eggNOG-mapper were further validated by BLASTp
697 against UniProt. Only sequences with ≥60% identity, ≥90% query coverage and E-

698 values <1 × 10^{-5} were retained[112]. Polyphenol-related genes were annotated using
699 CAMPER v1.0.13 with default parameters[113].

700 To quantify the abundance of functional genes, predicted proteins from SMGC genomes
701 were clustered at 95% amino acid identity using CD-HIT (v4.8.1; -c 0.95 -aS 0.9 -g 1 -
702 d 0)[114]. Metagenomic clean reads were then mapped to the non-redundant MAG gene
703 catalogue with Salmon (v1.10.3; -validateMappings -meta)[115] in mapping-based mode,
704 and gene abundance was expressed as genes per million (GPM). For each BGC, the
705 metagenomic abundance was estimated as the abundance of the cluster's core gene
706 sequence, quantified with the same Salmon workflow and reported in GPM.

707 **Phylogenetic analyses**

708 Phylogenomic trees of bacterial and archaeal genomes were constructed using
709 concatenated alignments of 120 bacterial and 53 archaeal conserved marker genes
710 identified using GTDB-Tk (v2.4.1)[116]. Aligned marker genes were concatenated to
711 generate genome-scale alignments and phylogenetic trees were inferred using
712 FastTree[117] (v2.1.11) under default parameters. Protein phylogenetic trees were
713 constructed for selected functional genes (e.g., AlkB). Reference sequences of AlkB
714 were retrieved from KEGG and UniProt databases. Protein sequences were aligned and
715 phylogenetic inference was performed using IQ-TREE (v2.3.3)[118] with model selection
716 (−m MFP) and branch support estimated using 1000 ultrafast bootstrap replicates and
717 SH-aLRT tests. All phylogenetic trees were visualized using iTOL (v7)[119].

718 **Protein structure prediction and structural comparison**

719 Representative protein sequences of selected functional genes (e.g., AlkB and plastic-
720 degrading enzymes) were subjected to structural prediction using AlphaFold3
721 (https://alphafoldserver.com/)[120]. Predicted protein structures were evaluated using the
722 predicted template modeling (pTM) score, and representative models were selected for
723 structural comparison. Pairwise structural alignments between predicted structures and
724 experimentally resolved reference proteins were performed using TM-align
725 (v20170708; default parameters)[121], and structural similarity was evaluated using TM-
726 scores. Protein structures and alignments were visualized using PyMOL (v3.0.0)[122].

727 **RNA extraction and metatranscriptomic sequencing**

728 Total RNA was extracted using the RNeasy PowerSoil Total RNA Kit (QIAGEN) from
729 72 water samples, including 46 microbial metatranscriptome samples collected on 0.22
730 μm filters and 26 additional virus-enriched samples obtained by pre-filtration through
731 0.22 μm membranes, ferric chloride flocculation (0.1%, 30 min), and subsequent
732 recovery of virus-enriched material prior to RNA extraction. Residual genomic DNA
733 was removed by DNase treatment. Ribosomal RNA was depleted using a probe-based
734 magnetic bead depletion workflow. The enriched RNA was reverse-transcribed, and
735 cDNA libraries were prepared using the NEBNext Ultra II RNA Library Prep Kit
736 following an Illumina-compatible protocol. Libraries were sequenced by Biozeron
737 Biological Technology Co., Ltd. on the MGI-T7 platform in paired-end mode (2 × 150
738 bp).

739 **Metatranscriptomic quantification of microbial and viral activity**

740 Raw metatranscriptomic reads were quality-filtered using the Read_QC module of
741 MetaWRAP (v1.3.2; -skip-bmtagger)[85]. Ribosomal RNA sequences were removed
742 using SortMeRNA (v4.3.7) with default settings[123]. Transcript abundances of
743 prokaryotic functional genes and BGCs were determined by mapping clean reads to the
744 non-redundant MAG gene catalogue and core biosynthetic genes using Salmon
745 (v1.10.3; -validateMappings -meta)[115]. Gene transcript abundance was expressed as
746 transcripts per million (TPM). For viral transcriptional activity, metatranscriptomic
747 reads were mapped to vOTU sequences using Bowtie2 (v2.5.4) with default
748 parameters[124]. SAM files were converted to BAM, sorted and indexed with SAMtools
749 (v1.20). Alignments were filtered with CoverM (v0.7.0)[89], retaining only reads aligned
750 over ≥90% of their length with ≥95% nucleotide identity, as in a previous study[125].
751 Coverage profiles of viral contigs were generated using BEDTools genomecov
752 (v2.31.1)[126]. To minimize spurious detection, only viral contigs with mapped reads
753 covering at least 70% of the contig length were retained. Average gene-level coverage
754 was then calculated using BEDTools coverage based on the filtered BAM files and

755 Prodigal GFF annotations. A viral population was considered transcriptionally detected
756 if at least one gene per 10 kb of viral genome had positive average coverage following
757 a previous study[125]. Final vOTU coverage was calculated as the mean coverage of all
758 genes within each viral population. Viral populations not meeting these detection
759 criteria were assigned zero coverage. Gene-level coverage values were linked to
760 functional annotations to assess AMG transcription.

761 **Statistical analysis**

762 Statistical analyses were performed in R (v4.2.3). Alpha diversity was calculated using
763 the Simpson index and compared among groups using Kruskal–Wallis tests, followed
764 by pairwise two-sided Wilcoxon rank-sum tests where appropriate. Gene abundance
765 and transcript abundance were compared among groups using the same approach.
766 Community dissimilarities were calculated using Bray–Curtis distances and visualized
767 by non-metric multidimensional scaling (NMDS) implemented in the vegan package.
768 Differences in community composition were assessed using analysis of similarities
769 (ANOSIM) with 999 permutations. Associations between environmental variables and
770 community composition were assessed using Mantel tests. Unless otherwise stated,
771 statistical significance was defined as $P < 0.05$.

772 **Data availability**

773 All metagenomic, metatranscriptomic and viromic sequencing data generated in this
774 study have been deposited in the National Genomics Data Center (NGDC) under
775 accession number PRJCA0589xx. The genomic sequences of the SMGC and SVC are
776 publicly available at https://doi.org/10.5281/zenodo.189340xx.

777 **Code availability**

778 The present study did not generate codes, and the mentioned tools used for the data
779 analysis were applied with default parameters unless specified otherwise.

780 # References

781 1. Luijendijk, A. et al. The state of the world's beaches. *Scientific Reports* **8**, 6641 (2018).

2. Wei, L. et al. Rapid nitrogen removal in sandy beaches driven by periodic tidal inundations. *Limnology and Oceanography* **71**, e70314 (2026).
3. Kim, K.H. et al. Spatial patterns of groundwater biogeochemical reactivity in an intertidal beach aquifer. *Journal of Geophysical Research: Biogeosciences* **122**, 2548-2562 (2017).
4. Heiss, J.W., Post, V.E.A., Laattoe, T., Russoniello, C.J. & Michael, H.A. Physical controls on biogeochemical processes in intertidal zones of beach aquifers. *Water Resources Research* **53**, 9225-9244 (2017).
5. Chen, K., Guo, Z., Zhan, Y., Roden, E.E. & Zheng, C. Heterogeneity in Permeability and Particulate Organic Carbon Content Controls the Redox Condition of Riverbed Sediments at Different Timescales. *Geophysical Research Letters* **51**, e2023GL107761 (2024).
6. Martínez, A. et al. Human access impacts biodiversity of microscopic animals in sandy beaches. *Communications Biology* **3**, 175 (2020).
7. Archana, A., Francis, C.A. & Boehm, A.B. The beach aquifer microbiome: research gaps and data needs. *Frontiers in Environmental Science* **9**, 653568 (2021).
8. Whitman, R.L. et al. Microbes in beach sands: integrating environment, ecology and public health. *Reviews in Environmental Science and Bio/Technology* **13**, 329-368 (2014).
9. Boehm, A.B., Yamahara, K.M. & Sassoubre, L.M. Diversity and transport of microorganisms in intertidal sands of the California Coast. *Applied and Environmental Microbiology* **80**, 3943-3951 (2014).
10. Miksch, S. et al. Bacterial communities in temperate and polar coastal sands are seasonally stable. *ISME Communications* **1**, 29 (2021).
11. Okamoto, N., Keeling, P.J., Leander, B.S. & Tai, V. Microbial communities in sandy beaches from the three domains of life differ by microhabitat and intertidal location. *Molecular Ecology* **31**, 3210-3227 (2022).
12. Kessler, A.J. et al. Bacterial fermentation and respiration processes are uncoupled in anoxic permeable sediments. *Nature Microbiology* **4**, 1014-1023 (2019).
13. Li, J., Zheng, Y. & Achal, V. Impact of anthropogenic activities on bacterial community diversity in a coastal city: A case study from Shantou. *Continental Shelf Research* **289**, 105450 (2025).
14. Brick, S., Niggemann, J., Reckhardt, A., Könneke, M. & Engelen, B. Interstitial microbial communities of coastal sediments are dominated by Nanoarchaeota. *Frontiers in Microbiology* **16**, 1532193 (2025).
15. Chen, Y.J. et al. Hydrodynamic disturbance controls microbial community assembly and biogeochemical processes in coastal sediments. *The ISME Journal* **16**, 750-763 (2022).
16. Carducci, A. et al. Virus contamination and infectivity in beach environment: Focus on sand and stranded material. *Marine Pollution Bulletin* **185**, 114342 (2022).
17. Geng, X. et al. Geochemical fluxes in sandy beach aquifers: Modulation due to major physical stressors, geologic heterogeneity, and nearshore morphology. *Earth-Science Reviews* **221**, 103800 (2021).
18. Magalhães, E.A., De Jesus, H.E., Pereira, P.H.F., Gomes, A.S. & Santos, H.F.D. Beach sand plastispheres are hotspots for antibiotic resistance genes and potentially pathogenic bacteria even in beaches with good water quality. *Environmental Pollution* **344**, 123237 (2024).
19. Selvarajan, R., Sibanda, T., Ullah, H. & Abia, A.L.K. Beach sand mycobiome: The silent threat of pathogenic fungi and toxic metal contamination for beachgoers. *Marine Pollution Bulletin* **198**, 115895 (2024).

20. Niu, Y. et al. Tidal dynamics regulates potential coupling of carbon-nitrogen-sulfur cycling microbes in intertidal flats. *Science of The Total Environment* **899**, 165663 (2023).
21. Adriaenssens, E.M. et al. Tracing the fate of wastewater viruses reveals catchment-scale virome diversity and connectivity. *Water Research* **203**, 117568 (2021).
22. Singleton, C.M. et al. The Microflora Danica atlas of Danish environmental microbiomes. *Nature* **649**, 971-981 (2025).
23. Ji, M. et al. Biodiversity of mudflat intertidal viromes along the Chinese coasts. *Nature Communications* **15**, 8611 (2024).
24. Bowers, R.M. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology* **35**, 725-731 (2017).
25. Srinivas, P., Peterson, S.B., Gallagher, L.A., Wang, Y. & Mougous, J.D. Beyond genomics in Patescibacteria: A trove of unexplored biology packed into ultrasmall bacteria. *Proceedings of the National Academy of Sciences* **121**, e2419369121 (2024).
26. Li, L. et al. Globally distributed Myxococcota with photosynthesis gene clusters illuminate the origin and evolution of a potentially chimeric lifestyle. *Nature Communications* **14**, 6450 (2023).
27. Borton, M.A. et al. A functional microbiome catalogue crowdsourced from North American rivers. *Nature* **637**, 103-112 (2025).
28. Ma, B. et al. A genomic catalogue of soil microbiomes boosts mining of biodiversity and genetic resources. *Nature Communications* **14**, 7318 (2023).
29. Chen, J. et al. Global marine microbial diversity and its potential in bioprospecting. *Nature* **633**, 371-379 (2024).
30. Xiao, X. et al. Microbial ecosystems and ecological driving forces in the deepest ocean sediments. *Cell* **188**, 1363-1377.e1369 (2025).
31. Delleuze, M. et al. Habitat specificity modulates the bacterial biogeographic patterns in the Southern Ocean. *FEMS Microbiology Ecology* **100**, fiae134 (2024).
32. Koonin, E.V., Krupovic, M. & Yutin, N. Evolution of double-stranded DNA viruses of eukaryotes: from bacteriophages to transposons to giant viruses. *Annals of the New York Academy of Sciences* **1341**, 10-24 (2015).
33. Yutin, N., Bäckström, D., Ettema, T.J.G., Krupovic, M. & Koonin, E.V. Vast diversity of prokaryotic virus genomes encoding double jelly-roll major capsid proteins uncovered by genomic and metagenomic sequence analysis. *Virology Journal* **15**, 67 (2018).
34. Gregory, A.C. et al. Marine DNA viral macro-and microdiversity from pole to pole. *Cell* **177**, 1109-1123.e1114 (2019).
35. Wu, Z. et al. Unveiling the unknown viral world in groundwater. *Nature Communications* **15**, 6788 (2024).
36. Ma, B. et al. Biogeographic patterns and drivers of soil viromes. *Nature Ecology & Evolution* **8**, 717-728 (2024).
37. An, L. et al. Global diversity and ecological functions of viruses inhabiting oil reservoirs. *Nature Communications* **15**, 6789 (2024).
38. Brandão, J. et al. Strategies for monitoring microbial life in beach sand for protection of public health. *International Journal of Environmental Research and Public Health* **20**, 5710 (2023).
39. 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).

40. Yang, T. et al. Diversity and evolution of prokaryotic viral lytic proteins. *The ISME Journal* **19**, wraf200 (2025).
41. Ghaly, T.M. et al. Exploring virus-host-environment interactions in a chemotrophic-based underground estuary. *Environmental Microbiome* **19**, 9 (2024).
42. 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**, e3002083 (2023).
43. Bignaud, A. et al. Phages with a broad host range are common across ecosystems. *Nature Microbiology* **10**, 2537-2549 (2025).
44. Coelho, L.P. et al. Towards the biogeography of prokaryotic genes. *Nature* **601**, 252-256 (2022).
45. Salazar, G. et al. Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. *Cell* **179**, 1068-1083.e1021 (2019).
46. Han, Y. et al. A comprehensive genomic catalog from global cold seeps. *Scientific Data* **10**, 596 (2023).
47. Bahram, M. et al. Structure and function of the global topsoil microbiome. *Nature* **560**, 233-237 (2018).
48. Liu, Y. et al. A genome and gene catalog of glacier microbiomes. *Nature Biotechnology* **40**, 1341-1348 (2022).
49. Leung, P.M. et al. Aridity drives global convergence of desert microbiomes and biogeochemical activities. *bioRxiv* https://doi.org/10.1101/2025.11.29.691269 (2025).
50. Siles, J.A. et al. Distribution of phosphorus cycling genes across land uses and microbial taxonomic groups based on metagenome and genome mining. *Soil Biology and Biochemistry* **174**, 108826 (2022).
51. Whyte, L. Prevalence of alkane monooxygenase genes in Arctic and Antarctic hydrocarbon-contaminated and pristine soils. *FEMS Microbiology Ecology* **41**, 141-150 (2002).
52. Long, H. et al. Diversity of crude oil-degrading bacteria and alkane hydroxylase (alkB) genes from the Qinghai-Tibet Plateau. *Environmental Monitoring and Assessment* **189** (2017).
53. Wang, L., Wang, W., Lai, Q. & Shao, Z. Gene diversity of CYP153A and AlkB alkane hydroxylases in oil-degrading bacteria isolated from the Atlantic Ocean. *Environmental Microbiology* **12**, 1230-1242 (2010).
54. Liu, Q. et al. Broad-spectrum hydrocarbon-degrading microbes in the global ocean metagenomes. *Science of The Total Environment* **926**, 171746 (2024).
55. Benito Merino, D., Lipp, J.S., Borrel, G., Boetius, A. & Wegener, G. Anaerobic hexadecane degradation by a thermophilic Hadarchaeon from Guaymas Basin. *The ISME Journal* **18** (2024).
56. Dong, X. et al. Thermogenic hydrocarbon biodegradation by diverse depth-stratified microbial populations at a Scotian Basin cold seep. *Nature Communications* **11**, 5825 (2020).
57. Acosta, D.J. et al. Plastic degradation by enzymes from uncultured deep sea microorganisms. *The ISME Journal* **19** (2025).
58. Chai, B. et al. Microplastic contamination on the beaches of south China. *Frontiers in Marine Science* **9** (2022).
59. Tournier, V. et al. An engineered PET depolymerase to break down and recycle plastic bottles. *Nature* **580**, 216-219 (2020).
60. Wei, R., Westh, P., Weber, G., Blank, L.M. & Bornscheuer, U.T. Standardization guidelines and future trends for PET hydrolase research. *Nature Communications* **16**, 4684 (2025).
61. Delacuvellerie, A., Cyriaque, V., Gobert, S., Benali, S. & Wattiez, R. The plastisphere in marine

ecosystem hosts potential specific microbial degraders including Alcanivorax borkumensis as a key player for the low-density polyethylene degradation. *Journal of Hazardous Materials* **380**, 120899 (2019).

62. Denaro, R. et al. Marine hydrocarbon-degrading bacteria breakdown poly(ethylene terephthalate) (PET). *Science of The Total Environment* **749**, 141608 (2020).
63. Whitfield, C., Wear, S.S. & Sande, C. Assembly of Bacterial Capsular Polysaccharides and Exopolysaccharides. *Annual Review of Microbiology* **74**, 521-543 (2020).
64. Pflughoeft, K.J., Kierek, K. & Watnick, P.I. Role of ectoine in Vibrio cholerae osmoadaptation. *Applied and Environmental Microbiology* **69**, 5919-5927 (2003).
65. Sagot, B. et al. Osmotically induced synthesis of the dipeptide N-acetylglutaminylglutamine amide is mediated by a new pathway conserved among bacteria. *Proceedings of the National Academy of Sciences* **107**, 12652-12657 (2010).
66. Cimermancic, P. et al. Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. *Cell* **158**, 412-421 (2014).
67. Han, P. et al. Unveiling unique microbial nitrogen cycling and nitrification driver in coastal Antarctica. *Nature Communications* **15**, 3143 (2024).
68. Perry, E.K., Meirelles, L.A. & Newman, D.K. From the soil to the clinic: the impact of microbial secondary metabolites on antibiotic tolerance and resistance. *Nature Reviews Microbiology* **20**, 129-142 (2022).
69. Auer, F., Greskowiak, J., Meyer, R., Reckhardt, A. & Holtappels, M. The beach bioreactor: unraveling the anomaly of intensive remineralization above a deep oxycline. *Environmental Science & Technology* **59**, 19835-19845 (2025).
70. Wagner, M. et al. Microhabitat partitioning between sympatric intertidal fish species highlights the importance of sediment composition in gravel beach conservation. *Ecology and Evolution* **13**, e10302 (2023).
71. Ji, M. et al. Temporal turnover of viral biodiversity and functional potential in intertidal wetlands. *npj Biofilms and Microbiomes* **10**, 48 (2024).
72. Wang, Y. et al. Nutrient-dependent viral lifestyle shifts drive microbial turnover and soil organic carbon accrual. *European Journal of Soil Science* **76**, e70220 (2025).
73. Wang, F. et al. Emerging contaminants: a one health perspective. *The Innovation* **5**, 100612 (2024).
74. Shin, B., Bociu, I., Kolton, M., Huettel, M. & Kostka, J.E. Succession of microbial populations and nitrogen-fixation associated with the biodegradation of sediment-oil-agglomerates buried in a Florida sandy beach. *Scientific Reports* **9**, 19401 (2019).
75. Yang, Y. et al. Effects of oxygen availability on mycobenthic communities of marine coastal sediments. *Scientific Reports* **13**, 15218 (2023).
76. Zhang, X., Dong, Z., Zhang, S., Ma, J. & Liu, S. Microplastic biofilm as hotspots of antibiotic resistance genes and potential pathogens. *npj Biofilms and Microbiomes* **12**, 24 (2026).
77. Shi, J. et al. A 39-year high resolution wave hindcast for the Chinese coast: Model validation and wave climate analysis. *Ocean Engineering* **183**, 224-235 (2019).
78. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* **30**, 2114-2120 (2014).
79. Shaw, J. & Yu, Y.W. Rapid species-level metagenome profiling and containment estimation with sylph. *Nature Biotechnology* **43**, 1348-1359 (2025).

80. Li, D., Liu, C.M., Luo, R., Sadakane, K. & Lam, T.W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics (Oxford, England)* **31**, 1674-1676 (2015).
81. Kang, D.D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
82. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* https://doi.org/10.48550/arXiv.1303.3997 (2013).
83. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078-2079 (2009).
84. Pan, S., Zhao, X.-M. & Coelho, L.P. SemiBin2: self-supervised contrastive learning leads to better MAGs for short- and long-read sequencing. *Bioinformatics (Oxford, England)* **39**, i21-i29 (2023).
85. Uritskiy, G.V., DiRuggiero, J. & Taylor, J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).
86. 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. *The ISME Journal* **11**, 2864-2868 (2017).
87. Chklovski, A., Parks, D.H., Woodcroft, B.J. & Tyson, G.W. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nature Methods* **20**, 1203-1212 (2023).
88. Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P. & Parks, D.H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics (Oxford, England)* **36**, 1925-1927 (2020).
89. Aroney, S.T.N. et al. CoverM: read alignment statistics for metagenomics. *Bioinformatics (Oxford, England)* **41**, btaf147 (2025).
90. Camargo, A.P. et al. Identification of mobile genetic elements with geNomad. *Nature Biotechnology* **42**, 1303-1312 (2024).
91. Guo, J. et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 (2021).
92. Nayfach, S. et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature Biotechnology* **39**, 578-585 (2021).
93. Shaffer, M. et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Research* **48**, 8883-8900 (2020).
94. Berman, H.M. The protein data bank. *Nucleic Acids Research* **28**, 235-242 (2000).
95. Van Kempen, M. et al. Fast and accurate protein structure search with Foldseek. *Nature Biotechnology* **42**, 243-246 (2024).
96. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410 (1990).
97. Coutinho, F.H. et al. RaFAH: host prediction for viruses of Bacteria and Archaea based on protein content. *Patterns* **2**, 100274 (2021).
98. Galiez, C., Siebert, M., Enault, F., Vincent, J. & Söding, J. WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics (Oxford, England)* **33**, 3113-3114 (2017).
99. Wang, W. et al. A network-based integrated framework for predicting virus–prokaryote

interactions. *NAR Genomics and Bioinformatics* **2** (2020).

100. Lu, C. et al. Prokaryotic virus host predictor: a Gaussian model for host prediction of prokaryotic viruses in metagenomics. *BMC Biology* **19**, 5 (2021).

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

102. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology* **35**, 1026-1028 (2017).

103. Cantalapiedra, C.P., Hernandez-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular biology and evolution* **38**, 5825-5829 (2021).

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

105. Khot, V. et al. CANT-HYD: a curated database of phylogeny-derived hidden Markov models for annotation of marker genes involved in hydrocarbon degradation. *Frontiers in Microbiology* **12**, 764058 (2022).

106. Sankara Subramanian, S.H., Balachandran, K.R.S., Rangamaran, V.R. & Gopal, D. RemeDB: tool for rapid prediction of enzymes involved in bioremediation from high-throughput metagenome data sets. *Journal of Computational Biology* **27**, 1020-1029 (2020).

107. Blin, K. et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Research* **51**, W46-W50 (2023).

108. Daugberg, A.O.H. et al. epsSMASH uncovers exopolysaccharide biosynthetic gene clusters in environmental and human microbiomes. *bioRxiv* https://doi.org/10.64898/2025.12.21.693542 (2025).

109. Alcock, B.P. et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Research* **51**, D690-d699 (2023).

110. Arango-Argoty, G. et al. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome* **6**, 23 (2018).

111. Ugarcina Perovic, S. et al. argNorm: normalization of antibiotic resistance gene annotations to the Antibiotic Resistance Ontology (ARO). *Bioinformatics (Oxford, England)* **41**, btaf173 (2025).

112. Liu, R. et al. Extensive halogenated organic compound reservoirs and active microbial dehalogenation in Mariana Trench sediments. *The ISME Journal* **19** (2025).

113. McGivern, B.B. et al. Microbial polyphenol metabolism is part of the thawing permafrost carbon cycle. *Nature Microbiology* **9**, 1454-1466 (2024).

114. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics (Oxford, England)* **22**, 1658-1659 (2006).

115. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**, 417-419 (2017).

116. Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P. & Parks, D.H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics (Oxford, England)* **38**, 5315-5316 (2022).

117. Price, M.N., Dehal, P.S. & Arkin, A.P. FastTree: computing large minimum evolution trees with

profiles instead of a distance matrix. *Molecular Biology and Evolution* **26**, 1641-1650 (2009).
118. Nguyen, L.-T., Schmidt, H.A., Von Haeseler, A. & Minh, B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**, 268-274 (2015).
119. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research* **52**, W78-W82 (2024).
120. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493-500 (2024).
121. Zhang, Y. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research* **33**, 2302-2309 (2005).
122. Schrödinger, L.L.C. The PyMOL Molecular Graphics System. (2015).
123. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics (Oxford, England)* **28**, 3211-3217 (2012).
124. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-359 (2012).
125. Emerson, J.B. et al. Host-linked soil viral ecology along a permafrost thaw gradient. *Nature Microbiology* **3**, 870-880 (2018).
126. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* **26**, 841-842 (2010).

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2022YFC3106100), the National Natural Science Foundation of China (U22A2058), and the Scientific Research Foundation of the Third Institute of Oceanography, Ministry of Natural Resources (2025035). We thank Jixiang Zheng for coordination of field campaigns and Haitao Xu for field safety oversight. We are grateful to colleagues who assisted with field sampling and logistical support, including Chengtao Wang, Zheyu Xiao, Jiahao Liu, Zhanyang Qi, Enquan Zhang, Chengpeng Li, Zehao Wang, Tengkai Chen, Weiming Wu, Zhuoyun Chen, Rupeng Du, Yongxin Shen, Zhipeng Zhu and Xiyu Pu.

Author contributions

F.C. conceived the study, supervised the project, secured funding and contributed to manuscript writing. X.D. contributed to study design, performed analyses and led manuscript writing. Y.H. and C.Z. conducted the multi-omics analyses and contributed to figure preparation and manuscript writing. H.Q. led the geological characterization and morphodynamic analyses. L.W. and Z.P. conducted the physicochemical

measurements. Y.C., Zr.L., Zj.L. and X.G. assisted with multi-omics analyses and figure preparation. S.Z., J.L., S.L., C.R. and L.Z. contributed to data interpretation and manuscript revision. All authors discussed the results and approved the final manuscript.

Competing interests

The authors declare no competing interests.

1087 ## Figure legends

1088 **Figure 1. Study sites, sampling design and environmental gradients across sandy**
1089 **beach ecosystems along the Chinese coastline. (a)** Locations of 18 sandy beaches
1090 (18°–40°N) spanning tropical, subtropical and temperate zones, including 15
1091 recreational beaches and three natural reference sites. Circular charts indicate the
1092 numbers of water and sediment samples collected at each site. **(b)** Conceptual cross-
1093 shore sampling framework showing the supratidal, high-intertidal, mid-intertidal, low-
1094 intertidal and subtidal zones, together with the positions of surface sediment and
1095 seawater sampling. MHW, mean high water; MSL, mean sea level; MLW, mean low
1096 water. **(c)** Beachface slope variation among sites. **(d)** Cross-shore geomorphological
1097 characteristics across beaches, including dry-beach width (orange bars) and intertidal
1098 width (brown bars). **(e)** Tidal range (purple bars) and significant wave height (blue bars)
1099 across study sites. **(f)** Surface sediment grain-size composition (gravel, sand and mud
1100 fractions) and sediment textural parameters, including mean grain size, sorting and
1101 skewness. **(g)** Water physicochemical variables, including temperature, pH and
1102 dissolved nutrients. **(h)** Metal concentrations (Fe, Cu, Zn, Pb and Cr) in sediments
1103 across sites.

1104 **Figure 2. Composition, diversity and environmental drivers of prokaryotic and**
1105 **viral communities across sandy beaches. (a, b)** Taxonomic composition of
1106 prokaryotic (phylum level) **(a)** and viral (class level) **(b)** communities across beaches,
1107 with samples grouped by tidal zone within each site (supratidal, high-intertidal, mid-
1108 intertidal, low-intertidal and subtidal). Sites are ordered latitudinally from north to south
1109 along the Chinese coastline. Asterisks denote natural reference beaches. **(c, d)**
1110 Prokaryotic α-diversity (Simpson index) across tidal zones **(c)** and β-diversity
1111 visualized by non-metric multidimensional scaling (NMDS) based on Bray–Curtis
1112 dissimilarities **(d)**. **(e, f)** Viral α-diversity (Simpson index) across tidal zones **(e)** and β-
1113 diversity visualized by NMDS based on Bray–Curtis dissimilarities **(f)**. **(g–l)**
1114 Comparisons of α-diversity among environmental groupings, including matrix type
1115 (sand versus water), geographic region (north versus south), and beach usage (natural
1116 versus recreational), for prokaryotic **(g, i, k)** and viral **(h, j, l)** communities. **(m, n)**
1117 Relationships between environmental variables and prokaryotic and viral community
1118 composition for sediment-associated geochemical variables **(m)** and water

physicochemical variables together with selected elemental variables **(n)**. Heatmaps show Pearson correlations among environmental variables, and links indicate significant Mantel correlations between environmental factors and community composition. For boxplots, center lines indicate medians, boxes indicate interquartile ranges, and whiskers indicate 1.5 × interquartile ranges. α-diversity differences were assessed using Kruskal–Wallis tests or two-sided Wilcoxon rank-sum tests, as appropriate. β-diversity differences were evaluated using ANOSIM with 999 permutations. Exact statistics are provided in **Tables S10–S15**.

Figure 3. Taxonomic composition and genomic novelty of the Sandbeach Microbiome Genome Catalogue (SMGC). (a) Taxonomic distribution of bacterial metagenome-assembled genomes (MAGs) across taxonomic ranks from phylum to family based on GTDB r226. **(b)** Maximum-likelihood phylogenomic tree of archaeal MAGs reconstructed from 120 concatenated single-copy marker genes. Outer rings indicate phylum affiliation, genome size, GC content and genome completeness. **(c)** Taxonomic novelty relative to GTDB r226, showing the numbers of annotated and novel bacterial and archaeal lineages at each taxonomic rank. **(d)** Contribution of SMGC genomes to phylogenetic diversity within the ten most abundant bacterial phyla. Values indicate the percentage increase in phylogenetic diversity relative to the corresponding reference genome set. **(e)** Relative abundance of novel species-representative genomes (SRGs) across habitat types (water versus sand) and geographic regions (south versus north), shown separately for habitat and geographic comparisons. Boxplots show medians (center lines), interquartile ranges (boxes), and 1.5 × interquartile-range whiskers. Statistical significance was assessed using two-sided Wilcoxon rank-sum tests. **(f)** Comparison of species-level novelty between the SMGC and genome catalogues from Sandbeach, Freshwater, GOMC, Soil, Cold seep and Trench ecosystems. The number of genomes in each catalogue is indicated in white.

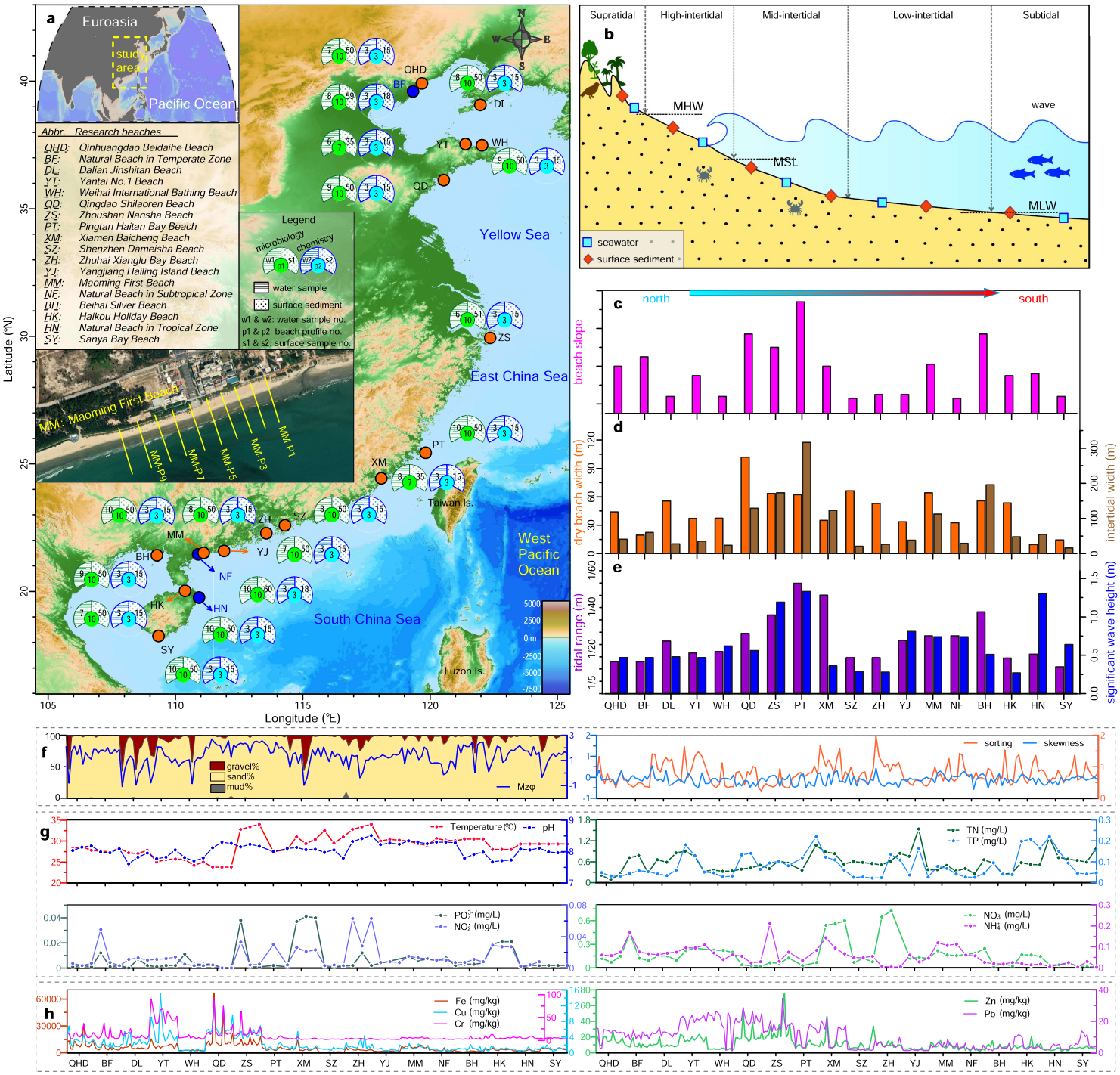
Figure 4. Taxonomic composition, activity, structural diversity and host associations of viruses in the Sandbeach Virome Catalogue (SVC). (a) Taxonomic annotation of viral operational taxonomic units (vOTUs) in the SVC. The left panel shows the proportions of classified and unclassified vOTUs. The middle panel displays the family-level composition of classified vOTUs, and the right panel summarizes the relative proportions of annotated and unannotated vOTUs across major taxonomic

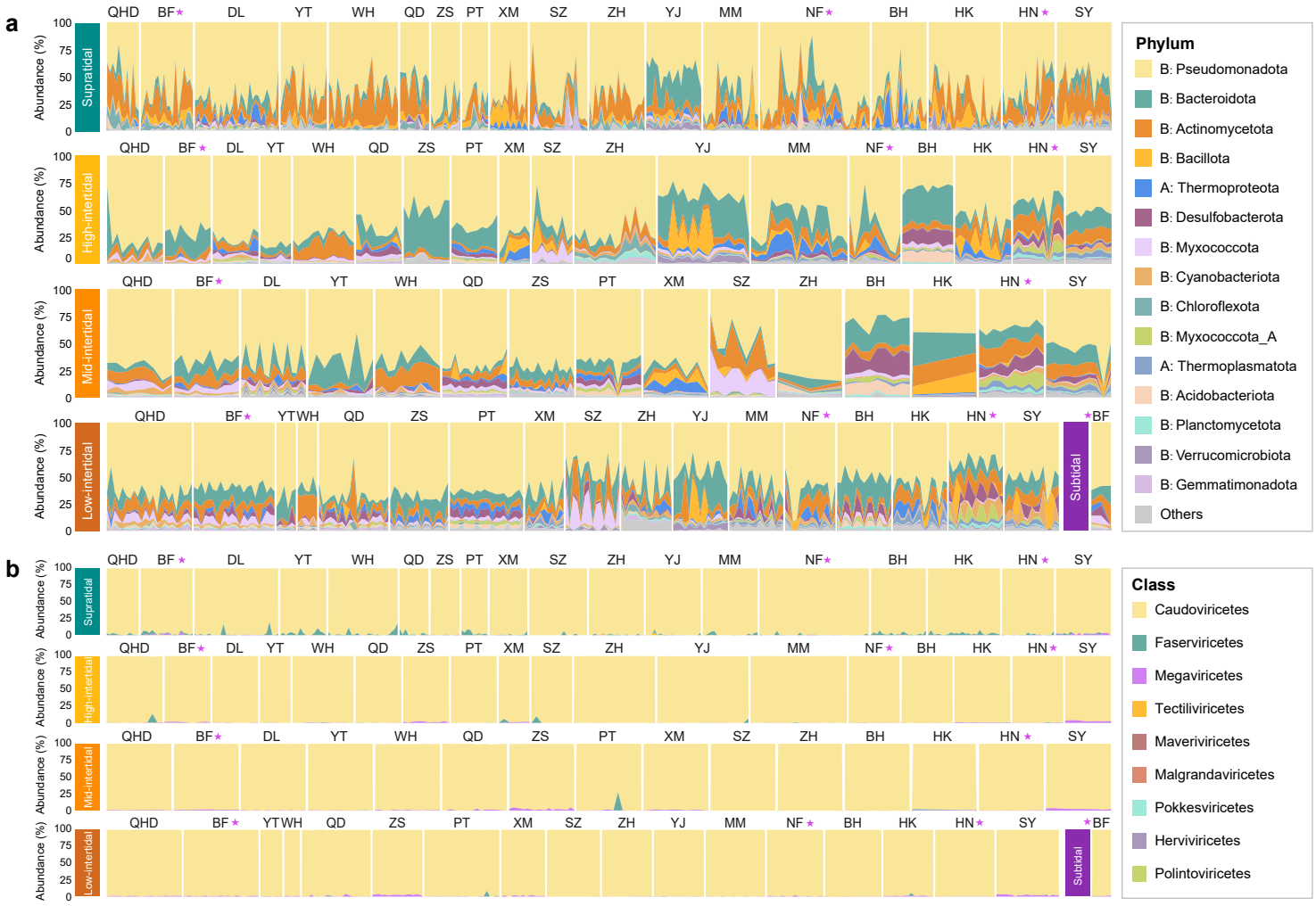
1151 ranks from realm to family. **(b)** Transcript-derived coverage of active vOTUs in water
1152 samples from the mid- and low-intertidal zones. **(c)** Structural diversity of major capsid
1153 proteins (MCPs). Predicted MCP models were aligned to representative reference
1154 structures, revealing double jelly-roll and HK97-like folds. **(d)** Overlap between the
1155 SVC and virome catalogues from other environments, showing shared and habitat-
1156 specific vOTUs together with total catalogue sizes. Similarity thresholds used for cross-
1157 catalogue comparisons are described in the Methods. (**e)** Sankey diagram summarizing
1158 predicted virus–host linkages from prokaryotic domain, phylum and family to viral
1159 class. **(f)** Comparison between host phylum abundance and the number of linked
1160 vOTUs. Blue bars indicate mean host abundance (RPKM) across sandy beach samples,
1161 and the orange line indicates the number of vOTUs predicted to infect each host phylum.
1162 Detailed taxonomic, structural and virus–host association data are provided in **Tables**
1163 **S19–S23**.

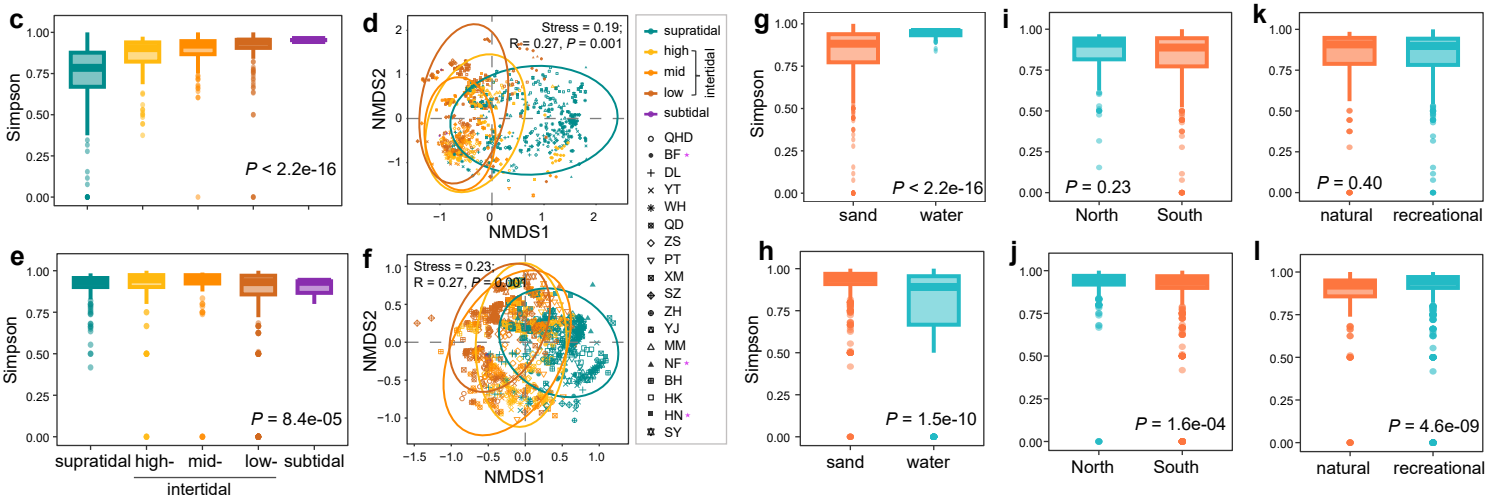
1164 **Figure 5. Metabolic differentiation of sandy beach microbial and viral**
1165 **communities across tidal zones. (a)** Conceptual schematic summarizing metabolic
1166 zonation across sandy beach habitats, highlighting the predominance of heterotrophic
1167 processes in supratidal sediments and the increasing importance of redox metabolism
1168 and autotrophic pathways toward the subtidal zone. For summary visualization, habitats
1169 are grouped into supratidal, intertidal and subtidal zones. **(b–g)** Transcript abundances
1170 (TPM) of functional marker genes involved in heterotrophy **(b)**, central carbon
1171 metabolism **(c)**, redox processes **(d)**, aerobic respiration **(e)**, autotrophy **(f)** and
1172 phosphorus acquisition **(g)**. Transcripts were quantified from metatranscriptomes
1173 generated from water samples collected in the mid- and low-intertidal zones. Colors
1174 denote tidal position. Boxplots show medians (center lines), interquartile ranges (boxes)
1175 and 1.5 × interquartile-range whiskers. Statistical differences were evaluated using
1176 Kruskal–Wallis tests, with significance indicated as ns, $*P < 0.05$, $**P < 0.01$ and $***P$
1177 < 0.001. In panel b, PME denotes pectinesterase. **(h)** Schematic overview of viral
1178 auxiliary metabolic genes detected in sandy beach viromes, showing potential viral
1179 modulation of host pathways related to photosynthesis, oxidative phosphorylation,
1180 carbon metabolism, phosphorus metabolism and polysaccharide degradation. Dots
1181 indicate gene-level and transcriptome-level evidence, as shown in the panel.

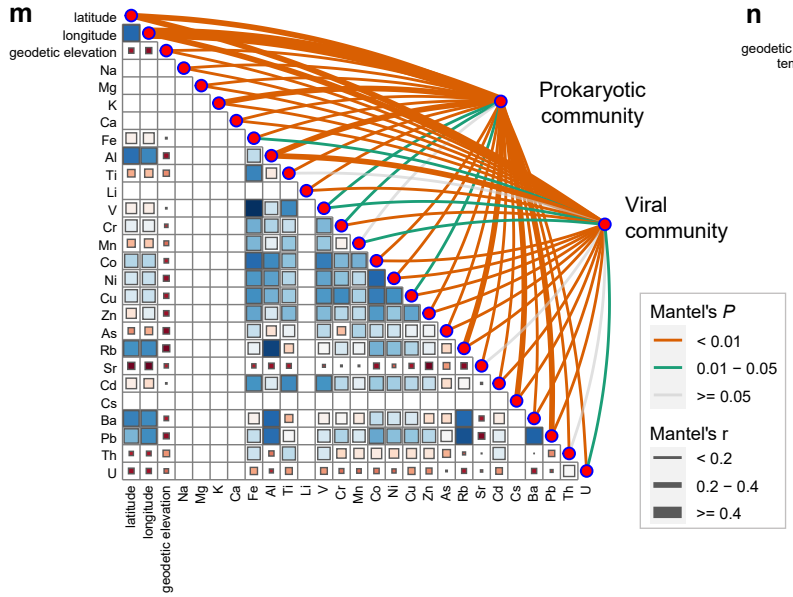
1182 **Figure 6. Sandy beach microbiomes encode pollutant transformation, secondary**

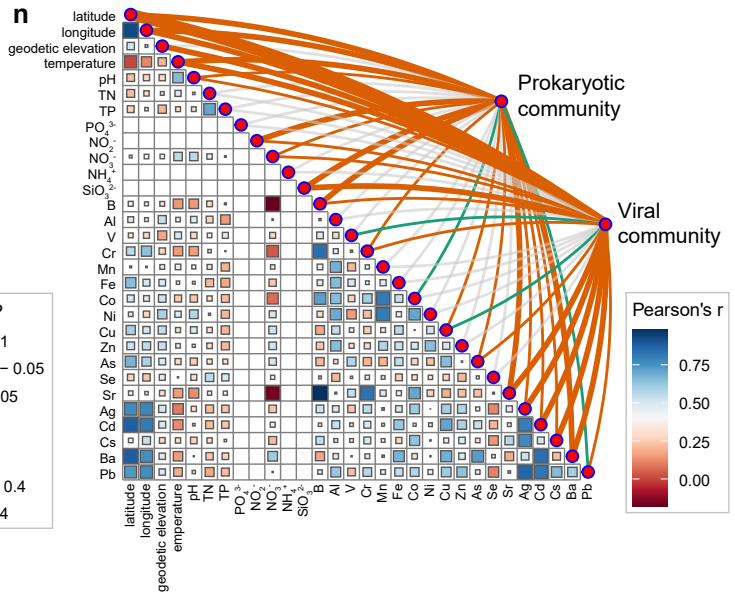
1183 **metabolism and antibiotic resistance functions. (a)** Functional profile of
1184 hydrocarbon-degradation genes in the SMGC, grouped by substrate class and
1185 degradation mode. **(b)** Average abundance of hydrocarbon-degradation genes across
1186 tidal zones, expressed as $\log_2$-transformed genes per million (GPM). **(c)** Representative
1187 hydrocarbon-degradation gene clusters in two MAGs, illustrating co-localized
1188 pathways for oxidation of multiple hydrocarbon classes. **(d)** Average abundance of
1189 plastic-degrading enzyme classes across tidal zones, expressed as $\log_2$-transformed
1190 GPM. **(e)** Structural alignments of representative plastic-degrading enzymes with
1191 experimentally characterized reference proteins. Catalytic residues, reference
1192 identifiers and TM-scores are shown. **(f)** Distribution of biosynthetic gene cluster (BGC)
1193 classes across major bacterial and archaeal phyla in the SMGC. **(g)** Gene and transcript
1194 abundances of selected BGC-associated product classes and adaptive metabolites,
1195 expressed as $\log_{10}$-transformed GPM and TPM, respectively. **(h)** Composition of
1196 antibiotic resistance genes (ARGs) in the SMGC, including ARG counts by resistance
1197 class and the compositions of archaeal and bacterial ARG repertoires. Detailed
1198 annotations are provided in **Table S36**.

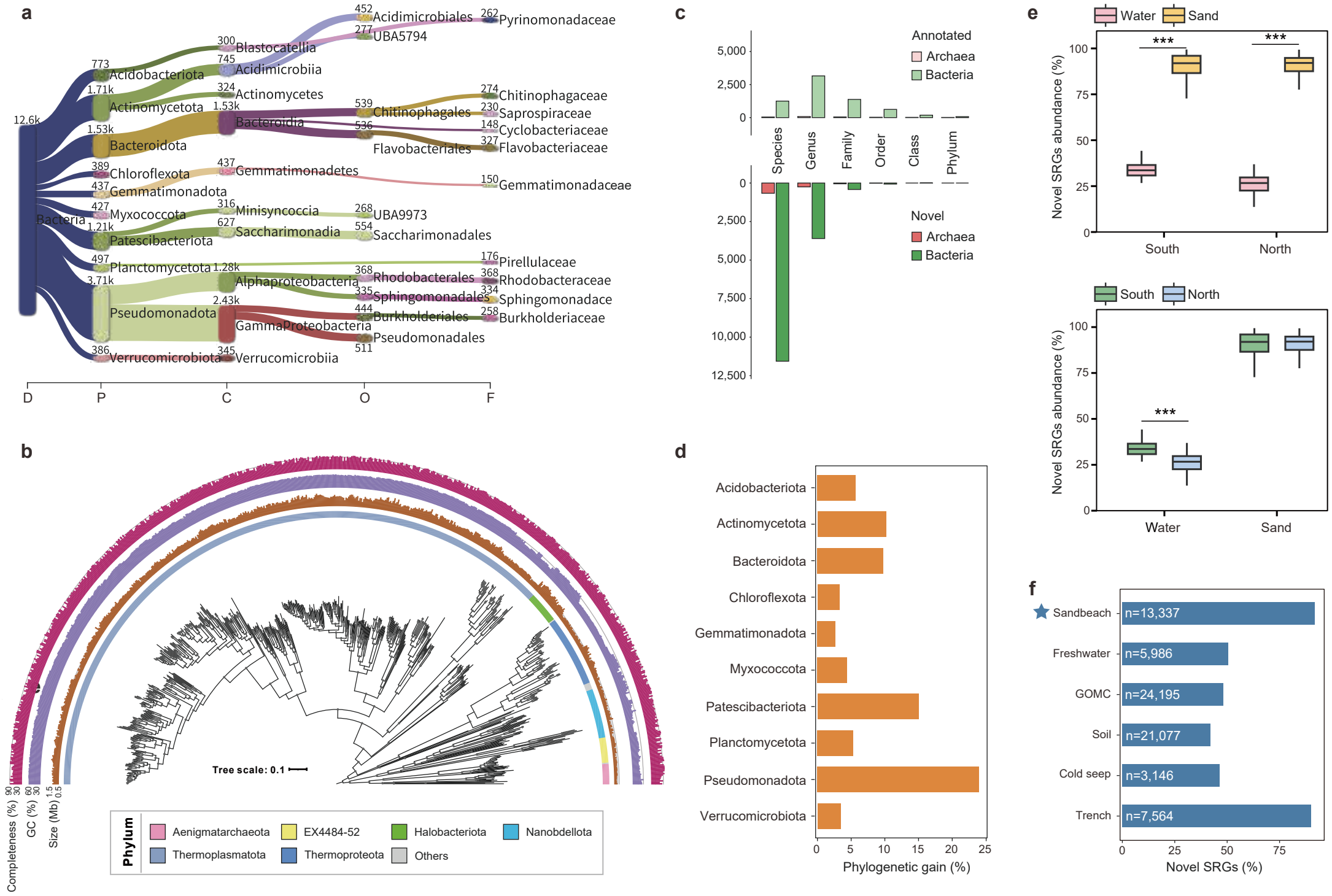

a
Euroasia
study area
Pacific Ocean
Abbr. Research beaches
QHD: Qinhuangdao Beidaihe Beach
BF: Natural Beach in Temperate Zone
DL: Dalian Jinshitan Beach
YT: Yantai No.1 Beach
WH: Weihai International Bathing Beach
QD: Qingdao Shilaoren Beach
ZS: Zhoushan Nansha Beach
PT: Pingtan Haitan Bay Beach
XM: Xiamen Baicheng Beach
SZ: Shenzhen Dameisha Beach
ZH: Zhuhai Xianglu Bay Beach
YJ: Yangjiang Hailing Island Beach
MM: Maoming First Beach
NF: Natural Beach in Subtropical Zone
BH: Beihai Silver Beach
HK: Haikou Holiday Beach
HN: Natural Beach in Tropical Zone
SY: Sanya Bay Beach
Legend
microbiology
chemistry
water sample
surface sediment
w1 & w2: water sample no.
p1 & p2: beach profile no.
s1 & s2: surface sample no.
MM: Maoming First Beach
MM-P9
MM-P7
MM-P5
MM-P3
MM-P1
Yellow Sea
East China Sea
South China Sea
West Pacific Ocean
Taiwan Is.
Luzon Is.
Latitude (°N)
Longitude (°E)
b
Supratidal
High-intertidal
Mid-intertidal
Low-intertidal
Subtidal
MHW
MSL
MLW
wave
seawater
surface sediment
c
north
south
beach slope
d
dry beach width (m)
intertidal width (m)
e
tidal range (m)
significant wave height (m)
QHD BF DL YT WH QD ZS PT XM SZ ZH YJ MM NF BH HK HN SY
f
gravel%
sand%
mud%
Mzφ
sorting
skewness
g
Temperature (°C)
pH
TN (mg/L)
TP (mg/L)
PO_4^{3-} (mg/L)
NO_2^- (mg/L)
NO_3^- (mg/L)
NH_4^+ (mg/L)
h
Fe (mg/kg)
Cu (mg/kg)
Cr (mg/kg)
Zn (mg/kg)
Pb (mg/kg)

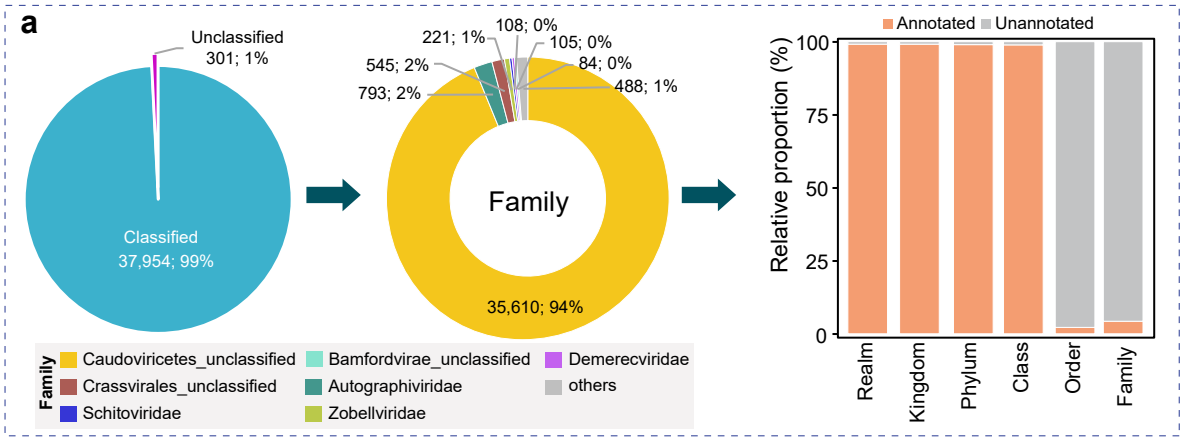

a
b
Supratidal
High-intertidal
Mid-intertidal
Low-intertidal
Subtidal
Abundance (%)
QHD BF DL YT WH QD ZS PT XM SZ ZH YJ MM NF BH HK HN SY
Phylum
B: Pseudomonadota
B: Bacteroidota
B: Actinomycetota
B: Bacillota
A: Thermoproteota
B: Desulfobacterota
B: Myxococcota
B: Cyanobacteriota
B: Chloroflexota
B: Myxococcota_A
A: Thermoplasmatota
B: Acidobacteriota
B: Planctomycetota
B: Verrucomicrobiota
B: Gemmatimonadota
Others
Class
Caudoviricetes
Faserviricetes
Megaviricetes
Tectiliviricetes
Maveriviricetes
Malgrandaviricetes
Pokkesviricetes
Herviviricetes
Polintoviricetes

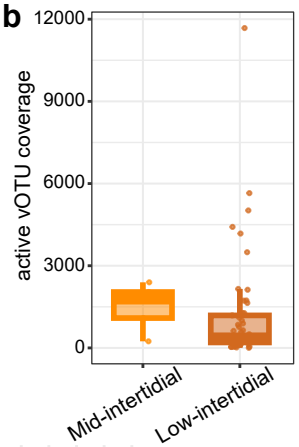

c
d
e
f
g
h
i
j
k
l
Simpson
NMDS1
NMDS2
Stress = 0.19; R = 0.27, P = 0.001
Stress = 0.23; R = 0.27, P = 0.001
P < 2.2e-16
P = 8.4e-05
P = 1.5e-10
P = 1.6e-04
P = 4.6e-09
P = 0.23
P = 0.40
supratidal
high-
mid-
low-
subtidal
intertidal
sand
water
North
South
natural
recreational
supratidal
high
mid
low
intertidal
subtidal
QHD
BF
DL
YT
WH
QD
ZS
PT
XM
SZ
ZH
YJ
MM
NF
BH
HK
HN
SY

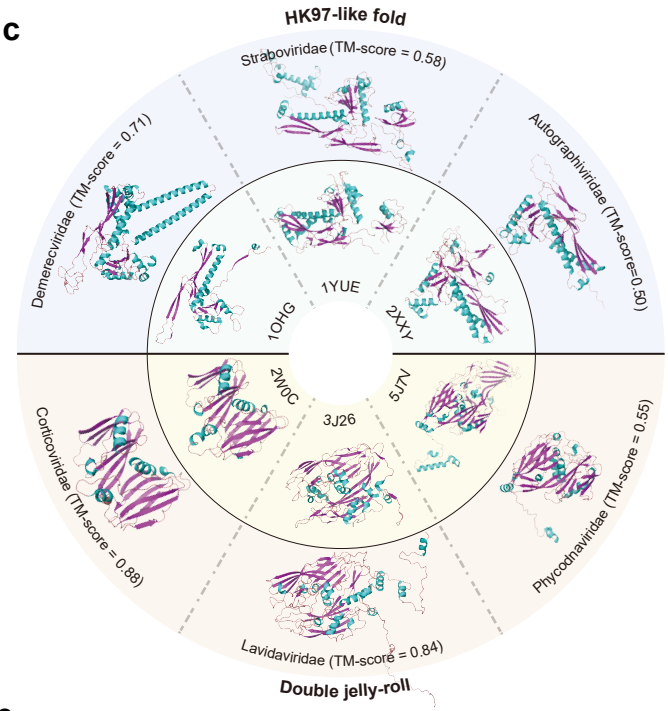

m
latitude
longitude
geodetic elevation
Na
Mg
K
Ca
Fe
Al
Ti
Li
V
Cr
Mn
Co
Ni
Cu
Zn
As
Rb
Sr
Cd
Cs
Ba
Pb
Th
U
Prokaryotic community
Viral community
Mantel's P
< 0.01
0.01 – 0.05
>= 0.05
Mantel's r
< 0.2
0.2 – 0.4
>= 0.4

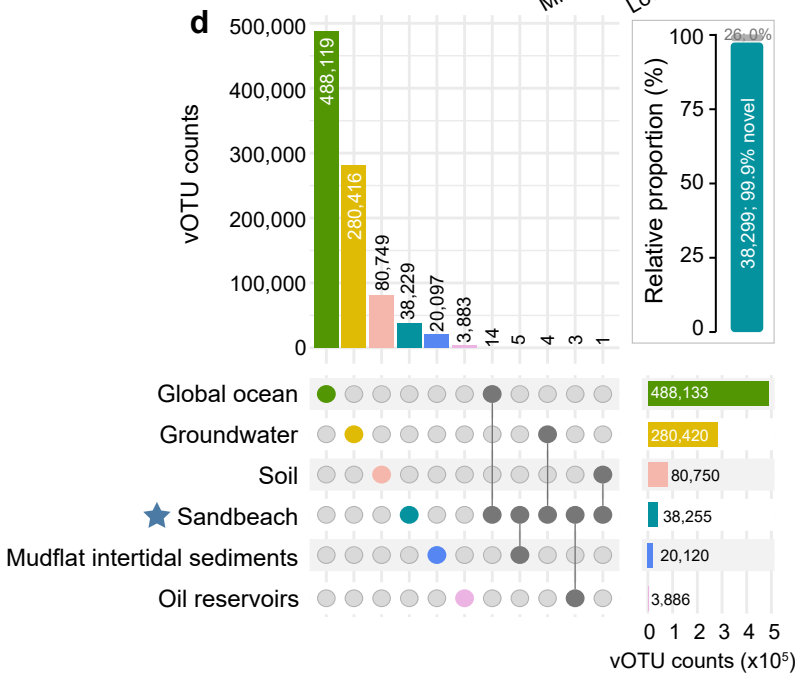

n
latitude
longitude
geodetic elevation
temperature
pH
TN
TP
Prokaryotic community
Viral community
Pearson's r
0.75
0.50
0.25
0.00

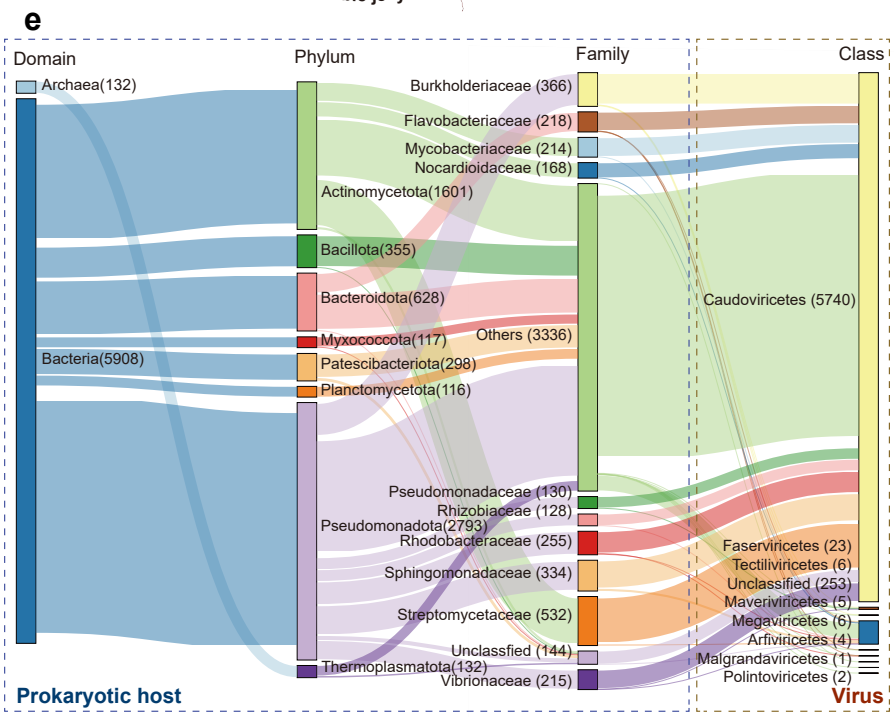

a
12.6k
Bacteria
773 Acidobacteriota
1.71k Actinomycetota
1.53k Bacteroidota
389 Chloroflexota
437 Gemmatimonadota
427 Myxococcota
1.21k Patescibacteriota
497 Planctomycetota
3.71k Pseudomonadota
386 Verrucomicrobiota
300 Blastocatellia
745 Acidimicrobiia
324 Actinomycetes
1.53k Bacteroidia
437 Gemmatimonadetes
316 Minisyncoccia
627 Saccharimonadia
1.28k Alphaproteobacteria
2.43k GammaProteobacteria
345 Verrucomicrobiia
452 Acidimicrobiales
277 UBA5794
539 Chitinophagales
536 Flavobacteriales
268 UBA9973
554 Saccharimonadales
368 Rhodobacterales
335 Sphingomonadales
444 Burkholderiales
511 Pseudomonadales
262 Pyrinomonadaceae
274 Chitinophagaceae
230 Saprospiraceae
148 Cyclobacteriaceae
327 Flavobacteriaceae
150 Gemmatimonadaceae
176 Pirellulaceae
368 Rhodobacteraceae
334 Sphingomonadace
258 Burkholderiaceae
D
P
C
O
F
b
Completeness (%)
GC (%)
Size (Mb)
Tree scale: 0.1
Phylum
Aenigmatarchaeota
EX4484-52
Halobacteriota
Nanobdellota
Thermoplasmatota
Thermoproteota
Others
c
Annotated
Archaea
Bacteria
Species
Genus
Family
Order
Class
Phylum
Novel
Archaea
Bacteria
d
Acidobacteriota
Actinomycetota
Bacteroidota
Chloroflexota
Gemmatimonadota
Myxococcota
Patescibacteriota
Planctomycetota
Pseudomonadota
Verrucomicrobiota
Phylogenetic gain (%)
e
Water
Sand
South
North
Novel SRGs abundance (%)
***
f
Sandbeach
Freshwater
GOMC
Soil
Cold seep
Trench
n=13,337
n=5,986
n=24,195
n=21,077
n=3,146
n=7,564
Novel SRGs (%)

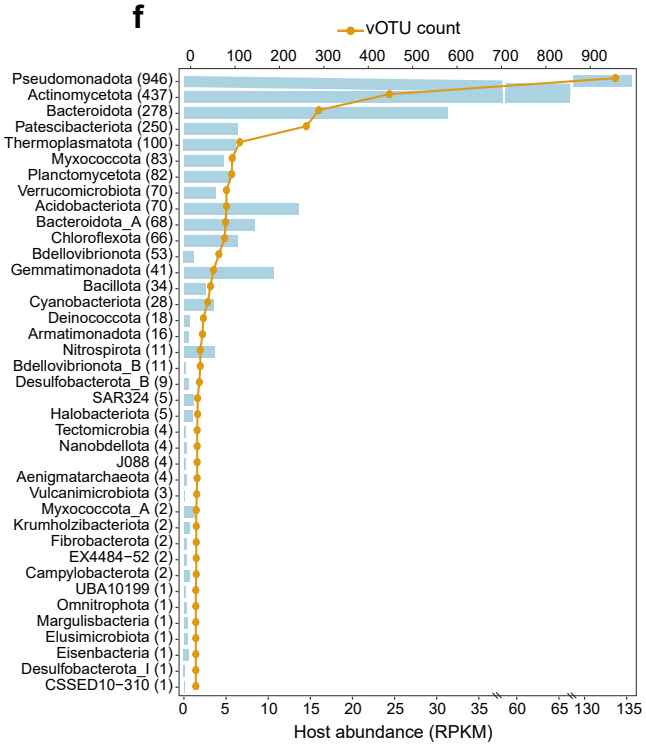

a
Unclassified
301; 1%
Classified
37,954; 99%
Family
35,610; 94%
221; 1%
108; 0%
105; 0%
545; 2%
84; 0%
793; 2%
488; 1%
Annotated
Unannotated
Relative proportion (%)
Realm
Kingdom
Phylum
Class
Order
Family
Family
Caudoviricetes_unclassified
Bamfordvirae_unclassified
Demerecviridae
Crassvirales_unclassified
Autographiviridae
others
Schitoviridae
Zobellviridae

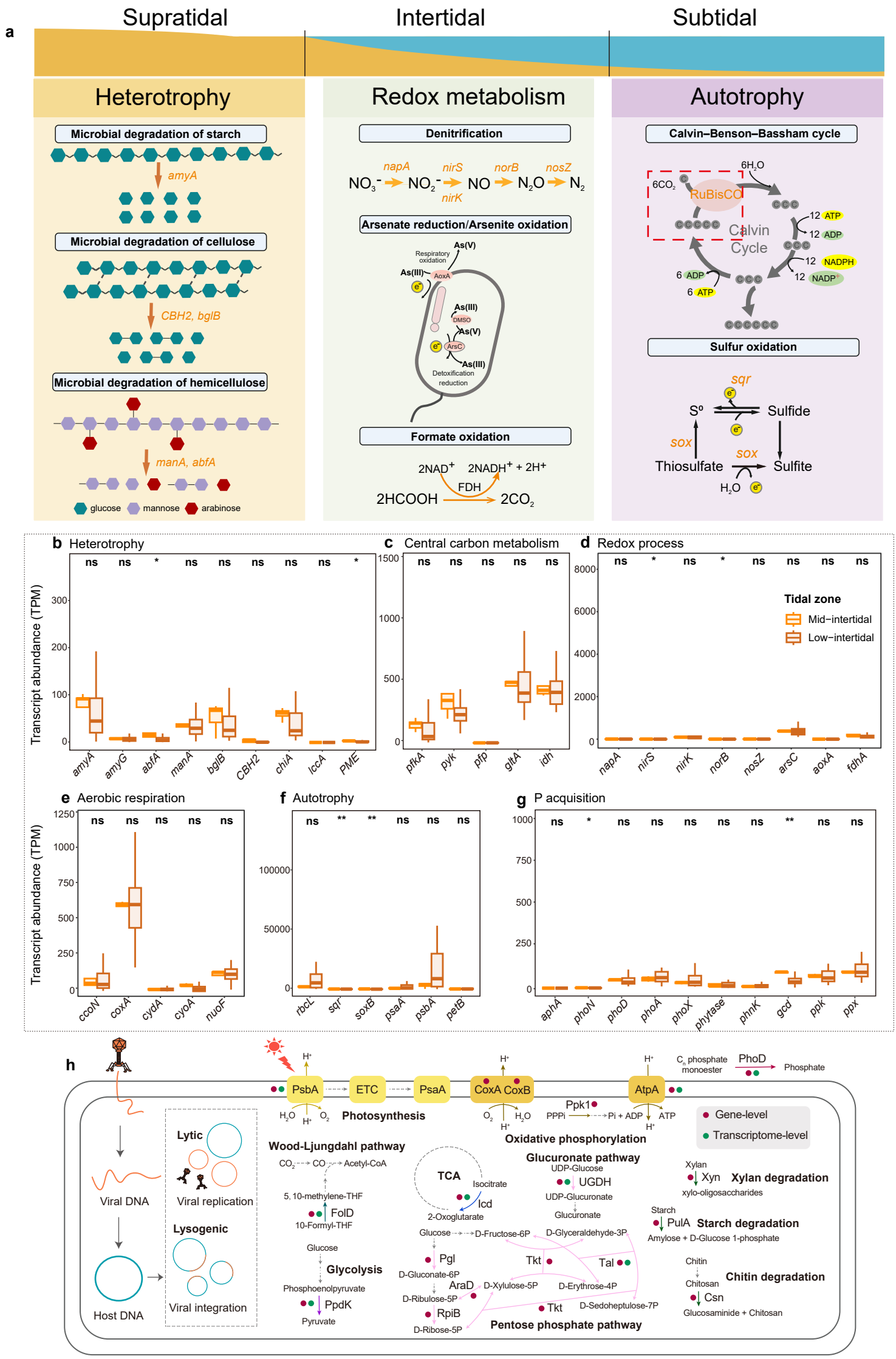

b
active vOTU coverage
Mid-intertidial
Low-intertidial

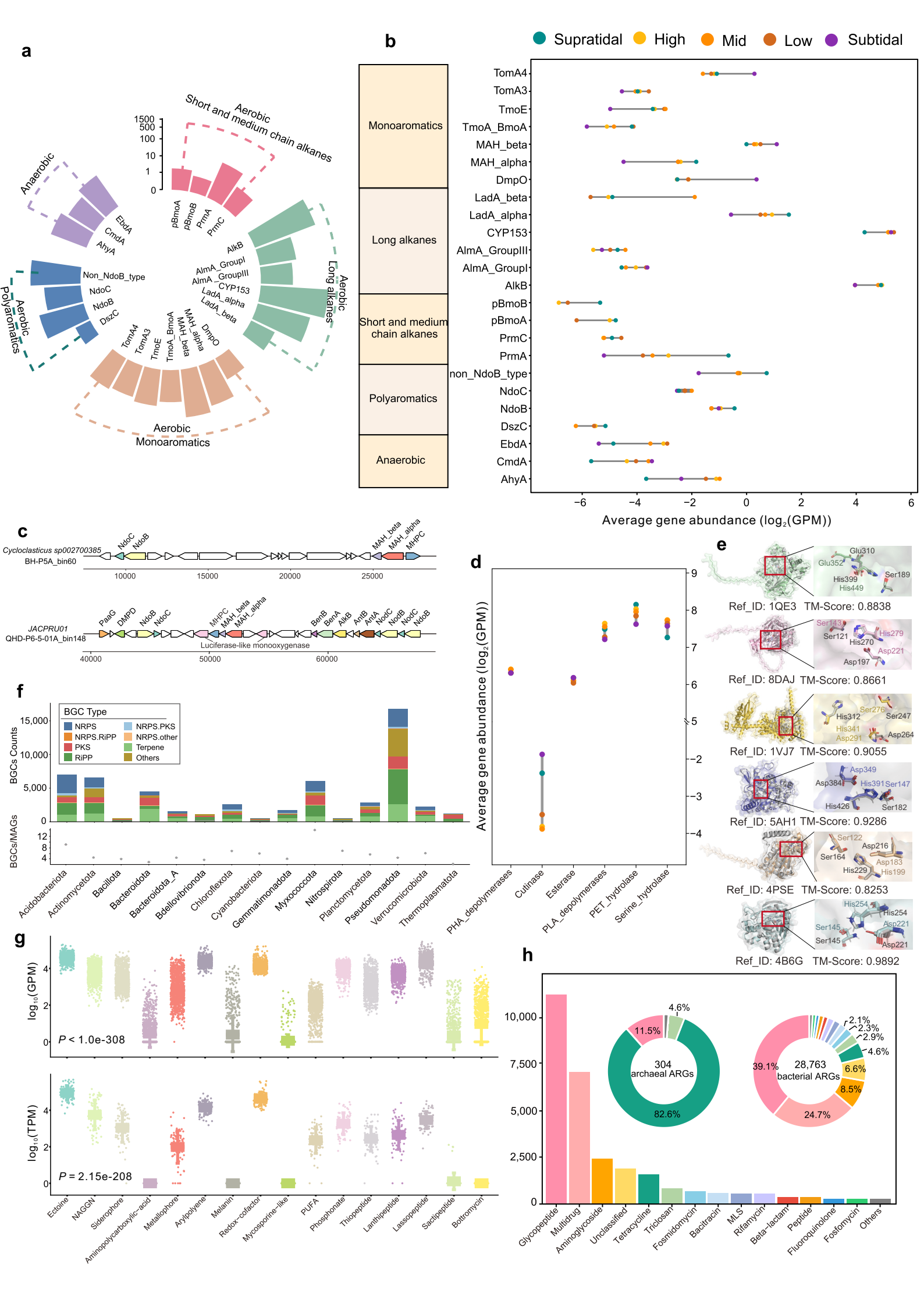

c
HK97-like fold
Strabovirdae (TM-score = 0.58)
Demerecviridae (TM-score = 0.71)
Autographiviridae (TM-score=0.50)
1YUE
1OHG
2XXY
2W0C
3J26
5J7N
Corticoviridae (TM-score = 0.88)
Phycodnaviridae (TM-score = 0.55)
Lavidaviridae (TM-score = 0.84)
Double jelly-roll


d
vOTU counts
488,119
280,416
80,749
38,229
20,097
3,883
14
5
4
3
1
Relative proportion (%)
26.0%
38,299; 99.9% novel
Global ocean
Groundwater
Soil
Sandbeach
Mudflat intertidal sediments
Oil reservoirs
488,133
280,420
80,750
38,255
20,120
3,886
vOTU counts (x10⁵)


e
Domain
Phylum
Family
Class
Archaea(132)
Bacteria(5908)
Burkholderiaceae (366)
Flavobacteriaceae (218)
Mycobacteriaceae (214)
Nocardioidaceae (168)
Actinomycetota(1601)
Bacillota(355)
Bacteroidota(628)
Myxococcota(117)
Patescibacteriota(298)
Planctomycetota(116)
Others (3336)
Pseudomonadaceae (130)
Rhizobiaceae (128)
Pseudomonadota(2793)
Rhodobacteraceae (255)
Sphingomonadaceae (334)
Streptomycetaceae (532)
Unclassfied (144)
Thermoplasmatota(132)
Vibrionaceae (215)
Caudoviricetes (5740)
Faserviricetes (23)
Tectiliviricetes (6)
Unclassified (253)
Maveriviricetes (5)
Megaviricetes (6)
Arfiviricetes (4)
Malgrandaviricetes (1)
Polintoviricetes (2)
Prokaryotic host
Virus


f
vOTU count
Pseudomonadota (946)
Actinomycetota (437)
Bacteroidota (278)
Patescibacteriota (250)
Thermoplasmatota (100)
Myxococcota (83)
Planctomycetota (82)
Verrucomicrobiota (70)
Acidobacteriota (70)
Bacteroidota_A (68)
Chloroflexota (66)
Bdellovibrionota (53)
Gemmatimonadota (41)
Bacillota (34)
Cyanobacteriota (28)
Deinococcota (18)
Armatimonadota (16)
Nitrospirota (11)
Bdellovibrionota_B (11)
Desulfobacterota_B (9)
SAR324 (5)
Halobacteriota (5)
Tectomicrobia (4)
Nanobdellota (4)
J088 (4)
Aenigmatarchaeota (4)
Vulcanimicrobiota (3)
Myxococcota_A (2)
Krumholzibacteriota (2)
Fibrobacterota (2)
EX4484-52 (2)
Campylobacterota (2)
UBA10199 (1)
Omnitrophota (1)
Margulisbacteria (1)
Elusimicrobiota (1)
Eisenbacteria (1)
Desulfobacterota_I (1)
CSSED10-310 (1)
Host abundance (RPKM)


a
Supratidal
Intertidal
Subtidal
Heterotrophy
Microbial degradation of starch
amyA
Microbial degradation of cellulose
CBH2, bglB
Microbial degradation of hemicellulose
manA, abfA
glucose
mannose
arabinose
Redox metabolism
Denitrification
napA
nirS
norB
nosZ
nirK
Arsenate reduction/Arsenite oxidation
Respiratory oxidation
AoxA
DMSO
ArsC
Detoxification reduction
Formate oxidation
FDH
Autotrophy
Calvin–Benson–Bassham cycle
RuBisCO
Calvin Cycle
Sulfur oxidation
sqr
sox
Sulfide
Thiosulfate
Sulfite
b Heterotrophy
c Central carbon metabolism
d Redox process
Tidal zone
Mid−intertidal
Low−intertidal
Transcript abundance (TPM)
e Aerobic respiration
f Autotrophy
g P acquisition
h
Viral DNA
Host DNA
Lytic
Viral replication
Lysogenic
Viral integration
PsbA
ETC
PsaA
CoxA CoxB
AtpA
Photosynthesis
Oxidative phosphorylation
Ppk1
PhoD
Phosphate
Gene-level
Transcriptome-level
Wood-Ljungdahl pathway
TCA
Glucuronate pathway
Xylan degradation
Starch degradation
Glycolysis
Chitin degradation
Pentose phosphate pathway


a
Aerobic
Short and medium chain alkanes
Aerobic
Long alkanes
Aerobic
Monoaromatics
Aerobic
Polyaromatics
Anaerobic
pBmoA
pBmoB
PrmA
PrmC
AlkB
AlmA_GroupI
AlmA_GroupIII
CYP153
LadA_alpha
LadA_beta
MAH_alpha
MAH_beta
DmpO
TmoA_BmoA
TmoE
TomA3
TomA4
Non_NdoB_type
NdoC
NdoB
DszC
EbdA
CmdA
AhyA
b
Supratidal
High
Mid
Low
Subtidal
Monoaromatics
Long alkanes
Short and medium chain alkanes
Polyaromatics
Anaerobic
Average gene abundance (log2(GPM))
c
Cycloclasticus sp002700385
BH-P5A_bin60
JACPRU01
QHD-P6-5-01A_bin148
Luciferase-like monooxygenase
d
Average gene abundance (log2(GPM))
PHA_depolymerases
Cutinase
Esterase
PLA_depolymerases
PET_hydrolase
Serine_hydrolase
e
Ref_ID: 1QE3 TM-Score: 0.8838
Ref_ID: 8DAJ TM-Score: 0.8661
Ref_ID: 1VJ7 TM-Score: 0.9055
Ref_ID: 5AH1 TM-Score: 0.9286
Ref_ID: 4PSE TM-Score: 0.8253
Ref_ID: 4B6G TM-Score: 0.9892
f
BGC Type
NRPS
NRPS.RiPP
PKS
RiPP
NRPS.PKS
NRPS.other
Terpene
Others
BGCs Counts
BGCs/MAGs
Acidobacteriota
Actinomycetota
Bacillota
Bacteroidota
Bacteroidota_A
Bdellovibrionota
Chloroflexota
Cyanobacteriota
Gemmatimonadota
Myxococcota
Nitrospirota
Planctomycetota
Pseudomonadota
Verrucomicrobiota
Thermoplasmatota
g
log10(GPM)
P < 1.0e-308
log10(TPM)
P = 2.15e-208
Ectoine
NAGGN
Siderophore
Aminopolycarboxylic-acid
Metallophore
Arylpolyene
Melanin
Redox-cofactor
Mycosporine-like
PUFA
Phosphonate
Thiopeptide
Lanthipeptide
Lassopeptide
Sactipeptide
Bottromycin
h
304 archaeal ARGs
82.6%
11.5%
4.6%
28,763 bacterial ARGs
39.1%
24.7%
8.5%
6.6%
4.6%
2.9%
2.3%
2.1%
Glycopeptide
Multidrug
Aminoglycoside
Unclassified
Tetracycline
Triclosan
Fosmidomycin
Bacitracin
MLS
Rifamycin
Beta-lactam
Peptide
Fluoroquinolone
Fosfomycin
Others