RESEARCH | Article in Press

Recombination maintains metastable viral populations across global hypersaline ecosystems

Jose M. Haro-Moreno[1] · Juan J. Roda-Garcia[1] · Kaiyang Zheng[2] · Mario López-Pérez[1]

Received: 25 March 2026 / Accepted: 19 August 2026

Abstract

Background Viral populations in natural environments display extensive microdiversity, yet how they maintain genomic cohesion across space and time remains unresolved. Hypersaline crystallizer ponds, characterized by low host diversity and extremely high viral abundance, provide a simplified system to investigate viral population structure under strong environmental selection. Here, we applied long-read metagenomics to the cellular fraction (0.22–5 µm) to characterize viruses associated with *Haloquadratum walsbyi*, enabling direct investigation of actively infecting viral populations and their fine-scale population structure.

Results Long-read sequencing revealed extensive viral genomic diversity that was largely missed by assembly-based approaches, substantially expanding the diversity of *H. walsbyi*-associated viruses. Despite this diversity, viral populations from geographically distant hypersaline systems displayed near-identical population-level nucleotide identity and largely conserved gene content, indicating a globally preserved genomic backbone. In contrast, fine-scale analyses revealed extensive local microdiversity, with limited overlap of single-nucleotide polymorphisms among sites and variability concentrated in genes involved in host interaction and attachment. Consistently low pN/pS ratios indicated pervasive purifying selection, while elevated recombination signals were consistent with frequent genetic exchange among co-occurring viral lineages. Metatranscriptomic analyses further showed that these viral populations remain transcriptionally active throughout the year, with higher activity in winter, and environmental comparisons indicated that salinity acts as a selective pressure shaping viral microdiversity.

Conclusions We propose a "metastable cohesive viral cloud" model in which viral populations maintain global genomic coherence while continuously reshaping local genetic variation. In this framework, recombination and purifying selection contribute to preserving a conserved genomic backbone, whereas environmental filtering drives fine-scale diversification. This regime reconciles global connectivity with local microdiversity and may represent a strategy for persistence among dominant *H. walsbyi*-associated dsDNA viral populations inhabiting high-density microbial ecosystems.

Keywords Long-read metagenomics · Archaeal viruses · Hypersaline environments · Viral microdiversity · Haloquadratum walsbyi · Metastable populations · Viral recombination · Viral population structure

https://doi.org/10.1186/s40168-026-02533-3

Published online: 10 September 2026

Background

Environmental viral populations are not genetically homogeneous entities, but rather consortia of closely related lineages that coexist within populations and collectively structure genetic diversity [1, 2]. This microdiversity is now recognized as a pervasive feature of viral populations across ecosystems and is thought to play a central role in viral adaptation and persistence [3, 4]. However, the widespread presence of extensive microdiversity raises a fundamental question: how do viral populations retain genomic cohesion over space and time without fragmenting into divergent lineages?

Most mechanistic insights into viral microdiversity derive from laboratory virus–host models, in which diversification typically emerges from a single ancestral population under controlled conditions [5, 6]. These studies have highlighted mutation and antagonistic coevolution as primary drivers of diversification [7, 8]. However, they inherently limit ecological complexity and population size, constraining our ability to explain how large natural viral populations maintain both high diversity and genomic coherence in fluctuating environments [9, 10].

In contrast, recent environmental studies have revealed that viral populations exhibit structured microdiversity patterns shaped by ecological and physicochemical gradients, including depth, temperature, and salinity [1–4, 11–13]. These findings suggest that viral evolution in nature reflects the combined effects of dispersal, environmental selection, and host interactions. Nevertheless, the mechanisms that allow viral populations to remain globally cohesive while undergoing continuous local diversification remain poorly understood.

Addressing this question requires natural systems in which ecological complexity is reduced while environmental selection is strong and well defined. Hypersaline solar salterns provide such a system. These artificial coastal environments consist of interconnected ponds of increasing salinity, culminating in salt-saturated crystallizers where microbial diversity collapses while biological activity remains high [14, 15]. Crystallizer ponds sustain extremely dense microbial populations and the highest concentrations of virus-like particles reported for aquatic ecosystems, reaching up to 10^{10} viral-like particles (VLPs) mL^{-1} [16–18]. At near-saturating salinities (*ca.* 35% NaCl), communities are typically dominated by a single archaeal species, *Haloquadratum walsbyi*, which often represents the majority of the microbial biomass [19–23]. This ecological simplicity, combined with an extreme viral abundance, decouples host diversity from viral diversity and renders hypersaline crystallizers powerful natural models for exploring viral microdiversification and population structure under intense environmental selection.

Despite this ecological simplicity, the diversity and population structure of viruses associated with *H. walsbyi* remain poorly resolved, largely due to the difficulty in cultivating this host and the limitations of short-read metagenomics in reconstructing highly microdiverse viral populations [19, 24]. Consequently, current knowledge of *H. walsbyi*-associated viruses (HqrVs) derives almost exclusively from culture-independent approaches [25–27], including fosmid libraries [28] and short-read metagenomic analyses of *Haloquadratum*-enriched viromes [29], which provide limited resolution of population-level genomic structure.

Recent advances in third-generation sequencing technologies offer a powerful alternative to overcome these limitations. Long-read metagenomics enables the direct recovery of complete or near-complete viral genome fragments without reliance on assembly, providing access to highly variable genomic regions that are typically inaccessible using short-read approaches [30]. In this study, we apply long-read metagenomics to the cellular size fraction (0.22–5 μm) of the CR30 crystallizer pond from the Santa Pola solar salterns (Alicante, Spain). By targeting the cellular fraction, we focus on biologically active viral populations engaged in ongoing infections, capturing viral variants within host cells and providing a direct window into viral population structure and microevolution in situ [30, 31].

Using this approach, we test the hypothesis that long-read sequencing overcomes the limitations of second-generation metagenomics in resolving highly heterogeneous viral populations infecting *H. walsbyi*, thereby revealing previously inaccessible viral microdiversity and population-level structure. Our results suggest that recombination, rather than mutation alone, can be the dominant force maintaining cohesion in viral populations inhabiting high-density, low-complexity ecosystems, challenging classical views of viral population structure. This work

establishes hypersaline solar salterns as a powerful natural framework for dissecting viral microevolution under extreme environmental selection.

Results

Assessing viral diversity recovery from the cellular fraction

In order to evaluate the extent to which viral genomic diversity can be recovered from the cellular fraction of hypersaline environments, we analysed two samples collected from the Santa Pola CR30 crystallizer pond in two consecutive summers (2023 and 2024). Both samples were sequenced using Illumina short-read and PacBio Sequel IIe long-read technologies, enabling direct comparison of sequencing strategies applied to the same biological material. Based on these data, we generated three complementary datasets for downstream analyses: short-read assemblies (SRa), long-read assemblies (LRa), and unassembled, but quality-filtered, long reads (LR).

To compare the viral diversity recovered by each approach, we used the large terminase subunit (TerL) as a conserved viral marker. A total of 13,674 TerL sequences were identified across the three datasets, which clustered into 872 non-redundant groups at 90% identity, revealing extensive viral diversity within the cellular fraction (Fig. 1A). The incorporation of representative haloviruses that infect well-known archaeal and bacterial

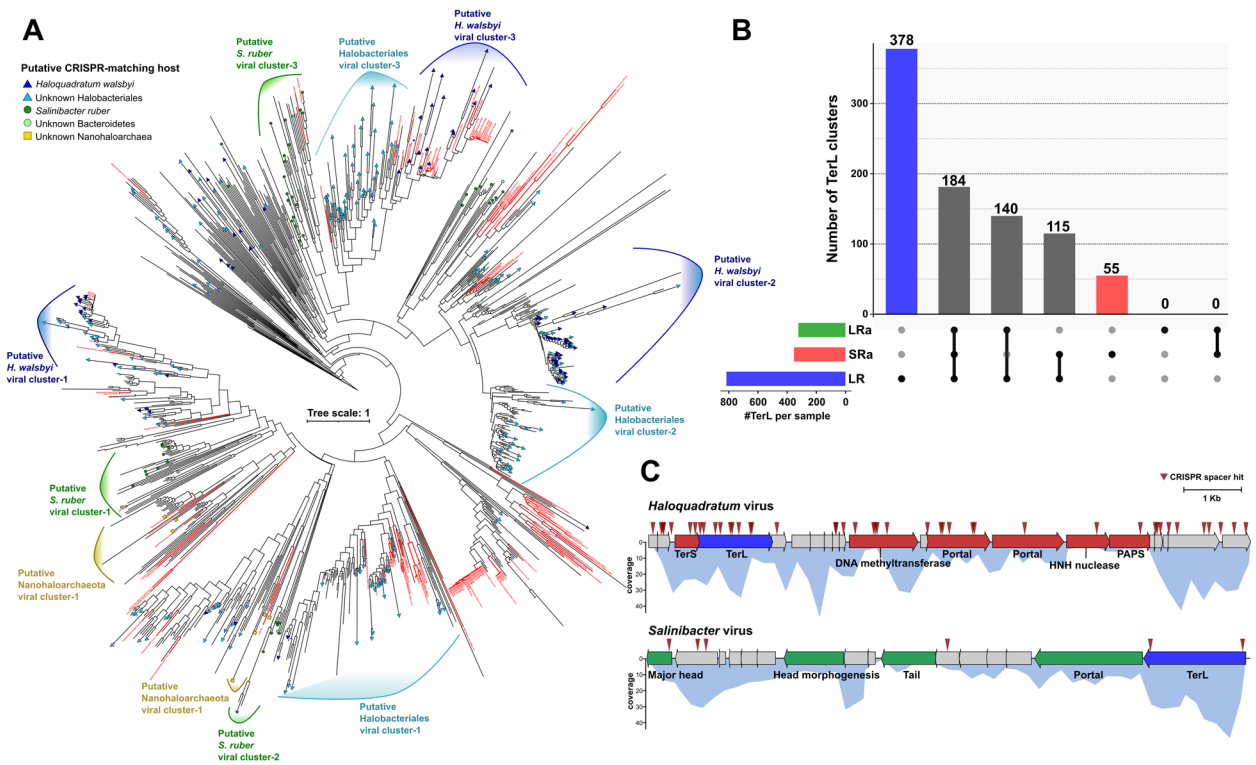

Fig. 1 Salterns viral diversity based on the large terminase subunit (TerL). **A** Maximum-likelihood phylogenetic tree of 872 TerL protein clusters (90% amino acid identity). Sequences and branches colored in red correspond to reference haloarchaeal viruses. Branches not labelled represent the TerL sequences retrieved in this study. Geometric shapes represent putative host-virus interactions based on CRISPR spacer matching. **B** Upset plot with the metagenomic origin (HiFi reads [LR], long-read [LRa], or short-read [SRa] assemblies for the 872 TerL clusters. **C** Distribution of CRISPR spacer matches along representative viral genomes associated with *Haloquadratum walsbyi* (top) and *Salinibacter ruber* (bottom). Arrows indicate predicted genes and red triangles mark CRISPR spacer matches; blue shading represents metagenomic read coverage

species [32] revealed a tendency to cluster closely, although numerous branches were composed solely of TerL sequences obtained in this study.

To infer putative hosts, the spacers from the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas systems were extracted and aligned, with an evidence score of ≥ 3 (see the "Materials and methods" section for a detailed description). Three hundred seventy-three TerL-containing sequences had at least one hit to a CRISPR spacer; however, because most CRISPR systems derived from HiFi reads, their small read size (~ 10 Kb) complicated their putative host taxonomic assignment, and thus many were classified as unknown Halobacteriales (Fig. 1A). Nevertheless, it was possible to assign sequences to putatively infecting members of *H. walsbyi* and *S. ruber*, two of the most abundant species. For them, three large clusters appeared distributed throughout the phylogenetic tree (Fig. 1A).

As shown in Fig. 1B, TerL sequences derived from LR were found in the majority of clusters (94%), whereas sequences derived from LRa and SRa accounted for approximately 37% and 40%, respectively. Consistent with this pattern, the LR dataset contributed, alone, with 378 TerL clusters (~ 43%), while Illumina SRa assemblies recovered only 55, despite a sequencing effort 2.5 times larger. These results demonstrate that assembly-based approaches substantially underestimate the diversity of TerL-containing viral populations in hypersaline cellular fractions, whereas long-read sequencing captures a broader spectrum of these viral lineages.

It should be noted that this analysis is based on the TerL as a conserved viral marker and therefore primarily captures tailed dsDNA viruses and related groups encoding terminase-mediated genome packaging systems. Consequently, other hypersaline viral lineages lacking TerL homologues, including pleolipoviruses, filamentous viruses, and spindle-shaped viruses, are not specifically represented in this dataset. Therefore, the diversity patterns described here should be interpreted as reflecting the diversity of TerL-containing viral populations recovered from the cellular fraction rather than the entirety of hypersaline viral diversity.

Finally, we examined the distribution of CRISPR spacer matches along representative viral genomes linked to *H. walsbyi* (cluster-2) and *S. ruber* (cluster-3). Most spacers were located within conserved viral genes, including terminase subunits and DNA methyltransferases, as well as highly covered conserved proteins of unknown function (Fig. 1C). In contrast, genes located in hypervariable regions, including HNH nucleases, phosphoadenosine 5′-phosphosulphate synthases, and structural proteins such as tail or head components, showed few or no spacer matches.

Recovery of viral genomic sequences reveals a strong bias toward Haloquadratum-associated viruses in long reads

Following the confirmation that the cellular fraction retained a high level of viral diversity, we next sought to recover long viral genomic sequences. As outlined in the Materials and Methods section, we applied a standardized workflow in which viral sequences longer than 15 Kb from SRa, LRa, and LR were retained, while redundancy was removed by dereplication.

This approach yielded 147 dereplicated viral genomic sequences (Supplementary Table S1), spanning from 15.2 to 138.2 Kb and a broad GC range (~ 32–66%). Most sequences were classified within Duplodnaviria/Caudoviricetes, with many assigned to Thumleimavirales and frequently to the family Hafunaviridae. Host-prediction analysis revealed the dominance of HqrVs among them, with 68 linked to *H. walsbyi* (Supplementary Table S1). Notably, all HqrVs were recovered exclusively from the LR dataset, indicating that assembly-based approaches fail to capture a substantial fraction of these highly microdiverse populations. Moreover, ten viral genomes were linked to *Salinibacter ruber*, all of which were likewise recovered from the LR dataset (Supplementary Table S1 and Fig. S1). Lastly, a small number was linked to other archaeal taxa, including *Halorubrum* and *Halogranum* (Supplementary Table S1). Together, these results demonstrate that long-read sequencing preferentially recovers viral genomes infecting the dominant host, *Haloquadratum*, revealing a strong ecological bias that reflects the underlying community structure.

Novelty and genomic context of Haloquadratum-associated viruses inferred by gene-sharing networks

To assess the novelty of the recovered HqrVs, viral genomes were integrated into a gene-sharing network using vConTACT3 [33], together with publicly available *H. walsbyi* viral sequences (Supplementary Table S2) [28, 29]. The resulting network was used to define operational genomic groups based on shared protein content, hereafter referred to as genomic groups (G1 to G15; Supplementary Tables S1 and S2). These groups were used for comparative and ecological analyses and should not be interpreted as formal taxonomic ranks.

Across the network, a substantial fraction of our genomes grouped with previously described HqrVs, indicating continuity with known lineages recovered through fosmid libraries and enrichment-based approaches. For example, G5 contained multiple reference sequences from both the enrichment (MT7*) [29] and fosmid (JQ*) [28] datasets, and also included several of our genomes. Notably, G1, the largest group ($n = 29$), contained almost exclusively genomes from this study, except for a single fosmid-derived reference (JQ807258.1), highlighting previously undersampled diversity within the *H. walsbyi* virosphere.

Provisional genus assignments previously proposed for HqrVs [29] did not correspond consistently to the genomic groups defined by vConTACT3 (Supplementary Tables S1 and S2), nor to phylogenetic relationships inferred using the VIPtree approach (Supplementary Fig. S2). For instance, genomes assigned to the proposed viral genus *Polavirus* clustered within G6, whereas genomes previously classified as *Squarevirus* and *Walsbyivirus* clustered together within G5 (Supplementary Fig. S2). We therefore do not use G1–G15 as genus-level assignments, but as vConTACT-derived genomic groups for downstream comparative analyses.

We also identified several genomic groups (G7–G15) that showed limited connectivity to available references, consistent with divergent or previously uncharacterized lineages. Collectively, these results expand the known genomic landscape of HqrVs, revealing both continuity with previously described lineages and substantial novel diversity uncovered through long-read sequencing.

Geographic distribution and recruitment patterns of Haloquadratum-associated viral genomes

To assess the global distribution and ecological prevalence of *H. walsbyi* viral populations, we quantified genome recruitment across publicly available cellular metagenomes from hypersaline environments worldwide [34]. Viral genomes were considered present when recruitment values were ≥10 reads per genome kilobase per gigabase of metagenome (RPKG) and ≥70% genome coverage (Supplementary Table S3).

Consistent with previous studies, HqrVs' presence and abundance closely followed the distribution of *H. walsbyi* across hypersaline systems [34]. Recruitment patterns revealed a clear geographic structure. The strongest recruitment signals were observed in hypersaline systems from Spain, particularly in the Santa Pola solar salterns, and in a hypersaline inland lake in Argentina, where up to 98 viral genomes were detected, and individual genomes reached maximum recruitment values of ~ 146 RPKG (Supplementary Table S3). Additional, but consistently weaker, recruitment signals were detected in other Spanish hypersaline systems, including salterns from Mallorca and Lanzarote, as well as in inland hypersaline environments from the USA (Utah). Notably, across all analyzed metagenomes, no consistent association was detected between viral recruitment intensity and hypersaline system type, salinity, or ionic composition (Supplementary Table S3) [34].

Analysis at the genomic cluster level revealed distinct distribution–abundance patterns among viral lineages. Clusters G1 and G5 were the most widely distributed, being detected in eight metagenomes spanning six geographically distinct locations across Spain, Argentina, and the USA (Supplementary Table S3). These clusters consistently accounted for the highest cumulative recruitment across samples. In contrast, other clusters exhibited more restricted distributions. Cluster G6 was detected in only three metagenomes but showed high local recruitment where present, with maximum RPKG values exceeding 100 in individual samples from Argentina. Several genomic clusters (including G7, G8, and G10–G15) were not detected in any cellular metagenome, indicating

that these viral lineages are either rare, geographically restricted, or present at abundances below the detection limit of cellular metagenomic datasets.

Seasonal patterns of viral transcriptional activity

Having shown that HqrVs are highly abundant in the cellular fraction, we next examined their transcriptional activity using two publicly available metatranscriptomes from the Santa Pola crystallizer pond, collected in winter 2012 and summer 2014 [35].

Despite size-based filtration excluding free viral particles, viral transcripts were readily detected, consistent with ongoing active infections within the cellular fraction. In the two available metatranscriptomes analyzed here, viral transcriptional activity was markedly higher in winter than in summer, with mean expression values increasing from ~ 10–20 RPKG in summer to ~ 100–150 RPKG in winter (Fig. 2A). The most highly expressed viral genomes belonged predominantly to G5. Applying a conservative threshold of > 5 RPKG to define transcriptionally active genes, 96.5% of them were expressed in at least one of the two sampled seasons. Changes in gene expression between seasons, despite the detection of the same viral genomes in both samples suggest that differences are more consistent with changes in transcriptional output than with complete seasonal replacement of viral populations HqrVs within the analyzed datasets.

Functional annotation of seasonally expressed genes revealed that most winter-overexpressed genes belonged to poorly characterized functional categories. Among annotated genes, structural proteins were the most represented category (13.7%), followed by replication-related genes (3.5%) and genes involved in virus–host interactions (1.1%) (Fig. 2A). Structural and replication genes exhibited strong winter-specific upregulation, while virus–host interaction genes were expressed at lower absolute levels but displayed a comparable seasonal bias.

Intra-sample viral population diversity across sequencing fractions

Given the extensive recruitment of HqrVs across cellular metagenomes from hypersaline systems worldwide and viromes generated exclusively from the Santa Pola solar salterns across a range of salinities [29] (Supplementary Table S3), we quantified intra-population genetic diversity using nucleotide diversity (π) and complementary single-nucleotide polymorphisms (SNPs)-based metrics.

Across the dataset, viral genomes exhibited moderate to high intrapopulation π, with mean values typically ranging from ~ 0.02 to 0.03 and reaching > 0.04 in some cases, indicating the coexistence of multiple closely related variants within each viral population rather than clonal dominance. This interpretation was further supported by average nucleotide identity of recruited reads (ANIr) estimates for the two most abundant and widespread genomic clusters, G1 and G5, calculated from three metagenomes and three Santa Pola viromes showing high recruitment. Mean ANIr values showed measurable variation within samples (G1: metagenomes $95.44 \pm 0.24\%$, viromes $94.98 \pm 0.13\%$; G5: metagenomes $95.11 \pm 0.21\%$, viromes $94.75 \pm 0.09\%$), confirming the presence of genetically heterogeneous, non-clonal viral populations.

For genomes detected in both sequencing fractions (metagenome and virome), π was slightly but consistently higher in viromes (mean 0.0279) than in metagenomes (mean 0.0277), resulting in a small yet reproducible shift (median $\Delta\pi \approx 0.0017$; viromes − metagenomes) that was statistically significant (paired Wilcoxon test, $p \approx 2.3 \times 10^{-8}$). Although differences were small, the consistent increase in diversity within viromes suggests that free viral particles capture a slightly broader fraction of standing genetic variation. Diversity patterns were broadly conserved across genomic clusters, although G1 (median ~ 0.0316) exhibited slightly higher π values than G5 (median ~ 0.0271), while G3 showed the lowest diversity (π typically around ~ 0.02) (Fig. 2B). Differences between fractions were generally small and centered near zero, although a few genomes exhibited larger positive shifts in viromes. The largest positive shifts were observed for a few genomes, such as VG-LRSP088 (G9; $\Delta\pi \approx 0.00545$) and VG-LRSP097 (G3; $\Delta\pi \approx 0.00406$), while rare cases showed modestly higher diversity in metagenomes (e.g., VG-LRSP081 (G3; $\Delta\pi \approx -0.00211$) and VG-LRSP065 (G1; $\Delta\pi \approx -0.00199$)) (Fig. 2B).

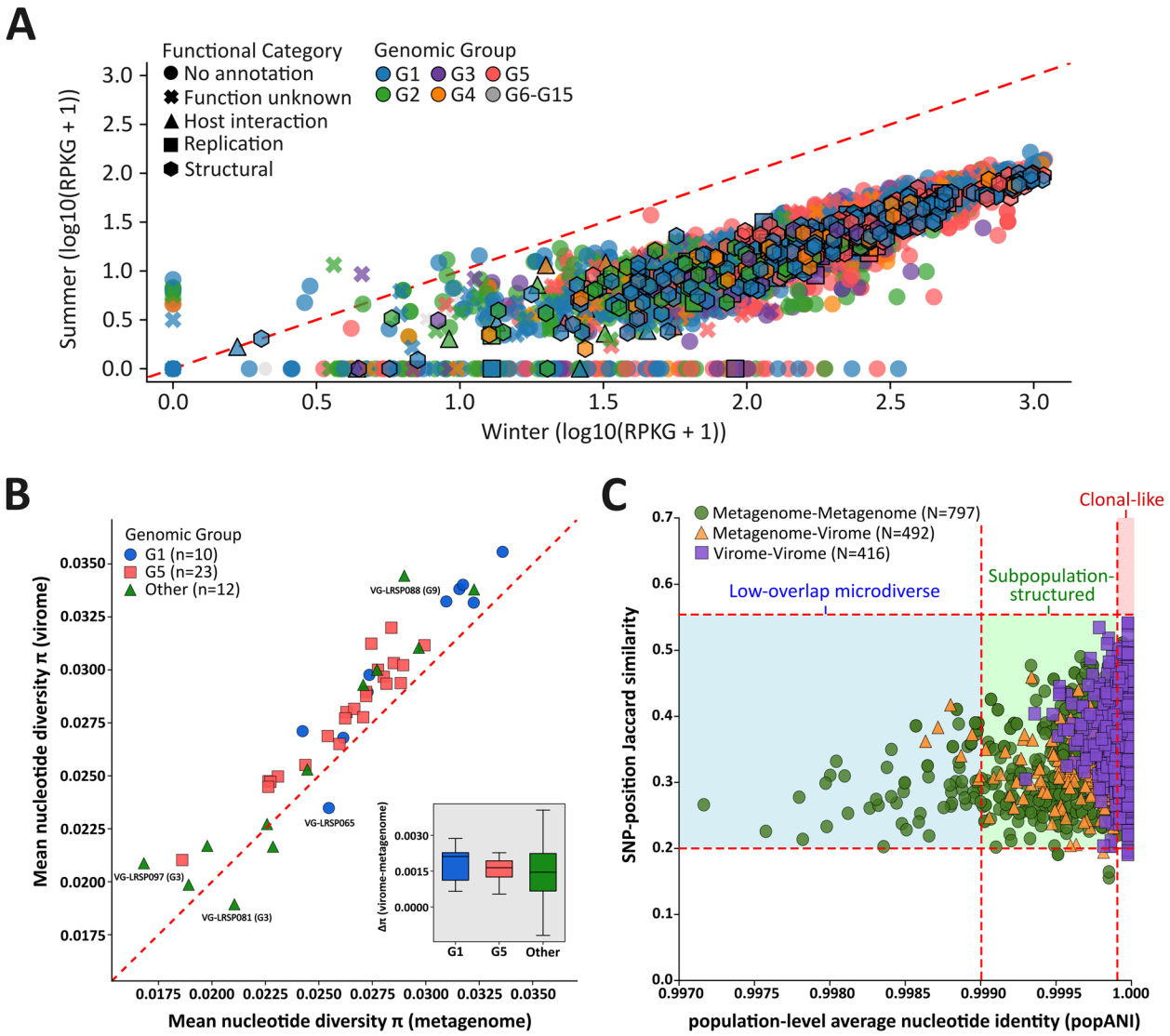

Fig. 2 Intra- and inter-population diversity of *Haloquadratum*-associated viral populations. **A** Seasonal patterns of viral gene expression. Scatter plot comparing viral gene expression levels in winter and summer metatranscriptomes from Santa Pola crystallizer ponds. Each point represents a viral gene, plotted as $\log_{10}(\text{RPKG}+1)$. Colors indicate vConTACT-derived genomic groups (G1–G5; remaining groups in grey), and point shapes denote functional categories. Genes involved in structural functions, replication, and virus–host interactions are highlighted with a black outline. The dashed diagonal line ($y=x$) indicates equal expression between seasons; points above or below the line correspond to higher expression in summer or winter, respectively. Genes with expression < 5 RPKG in both seasons were excluded. **B** Comparison of mean nucleotide diversity (π) estimated from metagenomic and viromic fractions for viral genomes detected in both datasets. Each point represents one viral genome and is colored according to its vConTACT-derived genomic group (G1, G5, or other groups). The dashed line indicates a 1:1 relationship. Selected genomes showing larger positive or negative Δπ values (virome − metagenome) are labeled. The inset boxplot summarizes Δπ distributions by genomic group. **C** Population-level genomic similarity between viral populations based on population average nucleotide identity (popANI) and SNP-position Jaccard similarity. Each point represents a pairwise comparison between samples and is colored by comparison type: metagenome–metagenome (green), metagenome–virome (orange), and virome–virome (purple). Shaded regions delineate operational population states: Clonal-like populations (popANI ≥ 0.9999 and Jaccard ≥ 0.55), Subpopulation-structured populations (popANI ≥ 0.9990 and 0.20 ≤ Jaccard < 0.55), and Low-overlap microdiverse populations (popANI < 0.9990 and 0.20 ≤ Jaccard < 0.55)

SNP-density patterns were consistent with π estimates and further supported the presence of structured but genetically diverse viral populations. Across genomes, SNP density typically ranged from ~ 0.03 to 0.06 and exceeded 0.07 in some cases. SNP-density values were strongly correlated between metagenomes and viromes ($r \approx 0.868$) and showed a small but consistent shift toward higher values in viromes (median ΔSNP density ~ 0.00766). The most abundant and widespread genomic groups (G1 and G5) exhibited consistently higher SNP densities than more restricted lineages, indicating lineage-specific diversity regimes.

A globally distributed viral metapopulation structured by local microdiversity

While intrapopulation analyses revealed that HqrVs populations harbor substantial microdiversity within individual samples, an open question remains as to how these populations are structured across space. Specifically, it is unclear whether viral genomes detected in geographically distant hypersaline systems represent shared populations with common evolutionary histories, or instead they comprise distinct, locally differentiated populations. To address this, we examined population-level similarity across sites using genome-wide comparative metrics based on population-level average nucleotide identity (popANI) and SNP-position Jaccard similarity, defined as the proportion of polymorphic sites shared between two samples (see "Materials and methods" section).

Across all metagenome–metagenome comparisons ($n = 797$), popANI values were uniformly high (mean = 0.99966), indicating near-identical nucleotide composition. In contrast, Jaccard similarity values were lower and more variable (mean = 0.34, range = 0.16–0.51) (Fig. 2C), revealing marked differences in the genomic positions of SNPs among samples. This combination of extremely high nucleotide identity and variable SNP overlap suggests widespread population continuity coupled with fine-scale restructuring of genetic variation.

Based on these population metrics, most comparisons fell within a subpopulation-structured regime, characterized by very high nucleotide identity (popANI ≥ 0.9990) but only partial overlap in SNP positions (0.20 ≤ Jaccard < 0.55) (Fig. 2C). This pattern dominated metagenome–metagenome comparisons spanning large geographic distances, including intercontinental contrasts between Spain and Argentina, as well as comparisons between coastal hypersaline systems in Spain and inland hypersaline environments in the USA (Utah), where approximately 70–75% of pairwise comparisons were classified as subpopulation-structured. Despite near-identical nucleotide identity, these relationships exhibited limited sharing of polymorphic sites, consistent with parallel population histories shaped by local diversification rather than recent clonal exchange. At smaller geographic scales, higher levels of population connectivity were observed. Comparisons between Santa Pola (mainland Spain; the focal site of this study) and Mallorca (Balearic Islands, Spain) showed a more balanced distribution between clonal-like populations (popANI ≥ 0.9999 and Jaccard ≥ 0.55; 46.2%) and subpopulation-structured populations (53.8%), together with higher overall SNP-position overlap (Fig. 2C). This pattern indicates stronger population coupling at shorter geographic distances, even across marine–insular boundaries. Notably, even within Santa Pola, metagenomes sampled across different years exhibited a similar mixture of clonal-like and subpopulation-structured populations (47.6% and 51.5%, respectively), with Jaccard similarities typically around ~ 0.33 (Fig. 2C). These observations indicate persistent local microdiversity rather than temporal replacement by clonally expanding lineages, pointing to a stable coexistence of closely related viral variants through time.

Within Santa Pola, metagenome–virome comparisons conducted under comparable salinity conditions (Δsalinity ≤ 2%; $n = 492$) revealed very high population-level similarity between the cellular and viral fractions. Most viral populations were classified as either clonal-like (53.0%) or subpopulation-structured (44.9%) (Fig. 2C). Clonal-like comparisons were characterized by both extremely high nucleotide identity and strong overlap in SNP positions, whereas subpopulation-structured comparisons combined near-identical nucleotide identity with partial SNP overlap. These patterns demonstrate strong population continuity between the cellular and viral fractions while highlighting frequent fine-scale shifts in intrapopulation structure.

Comparisons between viromes from Santa Pola with different salinities revealed the clearest environmental signal. Although popANI remained extremely high across all salinity differences (mean ≈ 0.99993), Jaccard similarity

decreased progressively with increasing salinity distance. At moderate salinity differences (Δsalinity ≤ 10%), comparisons were dominated by clonal-like and subpopulation-structured relationships with relatively high SNP overlap (median Jaccard ~ 0.37–0.42) (Fig. 2C). At larger salinity differences (Δsalinity > 10%), Jaccard similarity declined toward ~ 0.30 and below, with low-overlap microdiverse populations (popANI < 0.9990 and 0.20 ≤ Jaccard < 0.55) becoming increasingly prevalent (Fig. 2C). This shift reflects growing divergence in the genomic distribution of polymorphisms as environmental distance increases, despite sustained high nucleotide identity.

Pangenome structure and global connectivity of Haloquadratum-associated viruses

Analyses of microdiversity revealed high sequence similarity within and across distantly hypersaline systems. However, whole-genome sequence alignments (Supplementary Fig. S3) and linear recruitment analyses (Supplementary Fig. S4) revealed regions of variable sequence coverage interspersed with conserved regions along the genomes, consistent with heterogeneous sequence conservation across genomic regions. To assess whether this sequence-level homogeneity is reflected in gene content, we first reconstructed the pangenome of the two most abundant and widespread groups (G1 and G5; see "Materials and methods" for details), and subsequently quantified the distribution of these gene families by recruiting metagenomic and viromic reads across samples from Santa Pola spanning three salinity ranges and from four geographically distant locations (Argentina, USA [Utah], and two sites in Spain: Santa Pola and Mallorca).

Viromes recovered a broader and more homogeneous gene repertoire than metagenomes, with an average of approximately 97% of pangenome genes detected per virome sample, compared to metagenomes, despite differences in salinity among samples (Supplementary Fig. S5). Across metagenomes, a substantial fraction of the pangenome was widely shared. In group G5, 150 of 218 genes (68.8%) were detected in all metagenomes, whereas in group G1, 116 of 199 genes (58.3%) showed the same ubiquitous distribution. In contrast, 54 genes (24.8%) in G5 and 64 genes (32.2%) in G1 displayed a restricted distribution across one or a few locations (Supplementary Fig. S5).

Genes detected across all metagenomes were predominantly annotated as virion structural proteins, replication-associated proteins, and other conserved viral functions. Conversely, genes with restricted distributions were enriched in hypothetical proteins and accessory functions. In both genomic groups, among annotated genes absent from all metagenomes (and in a few cases also absent from viromes), several corresponded to structural proteins annotated as tail fiber proteins (Supplementary Fig. S5). This pattern suggests that global genomic coherence is maintained at the level of core genes, while local variability is concentrated in accessory functions.

Pervasive purifying selection across viral populations

Viral microevolution was assessed on the highest-recruiting metagenomes (Santa Pola, Mallorca, and Argentina) and representative Santa Pola viromes across the salinity gradient (30–34% and ≥ 35% salinity) by read mapping and quantification of the ratio of nonsynonymous to synonymous polymorphisms (pN/pS). Genome-level pN/pS distributions were broadly overlapping across metagenomic and viromic datasets and across salinity conditions, with sample-level means spanning from 0.102 (Argentina metagenome) to 0.120 (Santa Pola virome, 30–34% salinity), and median values ranging between 0.099 and 0.113 (Supplementary Fig. S6).

The consistently low pN/pS values observed across samples are indicative of strong and efficient purifying selection acting on viral populations. Such low ratios are consistent with viral lineages that are well adapted to their local environments, in which nonsynonymous variants are removed over time. Together, these patterns are consistent with large, persistent viral populations evolving under long-term ecological stability rather than with recent adaptive radiations or demographic bottlenecks. Conversely, the enrichment of synonymous mutations may reflect elevated recombination rates, as commonly reported for double-stranded DNA viruses [36]. To quantify the contribution of recombination to viral microevolution, we estimated recombination parameters across G1 and G5 HqrVs populations, using samples selected for their high recruitment signals, including three cellular

metagenomes (Argentina, Mallorca and Santa Pola) and three Santa Pola viromes spanning a salinity gradient (20–24%, 30–34% and ≥ 35% salinity).

Viruses from group G1 exhibited consistently higher recombination relative to mutation (γ/μ) than those from group G5 across both sample types (Fig. 3A). In metagenomes, γ/μ averaged 4.37 ± 0.93 for G1 compared to 2.07 ± 0.76 for G5, while in viromes, values were 3.45 ± 0.86 and 2.47 ± 1.00, respectively. These estimates indicate that recombination is a dominant evolutionary force in G1 viruses, whereas G5 viruses display moderate to low recombination. The effective recombining pool (φ_pool) was also substantially larger in G1 than in G5. Mean φ_pool values for G1 were 0.34 ± 0.11 in metagenomes and 0.28 ± 0.09 in viromes, in contrast to 0.13 ± 0.03 and 0.16 ± 0.07 for G5 (Fig. 3A). This suggests that G1 viruses exchange genetic material within a broader and more diverse viral population, whereas recombination in G5 occurs within a more restricted genetic pool. In contrast, the genomic coverage of recombination (c) was comparable between groups, with values ranging from 0.63 to 0.68 across all datasets.

Selection patterns inferred from pN/pS variation across viromes along a salinity gradient

To quantify how selective pressures acting on viral genes vary along the hypersaline gradient, we analyzed viromes from Santa Pola spanning 20–35% salinity [29]. Selection was assessed at the gene level using the pN/pS ratio calculated for all genes encoded in HqrVs, supported by sufficient polymorphism (SNP > 5). Genes under relatively elevated selective pressure were defined as those exceeding the upper quartile of the global pN/pS distribution (Q3 = 0.14). A total of 1359 viral genes met these criteria and were included in the analysis.

Although the absolute differences in median pN/pS values were modest, pN/pS distributions differed significantly across the salinity gradient, with median values increasing from 20% (0.0721) to 25% (0.0814) and peaking at 30% salinity (0.0941), followed by a decline at 35% (0.0798) (Fig. 3B). All pairwise comparisons between salinity levels were statistically significant (Mann–Whitney *U* tests), and all remained significant after FDR correction (maximum FDR = 4.216×10^{-2}). This non-linear pattern indicates that viral genes experience the strongest selective pressures at intermediate–high salinity, rather than at the salinity extreme.

Genes above the Q3 threshold exhibited two dominant response patterns along the gradient (Fig. 3C). One subset of genes (Type 1; 12.7% of the total) showed increasing pN/pS values from 20% salinity to 30% followed by a sharp decline at 35%. A second subset (Type 2; 23.7%) displayed low pN/pS values at 20–30% salinity and a marked increase at 35%. The remaining 296 genes (63.7%) did not follow either trajectory, indicating additional, heterogeneous responses to salinity. Type 1 genes were enriched in annotations related to genome packaging, virion assembly and recombination, including terminase subunits and capsid-associated proteins (Supplementary Fig. S7A). Type 2 genes were dominated by annotations associated with structural and host-interaction functions, including tail-associated proteins and a viral DNA methyltransferase. For example, the DNA methyltransferase gene VG-LRSP086_16 showed pN/pS values of 0.086, 0.135, 0.052 and 0.258 across 20%, 25%, 30% and 35% salinity, respectively (Supplementary Fig. S7B). Stratification by genomic groups revealed that most dominant groups, G1 and G5, exhibited their highest median pN/pS at 30% salinity, despite lineage-specific baselines (Supplementary Fig. S7C). Only a minority of groups peaked at 35% (G14 and G10) or showed mixed trajectories (G7 and G11).

Discussion

Our results reveal that viral populations associated with *H. walsbyi* maintain global genomic coherence despite extensive local microdiversity, defining a coherent evolutionary regime in hypersaline crystallizers. By leveraging long-read metagenomics of the cellular fraction, we were able to capture a substantial fraction of the diversity of TerL-containing *H. walsbyi*-associated viruses that remains largely inaccessible to assembly-based approaches, particularly for highly microdiverse viral populations such as HqrVs.

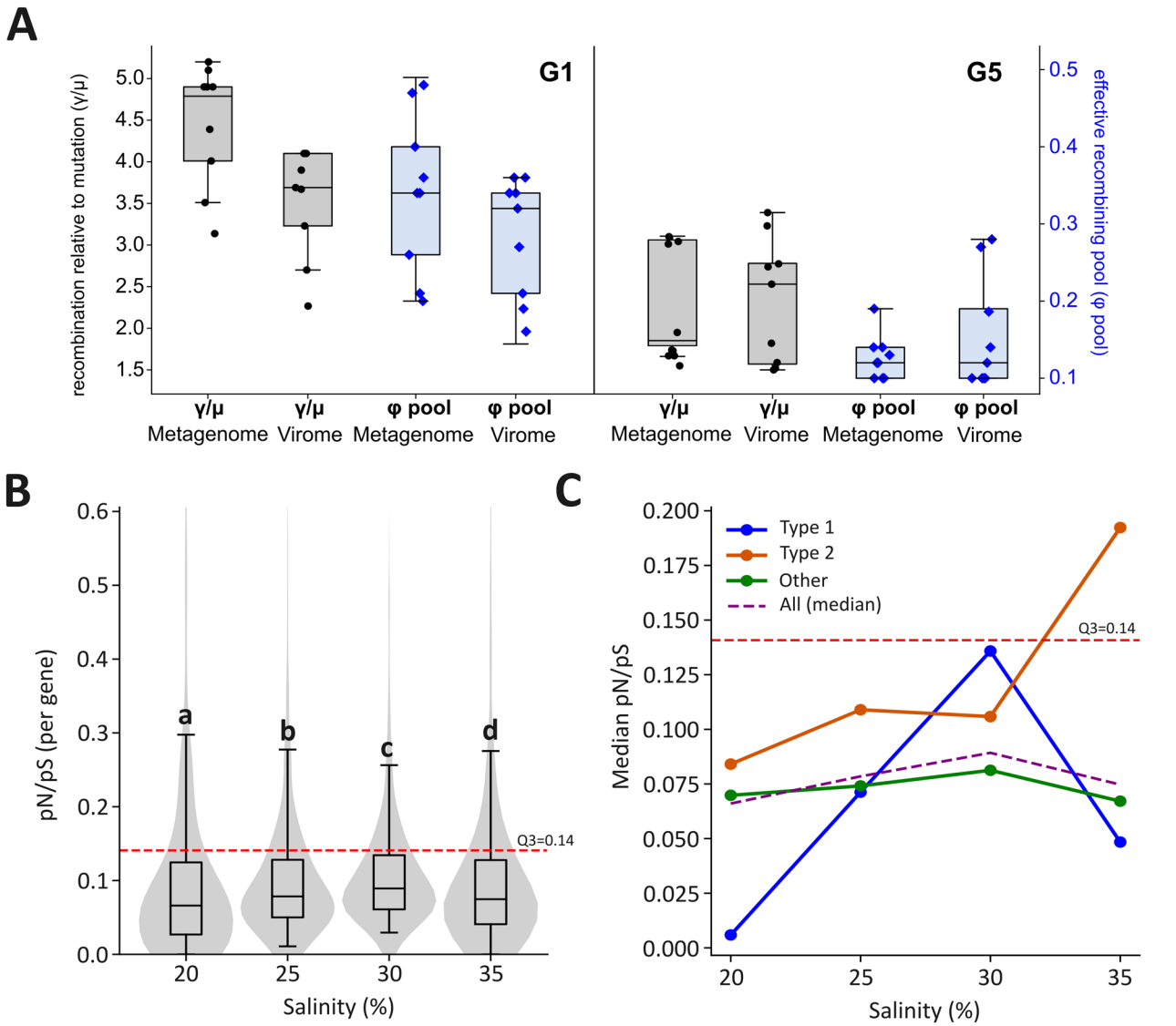

Fig. 3 Recombination, and salinity-dependent selection in *Haloquadratum*-associated viral populations. A. Recombination dynamics across viral genomic groups. Boxplots show estimates of the recombination-to-mutation ratio (γ/μ; left *y*-axis) and the effective recombining pool (φ_pool; right y-axis) for representative viruses from vConTACT-derived genomic groups G1 (VG-LRSP017, VG-LRSP064, JQ807258) and G5 (MT764228, VG-LRSP057, VG-LRSP091). Boxes represent the interquartile range, horizontal lines indicate medians, and points correspond to individual model estimates. **B** Gene-level pN/pS distributions across the salinity gradient. Violin plots show the distribution of gene-level pN/pS values across salinity classes (20–35%). Embedded boxplots indicate the median and interquartile range, with whiskers extending to the 5th and 95th percentiles. Different letters indicate significant differences between salinity groups (Mann–Whitney *U* tests with Benjamini–Hochberg correction, FDR < 0.05). All pairwise comparisons remained significant after correction (maximum FDR = 4.22×10^{-2}). The dashed horizontal line marks the upper quartile of the global pN/pS distribution (Q3 = 0.1408), used as a threshold for relatively elevated selective pressure. **C** Salinity-dependent pN/pS response patterns at the gene level. Median pN/pS trajectories across salinity classes (20–35%) for viral genes grouped by response pattern (type 1, type 2, and other). Lines connect median values across salinities, highlighting distinct selective regimes. Type 1 genes exhibit maximal pN/pS at intermediate–high salinity (30%), whereas Type 2 genes show increasing pN/pS toward the highest salinity (35%). The dashed horizontal line indicates the global upper quartile (Q3) of the pN/pS distribution

Despite this previously hidden diversity, HqrVs exhibit strong genomic coherence across geographically distant hypersaline systems. Viral populations share near-identical nucleotide identity across sites, indicating the presence

of a globally conserved genomic backbone. At the same time, population differentiation is evident as limited overlap in SNP positions between sites, revealing that diversification occurs primarily through local restructuring of standing variation rather than genome-wide divergence. This regime is consistent with a globally connected viral metapopulation in which dispersal is not limiting, but local environmental conditions shape fine-scale population structure. In this context, our results align with the ecological principle that "everything is everywhere, but the environment selects", originally proposed for microbial populations [37]. In HqrVs, global connectivity maintains genomic coherence, whereas environmental selection, particularly salinity, filters which variants persist locally. Importantly, this filtering acts on microdiversity rather than producing distinct, site-specific viral lineages.

Our findings contrast with classical virus–host coevolutionary models such as kill-the-winner [38]. In hypersaline crystallizers, host diversity is extremely low and *H. walsbyi* dominates persistently [39], violating the assumption of dynamic host replacement that underpins kill-the-winner dynamics. Although *H. walsbyi* dominates hypersaline crystallizers, it comprises coexisting subpopulations that differ in surface properties, including S-layer glycosylation patterns known to affect viral adsorption [40]. Our results are consistent with viral populations tracking this fine-scale host heterogeneity through intrapopulation diversification and recombination, rather than through host replacement or population collapse. Instead of recurring host–virus turnover, we observe large, stable viral populations evolving under long-term ecological constancy. Viral regulation in this system appears to operate through modulation of intrapopulation diversity and replication efficiency rather than through host population collapse.

Similarly, the evolutionary dynamics observed here differ from classical quasispecies models, which emphasize mutation-driven diversification [41]. The consistently low pN/pS values observed across samples indicate strong purifying selection, arguing against mutation-driven adaptive diversification. Instead, our data point to homologous recombination as a dominant force structuring viral populations. High recombination rates among co-circulating viral variants promote homogenization of the core genome while maintaining extensive intrapopulation variation [42]. This regime is consistent with theoretical and computational models describing metastable populations, in which large, persistent populations retain a stable genomic identity over long periods while diversity is continuously reshuffled at a fine scale [43].

The ecological context of hypersaline crystallizers provides the conditions that likely sustain this recombination-driven metastable regime. Extreme host densities favor frequent coinfection and intracellular recombination, while the strong conservation of genome synteny facilitates homologous recombination across large genomic regions [1, 44]. Together with efficient purifying selection, this allows deleterious mutations to be purged without reducing overall diversity. Salinity further modulates this evolutionary regime. Viral abundance, transcriptional activity and microdiversity peak at intermediate–high salinities, while more extreme salinities are associated with stronger, more directional selection. Because the dominant host remains largely constant along this gradient, these patterns are best explained by direct physicochemical selection acting on viral populations rather than by host turnover. Seasonal transcriptomic data further indicate year-round persistence and suggest increased viral activity during winter, although additional temporal datasets will be required to determine whether this pattern is recurrent, supporting a model of stable but environmentally responsive viral populations.

An important limitation of our study is that recombination parameters could be robustly estimated only for the most abundant HqrV genomic groups, mainly G1 and G5. Therefore, the recombination-driven metastable model proposed here should be interpreted as applying to dominant HqrV populations inhabiting hypersaline crystallizers, rather than as a general model for all hypersaline viruses. Viral groups with lower abundance, restricted distributions, or insufficient recruitment may experience different balances between mutation, recombination, selection, and drift.

In summary, we propose that HqrVs in hypersaline crystallizers follow a "metastable cohesive viral cloud" model (Fig. 4). In this regime, global genomic coherence is maintained by frequent recombination and strong purifying selection, while local environmental filtering continuously reshapes intrapopulation diversity without leading to lineage replacement or adaptive sweeps. This evolutionary strategy is favored by the ecological simplicity

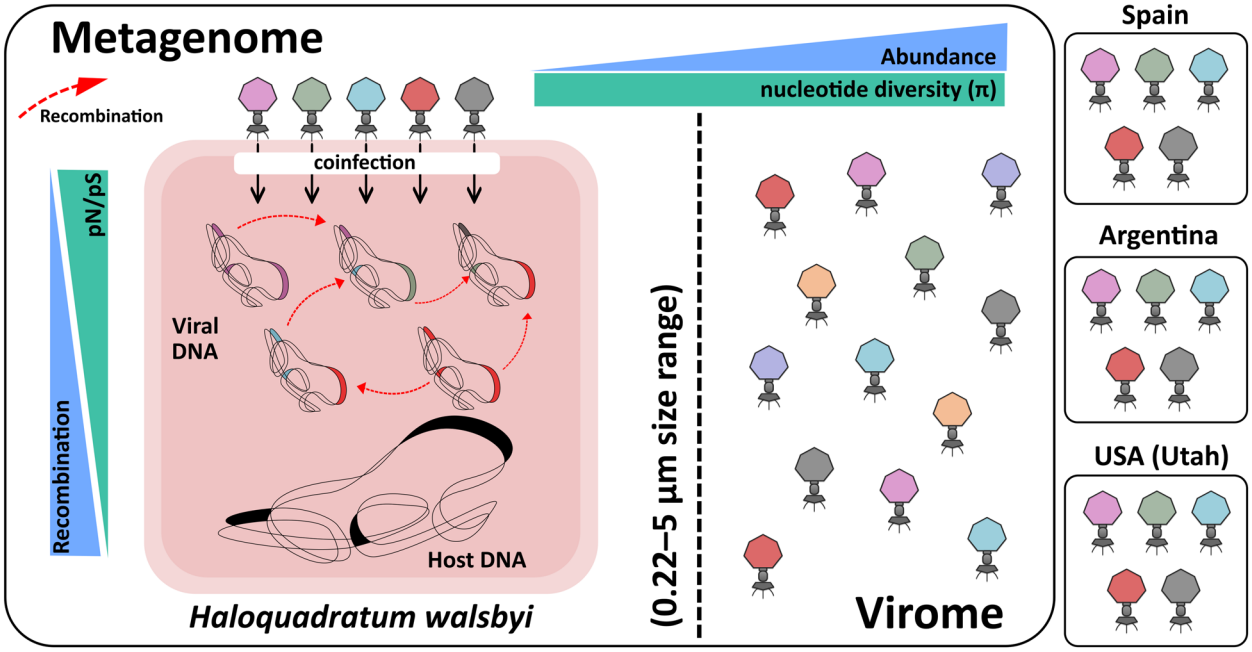

Fig. 4 Metastable cohesive viral cloud model for *Haloquadratum*-associated viruses in hypersaline crystallizers. Conceptual model illustrating the evolutionary regime of *Haloquadratum*-associated viruses in hypersaline crystallizers analysed in this study. High host densities promote frequent coinfection and superinfection within the cellular fraction (0.22–5 µm), facilitating homologous recombination among co-circulating viral variants. This process maintains a globally conserved genomic backbone while continuously reshuffling local microdiversity, consistent with high nucleotide diversity, low pN/pS ratios, and elevated recombination rates. Together, these dynamics are consistent with a metastable viral population in which recombination contributes to population cohesion despite ongoing local diversification

and physicochemical stability of hypersaline crystallizers, where extreme host densities promote coinfection and recombination while enabling long-term population persistence. Although recombination parameters could only be robustly estimated for the most abundant genomic groups (G1 and G5), our results suggest that recombination is a major force contributing to population cohesion in dominant HqrV populations inhabiting hypersaline crystallizers. By revealing this recombination-driven metastability through long-read metagenomics, our study provides a framework for investigating how recombination, selection, and environmental filtering interact to shape the evolution of archaeal viruses in extreme environments.

Materials and methods

Sampling site and DNA extraction

Two water samples were collected from the CR30 crystallizer pond of the Bras del Port solar salterns, Santa Pola, Alicante, Spain (38°11′43.0″N, 0°35′30.7″W) during consecutive years, in July 2023 and July 2024. For each sampling event, approximately 15–20 L of brine were filtered until filter saturation was reached. Salinity was measured in situ using a handheld refractometer, yielding 35.9% (359 g kg^{-1}) at 27 °C. Samples were sequentially filtered through 20, 5, and 0.22 µm pore filter polycarbonate filters (Millipore) and immediately stored at – 80 °C until processing. DNA extraction was performed from the 0.22 µm filter following the MagAttract Purification Kit protocol (QIAGEN) according to the manufacturer's instructions.

Sequencing, read processing, and metagenome assembly

Samples were sequenced using PacBio Sequel IIe and Illumina Novaseq X platforms (Macrogen, Republic of Korea) (Supplementary Table S4). Raw Illumina reads were trimmed using Trimmomatic v0.39 [45] with parameters. PacBio Sequel IIe reads were provided by the manufacturer as HiFi reads (phred score of 20 or higher), but further processing was applied. Specifically, using NanoFilt [46], sequences larger than 1,000 bp and with a *Q* value ≥ 30 were kept. Short-read assemblies (SRa) were generated individually for each sample using metaSPAdes v3.15 [47]. Assembly parameters included –meta –threads 64 –memory 950 -k 49,59,69,79,89,99. Long-read assemblies (LRa) were generated using flye [48], with the meta option and ten rounds of post-polishing (Supplementary Table S4).

Viral sequence recovery and curation

Viral sequences were identified from LRa and LR datasets ≥ 15 Kb using VIBRANT v1.2.0 [49] with default parameters. Redundant nucleotide sequences were deduplicated at 99% identity using CD-HIT v4.8.1 [50]. Contig quality was assessed with CheckV v1.0.1 [51]**.** Potential viral hosts and lifestyle inference were predicted using iPHoP v1.3.2 [52], complemented with Phabox v2.0 [53], and geNomad v1.8.0 [54]. Host predictions were further evaluated using protein-based phylogenetic trees generated with VIPtree [55], including relevant viral reference genomes. Taxonomic assignment and genomic clustering were performed using vConTACT3 [33]. Genomic clustering was performed using vConTACT3 to facilitate comparative analyses. The resulting clusters (G1–G15) were used as operational genomic groups to organize related viral sequences and do not represent formal International Committee on Taxonomy of Viruses (ICTV) taxonomic assignments. Because many recovered sequences were incomplete, ICTV genus-level classification based on intergenomic similarity was not attempted. For comparative genomic analyses, all publicly available viral genomes reported as associated with *H. walsbyi* were retrieved from public databases (Supplementary Table S2). Genomes MT764231.1, MT764232.1, MT764233.1 and MT764235.1 were excluded because they displayed GC contents (from 56.3 to 66.7%) substantially higher than those observed for both *H. walsbyi* (approximately 48%) and the remaining *Haloquadratum*-associated viral genomes (typically 43–45% GC). In addition, genomes JQ807259.2 and JQ807261.2 were excluded because they did not cluster with the main HqrVs in ViPTree protein-based phylogenies and showed phylogenetic placements outside the major halophilic viral clades. These genomes were therefore excluded from downstream analyses to maintain a conservative dataset of putative *Haloquadratum*-associated viruses (Supplementary Table S2).

Recovery and phylogeny of TerL protein sequences from LR and metagenomic assemblies

To capture most of the viral diversity within the solar saltern metagenomes, a hidden Markov model (HMM) dataset of the large terminase subunit TerL was made, combining the HMM profiles from the Prokaryotic Virus Orthologous Groups (pVOG) database [56], the Virus Orthologous Groups database (VOGDB, release 229) [57] and a custom HMM containing the TerL sequences from *bonafide* haloviruses downloaded from the NCBI. Metagenomic proteins from LR, SRa, and LRa sequences ≥ 5 Kb long were predicted using prodigal [58] and later screened for putative TerL hits using hmmscan [59]. Hits with a bitscore ≥ 50 and an e-value ≤ 10^{-10} were extracted, clustered using CD-HIT (90% identity, global alignment), and aligned using MUSCLE v5 [60]. The resulting alignment was trimmed using trimAl [61], and a maximum-likelihood phylogenetic tree was created using IQ-TREE v3 [62] with the VT + F + R10 model as the best fitted model and 5000 ultrafast bootstraps. iTOL [63] was used to visualize the tree.

Screening of CRISPR-Cas systems from LR and metagenomic assemblies

Nucleotide sequences ≥ 5 Kb from LR, SRa, and LRa datasets were screened for putative CRISPR systems using minCED (https://github.com/ctSkennerton/minced), with default options. The resulting CRISPR systems and associated Cas proteins were further confirmed using CRISPRCasFinder [64] and CRISPRCasTyper [65], selecting CRISPR systems with or without Cas-associated proteins that had an evidence score of ≥ 3.

The taxonomic affiliation of nucleotide sequences containing a CRISPR system (i.e., assembled contigs and HiFi reads) was determined by aligning their predicted proteins against the NCBI non-redundant database using diamond [66]. At least half of the proteins must share the same taxon to be classified as such. Finally, to identify the putative hosts of the viral sequences containing a *terL* (see above), CRISPR spacers were aligned against them using BLASTn [67], with the task option "blastn-short". Only those matches that exhibited 100% sequence coverage of the spacer length and a maximum of one mismatch were designated as true positives.

Metagenomic read recruitment in environmental samples

Viromic [29] and metatranscriptomic [35] samples from Santa Pola were downloaded from the European Nucleotide Archive (ENA) under BioProject accessions PRJEB69609 and PRJNA633445, respectively. In addition, global metagenomic samples from hypersaline environments [39] were downloaded from the BioProject accession numbers PRJEB75750 and PRJEB45291. Viral genomes were used to recruit metagenomic reads via BLASTn v2.10.1 with a minimum nucleotide identity of 99% over ≥ 50 bp and a minimum genome coverage of 70%. Recruited values were normalized as RPKG.

Microdiversity analysis

Illumina reads were competitively mapped to a FASTA file containing all genomes using Bowtie2 [68] and analyzed with inStrain v1.8.0 [69] to estimate intrapopulation nucleotide diversity. Gene-level selective pressure was quantified using the pN/pS ratio. To ensure robust estimates, only viral genomes with more than five detected SNPs were retained for downstream analyses. In addition, only genomic positions supported by a minimum read depth of 10× were considered, and viral genomes were required to exhibit at least 80% breadth of coverage. These thresholds were applied to minimize stochastic effects associated with low coverage and to ensure reliable population-level microdiversity estimates.

Population-level genomic similarity analysis

Population-level similarity between viral populations was quantified using popANI and SNP-position Jaccard similarity. Both metrics were calculated using inStrain v1.8.0 in *compare* mode. inStrain *profile* was run using default parameters, requiring a minimum read mapping quality of 30, minimum base quality of 20, and a minimum genome breadth of coverage of 80%. Pairwise comparisons were restricted to genomic positions supported by a minimum read depth of 10 × in both samples. Based on the empirical distribution of these metrics and previous population-genomic frameworks, population comparisons were classified into operational categories in the popANI–Jaccard space: Clonal-like (popANI ≥ 0.9999 and Jaccard ≥ 0.55), Subpopulation-structured (popANI ≥ 0.9990 and 0.20 ≤ Jaccard < 0.55), Low-overlap microdiverse (popANI < 0.9990 and 0.20 ≤ Jaccard < 0.55), and Divergent populations (Jaccard < 0.20). These categories reflect increasing degrees of population similarity and internal structure and were used as descriptive tools rather than strict taxonomic assignments.

To assess gene family prevalence across the G1 and G5 groups, pangenome reconstruction was performed using PPanGGOLiN [70], applying a minimum sequence identity threshold of 80% and a minimum alignment coverage of 50%. Proteins were functionally annotated by comparison against HMM profiles using hmmscan. Searches were conducted against the pVOGs, VOGDB, and viral Families database (vFam release 227) [71].

Recombination analysis

High-quality trimmed Illumina metagenomic reads were aligned against representative HqrVs genomes using Bowtie2 v2.4.5 in *–sensitive-local* mode. Recombination analyses focused on viral populations belonging to genomic groups G1 and G5. Three representative genomes per group were selected based on their high abundance and broad distribution across samples: VG-LRSP017, VG-LRSP064, and JQ807258.1 for G1, whereas VG-LRSP057, VG-LRSP091, and MT764228 for G5. The resulting SAM files were converted and sorted into BAM files using SAMtools [72]. Homologous recombination parameters were estimated using mcorr [73], which infers recombination dynamics from correlations in synonymous substitutions along the genome. Specifically, we estimated the ratio of recombination to mutation (γ/μ), the effective recombining pool (φ_pool), and the fraction of the genome affected by recombination (c).

Statistical analysis

Salinity-dependent variation in gene-level pN/pS distributions was evaluated using Mann–Whitney *U* tests, with *p*-values adjusted for multiple comparisons using the Benjamini–Hochberg false discovery rate correction. Statistical analyses were performed in Python (v3.9), and adjusted p-values < 0.05 were considered significant.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1186/s40168-026-02533-3.

Acknowledgements We thank the owners of the "Bras del Port" saltern of Santa Pola (Alicante, Spain) for allowing access to the premises and to collect samples.

Authors' contributions MLP conceived the study. Environmental samples were collected by JRG and MLP. Data analysis was performed by JHM, JRG, KZ, and MLP. The manuscript was written by MLP and JHM. All authors reviewed and approved the final version.

Funding This research was funded by the project "MICRO3GEN" (PID2023-150293NB-I00), awarded by the Spanish Ministry of Economy, Industry and Competitiveness, and co-financed with FEDER funds, to MLP.

Data availability Illumina raw reads and PacBio CCS reads have been submitted to the European Nucleotide Archive (ENA) and are available under BioProject accession number PRJEB107280. In addition, Viral sequences are publicly available in Zenodo (https://doi.org/10.5281/zenodo.19218717).

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Gregory AC, Zayed AA, Conceição-Neto N, Temperton B, Bolduc B, Alberti A, et al. Marine DNA viral macro- and microdiversity from pole to pole. Cell. 2019;177:1109-1123.e14. https://doi.org/10.1016/j.cell.2019.03.040.
2. Coutinho FH, Rosselli R, Rodríguez-Valera F. Trends of microdiversity reveal depth-dependent evolutionary strategies of viruses in the Mediterranean. mSystems. 2019;4:1–17. https://doi.org/10.1128/mSystems.00554-19.
3. Zhong Z-P, Vik D, Rapp JZ, Zablocki O, Maughan H, Temperton B, et al. Lower viral evolutionary pressure under stable versus fluctuating conditions in subzero Arctic brines. Microbiome. 2023;11:174. https://doi.org/10.1186/s40168-023-01619-6.
4. Peng Y, Lu Z, Pan D, Shi L-D, Zhao Z, Liu Q, et al. Viruses in deep-sea cold seep sediments harbor diverse survival mechanisms and remain genetically conserved within species. ISME J Nature Publishing Group. 2023;17:1774–84. https://doi.org/10.1038/s41396-023-01491-0.
5. Buckling A, Rainey PB. Antagonistic coevolution between a bacterium and a bacteriophage. Proc Biol Sci. 2002;269:931–6. https://doi.org/10.1098/rspb.2001.1945.
6. Meyer JR, Dobias DT, Weitz JS, Barrick JE, Quick RT, Lenski RE. Repeatability and contingency in the evolution of a key innovation in phage lambda. Science. 2012;335:428–32. https://doi.org/10.1126/science.1214449.
7. Marston MF, Pierciey FJ, Shepard A, Gearin G, Qi J, Yandava C, et al. Rapid diversification of coevolving marine *Synechococcus* and a virus. Proc Natl Acad Sci U S A. 2012;109:4544–9. https://doi.org/10.1073/pnas.1120310109.
8. Avrani S, Lindell D. Convergent evolution toward an improved growth rate and a reduced resistance range in Prochlorococcus strains resistant to phage. Proceedings of the National Academy of Sciences. Proceed Nation Acad Sci. 2015;112:E2191–200. https://doi.org/10.1073/pnas.1420347112
9. Dennehy JJ. What can phages tell us about host-pathogen coevolution? Int J Evol Biol. 2012. https://doi.org/10.1155/2012/396165.
10. Hall AR, Scanlan PD, Morgan AD, Buckling A. Host-parasite coevolutionary arms races give way to fluctuating selection: bacteria-phage coevolutionary dynamics. Ecol Lett. 2011;14:635–42. https://doi.org/10.1111/j.1461-0248.2011.01624.x.
11. Burmeister AR, Lenski RE, Meyer JR. Host coevolution alters the adaptive landscape of a virus. Proc Biol Sci. 2016;283:20161528. https://doi.org/10.1098/rspb.2016.1528.
12. Koskella B, Brockhurst MA. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. FEMS Microbiology Reviews Narnia. 2014;38:916–31. https://doi.org/10.1111/1574-6976.12072.
13. Chen T, Xiong Y, Zhang J, Zhang Q, Wu J, Xu N, et al. Temporal dynamics, microdiversity, and ecological functions of viral communities during cyanobacterial blooms in Lake Taihu. npj Biofilms Microbiomes. Nature Publishing Group; 2025;11:178. https://doi.org/10.1038/s41522-025-00771-1
14. Viver T, Orellana LH, Díaz S, Urdiain M, Ramos-Barbero MD, González-Pastor JE, et al. Predominance of deterministic microbial community dynamics in salterns exposed to different light intensities. Environ Microbiol. 2019;21:4300–15. https://doi.org/10.1111/1462-2920.14790.
15. Konstantinidis KT, Viver T, Conrad RE, Venter SN, Rossello-Mora R. Solar salterns as model systems to study the units of bacterial diversity that matter for ecosystem functioning. Curr Opin Biotechnol. 2022;73:151–7. https://doi.org/10.1016/j.copbio.2021.07.028.
16. Santos F, Yarza P, Parro V, Meseguer I, Rosselló-Móra R, Antón J. Culture-independent approaches for studying viruses from hypersaline environments. Appl Environ Microbiol Am Soc Microbiol. 2012;78:1635–43. https://doi.org/10.1128/AEM.07175-11.
17. Guixa-Boixareu N, Calderón-Paz J, Heldal M, Bratbak G, Pedrós-Alió C. Viral lysis and bacterivory as prokaryotic loss factors along a salinity gradient. Aquat Microb Ecol. 1996;11:215–27. https://doi.org/10.3354/ame011215.
18. Dyall-Smith M, Tang S-L, Bath C. Haloarchaeal viruses: how diverse are they? Res Microbiol. 2003;154:309–13. https://doi.org/10.1016/S0923-2508(03)00076-7.
19. Antón J, Llobet-Brossa E, Rodríguez-Valera F, Amann R. Fluorescence *in situ* hybridization analysis of the prokaryotic community inhabiting crystallizer ponds. Environ Microbiol. 1999;1:517–23. https://doi.org/10.1046/j.1462-2920.1999.00065.x.
20. Oh D, Porter K, Russ B, Burns D, Dyall-Smith M. Diversity of Haloquadratum and other haloarchaea in three, geographically distant. Australian saltern crystallizer ponds Extremophiles. 2010;14:161–9. https://doi.org/10.1007/s00792-009-0295-6.
21. Mora-Ruiz MdelR, Cifuentes A, Font-Verdera F, Pérez-Fernández C, Farias ME, González B, et al. Biogeographical patterns of bacterial and archaeal communities from distant hypersaline environments. Syst Appl Microbiol. 2018;41:139–50. https://doi.org/10.1016/j.syapm.2017.10.006.
22. Bolhuis H, Palm P, Wende A, Falb M, Rampp M, Rodriguez-Valera F, et al. The genome of the square archaeon *Haloquadratum walsbyi*: life at the limits of water activity. BMC Genomics. 2006;7:169. https://doi.org/10.1186/1471-2164-7-169.

23. Elevi Bardavid R, Khristo P, Oren A. Interrelationships between *Dunaliella* and halophilic prokaryotes in saltern crystallizer ponds. Extremophiles. 2008;12:5–14. https://doi.org/10.1007/s00792-006-0053-y.
24. Burns DG, Janssen PH, Itoh T, Kamekura M, Li Z, Jensen G, et al. *Haloquadratum walsbyi* gen. nov., sp. nov., the square haloarchaeon of Walsby, isolated from saltern crystallizers in Australia and Spain. Int J Syst Evol Microbiol. 2007;57:387–92. https://doi.org/10.1099/ijs.0.64690-0.
25. Santos F, Yarza P, Parro V, Briones C, Antón J. The metavirome of a hypersaline environment. Environ Microbiol. 2010;12:2965–76. https://doi.org/10.1111/j.1462-2920.2010.02273.x.
26. Sime-Ngando T, Lucas S, Robin A, Tucker KP, Colombet J, Bettarel Y, et al. Diversity of virus–host systems in hypersaline Lake Retba. Senegal Environmental Microbiology. 2011;13:1956–72. https://doi.org/10.1111/j.1462-2920.2010.02323.x.
27. Emerson JB, Thomas BC, Andrade K, Allen EE, Heidelberg KB, Banfield JF. Dynamic viral populations in hypersaline systems as revealed by metagenomic assembly. Appl Environ Microbiol. 2012;78:6309–20. https://doi.org/10.1128/AEM.01212-12.
28. Garcia-Heredia I, Martin-Cuadrado AB, Mojica FJM, Santos F, Mira A, Antón J, et al. Reconstructing viral genomes from the environment using fosmid clones: the case of haloviruses. PLoS ONE. 2012. https://doi.org/10.1371/journal.pone.0033802.
29. Villamor J, Ramos-Barbero MD, Moreno-Paz M, Villena-Alemany C, Martínez-García M, Parro V, et al. Novel viruses of *Haloquadratum walsbyi* expand the known archaeal virosphere of hypersaline environments. ISME J. 2025;19:wraf149. https://doi.org/10.1093/ismejo/wraf149.
30. Zaragoza-Solas A, Haro-Moreno JM, Rodriguez-Valera F, López-Pérez M. Long-read metagenomics improves the recovery of viral diversity from complex natural marine samples. mSystems. 2022;7:e0019222. https://doi.org/10.1128/msystems.00192-22.
31. López-Pérez M, Haro-Moreno JM, Gonzalez-Serrano R, Parras-Moltó M, Rodriguez-Valera F. Genome diversity of marine phages recovered from Mediterranean metagenomes: size matters. PLoS Genet. 2017;13:e1007018. https://doi.org/10.1371/journal.pgen.1007018.
32. Liu Y, Demina TA, Roux S, Aiewsakun P, Kazlauskas D, Simmonds P, et al. Diversity, taxonomy, and evolution of archaeal viruses of the class Caudoviricetes. Suttle C, editor. PLoS Biol. 2021;19:e3001442. https://doi.org/10.1371/journal.pbio.3001442
33. Bolduc B, Zablocki O, Turner D, Bin Jang H, Guo J, Adriaenssens EM, et al. Machine learning enables scalable and systematic hierarchical virus taxonomy. Nat Biotechnol. Nature Publishing Group; 2025;1–10. https://doi.org/10.1038/s41587-025-02946-9
34. Viver T, Gago JF, Bustos-Caparros E, Aldeguer-Riquelme B, Rodriguez-R LM, Ramirez AS, et al. Metagenomics reveal allopatric speciation and higher connectivity among coastal vs. inland hypersaline lakes and solar salterns. Microbiol. 2025 Cited 2026 Mar 20. https://doi.org/10.1101/2025.10.01.679725
35. Rosselli R, López-Pérez M, Martin-Cuadrado A-B, Rodriguez-Valera F, Bolhuis H. Differences in gene expression patterns between cultured and natural Haloquadratum walsbyi ecotypes. Front Microbiol. 2022;13:1044446. https://doi.org/10.3389/fmicb.2022.1044446.
36. Pérez-Losada M, Arenas M, Galán JC, Palero F, González-Candelas F. Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. Infect Genet Evol. 2015;30:296–307. https://doi.org/10.1016/j.meegid.2014.12.022.
37. Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, et al. Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol. 2006;4:102–12. https://doi.org/10.1038/nrmicro1341.
38. Thingstad TF. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol Oceanogr. 2000;45:1320–8. https://doi.org/10.4319/lo.2000.45.6.1320.
39. Bustos-Caparros E, Viver T, Gago JF, Avontuur JR, Amiour S, Baxter BK, et al. Global dominance of *Haloquadratum walsbyi* by a single genomovar with distinct gene content and viral cohorts from close relatives. ISME J. 2025;19:wraf165. https://doi.org/10.1093/ismejo/wraf165.
40. López-Pérez M, Martin-Cuadrado AB, Rodriguez-Valera F. Homologous recombination is involved in the diversity of replacement flexible genomic Islands in aquatic prokaryotes. Frontiers in Genetics Frontiers Media SA. 2014;5:1–1. https://doi.org/10.3389/fgene.2014.00147.
41. Domingo E, Sheldon J, Perales C. Viral Quasispecies Evolution. Microbiol Mol Biol Rev. 2012;76:159–216. https://doi.org/10.1128/MMBR.05023-11.
42. Cordero OX, Polz MF. Explaining microbial genomic diversity in light of evolutionary ecology. Nature Reviews Microbiology Nature Publishing Group. 2014;12:263–73. https://doi.org/10.1038/nrmicro3218.
43. López-Pérez M, Haro-Moreno JM, Coutinho FH, Martinez-Garcia M, Rodriguez-Valera F. The evolutionary success of the marine bacterium SAR11 analyzed through a metagenomic perspective. mSystems. 2020. https://doi.org/10.1128/mSystems.00605-20.
44. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature. 2016;537:689–93. https://doi.org/10.1038/nature19366.

45. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20. https://doi.org/10.1093/bioinformatics/btu170.
46. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018;34:2666–9. https://doi.org/10.1093/bioinformatics/bty149.
47. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol. 2012;19:455–77. https://doi.org/10.1089/cmb.2012.0021.
48. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods. 2020;17:1103–10. https://doi.org/10.1038/s41592-020-00971-x.
49. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome. 2020;8:90. https://doi.org/10.1186/s40168-020-00867-0.
50. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics Oxford University Press. 2010;26:680–2. https://doi.org/10.1093/bioinformatics/btq003.
51. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol. 2021;39:578–85. https://doi.org/10.1038/s41587-020-00774-7.
52. Roux S, Camargo AP, Coutinho FH, Dabdoub SM, Dutilh BE, Nayfach S, et al. iPHoP: an integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. PLoS Biol. 2023;21:e3002083. https://doi.org/10.1371/journal.pbio.3002083.
53. Shang J, Peng C, Liao H, Tang X, Sun Y. PhaBOX: a web server for identifying and characterizing phage contigs in metagenomic data. Bioinformatics Adv. 2023;3:vbad101. https://doi.org/10.1093/bioadv/vbad101.
54. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, et al. Identification of mobile genetic elements with geNomad. Nat Biotechnol Nature Publishing Group. 2024;42:1303–12. https://doi.org/10.1038/s41587-023-01953-y.
55. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. ViPTree: the viral proteomic tree server. Valencia A, editor. Bioinformatics. 2017;33:2379–80. https://doi.org/10.1093/bioinformatics/btx157
56. Grazziotin AL, Koonin EV, Kristensen DM. Prokaryotic virus orthologous groups (pVOGs): a resource for comparative genomics and protein family annotation. Nucleic Acids Res. 2017;45:D491–8. https://doi.org/10.1093/NAR/GKW975.
57. Trgovec-Greif L, Hellinger HJ, Mainguy J, Pfundner A, Frishman D, Kiening M, et al. VOGDB—database of virus orthologous groups. Viruses. Multidisciplinary Digital Publishing Institute; 2024;16:1191. https://doi.org/10.3390/v16081191
58. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119. https://doi.org/10.1186/1471-2105-11-119.
59. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195.
60. Edgar RC. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. Nat Commun. Nature Publishing Group; 2022;13:6968. https://doi.org/10.1038/s41467-022-34630-w
61. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3. https://doi.org/10.1093/bioinformatics/btp348.
62. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74. https://doi.org/10.1093/molbev/msu300.
63. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;52:W78-82. https://doi.org/10.1093/nar/gkae268.
64. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 2018;46:W246–51. https://doi.org/10.1093/nar/gky425.
65. Russel J, Pinilla-Redondo R, Mayo-Muñoz D, Shah SA, Sørensen SJ. CRISPRCasTyper: automated identification, annotation, and classification of CRISPR-Cas loci. CRISPR J. 2020;3:462–9. https://doi.org/10.1089/crispr.2020.0059.
66. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12:59–60. https://doi.org/10.1038/nmeth.3176.
67. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402. https://doi.org/10.1093/nar/25.17.3389.
68. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9. https://doi.org/10.1038/nmeth.1923.
69. Olm MR, Crits-Christoph A, Bouma-Gregson K, Firek BA, Morowitz MJ, Banfield JF. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. Nat Biotechnol Nature Publishing Group. 2021;39:727–36. https://doi.org/10.1038/s41587-020-00797-0.
70. Gautreau G, Bazin A, Gachet M, Planel R, Burlot L, Dubois M, et al. PPanGGOLiN: Depicting microbial diversity via a partitioned pangenome graph. PLOS Computational Biology. Public Library of Science; 2020;16:e1007732.

71. Skewes-Cox P, Sharpton TJ, Pollard KS, DeRisi JL. Profile hidden Markov models for the detection of viruses within metagenomic sequence data. PLoS ONE. 2014;9:e105067. https://doi.org/10.1371/journal.pone.0105067.
72. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352.
73. Lin M, Kussell E. Inferring bacterial recombination rates from large-scale sequencing datasets. Nature Methods Nature Publishing Group. 2019;16:199–204. https://doi.org/10.1038/s41592-018-0293-7.

Authors and Affiliations

Jose M. Haro-Moreno[1] · Juan J. Roda-Garcia[1] · Kaiyang Zheng[2] · Mario López-Pérez[1]

✉ Mario López-Pérez
mario.lopezp@umh.es

Jose M. Haro-Moreno
moreno.jose@umh.es

Juan J. Roda-Garcia
jroda@umh.es

Kaiyang Zheng
zhengkaiyang@stu.ouc.edu.cn

[1] Microbial Genomics and Evolution Group, División de Microbiología, Universidad Miguel Hernández, Apartado 18, San Juan, Alicante 03550, Spain

[2] College of Marine Life Sciences, Institute of Evolution and Marine Biodiversity, Ocean University of China, Qingdao, China