Hawaiian Geothermal Fumaroles Contain Diverse and Novel Viruses

Pia Sen[1], Lausanne Lee Oliver[1], Kira S. Makarova[2], Yuri I. Wolf[2], Christina Pavloudi[1,3], Max Shlafstein[1], Jimmy H. Saw[1]

[1]Department of Biological Sciences, George Washington University, Washington DC, USA

[2]Division of Intramural Research, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

[3]European Marine Biological Resource Centre (EMBRC-ERIC), Paris, France

Co-Corresponding authors: Pia Sen (pia.sen.extremophiles@gmail.com) & Jimmy H. Saw (jsaw@email.gwu.edu)

Key Words: geothermal, volcanic, fumarole, bacteriophages, biofilms, biogeography

Abstract

Microbial communities of geothermal habitats are central to understanding the evolution of life on Earth. Metagenomics has provided insight into the role of viruses in shaping microbial diversity of complex environments. However, identification of novel viruses is constrained by lack of marker genes and low nucleotide similarities between related viral taxa. While microbial and viral diversity have been explored in terrestrial hot springs and hydrothermal vent systems, other volcanic features remain underexplored. Fumaroles (steam vents) are geothermal features that heat groundwater with magma, releasing steam and volcanic gases such as CO2 and H2S. Comparatively physicochemically dynamic to hot springs, fumarole temperatures and gas emissions rapidly fluctuate with volcanic activity. Here, we describe viruses identified metagenomically from microbial mats hosted near basaltic fumaroles on the Big Island of Hawai`i. To our knowledge, this is the first systematic survey of fumarole viruses. Our utilization

of a sensitive profile-based approach for identification reveals high viral diversity in fumaroles, resulting in estimation of two undescribed order-level clades of Caudoviricetes (tailed phages). Viral metabolic genes provide evidence of viral-mediated adaptation of microbes to fumarole conditions. We describe patterns of viral diversity that diverge from the Bank model of viral ecology, hinting at viral dispersal between biofilms and high viral richness and evenness. Lastly, we provide a description of the first terrestrial geothermal environment dominated by Microviridae, previously only described in viral communities of deep ocean hydrothermal vents. This study offers important findings for exploration of viral ecology in extreme environments.

Introduction

Fumaroles (terrestrial steam vents) are the most common, yet least explored geothermal feature on Earth. Fumaroles are formed when steam and volcanic gases exit cracks in basalt deposits (Wall et al., 2015). Temperature and volcanic gas emissions fluctuate with volcanic activity and geological context (Prescott et al., 2022). Located on the volcanically active Big Island of Hawai`i, the East Rift Zone (ERZ) is a geologically active region in the vicinity of Hawaii Volcanoes National Park. Along Hawaii Route 130 and between the towns of Pahoa and Kalapana, there are many physicochemically diverse steam vents, which produce steam that is mainly composed of water. These fumaroles occur on two lava flows with a wide age range (64-400 years old) and host phototrophic biofilms dominated by *Cyanobacteria*, *Chloroflexota*, *Deinococcota*, and *Alphaproteobacteria* (Prescott et al., 2022). Previous investigation of these fumaroles reveals microbial succession patterns that are dependent on early colonizing phototrophs with Actinobacteria and Proteobacteria becoming more abundant in the biofilms as geothermal activity decreases in the system (King, 2003; Prescott et al., 2022). While microbial diversity of other geothermal systems is shaped by both interspecific competition and habitat filtering, microbial diversity of the steam vent biofilms is primarily explained by interspecific

46 competition which results in an over dispersed microbial community. Functional annotation of
47 microbial taxa from fumaroles suggests that photosynthesis, CO_2 fixation, and central carbon
48 metabolism are the primary biogeochemical processes in the biofilms. The presence of taxa
49 associated with nitrogen and sulfur cycling indicated these processes may be relevant to biofilm
50 biogeochemical cycling as well. Metals and minerals are made bioavailable in fumaroles by
51 microbial alteration of basaltic substrate, while CO_2 and H_2S emissions from the vent stimulate
52 carbon fixation and sulfur oxidation (Wall et al., 2015).

53 Microbial viruses modulate functional microbial diversity through one of three main
54 actions: top-down predation of microbial hosts, transduction, or metabolic rewiring during
55 infection which contributes to metabolic reprogramming of hosts on an ecosystem-level scale
56 (Tian et al., 2024; Zimmerman et al., 2020). The Bank model of viral ecology describes
57 ecosystems where a fraction of viruses are active and abundant, while a myriad of other viruses
58 are inactive and present in low abundances until a shift in environmental conditions (e.g.
59 seasonal or diel intervals) giving rise to new dominant host-virus pairs as observed in the
60 succession patterns of environmental viral populations (Breitbart & Rohwer, 2005; Ji et al.,
61 2024; Liang et al., 2019). The Bank model predicts that a viral community high in vOTU richness
62 will be low in vOTU evenness. The density dependent 'Kill-the-Winner' (KtW) paradigm
63 describes virulent viruses infecting the most abundant or fastest growing hosts, while the
64 'Piggyback-the-Winner' paradigm describes temperate viruses increasing in abundance with
65 rates of lysogeny (Hu et al., 2025). While relative abundance cannot be taken as confirmation of
66 viral activity, it is a metric that provides a good indicator of active viral populations (Liang et al.,
67 2019; Zhou et al., 2025).

68 Viral biogeography cannot be separated from the biogeography of their microbial hosts,
69 but is controlled by an interplay of viral characteristics (life cycle, burst and virion size) and
70 abiotic factors. Geothermal systems are good models for exploring viral biogeography because
71 their microbial communities are distinct and endemic islands across a broad range of

thermophily (Castelán-Sánchez et al., 2020; Langwig et al., 2025; Power et al., 2018; Salgado et al., 2025; Teske & Reysenbach, 2015; Zhou et al., 2022). Viral dispersal in geothermal systems is limited and subject to distance-decay and strong habitat clustering patterns, but only within hot spring fields (Castelán-Sánchez et al., 2020; Cheng et al., 2022; Hwang et al., 2023; Salgado et al., 2025; Thomas et al., 2021). Microbial mats (biofilms) in geothermal systems are low mobility environments with highly specific host-virus pairs (Castelán-Sánchez et al., 2020; Jarett et al., 2020; Salgado et al., 2025).

First, we seek to describe patterns of viral dispersal and diversity across fumarole microbial populations. While microbial dispersal is known to vary across physicochemical profiles, geothermal systems of hot springs and hydrothermal vents have consistently been described as viral and microbial “islands”, with limited movement between both geothermal sites and smaller ranges of specific geothermal features within the broader system (Langwig et al., 2025; Prescott et al., 2022; Salgado et al., 2025). In such systems with high viral endemism and low dispersal, highly specific host-virus pairing is observed (Salgado et al., 2025). We gain insight into dispersal dynamics with viral read mapping and calculations of vOTU relative abundance. We predict host-virus pairing by matching microbial CRISPR spacers to viral genomes. If we observe CRISPR-based evidence of abundant vOTUs in an ‘arms race’ with a specific microbial population (many spacer matches to a single vOTU) we expect highly specific host-virus pairing and that there will be corresponding fluctuations in abundances of these viral and microbial populations across samples (Castledine et al., 2022). This would reflect lytic fumarole viral populations following the Bank model and adhering to the ‘Kill-the-Winner’ hypothesis, which describes a boom-bust dynamic where viruses target the dominant microbial population, triggering a proportional increase in vOTU abundance as viruses and their hosts engage in a coevolutionary arms race (Xue & Goldenfeld, 2017). This would reflect lytic fumarole viral populations following the Bank model and adhering to the ‘Kill-the-Winner’ hypothesis, which describes a boom-bust dynamic where viruses target the dominant microbial

population, triggering a proportional increase in vOTU abundance as viruses and their hosts engage in a coevolutionary arms race (Bucher & Czyż, 2024). Third, we hypothesize that the viruses within the fumarole biofilms will contain genomic evidence of viral capability to facilitate microbial adaptation in a dynamic and extreme environment. We predict that viruses will contain auxiliary metabolic genes that could increase host fitness in volcanic systems.

In this study, we test these hypotheses with metagenomic identification and characterization of viruses in Pahoa fumarole biofilms from three different geothermal locations on the Big Island of Hawai`i (steam wall and steam pit within the East Rift Zone (ERZ), and the Big Ell cave in the Kilauea caldera). We use a sensitive and profile-based approach to identify and classify vOTUs, many of which are distantly related to known viruses. Further, we characterize viral genomes to identify shared predicted metabolic traits with viruses previously described in other early Earth analog geothermal environments. We find that ERZ fumaroles are a unique system for viral ecology, and demonstrate high viral dispersal and phylogenetic novelty. The presence of auxiliary metabolic gene (AMG) candidates demonstrates potential for viral influence on host metabolism. Our study provides insight into the diversification of early life and patterns of viral and microbial succession (Godin et al., 2023; Hadland et al., 2024).

Results

Sampling and retrieval of viruses from metagenome

Biofilm samples from ERZ fumaroles are described in previous studies (Prescott et al., 2022; J. H. Saw et al., 2026). Briefly, biofilms that were sampled were several centimeters thick and grew on the surface of volcanic rock, directly where steam was being released. These biofilms do not exhibit clear segmentation of layers as observed in some microbial mats, and display a variety of phenotypes and colors (J. H. Saw et al., 2026). Most Gloeobacter-dominated biofilms are purple with patches of dark green, while other biofilms are green, yellow, brown,

and grey. While some biofilms have a visible extracellular polysaccharide matrix (EPS) layer and appear wet, other biofilms are dry or otherwise textured in appearance. While the steam pit and steam wall are separated by ~100 meters, the Big Ell cave in the Kilauea Caldera is ~35 km from the steam vents along Hawaii Route 130 (J. H. Saw et al., 2026).

WGS raw sequences from related biofilm features were co-assembled for related features (20 metagenomes from the 46 collected samples) with the ViralAssembly module of metaviralSPAdes, producing 400.15 Mbp (12,253 contigs) of assembled sequence data from the metaviralSPAdes pipeline (Antipov et al., 2020; J. H. Saw et al., 2026). 112.97 Mbp of the assembled sequence data was generated from Big Ell cave in the Kilauea Caldera (Hawaii Volcanoes National Park) with 17.87 Mbp (452 contigs) of the sequence data from geothermal biofilm features within the cave, and 95.11 Mbp (3803 contigs) of sequence data from adjacent non-geothermal volcanic soil samples. 243.97 Mbp (6596 contigs) of sequence data was assembled from exposed steam wall biofilms, and 43.21 Mbp (1402 contigs) of sequence data was assembled from the steam vents along Hawaii Route 130. From these assembled contigs, the ViralVerify module of metaviralSPAdes identified 754 contigs (6.33 Mbp) as either confidently (n=76) or uncertainly (n=678) viral (Supplemental Material S1). These contigs were then used in downstream analysis. VirSorter2 (Guo et al., 2021) identified 119.46 Mbp (n=16,002) of candidate viral contigs. From there, the VirSorter2 contigs (annotated with DRAM-V) were filtered for PFAM clan identifiers for the capsid genes (Phage-Coat: CL0373 and Inovirus-Coat: CL0371) producing 14.29 Mbp of candidate viruses (802 contigs) which were used for downstream analysis.

The accuracy of VirSorter2 filtering by PFAM clan was assessed by screening the remaining 15200 contigs identified as candidate viruses by VirSorter2 for the terminase large subunit gene (TerL) or another packaging gene (Supplemental Material S2). 320 contigs (~2% of filtered VirSorter2 output) contained TerL and were manually assessed for viral identity, with upstream and downstream regions searched for other viral components. Of these contigs, 116

146 (0.76% of filtered contigs) were determined to be viral (containing both head and packaging
147 genes). With geNomad (Camargo et al., 2024) we identified 3.41 Mbp (n=107) viral candidates
148 (Supplemental Material S3). When final viral designations were made and vOTUs were
149 generated, 24 vOTUs were identified by all of the tools used: VirSorter2, ViralVerify, and
150 geNomad (Supplemental Figure 01). 15 vOTUs were identified by both VirSorter2 and
151 ViralVerify, and a single vOTU was identified by both ViralVerify and geNomad. The majority of
152 vOTUs (n=313) were identified from VirSorter2. 6 vOTUs were only identified by ViralVerify, and
153 5 vOTUs were identified only by geNomad.

Remote homology-based identification and network analysis reveals diversifying vOTUs

154 569 viral contiguous sequences (contigs) were identified from 20 fumarole sample
155 metagenomes (Supplemental Material S4). From these contigs, we performed clustering by
156 community standards (Roux et al., 2019) and sorted viral sequences into 383 species-rank level
157 vOTUs (viral populations). 51% of identified fumarole vOTUs are identified as Caudoviricetes
158 (195 vOTUs), indicating that microbial viruses make up a significant portion of the
159 metagenomically retrieved dsDNA viruses in fumarole biofilms. 46% of vOTUs (180 vOTUs)
160 cannot be placed into realms by vConTACT3 v.3.0 despite verification of structural viral genes
161 on each contig. These unclassified vOTUs may have ecological significance, making up at least
162 10% of viral relative abundance in the majority (17 out of 20) of samples (Figure 01).
163 Of the 180 vOTUs which could not be placed into a realm with network analysis, 100 of
164 the ‘no realm’ vOTUs were missed by geNomad and filtered out in the VirSorter2 pipeline (Tian
165 et al., 2024) during the CheckV step (Nayfach et al., 2021) suggesting stringent pipelines may
166 miss some of the ecologically significant and highly divergent viruses (Camargo et al., 2024;
167 Liang et al., 2019). The vOTUs were placed into genus-rank viral clusters (VCs) established by
168 vConTACT3 v.3.0 and were visualized in Cytoscape (v3.10.2) with and without all viruses from

169 NCBI RefSeq (Goldfarb et al., 2024). Clusters were filtered for quality by topological and
170 clustering coefficients (>0.85). This network yielded 230 distinct topological groups, including
171 subclusters, nested subclusters, and singletons within the network (Figure 01). Of these
172 topological formations, 20 VCs form distinct subclusters (clusters of viral genomes with three or
173 more genomes without links to the rest of the network). Multiple distinct viral populations
174 (vOTUs) within a single VC provides evidence of genomic variation and diversification within the
175 genus (Bolduc et al., 2025). High 'Closeness Centrality' values support the relationships
176 observed within network clusters and prevent incomplete viral genomes from inflating network
177 representation of genomic distance between vOTUs (Nayfach et al., 2021; Su et al., 2014).
178 Resulting network clusters made up entirely of distinct fumarole vOTUs indicates that related
179 viral populations share many genes with each other while still providing evidence of
180 diversification between related vOTUs.

181 While 61% of Caudoviricetes clusters contain one vOTU, six clusters have between 8-24
182 different vOTUs in a single cluster (Figure 01). One highly abundant cluster, Caudoviricetes
183 (Cluster B), makes up to 20% of viral relative abundance in three different biofilm samples from
184 the steam wall. This cluster contains 20 unique vOTUs. The high abundance of related vOTUs
185 is evidence of potential speciation occurring in abundant viral populations (Liang et al., 2019).
186 There is also diversity observed within VCs of viruses that cannot be placed into a realm
187 confidently by vConTACT3, providing tentative evidence of diversification within highly novel
188 viral genera: out of the 21 VCs (three or more viral genomes within a cluster) which could not be
189 realm placed, 11 of the 'no realm' filtered VCs contain more than a single vOTU (Supplemental
190 Material S5). All of these 11 VCs are internally coherent with high connectivity measured by the
191 NetworkAnalyzer plugin (Su et al., 2014).

Phylogenetic estimation of viral novelty

Fumaroles contain new candidate Caudoviricetes orders

192 Out of the 196 vOTUs identified as Caudoviricetes by VContact3 assignment, 179
193 Caudoviricetes vOTUs contained a terminase large subunit (TerL) gene, and thus, were
194 included in phylogenetic analysis with maximum likelihood (91% of Caudoviricetes vOTUs).
195 Using 320 first neighbors of fumarole viruses from ICTV MSL40.v1, only 3% of these references
196 could be placed into an existing order. Less than half (40%) of these references could be placed
197 into previously identified Caudoviricetes families. The lack of classification capability for the
198 reference viruses which are most closely related to fumarole vOTUs is evidence of fumarole
199 viral novelty. Many fumarole viruses form distinct clades, with references often paraphyletic or
200 sister taxa to fumarole viruses, suggesting that steam vent viruses expand the diversity of
201 Caudoviricetes. The resulting maximum likelihood tree provides evidence of phylogenetic
202 diversification and potentially speciation within Caudoviricetes clades. There is evidence of
203 recent diversification within at least five clades that are almost entirely made up of fumarole
204 viruses and contain 130 vOTUs (Figure 02). These species-rank vOTUs meet MIUViG
205 standards necessary for taxonomic proposals (Roux et al., 2019).
206 We put forward two Caudoviricetes order-level clades containing 30 and 56 unique
207 fumarole vOTUs respectively. These clades expand the diversity of dsDNA viruses and enable
208 classification of three previously identified Caudoviricetes families which have not been
209 assigned to orders previously. The first clade, which we call *Kilaueavirales*, contains the
210 *Casjensviridae* and *Herelleviridae* families. The *Casjensviridae* family is made up of
211 flagellotrophic phages which frequently infect Enterobacteriaceae and can be found in high
212 abundances in wastewater (Chakraborty et al., 2024) while *Herelleviridae* viruses infect the
213 Bacillota phylum of bacteria (Barylski et al., 2020). Clade 1 (*Kilaueavirales*) contains a family-
214 level clade of 30 fumarole virus vOTUs (40 contigs) and 17 previously described RefSeq

Caudoviricetes species which had not previously been assigned to viral families. These unassigned reference viruses have hosts such as *Xanthomonas*, *Burkholderia*, and *Enterobacter* (Black et al., 2025). This family-level clade of 30 fumarole vOTUs and 20 RefSeq viruses is monotypic with low support (support=60%) to the *Faecalibacterium* phage, *Eponavirus epona.* This family-level clade and the *Casjensviridae* family together form a well supported clade (support=100%) that is a sister group to family *Herelleviridae* (support=98.9%). Together, this resulting clade is Clade 1 (*Kilaueavirales*) (Supplemental Figure 02).

Clade 2 (*Pahoavirales*) contains 56 fumarole vOTUs, some of which make up clades which are sister groups to reference viruses (support=90.5%). One sister group within the *Pahoavirales* clade contains a fumarole virus with a monotypic relationship (support=89.1%) to RefSeq *Ralstonia*, *Agrobacterium*, and *Myxococcus* viruses. A sister group to this clade contains two sister subclades (support=95.8%). In one of these sister subclades, there are 24 fumarole virus vOTUs and viruses from the *Mesyanzhinovviridae* family are shallow-branching taxa close to tree tips. The second subclade (which is a sister group to the first subclade containing *Mesyanzhinovviridae)* is mostly made up of fumarole viruses (32 fumarole vOTUs), but also places two reference viruses (*Dinoroseobacter* and *Microbacterium* RefSeq phages) within the subclade (support=100%).The shallow branching relationships observed in the second sister subclade of *Pahoavirales* suggests recent divergence of closely related viruses without many close relatives (Supplemental Figure 02).

In addition to the 86 Caudoviricetes vOTUs which belong to two previously undescribed order-level clades, we observe other clades of Caudoviricetes fumarole viruses with some relation to previously identified viral families (Supplemental Figure 03). One fumarole vOTU is estimated to be within the same clade of Caudoviricetes incertae sedis viral family *Hodgkinviridae* (support=75.1%), which is a family of *Microbacterium* phages (Turner et al., 2025). We estimate 19 fumarole vOTUs (25 viral contigs) to be a sister group to the *Hodgkinviridae* family (support=93.7%). Six fumarole vOTUs are a sister group to the

241 *Straboviridae* family (support=87.8%) (Turner et al., 2025). Five fumarole Caudoviricetes vOTUs
242 are estimated to form a clade with the incertae sedis *Peduoviridae* family (support=85.9%). Two
243 fumarole vOTUs are part of a clade with the *Zobellviridae* which is an incertae sedis
244 Caudoviricetes family (support=96.5%). One fumarole vOTU forms a clade with the
245 Caudoviricetes incertae sedis family *Saffermanviridae* (support=95.6%). The relationships of
246 fumarole Caudoviricetes to numerous Caudoviricetes incertae sedis families provides evidence
247 that the inclusion of these viruses in phylogenetic estimation expands the diversity of dsDNA
248 tailed viruses.

Estimation of diverse Singelaviria and Varidnaviria

249 Viruses within the Singelaviria realm (class of Laserviricetes) and the Varidnaviria realm
250 (class of Tectiliviricetes) are assigned to a single class category of
251 'Laserviricetes|Tectiliviricetes' by vConTACT3. Tectiliviricetes phylogeny was estimated with a
252 maximum likelihood tree that was constructed using the packaging ATPase gene for six
253 different fumarole Tectiliviricetes vOTUs (11 viral contigs). Viral species from all five genera
254 within the *Tectiviridae* family were used as references. The resulting phylogeny shows six
255 distinct fumarole Tectiliviricetes clades (Supplemental Figure 4). One vOTU is a sister group
256 with weak support (support = 52.5%) to the Alphatectivirus genus, which contains only the two
257 species of PRD1 and PRD4. Alphatectivirus is the only genus within the Tectiliviricetes class
258 that is known to contain plasmid dependent phages (PDPs). This vOTU may be a taxon within
259 Alphatectivirus, which is potentially worthy of further investigation when exploring the
260 phylogenetic distribution of related ecological characteristics such as plasmid dependence
261 (Bignaud et al., 2025). Two fumarole Tectiliviricetes vOTUs which are closely related and have
262 almost identical packaging ATPase sequences to each other on the amino acid level are a sister
263 group to the clade containing *Deltatectivirus* and *Epsilontectivirus* (support=93.8%).

An independent phylogenetic estimation with maximum likelihood was performed for Laserviricetes using the packaging ATPase gene. This phylogeny contains a single fumarole vOTU (Supplemental Figure 04) which is a sister taxon with strong support to the *Hukuchivirus* genus (support=100%). The *Hukuchivirus* genus is made up entirely of viruses infecting the thermophilic genus, *Thermus*. This vOTU could potentially be considered a member of a new genus within the family of *Matshushitaviridae*.

Microviridae form a poorly resolved clade with phiX174

A class-level maximum likelihood phylogeny of Malgrandaviricetes was built using the amino acid sequences of the major capsid protein (MCP) to classify the single fumarole *Microviridae* vOTU. The resulting phylogeny produced a poorly resolved clade, resulting in an apparent polytomy (support=100%) with Enterobacterial coliphage phiX174 (Supplemental Figure 05). These fumarole viruses and phiX174 are paraphyletic to the other genera of microviruses in the *Microviridae* family. The high genome conservation of phiX-like viruses (and more broadly, the *Bullavirinae* subfamily) (Kirchberger & Ochman, 2023) is potentially explanatory of an apparent polytomy with phiX174 with fumarole *Microviridae*, and suggests that phylogenetics is insufficient to capture the diversity of fumarole microviruses.

Comparative genomics of abundant Microviridae

A single vOTU of Microviridae was identified, with some genetic variation (subclustering) observed between phylogenetically identical Microviridae genomes of similar length (Figure 01). The relative abundance of Microviridae ranges from 22-82% across samples of varied microbial composition and biofilm phenotype (Figure 01). In almost half of biofilm samples, Microviridae makes up over half of the viral community, and is most abundant in moderately geothermal biofilm samples which have overall high viral richness and evenness. The high abundance of Microviridae in moderately geothermal biofilms is surprising because, to our knowledge, there

286 has never been a terrestrial environment reported which displays high abundances of
287 Microviridae. However, abundant Microviridae has been reported from viromic surveys of
288 viruses in hydrothermal vent sediment, as well as benthic sediment (Cheng et al., 2022;
289 Yoshida et al., 2018). Despite the more pronounced prevalence of Microviridae in moderately
290 geothermal biofilms, mesothermal biofilms also display a higher abundance of Microviridae
291 (>28% viral relative abundance) than both other terrestrial environments and mesothermal
292 volcanic soil collected adjacent to geothermal features from the Kilauea caldera, suggesting
293 Microviridae abundance is not simply predictable by measured temperature.

294 The low abundance of *Enterobacteria* in fumarole biofilms (Metagenome Assembled
295 Genomes (MAGs) and Amplicon Sequence Variants (ASVs) suggest that these Microviridae
296 have non-Enterobacterial hosts, and that the apparent polytomy of fumarole viruses with phiX-
297 174 could be explained by low phylogenetic resolution rather than functional similarity between
298 environmental *Microviridae* and phiX-174 (J. H. Saw et al., 2026). Our observation of viruses
299 from the Bullavirinae subfamily in fumaroles is intriguing. While lab models of phiX-like viruses
300 are heavily studied, the subfamily of Bullavirinae is considered extremely rare in nature and
301 occupies a specialist niche as a lytic phage with limited capability for uptake of genomic content
302 (Kirchberger et al., 2022).

303 Fumarole microviruses were characterized genomically with comparison of whole
304 genomes (amino acid sequences) with RefSeq *Microviridae*. Network analysis with vConTACT3
305 demonstrates consistency in length of reference and fumarole *Microviridae* (phiX174 genome
306 size is 5.386 Kb, fumarole *Microviridae* are 5.463 Kb), supporting that the potential for genome
307 completion issues does not falsely inflate the degree of genetic difference between fumarole
308 and reference *Microviridae*. Fumarole *Microviridae* are located within the same nested
309 subcluster as viruses from the Bullavirinae subfamily, providing evidence of genetic similarity
310 between fumarole viruses and Bullavirinae viruses beyond simply phylogenetic association. This
311 overall genome conservation may be explained by limited genomic uptake capabilities of

Microviridae, which occupies a specialist niche as a lytic phage, constraining genetic rearrangement and mosaicism (Kirchberger & Ochman, 2023). Despite this, fumarole and RefSeq Microviridae still vary widely with pairwise similarity as low as 40-60% between viruses (Figure 03). In addition to low overall genome similarity, there is evidence of syntenic changes between fumarole Microviridae, with differences from RefSeq Bullavirinae species (Supplemental Figure 06).

The low pairwise similarities and syntenic rearrangement we observe between fumarole *Microviridae* is highly unexpected. While innovations do occur in the *Microviridae* genome (gene overprinting, recombination) the overall high syntenic conservation of the *Microviridae* family is thought to reflect strong evolutionary constraints for transcriptional timing of *Microviridae* viruses (Kirchberger & Ochman, 2023).

Biogeography and diversity of vOTUs

Most vOTUs are shared between biofilms

All viral contigs were mapped to samples and normalized by calculating reads-per-kilobase-per-million (RPKM). Contigs were summed for each vOTU and VC for alpha-diversity analysis and visualization respectively (Roux et al., 2019; ter Horst et al., 2021). 99.7% of vOTUs (n=382) are shared between biofilm samples (all but one vOTU). The observation that most vOTUs are not unique to individual biofilms or sample sites is consistent with predictions by the Bank model (Breitbart & Rohwer, 2005). However, evidence of viral dispersal is unexpected within the physicochemical context of spatially separated and geothermally-active fumaroles that give way to harsh conditions (J. H. Saw et al., 2026). From hyperthermophilic hydrothermal vents to moderately geothermal hot springs, both microbes and viruses are found to be overwhelmingly endemic (Jarett et al., 2020; Langwig et al., 2025; Salgado et al., 2025). Both Hawaiian fumarole-hosted microbial communities (Prescott et al., 2022) and viral

communities show evidence of selective pressure associated with the geothermal activity. Within Big Ell samples, there are few shared vOTUs between the rock-hosted thermal biofilms and volcanic soil, collected directly adjacent to the cave wall. No vOTU which is found in over 5% abundance within thermal biofilms has an abundance over 1% in any non-thermal volcanic soil sample (Figure 01). The vOTU communities of thermal biofilms is suggestive of selection, rather than random vOTU dispersal (Liang et al., 2019).

Yet, while microbial communities of fumaroles form “islands” with distinct communities between biofilms of different thermal features (Prescott et al., 2022), shared vOTUs are observed in unexpectedly high abundances across samples. 9.5% of vOTUs (n=36) are shared between samples with individual population relative abundance over 1% (Figure 01). 5% of identified vOTUs (n=19) are shared between geothermal features and sites in abundances over 2%. The abundance of shared vOTUs between fumaroles is inconsistent with descriptions of distance decay observed in other physicochemically diverse geothermal systems, which are harsh to varied extents. For hot spring and hydrothermal vent systems, viral community dissimilarity increases with geographic distance, and fewer shared vOTUs are observed between distant geothermal fields (regions with localized and relatively high heat flow). Though the seed-bank model predicts that most of the shared vOTUs will be found in <2% relative abundance within a given community at any point in time, (Hevroni et al., 2020) we observe individual fumarole-hosted viral populations that are non-dominant and shared between samples can make up to 10% of a given biofilm’s viral relative abundance, representing a significant variation from the Bank model.

Perhaps as a consequence of dispersal limitation, specific host virus pairs are often observed in high abundances within geothermal systems (Guajardo-Leiva et al., 2021a; Salgado et al., 2025; R. Wu et al., 2023). This has been described as particularly prevalent in phototrophic thermal biofilms of hot springs, where matrix immobilization of microbes and their viruses result in dispersal limitations and facilitates tightly coupled host-virus pairs (Jarett et al.,

360 2020). While it is difficult to determine microbial and viral mobility metagenomically, CRISPR-
361 based analysis provides evidence of coevolutionary relationships between microbes and their
362 viruses in line with the Red Queen hypothesis (Common et al., 2019; Koonin & Wolf, 2015; R.
363 Wu et al., 2023). For the 10% of recovered fumarole vOTUs which were host matched (n=39),
364 specialized host-virus relationships were determined by the presence of more than three unique
365 spacer matches between a MAG and a vOTU. Out of the 21% of medium to high quality MAGs
366 (n=74) which were successfully host matched, 11 CRISPR arrays showed evidence of microbial
367 coevolution with vOTUs (J. H. Saw et al., 2026). The most spacer matches (n=61) within a
368 single array comes from a Rhodomicrobiaceae MAG from an unclassified genus. MAGs from
369 the Chloroflexota families Roseiflexaceae (n=35) and JAJPGT01 (n=34) also provide CRISPR-
370 based evidence of an antagonistic coevolutionary relationship between specific vOTUs and their
371 hosts.

372 Antagonistic coevolution described by 'Kill-the-Winner' is inherently density dependent
373 (Hu et al., 2025; Salgado et al., 2025). Despite this, not a single of the 14 vOTUs which is
374 identified as part of a specialized host-virus pair can be found in viral relative abundance over
375 1%. The associated host phyla abundance also does not correspond to linked vOTU abundance
376 (Supplemental Material S6) (Prescott et al., 2022; J. H. Saw et al., 2026). This unexpected low
377 abundance is observed across microbial taxa with specialized pairing to vOTUs. Cyanobacterial
378 genus *Chlorogloeopsis* is highly abundant in steam wall biofilms, and demonstrates a multi-
379 spacer relationship with a highly novel paired vOTU which is below relative abundance of 1% in
380 all samples except for steam wall biofilm PDB5 (9.24% relative abundance) non-thermal
381 volcanic soil sample HW-02, where it is found in 15.3% relative abundance. These samples are
382 not in close proximity to each other and have distinct microbial community composition
383 suggesting that Hawaiian fumarole viruses and their hosts are not necessarily endemic, even
384 though these shared viruses participate in specialized host-virus pairings (Audrey et al., 2023).
385 These spacers are not always located in the same array. We observe spacers paired to a vOTU

386 can be distributed in multiple arrays across the host genome. This phenomena could be
387 evidence of repeated exposure to a particular vOTU, or be a result of multiple different CRISPR
388 systems conferring viral resistance (Shmakov et al., 2018). The distribution of vOTU resistance
389 across two arrays is observed for three Chloroflexota MAGs from the Chloroflexia and UBA6077
390 classes, which contain various Type I and Type III systems (Supplemental Material S7). These
391 arrays are not necessarily genomically clustered, with up to ~61 kb observed between these
392 array pairs.

393 While the majority of samples do not have high abundances of identified tightly-coupled
394 virus-host pairs, high relative abundance is observed for two Caudoviricetes vOTUs which are
395 both prophages belonging to MAGs of *Gloeobacter kilaueensis*. This cyanobacteria is abundant
396 only in steam pit samples, which is observed for these vOTUs as well. High abundance of
397 Gloeobacter prophages is only observed as expected in samples dominated by *Gloeobacter*
398 *kilaueensis*. Summed vOTUs occur in relative abundances that range from 17-51% in all but
399 one of the biofilm samples with high abundance of microbial *Gloeobacter kilaueensis*. These
400 two prophage vOTUs follow the directly proportional abundance trends predicted by the
401 "Piggyback-the-Winner" hypothesis which describes a microbial community in which co-
402 abundant vOTUs favor a switch to lysogeny as a viral life cycle strategy, and accordingly,
403 abundant viral populations are identified as prophages (Hu et al., 2025). The presence of
404 Gloeobacter prophage vOTUs is in line with this theory. One of these vOTUs is present in
405 double the abundance of the other vOTU across all samples (Supplemental Material S6). Unlike
406 many of the other vOTUs with high dispersal inferred (abundance in samples doesn't
407 correspond to either hosts or microbial taxa) these two vOTUs appear restricted in dispersal to
408 be consistent with host.

Viral alpha-diversity

Chao1 (species richness) and Pielou (species evenness) index values were plotted against each other for each sample (Figure 04). Moderately geothermal biofilms cluster together, regardless of location or microbial distribution within the sample. Relative to other samples, biofilms which are collected from moderately geothermal fumaroles (40-60℃) have high richness and evenness. The 17 thermal biofilm samples show a strong positive correlation between vOTU richness and evenness. On the plot of richness to evenness, non-thermal volcanic soil diversity is clearly partitioned from the fumarolic active biofilms, displaying low vOTU richness and high evenness. The low richness is striking when compared to biofilm diversity, because soil has been described as a system that generally has high viral richness and evenness comparative to the viral communities of other systems such as marine and groundwater (Roux & Emerson, 2022; ter Horst et al., 2021).

Richness and evenness plots exhibit a clear partitioning of diversity (high richness/evenness in moderately geothermal fumaroles) that suggests diverse viral communities may be characteristic of active fumaroles. The biofilms which display this high species richness and evenness are dominated by different microbial taxa, suggesting that microbial taxonomy alone is insufficient to explain viral diversity (Figure 04). Microbial diversity indexes of biofilms were assessed using ASVs (Prescott et al., 2022; J. H. Saw et al., 2026). Unlike with viruses, there is no visually evident clustering of temperature-associated diversity (Figure 04). The decoupling of microbial and viral diversity is observed across samples of varied temperature ranges and microbial compositions. The samples with high viral richness and evenness are composed of diverse phyla (Figure 04). Interestingly, these dominant phyla all are known to occupy varied niches and display varied metabolic traits, suggesting microbial composition alone is insufficient to be the primary explanatory driver of viral diversity within these biofilms (Prescott et al., 2022).

When the relationship between richness and temperature was assessed, a quadratic model was found to be the best fit to explain this relationship. While the 4-parameter cubic model had the lowest residual sum of squares (RSS) value (RSS=36846.4), the quadratic model had the lowest BICc value (BICc=161.8) indicating that the quadratic model is the best fit for describing the relationship between temperature and viral richness for non-thermal volcanic soil and fumarolic biofilms, ranging in temperature from mesophilic to highly geothermal (>70°C) (Prescott et al., 2022; J. H. Saw et al., 2026). These tests support the significance of a non-linear (quadratic) relationship between temperature and viral richness in the Hawaiian fumarole system, with the highest vOTU richness observed at ~50°C. The extent to which microbial life adheres to Humboldt's theory of a temperature-diversity gradient has only been explored in a handful of geothermal environments, with hot springs making up the majority of geothermal systems studied for this (Sharp et al., 2014). In some moderately geothermal circumneutral hot springs, the highest species richness of bacteria has been observed at 50°C, with increasing temperatures associated with decreased species richness (Castelán-Sánchez et al., 2020). While the bacterial diversity of fumarolic biofilms observed does not clearly follow this trend, viral diversity demonstrates a surprisingly consistent temperature-diversity gradient which is independent of microbial diversity (Prescott et al., 2022; Sharp et al., 2014). This indicates that moderately geothermal features of fumaroles are potential biological hot spots for viruses, with high biodiversity of microbial/viral life which is observed across different geothermal systems at similar temperatures.

The R-squared value (R-squared=0.7292) indicates temperature explains ~73% of variance in viral richness, and is determined to be significant with the p-value of the F-statistic (p=1.619e-05) (Figure 04). Normality of residuals was supported by QQ and density plots, and the Shapiro-Wilk test confirms that the quadratic model meets normality assumptions (W=0.94514, p-value=0.2992) (Supplemental Figure 07). Though vOTU evenness varies greatly between samples (from Pielou=0.21-0.6), temperature cannot explain the great variation in viral

459 population evenness (Supplemental Figure 07). While there is a significant relationship between
460 vOTU richness and temperature (consistent with the cluster of high viral richness and evenness
461 for moderately geothermal biofilm samples) the evenness of these samples cannot be described
462 as temperature dependent, or consistent with any specific phylogenetic distribution, indicating
463 complex biogeographic factors may explain viral diversity of biofilms.

CRISPR defends fumarole bacteria against a breadth of unknown elements

464 CRISPR spacer targets (n=302) were identified for arrays from 13 different microbial
465 phyla (Supplemental Material S7). Matches are found for 20% of all fumarole-assembled MAGs
466 (n=74) (J. H. Saw et al., 2026). The majority of spacer matches (n=120) are for arrays that come
467 from Chloroflexota MAGs (classes:UBA6077, Ktedonobacteria, and Anaerolineae). Despite this,
468 only 70% of the Chloroflexota targets identified are vOTUs (n=9). Additionally, Chloroflexota
469 spacer matches were identified for 1 plasmid (0.65% of all spacer matches) and 33 unknown
470 elements. Beyond the Chloroflexota phylum, spacer targets are not limited to viruses either.
471 Even though the majority of matches identified were to vOTUs targeted by spacers (74.2% of
472 spacer matches) viruses were only 44% of the matched target elements.
473 These unknown elements represent possible other miscellaneous mobile genetic
474 elements (conjugative/integrative elements, transposons, insertion sequences) (Tokuda &
475 Shintani, 2024) or chromosomally integrated elements (Shmakov et al., 2017). Even though the
476 majority of spacer targets are non-viral, these unknown element matches make up only 20.86%
477 of total spacer matches. This suggests that the microbial populations of fumaroles utilize
478 CRISPR to defend themselves against a wide breadth of unknown elements, and that CRISPR
479 may play a minimal role in any sustained antagonistic coevolution occurring between non-viral
480 elements and microbes in the fumarole biofilms. (Martynov et al., 2017). One exception to this,
481 however, is multiple spacer matches (n=8) between an unknown element and a Chloroflexota

482 MAG (class UBA6077). Seven of these matches to the unknown element are part of a Type IA
483 CRISPR system, while one spacer match is contained within a different array that is part of a
484 Type IE CRISPR system.

Identification of putative AMGs and accessory viral genes

AMG mediated adaptation to heavy metal toxicity and resource limitation

485 Secondary curation of baseline DRAM-v AMG summary hits (212 genes) was performed
486 to eliminate viral life cycle and host-associated genes from AMG candidates. This yielded 72
487 putative AMGs and 70 accessory (moron) genes associated with defense against mobile
488 genetic elements (Brüssow et al., 2004). The majority of AMG candidates and accessory
489 defense genes we identify are from viruses in fumarole biofilms - only one accessory defense
490 gene and one AMG candidate is observed in viruses assembled from non-thermal soil. Overall,
491 7% of vOTUs (n=27) identified from fumarole metagenomes contain AMG candidates. These
492 low AMG frequencies have been observed in other geothermal systems such as hydrothermal
493 vents (Langwig et al., 2025). The majority (74%) of AMG-carrying vOTUs are classified as
494 Caudoviricetes, and the remaining vOTUs with AMG candidates cannot be assigned to realms
495 (Figure 05). Five vOTUs contain AMG-candidates related to energy metabolism (boosting ATP-
496 generation). Three of these vOTUs encode NifU-like IscA (scaffold and iron carrier) which is a
497 gene involved in Fe-S cluster biosynthesis. One vOTU contains AMG candidate gene lactate
498 dehydrogenase (reversible enzyme that converts pyruvate to lactate to generate NAD+ for
499 glycolysis, or converts pyruvate to lactate, feeding the TCA cycle). Another vOTU contains
500 putative AMG succinate dehydrogenase cytochrome b558, which, to our knowledge, has only

501 been described as an AMG in giant virus family Mimiviridae (Blanc-Mathieu et al., 2021; Lamb
502 et al., 2024).
503 The majority of putative AMGs identified in fumarole viruses encode transporter-related
504 gene products (Supplemental Material S8). Transporter types include: divalent metal (DMT)
505 transporters, RND efflux pumps, ABC-transporters, and Major facilitator superfamily (MFS)
506 transporters. Ten fumarole vOTUs contain AMG candidates that can be linked to cellular
507 detoxification from heavy metals, such as metallothioneins, Zn2+/Cd2+ exporting ATPases, and
508 heavy metal efflux proteins including transporter TolC. The TolC transporter functions with efflux
509 of transition metals, including first row transition metals which are present in basalt (Farias et al.,
510 2015). The transporters observed in fumarole viruses have also been described in AMG
511 candidates from viruses in hydrothermal vents and heavy metal enriched soils with similar
512 elemental composition to basaltic surfaces described in the ERZ volcanic system (Prescott et
513 al., 2022; Sun et al., 2023; Thomas et al., 2021). Other fumarole virus transporters are
514 protective proteins (such as Fe-Mn superoxide dismutase) that neutralize reactive oxygen
515 species (ROS) which can also be generated by metal induced cellular stress. Fumarolic non-
516 transporter AMG candidates (such as rubrerythrin) also neutralize ROS, which usually confers
517 host survival in microaerophilic or anoxic environments (Cardenas et al., 2016).
518 Other AMG candidates relate to nutrient deficiency. One vOTU contains sporulation
519 maturation protein CgeB which is sometimes observed in Bacillus subtilis prophages where it is
520 part of the Cge operon which modifies the surface properties of the spore (Shuster et al., 2019).
521 One putative vOTU contains a carbohydrate ABC transporter permease (Trehalose transport
522 system permease protein SugB) which is involved in trehalose recycling within the cell (Pohane
523 et al., 2021). One fumarole vOTU (prophage of Gloeobacter kilaueensis) encodes a gene highly
524 similar to polyhydroxybutyrate depolymerase (phaZ). Polyhydroxybutyrate depolymerase is an
525 intracellular enzyme which breaks down polyhydroxyalkanoates (PHAs) or Polyhydroxybutyrate
526 (PHB) biopolymers which can be used as carbon sources. Eight unique vOTUs contain AMG

527 candidates related to carbon utilization. Two Caudoviricetes vOTUs contain five different carbon
528 utilization genes that could possibly be involved in PHA biosynthetic pathways, converting short-
529 chain carboxylic acids (such as acetate, propionate,lactate, and butyrate) into PHA (Yao et al.,
530 2025). Alternatively, these five genes could be involved in beta oxidation of fatty acids, which
531 have been described in the giant virus family Mimiviridae (Brahim Belhaouari et al., 2022). The
532 presence of these five AMG candidates in Caudoviricetes as observed in two fumarole
533 Caudoviricetes vOTUs is unusual because Caudoviricetes hosts are prokaryotic while
534 Mimiviridae hosts are eukaryotic (Duda & Teschke, 2019). Another putative AMG contained
535 within one of these Caudoviricetes vOTUs is gene 2-methylcitrate dehydratase (prpD) which
536 catabolizes the third step of the methylcitrate cycle for propionate catabolism, and has been
537 described as an AMG in hydrothermal vents (Zheng et al., 2021). Host linking with experimental
538 verification of transcriptional activity and gene product expression is necessary to ascertain if
539 these putative AMGs are involved in beta oxidation of fatty acids and lipid catabolism pathways,
540 or PHA biosynthesis pathways.

Accessory vOTU genes may confer resistance to non-self viruses

541 In addition to AMG candidates, 32 vOTUs (8% of fumarole derived vOTUs) contain
542 virally encoded defense genes, which can potentially provide antiviral defense for the microbe
543 (increasing host fitness), or facilitate superinfection exclusion against other viruses. 78% of the
544 defense-containing vOTUs (n=25) were classified as Caudoviricetes, and the remaining
545 defense-containing vOTUs (n=7) could not be assigned to a realm classification. Queosine
546 biosynthesis genes (7-Deazaguanine modification) and (cytosine-5)-methyltransferase (dcm)
547 were initially flagged by DRAM-v as metabolic AMG candidates. However, they were manually
548 reclassified as defense-related because that is more likely reflective of functional role (Hutinet et
549 al., 2019; Wang et al., 2024).

Fumarole viruses contain a range of defense and anti-defense related genes: anti-CRISPR, anti-Thoeris, and 7-Deazaguanine modification genes were annotated by DRAM-v. Virally encoded toxin/antitoxin (TA) genes from Type I-IV systems were observed in ~5% of total fumarole vOTUs (n=18) (Supplemental Material S8). For most of the vOTUs (n=10) containing TA systems, both toxin and antitoxin genes of the same system type are positioned with no more than five genes between components, suggesting that at least some fumarole viruses encode functional TA systems. Two vOTUs contain CRISPR-associated proteins. One vOTU (vOTU 189) does not have an identified host and encodes a minimal Type I-C CRISPR-Cas Cascade system, which contains only three unique Cas proteins: Cas5d, Cas8c, and Cas7. This minimal Cascade encodes an RNA-guided multi-subunit CRISPR-Cas surveillance complex which targets foreign DNA for destruction (O'Brien et al., 2022).

A different vOTU (vOTU 212) contains a minimal Type IV system (genes Cas6, Csf3, and Csf2) and an unassociated CRISPR array. Type IV systems are primarily encoded by plasmids and more occasionally observed on prophages (Moya-Beltrán et al., 2021). Type IV CRISPR systems are enigmatic in mechanism, and usually lack a CRISPR array. They are hypothesized to modify existing CRISPR-Cas activity of hosts which already contain CRISPR systems (Koonin & Makarova, 2017). Accordingly, vOTU 212 is targeted by spacers within a Chloroflexota (class UBA6077) MAG which contains both Type IA and Type IE systems (J. H. Saw et al., 2026). In addition to the Type IV effector proteins, vOTU 212 encodes a CRISPR array. The maintenance of both an array and Type IV system (which characteristically, lack associated CRISPR arrays) (Pinilla-Redondo et al., 2020) is suggestive of inter-viral competition which has resulted in the use of CRISPR as a method for viruses to prevent superinfection exclusion (Muscatt et al., 2022).

Discussion

Fumaroles contain novel and diversifying viruses

Our approach to viral identification that utilizes sensitive, profile-based methods (Steinegger et al., 2019) for annotation of structural viral gene components reveals novel viruses from fumarole biofilms and volcanic soil. Our approach identifies viruses using amino acid sequences encoding viral structure components (capsid-associated and packaging genes) which are highly divergent (Söding et al., 2005). Our recovery of phylogenetically novel viruses from biofilm metagenomes and our subsequent careful manual verification supports our workflow which utilizes remote homology protein prediction for identifying structural virion components. This can, in turn, be used for identification of viruses in a high-throughput processing pipeline with existing tools (Felipe Benites et al., 2024) without extensive false positives (Supplemental Material S4).

Our approach yields recovery of highly novel viruses (Bolduc et al., 2025) which we find to be sometimes highly abundant relative to other viral populations (Figure 01). This is predictive of either functional significance for these novel and abundant viruses, or represents significant population dormancy (Liang et al., 2019). Because viruses do not all share a single gene that is common to all species (Harris & Hill, 2021), there is not a way to confidently place all viruses of microbes on a single phylogenetic tree of life. Thus, it is not possible to phylogenetically estimate viruses which cannot be classified into an existing realm (Koonin & Wolf, 2012). Our assessment of viral diversification for these ‘no-realm’ viruses relies on gene-sharing network analysis, which provides a method for comparison of genomic distance between viruses which cannot be otherwise taxonomically placed. While uncertainty in phage genome completion (vOTUs of variable length within a cluster) and chimeric or mosaic assemblies can constrain comparisons of viral genome similarity, our utilization of NetworkAnalyzer metric of ‘Closeness

596 Centrality' (Su et al., 2014) minimizes the effects of incomplete viruses producing erroneous
597 inferences of genomic distance (Bolduc et al., 2017). Our observation of subclusters for these
598 highly novel viruses, which are not placed into a realm by vConTACT3 and do not exhibit gene
599 sharing with previously described viruses, is evidence that there may be genomic diversification
600 between these novel viruses. Yet, conclusions about associated viral speciation for these highly
601 novel populations are limited by our inability for phylogenetic estimation without realm
602 classification.

603 The majority of recovered viruses are from the Caudoviricetes class. This could possibly
604 be a result of detection facilitated by high representation of Caudoviricetes genomes within
605 databases of known viruses (Weinheimer et al., 2025). As a result, our conclusions about
606 evolutionary patterns in other viral classes are limited by the low genomic diversity of recovered
607 fumarole vOTUs from the Laserviricetes, Tectiliviricetes, and Malgrandaviricetes classes. Even
608 though Caudoviricetes are the most characterized phage class, our identification of fumarole
609 viruses significantly expands the diversity of the class. Out of the 179 Caudoviricetes vOTUs
610 included in phylogenetic estimation, 86 fumarole vOTUs belong to two previously undescribed
611 order-level clades within the Caudoviricetes class (Supplemental Figure 02). The 93 other
612 fumarole Caudoviricetes vOTUs have varied levels of relation to previously described viral
613 families. Phylogenetic estimation exhibits that many of the fumarole-derived Caudoviricetes
614 vOTUs are related to families which are primarily incertae sedis within the Caudoviricetes class
615 such as *Hodgkinviridae*, *Peduoviridae*, *Zobellviridae*, and *Saffermanviridae* (Supplemental
616 Figure 03). This suggests that the inclusion of novel viruses may help resolve the phylogenetic
617 positions of problematic taxa within the Caudoviricetes class. Our observation of distinct
618 Caudoviricetes fumarole vOTUs which have close phylogenetic relationships (Figure 02)
619 suggests not only genomic diversification, but also speciation, could be occurring between
620 fumarole Caudoviricetes populations.

Abundant Microviridae may be a signature of geothermal environments

Microviridae are known to be diverse and globally distributed across a variety of environments (sediment, marine, animal, insect, and humans) (Kirchberger et al., 2022; Kirchberger & Ochman, 2023; Roux et al., 2012). Despite this, there has never been a description of a terrestrial environment in which Microviridae is the most abundant viral population, or otherwise represents a substantial portion of viral abundance in terrestrial or marine (oceanic zone) environments. Benchmarking with mock communities representative of open ocean marine and freshwater environments are consistent with observations that ssDNA viruses make up a very small portion (<5% total DNA viruses) of total viral content, and are much less abundant than dsDNA viruses (Kirchberger et al., 2022; Kirchberger & Ochman, 2020). The other environment in which Microviridae has been described as highly abundant is in deep ocean hydrothermal vents sediments (Castelán-Sánchez et al., 2019; Cheng et al., 2022; He et al., 2017; Y. Wu et al., 2024). It is possible that viruses belonging to Microviridae can be adapted to thermal conditions as an ssDNA virus. Phage phiX-174 has been demonstrated to be most stable at 37 - 44°C and is predicted to be favored over dsDNA and RNA phages under thermal stress because of the increased access that ssDNA viruses have to mutational diversity (Greenrod et al., 2025). Alternatively, it is possible that Microviridae has the capability to circularize which would enable abundance at higher temperatures (Laanto et al., 2017). However, the ERZ fumaroles with highest Microviridae abundances (50-70°C) are over three times cooler than sediment of Southwest Indian Ocean hydrothermal vents (temperatures as high as 400°C), where Microviridae has been identified as the most abundant virus (Cheng et al., 2022).

Out of ERZ fumarole viral communities, moderately geothermal and high geothermal biofilms are described as having the highest Microviridae abundances out of all samples.

644 However, mesothermal fumarole biofilms are observed to have high Microviridae abundance
645 (22-28%) which is higher than described in other terrestrial and marine environments (Figure
646 01). This may be explained by the high abundance of Microviridae in mesothermal fumaroles
647 experiencing rapid fluctuations in fumarole temperature. Mesothermal fumaroles which harbor
648 high Microviridae abundance may reflect a lag in decreasing abundance of ssDNA microvirus
649 abundance as the biofilm communities adapt to lower temperatures. Volcanic activity of the ERZ
650 varies widely across small spatial scales and fluctuates rapidly within short time spans (minute
651 to minute), making it possible that the steam pit mesothermal fumaroles are dying geothermal
652 features, while the ~100 m adjacent steam wall is more geothermally active. Many fumaroles
653 are ephemeral and experience more variation in activity as they near the end of their
654 geothermal life span (Madonia, 2020).

655 Viromics is known to overestimate the abundance of ssDNA viruses, and the perception
656 of high abundance is often a sequencing artifact. Many viromics extraction approaches may
657 utilize Adaptase prior to linker amplification and produce a quantitative bias, leading to
658 erroneous predictions of microvirus abundance (Roux et al., 2016). Alternatively, many viromics
659 DNA preparation methods for low biomass viromics may destroy Microviridae DNA (Gregory et
660 al., 2019; Y. Wu et al., 2024). Our approach utilizes bulk metagenomics prior to sequencing,
661 which may explain the abundance of microviruses we observe. We hypothesize that microvirus
662 diversity is underpredicted in geothermal environments, and temperature may be a selective
663 pressure which favors ssDNA viruses (Roux et al., 2016). Studies of other terrestrial geothermal
664 systems have not yielded similar observations of dominant Microviridae. However, some of
665 these other systems (such as terrestrial hot springs) are often explored in the context of low
666 biomass sediment that requires amplification or alternative methods of extraction which may
667 destroy microviruses (Gregory et al., 2019). Alternatively, these geothermal systems are
668 investigated with methods of viromics that destroy ssDNA viral representation in the sample
669 (Roux et al., 2012). Fumarole biofilms may not be the only terrestrial system whose viral

670 community contains dominant *Microviridae*, but the high fumarole biofilm biomass which is
671 utilized for DNA extraction may reveal this characteristic of thermal viral communities which is
672 otherwise obscured by methodological decisions which are made when investigating the
673 microbial and viral diversity of other terrestrial thermal systems.

674 Fumarole *Microviridae* are highly novel with pairwise genome similarities as low as 40-
675 60% (Figure 03) as well as syntenic rearrangement between viruses within the fumarole biofilms
676 and compared to reference *Microviridae* (Supplemental Figure 06). However, there is a need for
677 comparative genomic approaches when analyzing highly conserved ssDNA viruses to make
678 conclusions about the degree of novelty introduced with fumarole *Microviridae*. Network
679 analysis with vConTACT3 demonstrates multiple *Microviridae* form a single subcluster which is
680 unexpected because clusters usually represent genus level rank (Bolduc et al., 2017, 2025).
681 This suggests that environmental *Microviridae* are highly conserved and/or significantly under-
682 described.

Viruses are widely shared between geothermal features of ERZ fumaroles

683 There is evidence that viral dispersal has consequences for fumarole microbial
684 evolution. Antagonistic coevolutionary relationships between hosts and viruses are often
685 described through a 'Kill-the-Winner' dynamic, where both host and virus engaging in an active
686 arms race are abundant relative to other viral and microbial populations. Relatedly, viral
687 dispersal is thought to be host-dependent (Langwig et al., 2025) and associated with microbial
688 abundance (Salgado et al., 2025). Despite these expectations, we observe many of the shared
689 vOTUs are only found in low abundances within samples. For microbes that appear to utilize
690 CRISPR as a defense mechanism and have identified arrays, vOTUs were determined to be
691 part of a 'specialized' host-virus pair if there were more than three spacers linking a vOTU to a
692 single microbial host (indicative of antagonistic coevolution). The 14 vOTUs which have

specialized pairings with microbial hosts are not found in viral relative abundance over 1% within any sample. This suggests that low abundance viruses and hosts are still antagonistically coevolving and dispersal. Host-dependence and abundance are insufficient factors to predict limited dispersal of viruses in fumaroles. However, it should be noted that many microbial taxa do not rely on CRISPR for viral defense, and viral novelty constrains capability for host identification of 90% of identified fumarole vOTUs (Tan et al., 2024). Yet, even when simply searching for patterns of concurrent variation between abundances of unmatched vOTUs and hosts, no high abundance concurrent variation that would suggest an expected 'Kill-the-Winner' trend was observed (Supplemental Material S6).

Most vOTUs being shared between fumarole biofilms is highly unexpected. For hot springs and hydrothermal vents, individual vents and pools are biological "islands" with distinct communities between features (Castelán-Sánchez et al., 2019). Microbial communities of Hawaiian fumarole vents exhibit distinct communities as well (Prescott et al., 2022). In hot spring and hydrothermal vent systems, viruses display a distance decay effect with fewer shared viral populations observed with increasing distance between features (Jarett et al., 2020). The majority of viruses being shared between samples is even more unexpected when the context of phototrophic biofilm matrixes is considered (Castelán-Sánchez et al., 2020; Jarett et al., 2020). Compared to other mediums of geothermal systems (sediment or water), geothermal biofilms with strict spatial stratification are expected to have even less dispersal, with deterministic selection occurring on viruses within the microscale of centimeters within the biofilm (Guajardo-Leiva et al., 2021a). This is particularly remarkable considering the context of this comparison for fumarole biofilms to hot spring phototrophic biofilms with similar biogeochemical cycling processes and organismal characteristics (Guajardo-Leiva et al., 2021b; Salgado et al., 2025).

The high viral dispersal that is suggested by many shared vOTUs could be explained by unique dispersal mechanisms of fumaroles that diverge from the methods of dispersal seen in

hot spring and hydrothermal vent systems. Steam (groundwater heated by magma) may be a mechanism for viral dispersal (Wall et al., 2015). Steam may enable biofilms to act as "extremophile collectors" which retain viruses within biofilms (Prescott et al., 2022; Wall et al., 2015). This suggests Hawaiian fumaroles in the ERZ have unique ecological and complex biogeographic characteristics with high viral dispersal informing these communities. By elucidating the metabolic and phylogenetic diversity of fumarole viruses, we have identified ERZ fumaroles on the Big Island of Hawai`i as a promising and unique early Earth analog environment to investigate viral ecology.

Gloeobacter prophage may increase host fitness

The steam pit sample site is dominated by *Gloeobacter*, despite the close proximity and presence of other oxygenic phototrophic Cyanobacteria (J. H. Saw et al., 2026) which are faster-growing and photosynthetically efficient compared to *Gloeobacter* species which contain reduced photosystems (Grettenberger, 2021). Two vOTUs (Supplemental Figure 08) identified as *Gloeobacter kilaueensis* prophages by global alignment to *Gloeobacter* MAGs (J. H. Saw et al., 2026; J. H. W. Saw et al., 2013) are highly abundant and co-vary with *G.kilaueensis* in abundance for all but one of the biofilm samples dominated by *G.kilaueensis* (Figure 01). The high abundance of both host and prophage suggests that these two vOTUs follow a "Piggyback-the-Winner" pattern of viral ecology (Hu et al., 2025). Across all vOTUs, these are the only host matched viruses which directly increase in abundance with corresponding host taxa (Supplemental Material S6).

The only sample site which contains *Gloeobacter* dominated biofilms is the steam pit. The steam wall sample site, located ~100 m from the pit, is dominated by other oxygenic phototrophic cyanobacteria and is more exposed to light than the shaded steam pit. *Gloeobacter* lacks a thylakoid membrane and is susceptible to light induced stress, resulting in metabolic disruption and cellular damage from oxidative stress (J. H. W. Saw et al., 2013). Our

743 recovery of these vOTUs is the first description of viruses known to infect the *Gloeobacter*
744 genus, which is of astrobiological significance due to its evolutionary role as an ancient
745 cyanobacterium (Prescott et al., 2022; J. H. W. Saw et al., 2013). *Gloeobacter kilaueensis* is a
746 highly conserved species with MAGs sharing 99% ANI (J. H. Saw et al., 2026; J. H. W. Saw et
747 al., 2013). The presence of multiple vOTUs within steam pit biofilms suggests these prophages
748 introduce genetic variation to the *Gloeobacter* host and facilitate genome evolution of the
749 *Gloeobacter kilaueensis* species. CRISPR arrays in *Gloeobacter* MAGs contain a spacer match
750 to a single vOTU of the genus-rank cluster. The match to a single vOTU within the genus could
751 indicate reduced reliance on CRISPR as defense against the prophage or that the host no
752 longer benefits from defense against vOTUs as they diversify in close proximity without
753 geographic partitioning.

754 In addition to susceptibility to oxidative stress, *Gloeobacter*'s energy inefficiency makes
755 its high relative abundance in biofilms highly unexpected. All membrane-based energy
756 metabolism occurs within a single layer of the cytoplasmic membrane, restricting surface area
757 and metabolic rates. *Gloeobacter*'s reduced photosystems, light sensitivity, and inefficient
758 RuBisCO (slow carboxylation rate) also limit growth of the genus compared to other oxygenic
759 *Cyanobacteria* (Raven & Sánchez-Baracaldo, 2021). Energetically, it is possible that the
760 putative carbon utilization AMG polyhydroxybutyrate depolymerase (*phaZ*) confers host fitness
761 for *Gloeobacter*, contributing to its dominance in steam pit biofilms and optimizing host carbon
762 utilization. Polyhydroxybutyrate depolymerase is an intracellular enzyme which breaks down
763 polyhydroxyalkanoates (PHAs) or polyhydroxybutyrate (PHB) biopolymers. The ability to utilize
764 PHB biopolymers may provide a mobilizable reserve of carbon and energy to *Gloeobacter*,
765 enabling efficient resource management (granules storing nutrients without altering internal
766 osmotic pressure or loss of compounds through leakage) and increasing prophage-mediated
767 *Gloeobacter* survival over other competitors. The vOTU containing the *phaZ* AMG candidate is

768 doubled in abundance compared to the other *Gloeobacter* prophage vOTU, suggesting that
769 there is a comparative fitness benefit to the prophage maintaining the putative AMG.
770 There is evidence of the Gloeobacter prophage evading host defenses. The presence of
771 what appears to be a self-targeting spacer shows that CRISPR systems of the host may prevent
772 prophage activation or lytic activity (Supplemental Material S7). Alternatively, the prophage may
773 carry an anti-CRISPR protein that inhibits host CRISPR-Cas systems (*Gloeobacter* MAGs
774 contain both Type I and Type III CRISPR-Cas systems). The biofilm sample observed which
775 had both high abundance of *Gloeobacter* (J. H. Saw et al., 2026) and a low abundance of the
776 *Gloeobacter*-linked vOTUs was collected from a sunlight-exposed area, with a lower light
777 intensity than the steam wall sample site. High-light stress in exposed biofilms may compromise
778 *Gloeobacter* cells with metabolic disruption or cellular damage, destabilizing prophage
779 maintenance and reducing their relative abundance in the sample (Figure 01). Oxidative stress
780 could also trigger prophage excision and loss (Middelboe et al., 2025). The substantial decrease
781 of prophage content in biofilms despite the host remaining abundant suggests that light-induced
782 stress in Gloeobacter may limit lysogeny (consistent with “Piggyback-the-Winner” models)
783 influencing viral turnover and potentially altering persistence of lysogenic viral populations.

Active fumaroles host diverse viral communities.

784 Soil has been described as a medium that contains one of the most diverse and complex
785 viral communities with high richness and evenness suggesting significant levels of viral activity
786 (Roux & Emerson, 2022), In particular, young volcanic soil is known to harbor incredibly diverse
787 microbial and viral communities as a result of the soil’s nutrient rich physicochemical properties
788 (Aguila-Torres et al., 2025; Huang et al., 2024; Mihai et al., 2023). Despite these previous
789 descriptions of high volcanic soil viral richness, we discover that the biofilms of active (thermal
790 and steam emitting) fumarole features contain communities with vOTU richness that is
791 significantly higher than vOTU richness of non-thermal volcanic soil from the Kilauea caldera

Big Ell cave (Figure 04). Further, some of the biofilms also contain high vOTU evenness that is comparable to the evenness of volcanic soil. Because biofilms with high viral richness and evenness do not contain a specific microbial phylogenetic profile, we suggest that the environmental context of Hawaiian fumaroles as a geothermal system contributes significantly to the overall viral diversity. There is not a clear distance decay effect or geographic partitioning of diversity indexes, indicating that localized biogeographic context is significant in explaining viral diversity. The structural properties of biofilms and microbial metabolism within fumarole features (increased nutrient availability through basaltic bioleaching, metabolism of emitted gases, and temperature) may all contribute to fumarole viral diversity.

The statistically significant relationship between temperature and viral richness has not been described previously in a geothermal system other than hot springs (Castelán-Sánchez et al., 2020; Sharp et al., 2014). The steady increase of viral diversity in fumaroles which peaks at 50°C and decreases in high geothermal biofilm samples emitting steam (>70°C) provides compelling evidence that viral diversity in some geothermal systems can adhere to a temperature-diversity gradient, and that biodiversity hotspots in geothermal environments are are not limited to “Humboldt’s spa” of hot springs (Sharp et al., 2014). It is known that temperature influences structural biofilm properties, regulating processes such as viral particle attachment and mineral precipitation (Simmons et al., 2020). Indeed, temperature dependent biomineralization has been described within moderately geothermal ranges (Kostešić et al., 2023). These structural alterations could potentially turn biofilms into physical refugia for virion particles, increasing the density of virion particles with immobilization and in turn increasing metagenomically identified viral diversity for species richness while potentially increasing species evenness under specific physicochemical conditions which influence biofilm properties (Decho & Gutierrez, 2017). The physical proximity of viruses within the biofilms exhibiting thermal-induced structural properties could allow these viruses to act as “gene highways” (Thomas et al., 2021) and facilitate microbial adaptation to the dynamic and harsh conditions of

818 a volcanic system (Supplemental Material S8). Even though moderately geothermal samples
819 have high viral richness and evenness and there is a unimodal relationship between
820 temperature and species richness, microbial diversity indexes of samples do not follow the
821 same clear trend (Figure 04).. The decoupling of fumarole virus and microbial diversity could
822 reflect temporal lag and community turnover of viruses, or dormant but inactive viral populations
823 retained at higher abundances than expected within biofilms.

824 If the viral richness and evenness within biofilms is representative of active viral
825 populations, this may facilitate increased opportunity for viral competition while increasing the
826 fitness conferred by superinfection exclusion (Muscatt et al., 2022; Pinilla-Redondo et al., 2020).
827 This would contribute to explaining diverse and frequent genomic observations of TA systems
828 (encoded in ~5% of fumarole vOTUs) and other defense related proteins in fumarole viruses
829 (Supplemental Material S8).

830

831 Accessory viral genes may aid adaptation to basaltic 832 surface and resource limitation

833 Biofilms contain the majority of putative AMGs. Of identified fumarole vOTUs, only 7% of
834 fumarole vOTUs contain AMG candidates which is consistent with observations of AMGs as
835 rare in other geothermal environments such as hydrothermal vents (Langwig et al., 2025).
836 However, more samples with broader geographic distribution are necessary to make
837 conclusions about AMG prevalence in systems with high microbial endemism anticipated
838 (Castelán-Sánchez et al., 2019; Langwig et al., 2025; Salgado et al., 2025; Thomas et al.,
839 2021). The presence of diverse AMG candidates seen within vOTUs suggests that these
840 viruses may function as a genomic reservoir for microbial hosts and facilitate their persistence in
841 adverse and nutrient limited conditions (Supplemental Material S8). Basalt hosted microbial
842 communities in hydrothermal vents are known to engage in bioleaching with enzymatic
843 dissolution of substrate in hydrothermal vents, releasing metals in the process (Bergsten et al.,

844 2021; Thomas et al., 2021). The presence of ten different vOTUs with putative AMG involved in
845 heavy metal detoxification would suggest that viruses play a significant role in facilitating
846 microbial adaptations to high metal concentrations released from the dissolution of basalt
847 (Figure 05). This explains the presence of detoxifying metal efflux and ROS neutralization
848 putative AMGs, which are often seen in hydrothermal vents and heavy metal contaminated soil
849 where similar processes may occur(Cheng et al., 2022; Farias et al., 2015). While most of these
850 transporters are predicted to facilitate heavy metal resistance, some fumarole AMGs candidates
851 likely confer fitness for the host in other ways. MFS transporters may alternatively play a role in
852 conferring antibiotic resistance for the host, while biopolymer transport genes are also observed.
853 Other AMG candidates increase host fitness in resource limited conditions, primarily with
854 genes that participate in carbon utilization pathways. PHA/PHB metabolism in particular may
855 play an important role in microbial response to adverse conditions. Polyhydroxybutyrate
856 depolymerase is an AMG candidate in a *Gloeobacter* prophage vOTU which is highly abundant
857 in steam pit samples (Supplemental Material S8). Two other vOTUs contain carbon utilization
858 putative AMGs which could be involved in either beta oxidation of fatty acids or PHA
859 biosynthetic pathways (conversion of short-chain carboxylic fatty acids). In resource limited
860 conditions with nutrient depletion, PHA and PHB utilization could provide microbes the ability to
861 sequester and mobilize carbon and energy sources, aiding survival in dynamic and volcanism
862 dependent physicochemical conditions of fumaroles. Chitin degrading genes (chitinase and
863 chitosanase) are contained by three different fumarole vOTUs and are AMG candidates.
864 Chitinase has been described as an AMG in prophage encoded chitinase within hydrothermal
865 vents, where the AMG is involved in substrate dependent lysis-lysogeny switches of viruses. It
866 is known that carbohydrate degradation capacity is a driver of microbial community structure in
867 hadal zones, offering a possible explanation for how these putative AMGs could confer host
868 fitness in nutrient limited and dynamic environments such as fumarole biofilms (Middelboe et al.,
869 2025). Putative AMGs related to nutrient limitation suggest viral AMGs confer host fitness with

AMGs related to resource acquisition and scavenging (Monier et al., 2012; Tian et al., 2024; Zheng et al., 2021).

Sporulation related AMG candidates may facilitate survival in adverse conditions via dormancy, which has primarily been observed as an AMG in low oxygen zones, heavy metal contaminated soil, and sub-freezing environments (Brown et al., 2022; Jurgensen et al., 2022; Sun et al., 2023). These putative AMGs may help their host by facilitating microbial sporulation when conditions are resource limited until nutrients are available for host reactivation. The selection for microbial survivability under resource limitation may incentivize acquisition of viruses with these AMG candidates. However, identification of viral metabolic genes is constrained by sequence length (contigs under 10 kb). Biochemical characterization or transcriptomic changes provide evidence for confident AMG identifications, and should be performed for any structurally novel predicted AMGs. This is particularly true for CAZyme genes which are poorly conserved on a structural level (Martin et al., 2025).

Methods

Sample collection and sequencing

Detailed methods on sample collection and sequencing have been described in two previous studies (Prescott et al., 2022; J. H. Saw et al., 2026). Briefly, a total of 46 biofilm and soil samples were collected in an expedition in 2019 from steam vent features found within Kilauea Caldera inside Hawaii Volcanoes National Park and within East Rift Zone along Hawaii Route 130. DNA from samples were extracted using ZymoBIOMICs mini prep kit, quantified with Qubit HS dsDNA assay, then sent for sequencing at two sequencing centers: ASGPB at University of Hawaii at Manoa for 16S amplicon sequencing and CQLS at Oregon State University for whole-genome shotgun (WGS) sequencing using Illumina MiSeq and HiSeq,

respectively. Some of the amplicon data has been analyzed in a previous study that excluded the cave soil samples from Big Ell (Prescott et al., 2022). BBTools script bbduk.sh (Bushnell, 2014) was used for quality control prior to assembly. Adapters were trimmed from the right ends of the reads, with reads shorter than 50 bp discarded (--minlen=50). Adapters were trimmed based on paired overlap (tbo=t) with low quality bases trimmed on both ends of reads (--qtrim=rl) with low quality bases discarded (--trimq=20).

Identifying viral sequences

The ViralAssembly module of metaviralSPAdes v3.15.5 (Antipov et al., 2020) was used for viral assembly and mismatch correction (k-mer lengths of 21, 33, 55, and 77). Assembled contigs were input into ViralVerify v1.1.0 (Antipov et al., 2020) (Naive Bayesian classifier on hmmsearch gene predictions), geNomad v1.11.2 (Camargo et al., 2024) (hybrid approach of alignment-free models and gene models with neural networks for classification), and VirSorter2 v2.2.4 (Guo et al., 2021) (multi-classifier with HMM search for viral hallmark genes) for initial identification. These tools employ a diversity of methods for viral identification (Antipov et al., 2020; Bolduc et al., 2025; Camargo et al., 2024; Guo et al., 2021). Viruses were then identified by identifying both a head and packaging gene in each candidate viral contig. Default parameters were used for ViralVerify with the ‘-hmm’ flag. geNomad was run with the ‘--end-to-end’ flag. As per community standards for VirSorter2, the default score of 0.5 (low-cutoff) was used for maximum sensitivity and filtering. CheckV v1.0.3 (Nayfach et al., 2021) was then used on extracted contigs (end_to_end option) to remove false positives and trim host regions of proviruses prior to running VirSorter2 again on trimmed sequences (Guo et al., 2021; Tian et al., 2024). Lastly for the VirSorter2 pipeline, a secondary screening approach was employed on domain architecture DRAM-v annotations searched for capsid associated proteins (Phage-coat PFAM clans CL0373 or CL0371). PFAM clans account for evolutionary relationships which

914 cannot be encompassed by a single HMM profile of DRAM-v (Finn et al., 2016; Kashif-Khan et
915 al., 2024).
916 Open reading frames (ORFs) were called in metagenomic ('-meta') mode with Prodigal
917 v2.6.3 (Hyatt et al., 2010). Gene calls were annotated with HHblits (Steinegger et al., 2019)
918 (uniprot_sprot_vir70 database) with flag '-n 3' iterations for remote homology prediction.
919 Annotations were filtered for E value (<1E-1) and probability (>50%). The first 10 hits for each
920 Coding DNA Sequence (CDS) were screened for head and packaging associated genes using
921 an automated python script which iterated over annotations by contig. A list of 33 gene ontology
922 (GO) terms identified as head or packaging genes was curated from the QuickGO display
923 (Binns et al., 2009). UniProtKB annotations which correspond to GO terms for virion assembly
924 and maturation were extracted and curated to exclude baseplate and tail gene products (The
925 UniProt Consortium, 2022). The resulting list of 332 UniProtKB codes were used for the Python
926 search script which flagged contigs containing both head and packaging associated gene
927 products which were then manually verified. All manual checks for gene functional prediction
928 were performed with a combination of BLAST(Camacho et al., 2009), UniProt BLAST, and
929 HHPred (Swissprot DB) (Söding et al., 2005).

Viral clustering

930 Viral sequences were also grouped into roughly genus level viral clusters (VCs) using
931 vConTACT3 v.3.0 with database (v220) domain restricted to "prokaryotes" (Bolduc et al., 2025).
932 Graph output (.cyjs) was imported into Cytoscape (v3.10.2) and the NetworkAnalyzer plugin
933 was used to evaluate the quality of clusters (Su et al., 2014). 'Clustering Coefficient' and
934 'Topological Coefficient' metrics were filtered to be equal to or greater than 0.85 (Tian et al.,
935 2024). Subclustering (virus cluster of three or more nodes without links to the rest of the
936 network) is observed for 20 VCs. Packaging and head related genes of viruses that were unable
937 to be taxonomically assigned into a realm by vConTACT3 v.3.0 were manually verified using

938 BLAST, UniProt BLAST, and remote homology prediction tool HHPred. Genes for major capsid
939 protein and TerL subunit were annotated (HHblits) and clustered (BLAST 2.14.1+) to assess
940 cluster success with viral sequences of variable length (variation > 1 kb).

Generating vOTUs

941 Viruses were clustered into species-rank vOTUs reflecting viral populations using
942 Minimum Information about an Uncultivated Virus Genome (MIUViG) standards with
943 recommended thresholds (95% ANI and 85% AF) for clustering parameters (Roux et al., 2019).
944 Contigs were first compared with nucleotide BLAST (v2.11.0+) and pairwise comparisons were
945 generated by blasting all sequences together ("-outfmt '6 std qlen slen' -max_target_seqs
946 10000"). Finally, clustering of viral sequences was carried out using scripts anicalc.py and
947 aniclust.py ("--min_ani 95 --min_tcov 85 --min_qcov 0") which are distributed as part of the
948 CheckV package (Nayfach et al., 2021).

949 Calculating vOTU diversity and abundance

950 Individual viral contigs were mapped using Bowtie2 (Langmead & Salzberg, 2012) with
951 the bowtie2-build function to create .sam files. The .sam files were converted to .bam files then
952 sorted, indexed, and counted (samtools idxstats) with SAMtools (Li et al., 2009). Relative
953 abundance was calculated with reads per kilobase per million mapped reads (RPKM) for each
954 viral contig and then summed for all contigs for each vOTU (Yan et al., 2025). vOTU richness
955 and evenness were calculated with the vegan package v2.7-2 (Oksanen, 2025) in R. Richness
956 was assessed for each sample by finding the Chao1 index using the sample vOTU relative
957 abundance table, which was calculated with the specnumber() function. Shannon diversity was
958 calculated for each sample with the diversity() function from the sample relative vOTU
959 abundance table. Pielou's evenness was calculated by dividing the Shannon index by the ln() of
960 Chao1. The relationship between the temperature of samples and vOTU diversity indexes was

961 assessed for significance with a non-parametric least squares regression, and then assessed
962 for overfitting with a Bayesian information criterion test with sample correction (BICc) due to
963 sample size (n=20). Linear, quadratic, and cubic models were attempted, and the lowest BIC
964 value indicated the best fit model. Normal distribution and constant variance of residuals was
965 supported with density and QQ-plots for both relationships (Supplemental Figure 07). Normality
966 for models of temperature and related diversity indexes was confirmed with the Shapiro-Wilk
967 test.

Phylogenetic classification of viruses

968 Phylogenetic estimation was guided by vConTACT3 v.3.0 (Bolduc et al., 2017, 2025).
969 Phylogenies were estimated using maximum likelihood independently for classes
970 Caudoviricetes, Malgrandaviricetes, Laserviricetes, and Tectiliviricetes. Caudoviricetes
971 references were chosen by selecting first neighbors ('first neighbors of select nodes' Cytoscape
972 v3.10.2) of vConTACT3 v.3.0 assigned fumarole Caudoviricetes (Shannon et al., 2003). For
973 Malgrandaviricetes, Laserviricetes, and Tectiliviricetes phylogenies, all known references were
974 selected from ICTV MSL40.v1 and ORFs were called with Prodigal v2.6.3 (Hyatt et al., 2010).
975 All references were annotated as described above with HHblits to identify marker genes. The
976 marker genes were TerL, major capsid, and packaging ATPase marker genes for the
977 Caudoviricetes, Malgrandaviricetes, Laserviricetes,Tecitiviricetes trees respectively with 1000
978 iterations of tree topology, branch lengths, and substitution model parameters (Grybchuk et al.,
979 2024; Katoh et al., 2002). Trimming of the multiple sequence alignment was performed with
980 trimAl v1.5.rev0 with a gap threshold of 30% (Capella-Gutiérrez et al., 2009). IQTree v2.4.0 was
981 used for phylogenetic reconstruction, with the evolutionary model Q.pfam+F+I+G4 selected as
982 the best model for *Caudoviricetes* (Nguyen et al., 2014). The model Q.pfam+G4 was used for
983 Malgrandaviricetes and *Laserviricetes* trees, and the Q.pfam+I+G4 model was used for

984 *Tectiliviricetes*. 1000 bootstrap estimates were used for branch support estimation. All
985 phylogenies were midpoint rooted and visualized with iTOL v7 (Letunic & Bork, 2006).

Host matching with CRISPR spacers

986 Putative host prediction was performed with CRISPR spacer matching. CRISPR-Cas
987 systems were identified from the binned nucleic acid sequences with <5% contamination using
988 the standalone tool of CRISPRCasFinder as a Singularity image (Couvin et al., 2018). High
989 confidence arrays were identified by using '-levelMin 3' parameter. Arrays adjacent to Cas
990 systems within 10,000 base pairs (bp) were identified with the '-ccc' parameter. Spacers were
991 extracted from high confidence arrays using minCED and clustered with a cutoff of 100%
992 identity using cd-hit-est (directionality assessed with parameter '-r 1') to dereplicate spacer
993 sequences (Bland et al., 2007; Li & Godzik, 2006). Targets with host matches were sorted into
994 three categories of mobile elements during the viral identification workflow: plasmid, viral, or
995 unknown mobile genetic element (MGEs). MGEs that did not have chromosomal, plasmid, or
996 viral indicators (structural genes) were characterized as other MGEs. Chromosomal fragments
997 were confidently excluded by scanning matches and eliminating targets with either entire arrays
998 or multiple shared spacers and neighboring sequences which were marked as 'suspicious'
999 matches (Supplemental Material S7). Viruses identified as viral and plasmid hits or uncertain
1000 viral and plasmid sequences (ViralVerify, VirSorter2, and geNomad) were aligned with BLAST
1001 (2.14.1+). Matches were filtered for ratio *nident* (identical bp matches) to *qlen* (spacer length) >
1002 0.89 to be considered operationally valid.
1003 Matches that could be plasmid-encoded arrays or chromosomal fragments were
1004 eliminated as follows (Pinilla-Redondo et al., 2021): CRISPRDetect was used to identify arrays
1005 in chromosomal arrays and target elements which were then aligned with BLAST (2.14.1+).
1006 Identical matches between chromosomal and plasmid arrays were discarded immediately. If
1007 there were multiple shared spacers and neighboring sequences (spacers and repeat element)

1008 the hit was designated as suspicious (Supplemental Material S9). Target sequences were
1009 searched for ORFs using Prodigal (v2.6.3) and .faa (amino acid sequence) and .sco (ORF
1010 coordinates) output files were used in conjunction with BLAST (2.14.1+) match coordinates to
1011 identify which ORF was targeted by spacers. Targeted ORFs (coding sequences (CDS) or non-
1012 coding ORFs) were annotated using BLAST, UniProtBLAST, and HHPred.

1013 Viral genome annotation (metabolic and accessory genes)

1014 Three categories of genes were systematically recovered from vOTUs: auxiliary
1015 metabolic genes (AMGs) that may have a role in host metabolism during the infection cycle,
1016 viral resistance genes, and viral defense system related genes. Viral resistance and defense
1017 genes are non-essential accessory genes (morons) which provide an adaptive advantage to
1018 either the phage or host (Kang et al., 2024). Annotation for functional gene product prediction
1019 was performed with DRAM-v v0.1.2 (Shaffer et al., 2020) on all viral sequences. Final vOTUs
1020 were processed with VirSorter2 v1.0.1 using the -prep-for-dramv flag then annotated with
1021 DRAM-v to identify viral metabolic genes. Gene curation was performed by filtering and
1022 manually scanning both the DRAM-v AMG summary (275 gene predictions) and total annotation
1023 output (20,666 genes) for AMGs and accessory genes. Gene product annotations were
1024 discarded for genes that did not have an auxiliary score <4, or had DRAM-v auxiliary flags of
1025 (V) viral structural component, (A) attachment, (P), viral specific MEROPS peptidase, or (F)
1026 'near end of contig' (ter Horst et al., 2021; Tian et al., 2024). To eliminate the candidate genes
1027 within host-like gene operons (possible miscalled proviral ends) all candidate genes were
1028 scanned for flanking viral genes using DRAM-v flag (V) for structural virion components,
1029 annotations such as portal, tail, baseplate, capsid, and structural or replicative genes with
1030 VOGDB flags (Xs;Xr). This left 212 putative AMG candidates (Supplemental Material S8).
1031 To be conservative in AMG prediction, secondary manual curation was performed on
1032 putative AMG candidates to eliminate genes that could be involved in viral life cycle processes

(ter Horst et al., 2021; Trubl et al., 2018). This led to removal of putative AMGs involved in nucleotide or amino acid metabolism (DRAM-v and KEGG-MAPPER pathways and modules). Putative AMGs were also discarded if their annotation was overly general (genes annotated as glycosyltransferases, peptidases, glycosylhydrolases, adenylyltransferases, methyltransferases, and ribosomal proteins). Attempts were made to obtain a more specific gene prediction with BLASTp before discarding (Martin et al., 2025; ter Horst et al., 2021). To predict accessory genes, a secondary manual curation was performed by searching non-metabolic annotations that did not have the (M) auxiliary flag in the DRAM-v master annotation spreadsheet output. Participation in life cycle pathways was assessed using methods described above, and resulting accessory gene candidates were manually inspected.

Prophage annotation and boundaries were performed by aligning viral contigs to MAGs with BLAST+ alignment (2.14.1+). Prophage boundaries were identified in viral contigs PHASTEST and manually checked for flanking viral genes on viral contigs using annotation methods described in the workflow for viral identification (Wishart et al., 2023). ORFs were called with Prodigal (v2.6.3) and annotated with Pharokka (v1.7.5) identifying phage orthologous groups (PHROGs) with local alignment (Bouras et al., 2022). Gene diagrams (genomic context determined with MMSeq2 and pyHMMer) were plotted using LoVis4u (Egorov & Atkinson, 2025; Steinegger & Söding, 2017). Proteomic similarity matrix for *Microviridae* viruses was built with MMSeq2 alignments through LoVis4u.

Acknowledgements

We would like to thank George Washington University's high-performance computing facility, Pegasus, for providing data storage, support, and computing power for metagenomic analyses. We would also like to thank Pietro Tardelli Canedo, Pooja Anilkumar, Arash Raeisbahrami, Zachary J. Watts, and Max Seledes for discussion, as well as the International Viromics

Workshop at Ohio State University which provided authors Pia Sen and Lausanne Lee Oliver foundational knowledge for viral metagenomic analysis. We also thank Keith A. Crandall, Sarah Johnson, and Gary Trubl for manuscript review and comments. We gratefully acknowledge the computing resources provided on the High-Performance Computing cluster operated by Research Technology Services at The George Washington University. K.S.M. and Y.I.W. were supported by the Intramural Research Program of the National Institutes of Health (NIH). The contributions of the NIH authors are considered Works of the United States Government. The findings and conclusions presented in this paper are those of the author and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

Data availability

All raw data used in this study have been deposited in the European Nucleotide Archive (ENA) under project accession number PRJEB52128. Phylogenetic IQ-Trees, viral candidate contigs, and final viral contigs are available on Zenodo at DOI: https://doi.org/10.5281/zenodo.19378070

Code availability

Scripts to run analyses and viral identification pipeline custom scripts are available at https://github.com/pia-sen/Hawaii_fumarole_viruses

Author contributions

J.H.S. designed and led the sampling and sequencing. P.S. and J.H.S. determined the scope of analysis for this project. P.S, L.O, C.P. and M.S. performed data analyses, K.S.M. and Y.I.W. supervised and guided development of the viral identification pipeline, P.S. L.O. and J.H.S wrote the manuscript, and all authors reviewed and approved the final version of this paper.

Sequencing of the samples was made possible by the startup funds provided to J.H.S by The George Washington University.

References

Aguila-Torres, P., González, M., Hernández, M., Aguado-Norese, C., Maldonado, J. E., Miranda, R. M., González-Stegmaier, R., Palma, D. E., Rojas, L. A., & Mellado, M. (2025). Characterization of Bacterial Communities in Volcanic Soil from Northern Patagonian Area of Chile. *Microorganisms*, *13*(11), 2519. https://doi.org/10.3390/microorganisms13112519

Antipov, D., Raiko, M., Lapidus, A., & Pevzner, P. A. (2020). MetaviralSPAdes: Assembly of viruses from metagenomic data. *Bioinformatics*, *36*(14), 4126–4129. https://doi.org/10.1093/bioinformatics/btaa490

Audrey, B., Cellier, N., White, F., Jacques, P.-É., & Burrus, V. (2023). A systematic approach to classify and characterize genomic islands driven by conjugative mobility using protein signatures. *Nucleic Acids Research*, *51*(16), 8402–8412. https://doi.org/10.1093/nar/gkad644

Barylski, J., Kropinski, A. M., Alikhan, N.-F., Adriaenssens, E. M., & ICTV Report Consortium. (2020). ICTV Virus Taxonomy Profile: Herelleviridae. *Journal of General Virology*, *101*(4), 362–363. https://doi.org/10.1099/jgv.0.001392

Bergsten, P., Vannier, P., Klonowski, A. M., Knobloch, S., Gudmundsson, M. T., Jackson, M. D., & Marteinsson, V. T. (2021). Basalt-Hosted Microbial Communities in the Subsurface of the Young Volcanic Island of Surtsey, Iceland. *Frontiers in Microbiology*, *12*, 728977. https://doi.org/10.3389/fmicb.2021.728977

Bignaud, A., Conti, D. E., Thierry, A., Serizay, J., Labadie, K., Poulain, J., Cheny, O., Colón-González, M., Debarbieux, L., Guerrero-Osornio, M., Helaine, S., Hill, P., Le Tinier, G., Millot, G. A., Morales, L., Parada, A., Riera, N., Iraola, G., Koszul, R., & Marbouty, M. (2025). Phages with a broad host range are common across ecosystems. *Nature Microbiology*, *10*(10), 2537–2549. https://doi.org/10.1038/s41564-025-02108-2

Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C., & Apweiler, R. (2009). QuickGO: A web-based tool for Gene Ontology searching. *Bioinformatics*, *25*(22), 3045–3046. https://doi.org/10.1093/bioinformatics/btp536

Black, E. J., Powell, C. S., Dempsey, D. M., Hendrickson, R. C., Mims, L. R., & Lefkowitz, E. J. (2025). Virus taxonomy: The database of the International Committee on Taxonomy of Viruses. *Nucleic Acids Research*, gkaf1159. https://doi.org/10.1093/nar/gkaf1159

Blanc-Mathieu, R., Dahle, H., Hofgaard, A., Brandt, D., Ban, H., Kalinowski, J., Ogata, H., & Sandaa, R.-A. (2021). A Persistent Giant Algal Virus, with a Unique Morphology, Encodes an Unprecedented Number of Genes Involved in Energy Metabolism. *Journal of Virology*, *95*(8), e02446-20. https://doi.org/10.1128/JVI.02446-20

Bland, C., Ramsey, T. L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N. C., & Hugenholtz, P. (2007). CRISPR Recognition Tool (CRT): A tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics*, *8*(1), 209. https://doi.org/10.1186/1471-2105-8-209

Bolduc, B., Jang, H. B., Doulcier, G., You, Z.-Q., Roux, S., & Sullivan, M. B. (2017). vConTACT: An iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ*, *5*, e3243. https://doi.org/10.7717/peerj.3243

Bolduc, B., Zablocki, O., Turner, D., Jang, H. B., Guo, J., Adriaenssens, E. M., Dutilh, B. E., & Sullivan, M. B. (2025). *Scalable and systematic hierarchical virus taxonomy with vConTACT3* (p. 2025.11.06.686974). bioRxiv. https://doi.org/10.1101/2025.11.06.686974

Bouras, G., Nepal, R., Houtak, G., Psaltis, A., Wormald, P.-J., & Vreugde, S. (2022). Pharokka: A fast scalable bacteriophage annotation tool. *Bioinformatics*, *39*(1). https://doi.org/10.1093/bioinformatics/btac776

Brahim Belhaouari, D., Pires De Souza, G. A., Lamb, D. C., Kelly, S. L., Goldstone, J. V., Stegeman, J. J., Colson, P., La Scola, B., & Aherfi, S. (2022). Metabolic arsenal of giant viruses: Host hijack or self-use? *eLife*, *11*, e78674. https://doi.org/10.7554/eLife.78674

Breitbart, M., & Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? *Trends in Microbiology*, *13*(6), 278–284. https://doi.org/10.1016/j.tim.2005.04.003

Brown, T. L., Charity, O. J., & Adriaenssens, E. M. (2022). Ecological and functional roles of bacteriophages in contrasting environments: Marine, terrestrial and human gut. *Current Opinion in Microbiology*, *70*, 102229. https://doi.org/10.1016/j.mib.2022.102229

Brüssow, H., Canchaya, C., & Hardt, W.-D. (2004). Phages and the Evolution of Bacterial Pathogens: From Genomic Rearrangements to Lysogenic Conversion.

Microbiology and Molecular Biology Reviews, *68*(3), 560–602. https://doi.org/10.1128/MMBR.68.3.560-602.2004

Bucher, M. J., & Czyż, D. M. (2024). Phage against the Machine: The SIE-ence of Superinfection Exclusion. *Viruses*, *16*(9), 1348. https://doi.org/10.3390/v16091348

Bushnell, B. (2014). *BBMap: A Fast, Accurate, Splice-Aware Aligner* (LBNL-7065E). Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA (US). https://www.osti.gov/biblio/1241166

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 421. https://doi.org/10.1186/1471-2105-10-421

Camargo, A. P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., Chain, P. S. G., Nayfach, S., & Kyrpides, N. C. (2024). Identification of mobile genetic elements with geNomad. *Nature Biotechnology*, *42*(8), 1303–1312. https://doi.org/10.1038/s41587-023-01953-y

Cardenas, J. P., Quatrini, R., & Holmes, D. S. (2016). Aerobic Lineage of the Oxidative Stress Response Protein Rubrerythrin Emerged in an Ancient Microaerobic, (Hyper)Thermophilic Environment. *Frontiers in Microbiology*, *7*, 1822. https://doi.org/10.3389/fmicb.2016.01822

Castelán-Sánchez, H. G., Lopéz-Rosas, I., García-Suastegui, W. A., Peralta, R., Dobson, A. D. W., Batista-García, R. A., & Dávila-Ramos, S. (2019). Extremophile deep-sea viral communities from hydrothermal vents: Structural and functional analysis. *Marine Genomics*, *46*, 16–28. https://doi.org/10.1016/j.margen.2019.03.001

Castelán-Sánchez, H. G., Meza-Rodríguez, P. M., Carrillo, E., Ríos-Vázquez, D. I., Liñan-Torres, A., Batista-García, R. A., Pérez-Rueda, E., Rojas-Ruíz, N. E., & Dávila-Ramos, S. (2020). The Microbial Composition in Circumneutral Thermal Springs from Chignahuapan, Puebla, Mexico Reveals the Presence of Particular Sulfur-Oxidizing Bacterial and Viral Communities. *Microorganisms*, *8*(11), 1677. https://doi.org/10.3390/microorganisms8111677

Castledine, M., Sierocinski, P., Inglis, M., Kay, S., Hayward, A., Buckling, A., & Padfield, D. (2022). Greater Phage Genotypic Diversity Constrains Arms-Race Coevolution. *Frontiers in Cellular and Infection Microbiology*, *12*. https://doi.org/10.3389/fcimb.2022.834406

Chakraborty, S., Rohit, A., Prasanthi, S. J., & Chauhan, A. (2024). A New Casjensviridae Bacteriophage Isolated from Hospital Sewage for Inactivation of Biofilms of Carbapenem

Resistant Klebsiella pneumoniae Clinical Isolates. *Pharmaceutics*, *16*(7), 904. https://doi.org/10.3390/pharmaceutics16070904

Cheng, R., Li, X., Jiang, L., Gong, L., Geslin, C., & Shao, Z. (2022). Virus diversity and interactions with hosts in deep-sea hydrothermal vents. *Microbiome*, *10*(1), 235. https://doi.org/10.1186/s40168-022-01441-6

Common, J., Morley, D., Westra, E. R., & van Houte, S. (2019). CRISPR-Cas immunity leads to a coevolutionary arms race between Streptococcus thermophilus and lytic phage. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1772), 20180098. https://doi.org/10.1098/rstb.2018.0098

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., Rocha, E. P. C., Vergnaud, G., Gautheret, D., & Pourcel, C. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Research*, *46*(Web Server issue), W246–W251. https://doi.org/10.1093/nar/gky425

Decho, A. W., & Gutierrez, T. (2017). Microbial Extracellular Polymeric Substances (EPSs) in Ocean Systems. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.00922

Duda, R. L., & Teschke, C. M. (2019). The amazing HK97 fold: Versatile results of modest differences. *Current Opinion in Virology, Virus Structure and Expression*, *36*, 9–16. https://doi.org/10.1016/j.coviro.2019.02.001

Egorov, A. A., & Atkinson, G. C. (2025). LoVis4u: A locus visualization tool for comparative genomics and coverage profiles. *NAR Genomics and Bioinformatics*, *7*(1), lqaf009. https://doi.org/10.1093/nargab/lqaf009

Farias, P., Espírito Santo, C., Branco, R., Francisco, R., Santos, S., Hansen, L., Sorensen, S., & Morais, P. V. (2015). Natural Hot Spots for Gain of Multiple Resistances: Arsenic and Antibiotic Resistances in Heterotrophic, Aerobic Bacteria from Marine Hydrothermal Vent Fields. *Applied and Environmental Microbiology*, *81*(7), 2534–2543. https://doi.org/10.1128/AEM.03240-14

Felipe Benites, L., Stephens, T. G., Van Etten, J., James, T., Christian, W. C., Barry, K., Grigoriev, I. V., McDermott, T. R., & Bhattacharya, D. (2024). Hot springs viruses at Yellowstone National Park have ancient origins and are adapted to thermophilic hosts. *Communications Biology*, *7*(1), 1–12. https://doi.org/10.1038/s42003-024-05931-1

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., & Bateman, A.

(2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279–D285. https://doi.org/10.1093/nar/gkv1344

Godin, P. J., Moore, C. A., Smith, C., & Moores, J. E. (2023). Absorption and Scattering of UV and Visible Light Through Simulated Martian Regoliths and Rock Samples. *Astrobiology*, *23*(3), 280–290. https://doi.org/10.1089/ast.2021.0184

Goldfarb, T., Kodali, V. K., Pujar, S., Brover, V., Robbertse, B., Farrell, C. M., Oh, D.-H., Astashyn, A., Ermolaeva, O., Haddad, D., Hlavina, W., Hoffman, J., Jackson, J. D., Joardar, V. S., Kristensen, D., Masterson, P., McGarvey, K. M., McVeigh, R., Mozes, E., … Murphy, T. D. (2024). NCBI RefSeq: Reference sequence standards through 25 years of curation and annotation. *Nucleic Acids Research*, *53*(D1), D243–D257. https://doi.org/10.1093/nar/gkae1038

Greenrod, S. T. E., Hector, T. E., Blazanin, M., Cazares, D., & King, K. C. (2025). Temperature as a Driver of Phage Ecology and Evolution. *Annual Review of Microbiology*, *79*(Volume 79, 2025), 497–522. https://doi.org/10.1146/annurev-micro-042424-040029

Gregory, A. C., Zayed, A. A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., Ardyna, M., Arkhipova, K., Carmichael, M., Cruaud, C., Dimier, C., Domínguez-Huerta, G., Ferland, J., Kandels, S., Liu, Y., Marec, C., Pesant, S., Picheral, M., Pisarev, S., … Sullivan, M. B. (2019). Marine DNA viral macro- and micro-diversity from pole to pole. *Cell*, *177*(5), 1109-1123.e14. https://doi.org/10.1016/j.cell.2019.03.040

Grettenberger, C. L. (2021). Novel Gloeobacterales spp. From Diverse Environments across the Globe. *mSphere*, *6*(4), 10.1128/msphere.00061-21. https://doi.org/10.1128/msphere.00061-21

Guajardo-Leiva, S., Santos, F., Salgado, O., Regeard, C., Quillet, L., & Díez, B. (2021a). Unveiling Ecological and Genetic Novelty within Lytic and Lysogenic Viral Communities of Hot Spring Phototrophic Microbial Mats. *Microbiology Spectrum*, *9*(3), e00694-21. https://doi.org/10.1128/Spectrum.00694-21

Guajardo-Leiva, S., Santos, F., Salgado, O., Regeard, C., Quillet, L., & Díez, B. (2021b). Unveiling Ecological and Genetic Novelty within Lytic and Lysogenic Viral Communities of Hot Spring Phototrophic Microbial Mats. *Microbiology Spectrum*, *9*(3), e00694-21. https://doi.org/10.1128/Spectrum.00694-21

Guo, J., Bolduc, B., Zayed, A. A., Varsani, A., Dominguez-Huerta, G., Delmont, T. O., Pratama, A. A., Gazitúa, M. C., Vik, D., Sullivan, M. B., & Roux, S. (2021). VirSorter2: A

multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome*, *9*(1), 37. https://doi.org/10.1186/s40168-020-00990-y

Hadland, N., Hamilton, C. W., & Duhamel, S. (2024). Young volcanic terrains are windows into early microbial colonization. *Communications Earth & Environment*, *5*(1), 114. https://doi.org/10.1038/s43247-024-01280-3

Harris, H. M. B., & Hill, C. (2021). A Place for Viruses on the Tree of Life. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.604048

He, T., Li, H., & Zhang, X. (2017). Deep-Sea Hydrothermal Vent Viruses Compensate for Microbial Metabolism in Virus-Host Interactions. *mBio*, *8*(4), 10.1128/mbio.00893-17. https://doi.org/10.1128/mbio.00893-17

Hevroni, G., Flores-Uribe, J., Béjà, O., & Philosof, A. (2020). Seasonal and diel patterns of abundance and activity of viruses in the Red Sea. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(47), 29738–29747. https://doi.org/10.1073/pnas.2010783117

Hu, C., Chen, X., Wei, W., Wallace, D., Liu, J., Zhang, Y., Zhang, L., Xu, D., Batt, J., Xiao, X., Shi, Q., Zheng, Q., Ma, R., Luo, T., Jiao, N., & Zhang, R. (2025). To kill or to piggyback: Switching of viral lysis-lysogeny strategies depending on host dynamics. *Science of The Total Environment*, *959*, 178233. https://doi.org/10.1016/j.scitotenv.2024.178233

Huang, Q., Yang, F., Cao, H., Cheng, J., Jiang, M., Li, M., Ni, H., & Xie, L. (2024). Comparison of Microbial Diversity of Two Typical Volcanic Soils in Wudalianchi, China. *Microorganisms*, *12*(4), 656. https://doi.org/10.3390/microorganisms12040656

Hutinet, G., Kot, W., Cui, L., Hillebrand, R., Balamkundu, S., Gnanakalai, S., Neelakandan, R., Carstens, A. B., Fa Lui, C., Tremblay, D., Jacobs-Sera, D., Sassanfar, M., Lee, Y.-J., Weigele, P., Moineau, S., Hatfull, G. F., Dedon, P. C., Hansen, L. H., & de Crécy-Lagard, V. (2019). 7-Deazaguanine modifications protect phage DNA from host restriction systems. *Nature Communications*, *10*(1), 5442. https://doi.org/10.1038/s41467-019-13384-y

Hwang, Y., Roux, S., Coclet, C., Krause, S. J. E., & Girguis, P. R. (2023). Viruses interact with hosts that span distantly related microbial domains in dense hydrothermal mats. *Nature Microbiology*, *8*(5), 946–957. https://doi.org/10.1038/s41564-023-01347-5

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*(1), 119. https://doi.org/10.1186/1471-2105-11-119

Jarett, J. K., Džunková, M., Schulz, F., Roux, S., Paez-Espino, D., Eloe-Fadrosh, E., Jungbluth, S. P., Ivanova, N., Spear, J. R., Carr, S. A., Trivedi, C. B., Corsetti, F. A., Johnson, H. A., Becraft, E., Kyrpides, N., Stepanauskas, R., & Woyke, T. (2020). Insights into the dynamics between viruses and their hosts in a hot spring microbial mat. *The ISME Journal*, *14*(10), 2527–2541. https://doi.org/10.1038/s41396-020-0705-4

Ji, M., Li, Y., Zhou, J., Song, W., Zhou, Y., Ma, K., Wang, M., Liu, X., Li, Y., Gong, X., & Tu, Q. (2024). Temporal turnover of viral biodiversity and functional potential in intertidal wetlands. *Npj Biofilms and Microbiomes*, *10*(1), 48. https://doi.org/10.1038/s41522-024-00522-8

Jurgensen, S. K., Roux, S., Schwenck, S. M., Stewart, F. J., Sullivan, M. B., & Brum, J. R. (2022). Viral community analysis in a marine oxygen minimum zone indicates increased potential for viral manipulation of microbial physiological state. *The ISME Journal*, *16*(4), 972–982. https://doi.org/10.1038/s41396-021-01143-1

Kang, Y., Wang, J., Zhu, C., Zheng, M., & Li, Z. (2024). Unveiling the genomic diversity and ecological impact of phage communities in hospital wastewater. *Journal of Hazardous Materials*, *477*, 135353. https://doi.org/10.1016/j.jhazmat.2024.135353

Kashif-Khan, N., Savva, R., & Frank, S. (2024). Mining metagenomics data for novel bacterial nanocompartments. *NAR Genomics and Bioinformatics*, *6*(1), lqae025. https://doi.org/10.1093/nargab/lqae025

King, G. M. (2003). Contributions of Atmospheric CO and Hydrogen Uptake to Microbial Dynamics on Recent Hawaiian Volcanic Deposits. *Applied and Environmental Microbiology*, *69*(7), 4067–4075. https://doi.org/10.1128/AEM.69.7.4067-4075.2003

Kirchberger, P. C., Martinez, Z. A., & Ochman, H. (2022). Organizing the Global Diversity of Microviruses. *mBio*, *13*(3), e00588-22. https://doi.org/10.1128/mbio.00588-22

Kirchberger, P. C., & Ochman, H. (2020). Resurrection of a global, metagenomically defined gokushovirus. *eLife*, *9*, e51599. https://doi.org/10.7554/eLife.51599

Kirchberger, P. C., & Ochman, H. (2023). Microviruses: A World Beyond phiX174. *Annual Review of Virology*, *10*(Volume 10, 2023), 99–118. https://doi.org/10.1146/annurev-virology-100120-011239

Koonin, E. V., & Makarova, K. S. (2017). Mobile Genetic Elements and Evolution of CRISPR-Cas Systems: All the Way There and Back. *Genome Biology and Evolution*, *9*(10), 2812–2825. https://doi.org/10.1093/gbe/evx192

Koonin, E. V., & Wolf, Y. I. (2012). Evolution of microbes and viruses: A paradigm shift in evolutionary biology? *Frontiers in Cellular and Infection Microbiology*, *2*, 119. https://doi.org/10.3389/fcimb.2012.00119

Koonin, E. V., & Wolf, Y. I. (2015). Evolution of the CRISPR-Cas adaptive immunity systems in prokaryotes: Models and observations on virus–host coevolution. *Molecular bioSystems*, *11*(1), 20–27. https://doi.org/10.1039/c4mb00438h

Kostešić, E., Mitrović, M., Kajan, K., Marković, T., Hausmann, B., Orlić, S., & Pjevac, P. (2023). Microbial Diversity and Activity of Biofilms from Geothermal Springs in Croatia. *Microbial Ecology*, *86*(4), 2305–2319. https://doi.org/10.1007/s00248-023-02239-1

Laanto, E., Mäntynen, S., De Colibus, L., Marjakangas, J., Gillum, A., Stuart, D. I., Ravantti, J. J., Huiskonen, J. T., & Sundberg, L.-R. (2017). Virus found in a boreal lake links ssDNA and dsDNA viruses. *Proceedings of the National Academy of Sciences*, *114*(31), 8378–8383. https://doi.org/10.1073/pnas.1703834114

Lamb, D. C., Goldstone, J. V., Belhaouari, D. B., Andréani, J., Farooqi, A., Allen, M. J., Kelly, S. L., La Scola, B., & Stegeman, J. J. (2024). Cytochromes b5 Occurrence in Viruses Belonging to the Order Megavirales. *Research Square*, rs.3.rs-5246363. https://doi.org/10.21203/rs.3.rs-5246363/v1

Langwig, M. V., Koester, F., Martin, C., Zhou, Z., Joye, S. B., Reysenbach, A.-L., & Anantharaman, K. (2025). Endemism shapes viral ecology and evolution in globally distributed hydrothermal vent ecosystems. *Nature Communications*, *16*, 4076. https://doi.org/10.1038/s41467-025-59154-x

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Liang, X., Wagner, R. E., Zhuang, J., DeBruyn, J. M., Wilhelm, S. W., Liu, F., Yang, L., Staton, M. E., Sherfy, A. C., & Radosevich, M. (2019). Viral abundance and diversity vary with depth in a southeastern United States agricultural ultisol. *Soil Biology and Biochemistry*, *137*, 107546. https://doi.org/10.1016/j.soilbio.2019.107546

Madonia, P. (2020). Variations of low temperature fumaroles as a tool for detecting changes in volcanic activity state: A brief overview. *Advances in Geosciences*, *52*, 97–103. https://doi.org/10.5194/adgeo-52-97-2020

Martin, C., Emerson, J. B., Roux, S., & Anantharaman, K. (2025). A call for caution in the biological interpretation of viral auxiliary metabolic genes. *Nature Microbiology*, *10*(9), 2122–2129. https://doi.org/10.1038/s41564-025-02095-4

Martynov, A., Severinov, K., & Ispolatov, I. (2017). Optimal number of spacers in CRISPR arrays. *PLOS Computational Biology*, *13*(12), e1005891. https://doi.org/10.1371/journal.pcbi.1005891

Middelboe, M., Traving, S. J., Castillo, D., Kalatzis, P. G., & Glud, R. N. (2025). Prophage-encoded chitinase gene supports growth of its bacterial host isolated from deep-sea sediments. *The ISME Journal*, *19*(1), wraf004. https://doi.org/10.1093/ismejo/wraf004

Mihai, R. A., Espinoza-Caiza, I. A., Melo-Heras, E. J., Cubi-Insuaste, N. S., Pinto-Valdiviezo, E. A., & Catana, R. D. (2023). Does the Mineral Composition of Volcanic Ashes Have a Beneficial or Detrimental Impact on the Soils and Cultivated Crops of Ecuador? *Toxics*, *11*(10), 846. https://doi.org/10.3390/toxics11100846

Monier, A., Welsh, R. M., Gentemann, C., Weinstock, G., Sodergren, E., Armbrust, E. V., Eisen, J. A., & Worden, A. Z. (2012). Phosphate transporters in marine phytoplankton and their viruses: Cross-domain commonalities in viral-host gene exchanges. *Environmental Microbiology*, *14*(1), 162–176. https://doi.org/10.1111/j.1462-2920.2011.02576.x

Moya-Beltrán, A., Makarova, K. S., Acuña, L. G., Wolf, Y. I., Covarrubias, P. C., Shmakov, S. A., Silva, C., Tolstoy, I., Johnson, D. B., Koonin, E. V., & Quatrini, R. (2021). Evolution of Type IV CRISPR-Cas Systems: Insights from CRISPR Loci in Integrative Conjugative Elements of Acidithiobacillia. *The CRISPR Journal*, *4*(5), 656–672. https://doi.org/10.1089/crispr.2021.0051

Muscatt, G., Cook, R., Millard, A., Bending, G. D., & Jameson, E. (2022). *Ecological and evolutionary patterns of virus-host interactions throughout a grassland soil depth profile* (p. 2022.12.09.519740). bioRxiv. https://doi.org/10.1101/2022.12.09.519740

Nayfach, S., Camargo, A. P., Schulz, F., Eloe-Fadrosh, E., Roux, S., & Kyrpides, N. C. (2021). CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature Biotechnology*, *39*(5), 578–585. https://doi.org/10.1038/s41587-020-00774-7

O'Brien, R. E., Bravo, J. P. K., Ramos, D., Hibshman, G. N., Wright, J. T., & Taylor, D. W. (2022). *Modes of inhibition used by phage anti-CRISPRs to evade type I-C Cascade* (p. 2022.06.15.496202). bioRxiv. https://doi.org/10.1101/2022.06.15.496202

Pinilla-Redondo, R., Mayo-Muñoz, D., Russel, J., Garrett, R. A., Randau, L., Sørensen, S. J., & Shah, S. A. (2020). Type IV CRISPR–Cas systems are highly diverse and

involved in competition between plasmids. *Nucleic Acids Research*, *48*(4), 2000–2012. https://doi.org/10.1093/nar/gkz1197

Pinilla-Redondo, R., Russel, J., Mayo-Muñoz, D., Shah, S. A., Garrett, R. A., Nesme, J., Madsen, J. S., Fineran, P. C., & Sørensen, S. J. (2021). CRISPR-Cas systems are widespread accessory elements across bacterial and archaeal plasmids. *Nucleic Acids Research*, *50*(8), 4315–4328. https://doi.org/10.1093/nar/gkab859

Pohane, A. A., Carr, C. R., Garhyan, J., Swarts, B. M., & Siegrist, M. S. (2021). Trehalose Recycling Promotes Energy-Efficient Biosynthesis of the Mycobacterial Cell Envelope. *mBio*, *12*(1), 10.1128/mbio.02801-20. https://doi.org/10.1128/mbio.02801-20

Power, J. F., Carere, C. R., Lee, C. K., Wakerley, G. L. J., Evans, D. W., Button, M., White, D., Climo, M. D., Hinze, A. M., Morgan, X. C., McDonald, I. R., Cary, S. C., & Stott, M. B. (2018). Microbial biogeography of 925 geothermal springs in New Zealand. *Nature Communications*, *9*(1), 2876. https://doi.org/10.1038/s41467-018-05020-y

Prescott, R. D., Zamkovaya, T., Donachie, S. P., Northup, D. E., Medley, J. J., Monsalve, N., Saw, J. H., Decho, A. W., Chain, P. S. G., & Boston, P. J. (2022). Islands Within Islands: Bacterial Phylogenetic Structure and Consortia in Hawaiian Lava Caves and Fumaroles. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.934708

Raven, J. A., & Sánchez-Baracaldo, P. (2021). Gloeobacter and the implications of a freshwater origin of Cyanobacteria. *Phycologia*, *60*(5), 402–418. https://doi.org/10.1080/00318884.2021.1881729

Roux, S., Adriaenssens, E. M., Dutilh, B. E., Koonin, E. V., Kropinski, A. M., Krupovic, M., Kuhn, J. H., Lavigne, R., Brister, J. R., Varsani, A., Amid, C., Aziz, R. K., Bordenstein, S. R., Bork, P., Breitbart, M., Cochrane, G. R., Daly, R. A., Desnues, C., Duhaime, M. B., … Eloe-Fadrosh, E. A. (2019). Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nature Biotechnology*, *37*(1), 29–37. https://doi.org/10.1038/nbt.4306

Roux, S., & Emerson, J. B. (2022). Diversity in the soil virosphere: To infinity and beyond? *Trends in Microbiology*, *30*(11), 1025–1035. https://doi.org/10.1016/j.tim.2022.05.003

Roux, S., Krupovic, M., Poulet, A., Debroas, D., & Enault, F. (2012). Evolution and Diversity of the Microviridae Viral Family through a Collection of 81 New Complete Genomes Assembled from Virome Reads. *PLoS ONE*, *7*(7), e40418. https://doi.org/10.1371/journal.pone.0040418

Roux, S., Solonenko, N. E., Dang, V. T., Poulos, B. T., Schwenck, S. M., Goldsmith, D. B., Coleman, M. L., Breitbart, M., & Sullivan, M. B. (2016). Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ*, *4*, e2777. https://doi.org/10.7717/peerj.2777

Salgado, O., Martínez, O., & Rivas, Y. (2025). Highly specific virus-host linkages in mesothermophilic and circumneutral hot springs microbial communities. *Research in Microbiology*, 104300. https://doi.org/10.1016/j.resmic.2025.104300

Saw, J. H., Shlafstein, M. D., Pavloudi, C., Monsalve, N., Prescott, R. D., Chain, P. S. G., Decho, A. W., & Donachie, S. P. (2026). *Amplicon and metagenomic data from fumarole-associated geothermal features of Hawai□i* (p. 2026.02.26.708227). bioRxiv. https://doi.org/10.64898/2026.02.26.708227

Saw, J. H. W., Schatz, M., Brown, M. V., Kunkel, D. D., Foster, J. S., Shick, H., Christensen, S., Hou, S., Wan, X., & Donachie, S. P. (2013). Cultivation and Complete Genome Sequencing of Gloeobacter kilaueensis sp. Nov., from a Lava Cave in Kīlauea Caldera, Hawai'i. *PLOS ONE*, *8*(10), e76376. https://doi.org/10.1371/journal.pone.0076376

Shaffer, M., Borton, M. A., McGivern, B. B., Zayed, A. A., La Rosa, S. L., Solden, L. M., Liu, P., Narrowe, A. B., Rodríguez-Ramos, J., Bolduc, B., Gazitúa, M. C., Daly, R. A., Smith, G. J., Vik, D. R., Pope, P. B., Sullivan, M. B., Roux, S., & Wrighton, K. C. (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Research*, *48*(16), 8883–8900. https://doi.org/10.1093/nar/gkaa621

Sharp, C. E., Brady, A. L., Sharp, G. H., Grasby, S. E., Stott, M. B., & Dunfield, P. F. (2014). Humboldt's spa: Microbial diversity is controlled by temperature in geothermal environments. *The ISME Journal*, *8*(6), 1166–1174. https://doi.org/10.1038/ismej.2013.237

Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K. V., & Koonin, E. V. (2018). Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(23), E5307–E5316. https://doi.org/10.1073/pnas.1803440115

Shmakov, S. A., Sitnik, V., Makarova, K. S., Wolf, Y. I., Severinov, K. V., & Koonin, E. V. (2017). The CRISPR Spacer Space Is Dominated by Sequences from Species-Specific Mobilomes. *mBio*, *8*(5), e01397-17. https://doi.org/10.1128/mBio.01397-17

Shuster, B., Khemmani, M., Nakaya, Y., Holland, G., Iwamoto, K., Abe, K., Imamura, D., Maryn, N., Driks, A., Sato, T., & Eichenberger, P. (2019). Expansion of the Spore Surface Polysaccharide Layer in Bacillus subtilis by Deletion of Genes Encoding Glycosyltransferases and Glucose Modification Enzymes. *Journal of Bacteriology*, *201*(19), e00321-19. https://doi.org/10.1128/JB.00321-19

Simmons, E. L., Bond, M. C., Koskella, B., Drescher, K., Bucci, V., & Nadell, C. D. (2020). Biofilm Structure Promotes Coexistence of Phage-Resistant and Phage-Susceptible Bacteria. *mSystems*, *5*(3), 10.1128/msystems.00877-19. https://doi.org/10.1128/msystems.00877-19

Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, *33*(suppl_2), W244–W248. https://doi.org/10.1093/nar/gki408

Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S. J., & Söding, J. (2019). HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics*, *20*, 473. https://doi.org/10.1186/s12859-019-3019-7

Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, *35*(11), 1026–1028. https://doi.org/10.1038/nbt.3988

Su, G., Morris, J. H., Demchak, B., & Bader, G. D. (2014). BIOLOGICAL NETWORK EXPLORATION WITH CYTOSCAPE 3. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]*, *47*, 8.13.1-8.13.24. https://doi.org/10.1002/0471250953.bi0813s47

Sun, M., Yuan, S., Xia, R., Ye, M., & Balcázar, J. L. (2023). Underexplored viral auxiliary metabolic genes in soil: Diversity and eco-evolutionary significance. *Environmental Microbiology*, *25*(4), 800–810. https://doi.org/10.1111/1462-2920.16329

Tan, C. C. S., van Dorp, L., & Balloux, F. (2024). The evolutionary drivers and correlates of viral host jumps. *Nature Ecology & Evolution*, *8*(5), 960–971. https://doi.org/10.1038/s41559-024-02353-4

ter Horst, A. M., Santos-Medellín, C., Sorensen, J. W., Zinke, L. A., Wilson, R. M., Johnston, E. R., Trubl, G., Pett-Ridge, J., Blazewicz, S. J., Hanson, P. J., Chanton, J. P., Schadt, C. W., Kostka, J. E., & Emerson, J. B. (2021). Minnesota peat viromes reveal terrestrial and aquatic niche partitioning for local and global viral populations. *Microbiome*, *9*, 233. https://doi.org/10.1186/s40168-021-01156-0

Teske, A., & Reysenbach, A.-L. (2015). Editorial: Hydrothermal microbial ecosystems. *Frontiers in Microbiology*, *6*, 884. https://doi.org/10.3389/fmicb.2015.00884

The UniProt Consortium. (2022). UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*(D1), D523–D531. https://doi.org/10.1093/nar/gkac1052

Thomas, E., Anderson, R. E., Li, V., Rogan, L. J., & Huber, J. A. (2021). Diverse Viruses in Deep-Sea Hydrothermal Vent Fluids Have Restricted Dispersal across Ocean Basins. *mSystems*, *6*(3), 10.1128/msystems.00068-21. https://doi.org/10.1128/msystems.00068-21

Tian, F., Wainaina, J. M., Howard-Varona, C., Domínguez-Huerta, G., Bolduc, B., Gazitúa, M. C., Smith, G., Gittrich, M. R., Zablocki, O., Cronin, D. R., Eveillard, D., Hallam, S. J., & Sullivan, M. B. (2024). Prokaryotic-virus-encoded auxiliary metabolic genes throughout the global oceans. *Microbiome*, *12*(1), 159. https://doi.org/10.1186/s40168-024-01876-z

Tokuda, M., & Shintani, M. (2024). Microbial evolution through horizontal gene transfer by mobile genetic elements. *Microbial Biotechnology*, *17*(1), e14408. https://doi.org/10.1111/1751-7915.14408

Trubl, G., Jang, H. B., Roux, S., Emerson, J. B., Solonenko, N., Vik, D. R., Solden, L., Ellenbogen, J., Runyon, A. T., Bolduc, B., Woodcroft, B. J., Saleska, S. R., Tyson, G. W., Wrighton, K. C., Sullivan, M. B., & Rich, V. I. (2018). Soil Viruses Are Underexplored Players in Ecosystem Carbon Processing. *mSystems*, *3*(5), e00076-18. https://doi.org/10.1128/mSystems.00076-18

Turner, D., Adriaenssens, E. M., Amann, R. I., Bardy, P., Bartlau, N., Barylski, J., Błażejak, S., Bouzari, M., Briegel, A., Briers, Y., Carrillo, D., Chen, X., Claessen, D., Cook, R., Crisci, M. A., Dechesne, A., Deptula, P., Dutilh, B. E., Ely, B., … ICTV Taxonomy Summary Consortium. (2025). Summary of taxonomy changes ratified by the International Committee on Taxonomy of Viruses (ICTV) from the Bacterial Viruses Subcommittee, 2025. *Journal of General Virology*, *106*(7), 002111. https://doi.org/10.1099/jgv.0.002111

Wall, K., Cornell, J., Bizzoco, R. W., & Kelley, S. T. (2015). Biodiversity hot spot on a hot spot: Novel extremophile diversity in Hawaiian fumaroles. *MicrobiologyOpen*, *4*(2), 267–281. https://doi.org/10.1002/mbo3.236

Wang, Y., Wang, C., Guan, Z., Cao, J., Xu, J., Wang, S., Cui, Y., Wang, Q., Chen, Y., Yin, Y., Zhang, D., Liu, H., Sun, M., Jin, S., Tao, P., & Zou, T. (2024). DNA methylation

activates retron Ec86 filaments for antiphage defense. *Cell Reports*, *43*(10). https://doi.org/10.1016/j.celrep.2024.114857

Weinheimer, A. R., Ha, A. D., & Aylward, F. O. (2025). Towards a unifying phylogenomic framework for tailed phages. *PLOS Genetics*, *21*(2), e1011595. https://doi.org/10.1371/journal.pgen.1011595

Wishart, D. S., Han, S., Saha, S., Oler, E., Peters, H., Grant, J. R., Stothard, P., & Gautam, V. (2023). PHASTEST: Faster than PHASTER, better than PHAST. *Nucleic Acids Research*, *51*(W1), W443–W450. https://doi.org/10.1093/nar/gkad382

Wu, R., Davison, M. R., Nelson, W. C., Smith, M. L., Lipton, M. S., Jansson, J. K., McClure, R. S., McDermott, J. E., & Hofmockel, K. S. (2023). Hi-C metagenome sequencing reveals soil phage–host interactions. *Nature Communications*, *14*, 7666. https://doi.org/10.1038/s41467-023-42967-z

Wu, Y., Wu, Z., Guo, L., Shao, J., Xiao, H., Yang, M., Deng, C., Zhang, Y., Zhang, Z., & Zhao, Y. (2024). Diversity and distribution of a prevalent Microviridae group across the global oceans. *Communications Biology*, *7*(1), 1377. https://doi.org/10.1038/s42003-024-07085-6

Xue, C., & Goldenfeld, N. (2017). Coevolution Maintains Diversity in the Stochastic ``Kill the Winner'' Model. *Physical Review Letters*, *119*(26), 268101. https://doi.org/10.1103/PhysRevLett.119.268101

Yao, J., Zeng, Y., Hong, X., Wang, M., Zhang, Q., Chen, Y., Gou, M., Xia, Z.-Y., & Tang, Y.-Q. (2025). Phages-bacteria interactions underlying the dynamics of polyhydroxyalkanoate-producing mixed microbial cultures via meta-omics study. *mSystems*, *10*(4), e00200-25. https://doi.org/10.1128/msystems.00200-25

Yoshida, M., Mochizuki, T., Urayama, S.-I., Yoshida-Takashima, Y., Nishi, S., Hirai, M., Nomaki, H., Takaki, Y., Nunoura, T., & Takai, K. (2018). Quantitative Viral Community DNA Analysis Reveals the Dominance of Single-Stranded DNA Viruses in Offshore Upper Bathyal Sediment from Tohoku, Japan. *Frontiers in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.00075

Zheng, X., Liu, W., Dai, X., Zhu, Y., Wang, J., Zhu, Y., Zheng, H., Huang, Y., Dong, Z., Du, W., Zhao, F., & Huang, L. (2021). Extraordinary diversity of viruses in deep-sea sediments as revealed by metagenomics without prior virion separation. *Environmental Microbiology*, *23*(2), 728–743. https://doi.org/10.1111/1462-2920.15154

Zhou, Z., St. John, E., Anantharaman, K., & Reysenbach, A.-L. (2022). Global patterns of diversity and metabolism of microbial communities in deep-sea hydrothermal vent deposits. *Microbiome*, *10*, 241. https://doi.org/10.1186/s40168-022-01424-7

Zhou, Z., Tran, P. Q., Martin, C., Rohwer, R. R., Baker, B. J., McMahon, K. D., & Anantharaman, K. (2025). Unravelling viral ecology and evolution over 20 years in a freshwater lake. *Nature Microbiology*, *10*(1), 231–245. https://doi.org/10.1038/s41564-024-01876-7

Zimmerman, A. E., Howard-Varona, C., Needham, D. M., John, S. G., Worden, A. Z., Sullivan, M. B., Waldbauer, J. R., & Coleman, M. L. (2020). Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nature Reviews Microbiology*, *18*(1), 21–34. https://doi.org/10.1038/s41579-019-0270-x

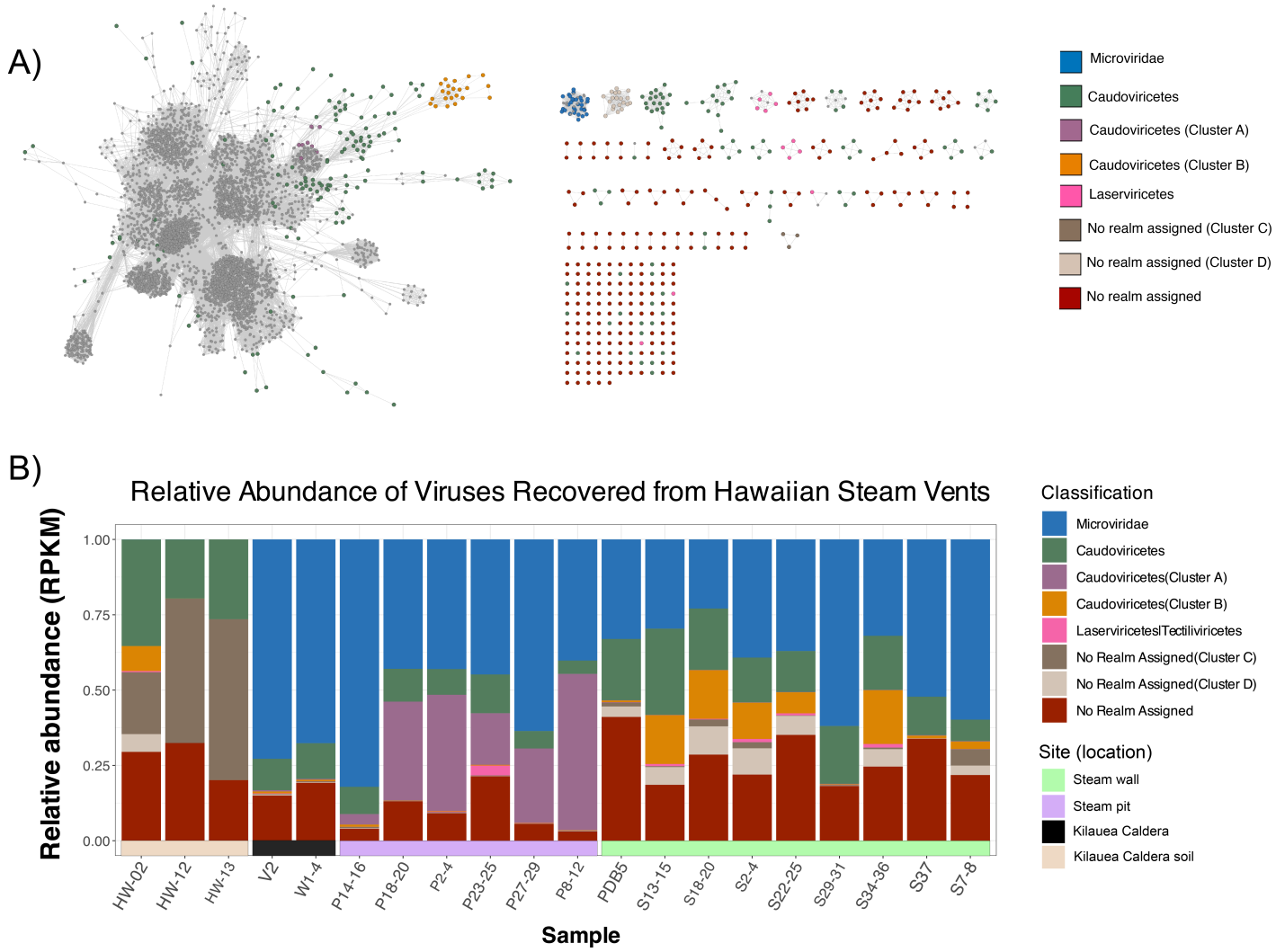


Figure 01. **A)** Gene-sharing network of confidently identified fumarole viral contigs visualized in Cytoscape. Nodes represent full viral sequences. Grey nodes are RefSeq viruses from the vConTACT3 v220 database, while colored nodes are fumarole viruses, grouped by vConTACT3 taxonomic predictions. Categories indicate the lowest confident taxonomic level assigned to each cluster; highly abundant clusters are colored independently of taxonomy. Edges represent the strength of gene-sharing between viruses, determined with protein clusters. All shown clusters have high clustering coefficients (>0.85), as calculated using the Cytoscape NetworkAnalyzer plugin. **B)** Relative abundance of identified viruses in all 20 samples, colored by cluster classification. Relative abundance (RPKM) was calculated for each individual viral contig and grouped by the taxonomic designation of clusters for visualization. Clusters found in notable abundance were colored and denoted independently, regardless of predicted taxonomy. Relative abundance of vOTUs used for statistical analysis is available as Supplemental Material S6.

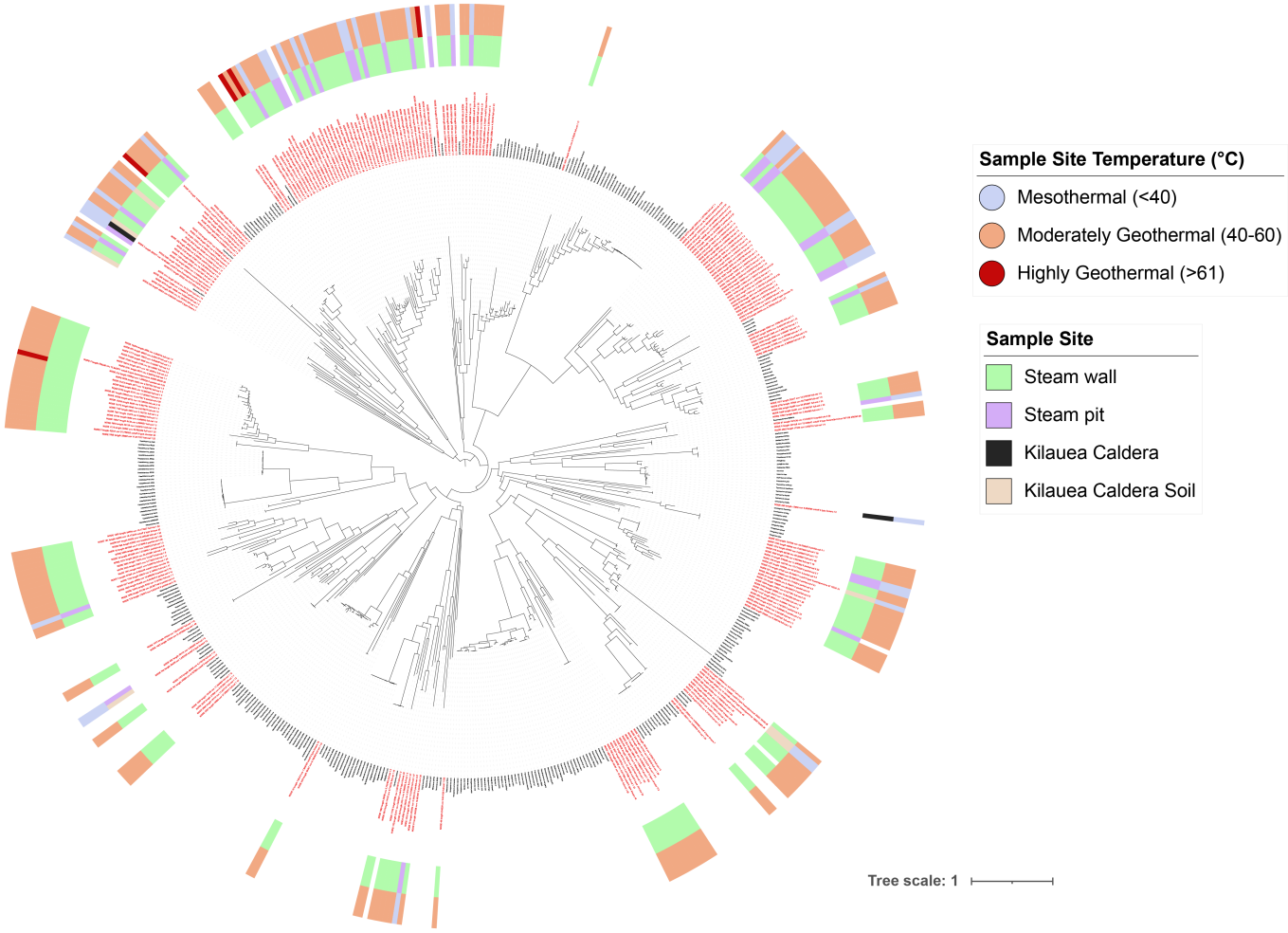


Figure 02. Maximum likelihood phylogenetic tree of Caudoviricetes terminase large subunit (TerL) amino acid gene sequences. Red node labels correspond to TerL sequences annotated in fumarole viral contigs. Black node labels correspond to RefSeq taxon which were determined to be first neighbors of fumarole viruses with Cytoscape plugin NetworkAnalyzer. Outmost ring refers to the temperature range of the sample from which the virus was metagenomically assembled. The middle ring refers to the sample site from which the corresponding sample was taken. All annotations of TerL in RefSeq and fumarole viruses were made with HHblits using parameters described in Methods.

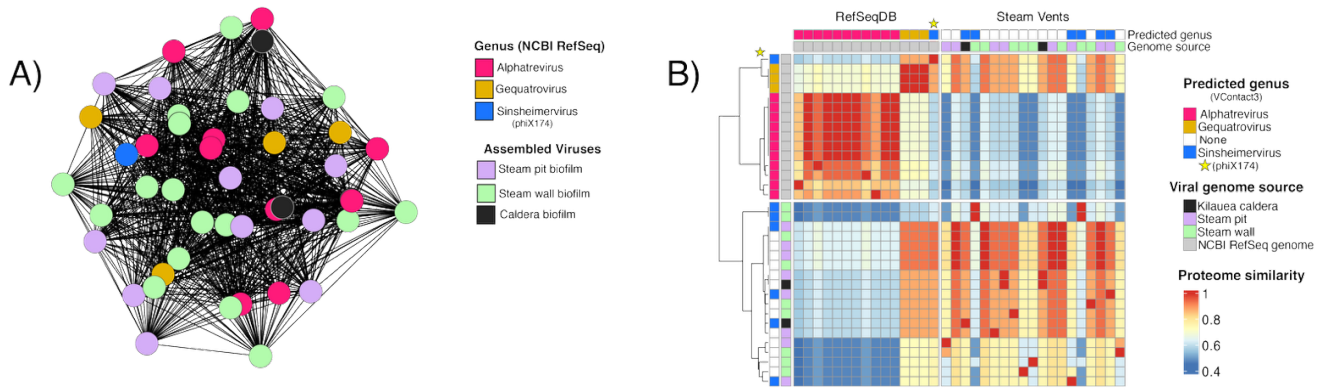


Figure 03. A) Subcluster of viral genomes within the family Microviridae. Network constructed using vConTACT3 and visualized in Cytoscape. Reference nodes are colored by genus (RefSeq and vConTACT3 predictions), while fumarole viruses are colored by sample site. Edges represent gene-sharing strength. The blue node (phiX174) forms a weakly resolved clade with fumarole viruses in the maximum-likelihood phylogeny based on major capsid protein amino acid sequences. **B)** Pairwise genome comparisons of RefSeq and fumarole Microviridae using local alignment (MMseqs2). The top row and left column indicate predicted genus assignments (vConTACT3) for fumarole and RefSeq viruses. The top-left quadrant shows RefSeq–RefSeq comparisons; the top-right shows fumarole–RefSeq comparisons (RefSeq colored by genus). The second row (genome source) is colored by fumarole sample site. The star denotes comparisons involving the phiX174 RefSeq genome.

Alpha Diversity of Fumarole Viruses

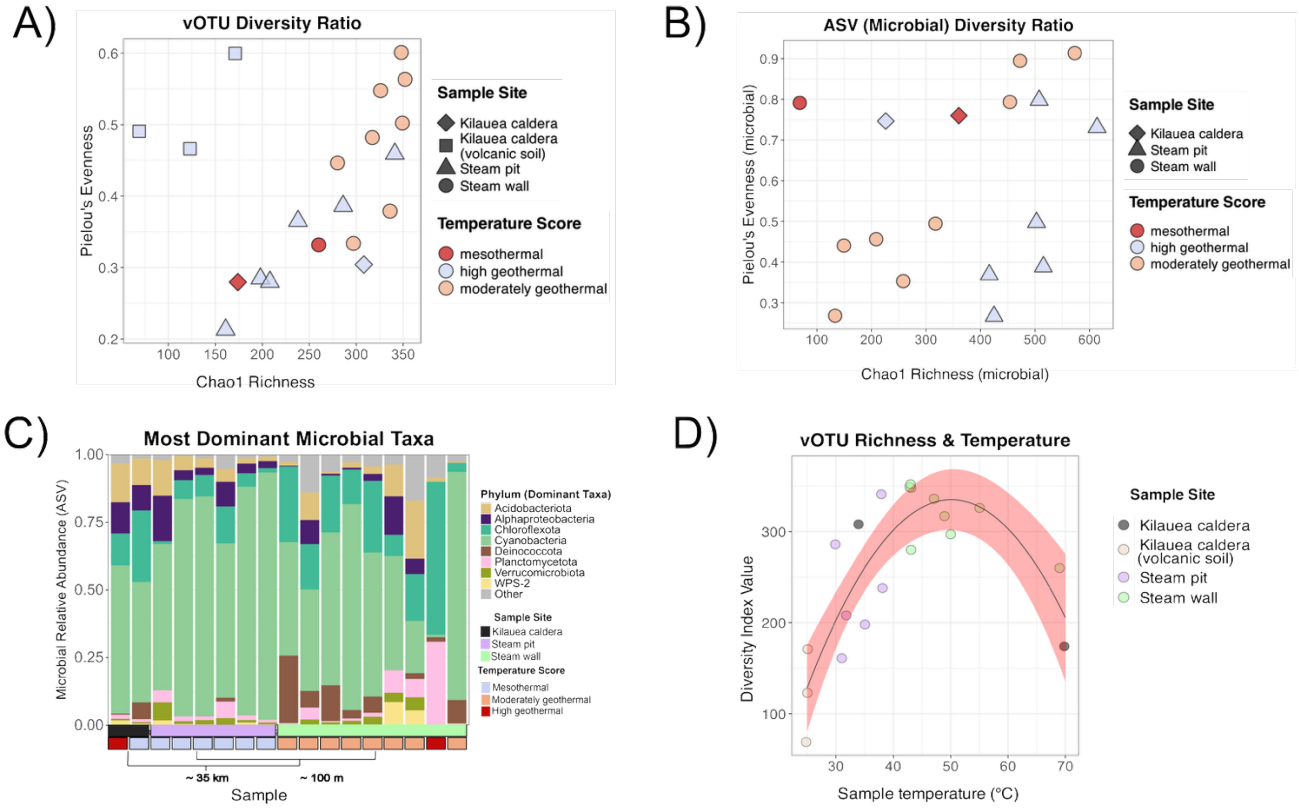


Figure 04. A) Richness (Chao1) and evenness (Pielou) of fumarole vOTUs calculated from RPKM-based relative abundances (Supplemental Material S6). Each point represents a sample. Shape indicates sample site, and color indicates temperature at time of collection. **B)** Richness and evenness of microbial ASV relative abundances. Each point represents a sample; color and shape denote sample characteristics as described in (A). **C)** Relative abundance of the most abundant bacterial taxa across samples based on ASVs. The top five phyla per sample are shown and used for visualization. The second row indicates sample site, and the bottom row indicates temperature at collection. The phylogram shows spatial relationships among sites: the steam pit and steam wall are ~100 m apart, while the Kīlauea caldera cave is ~35 km from both. **D)** Sample temperature versus vOTU richness (Chao1). Each point represents a sample, colored by site. vOTU richness follows a temperature-dependent pattern modeled by a polynomial (quadratic) regression. Richness increases with temperature up to ~50°C and declines thereafter.

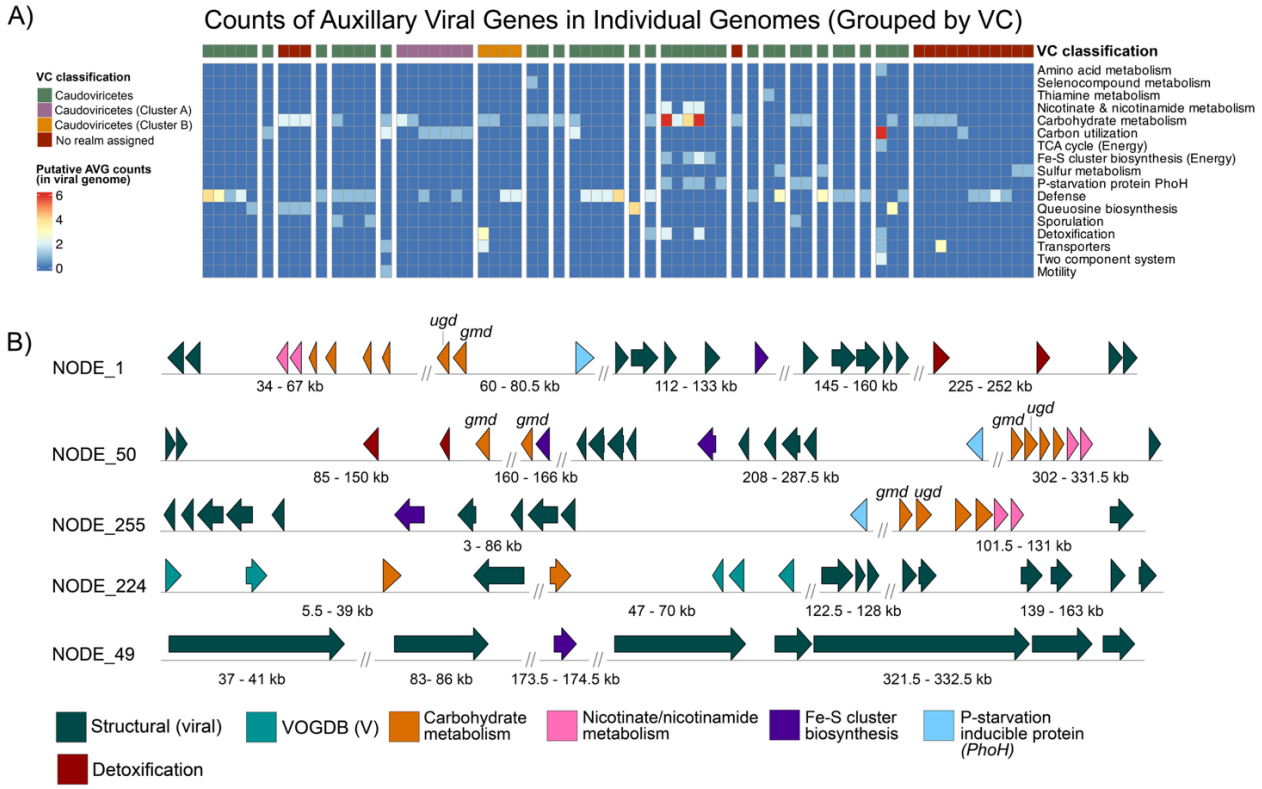
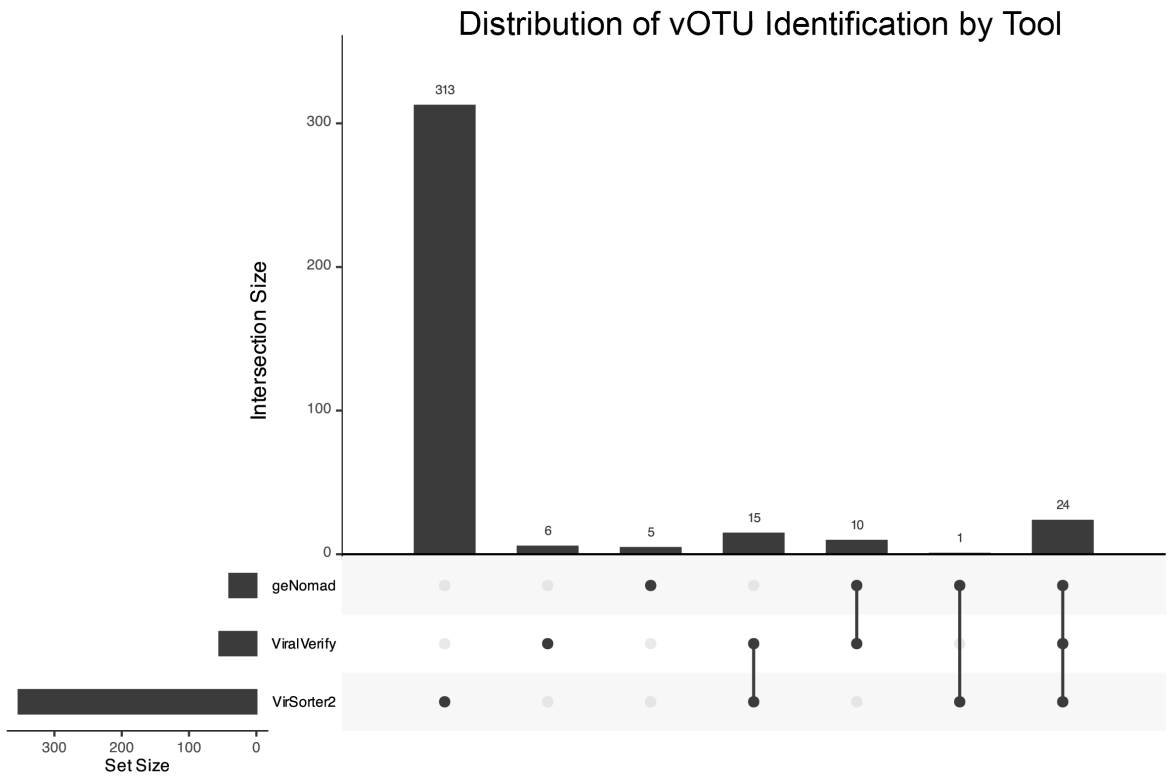
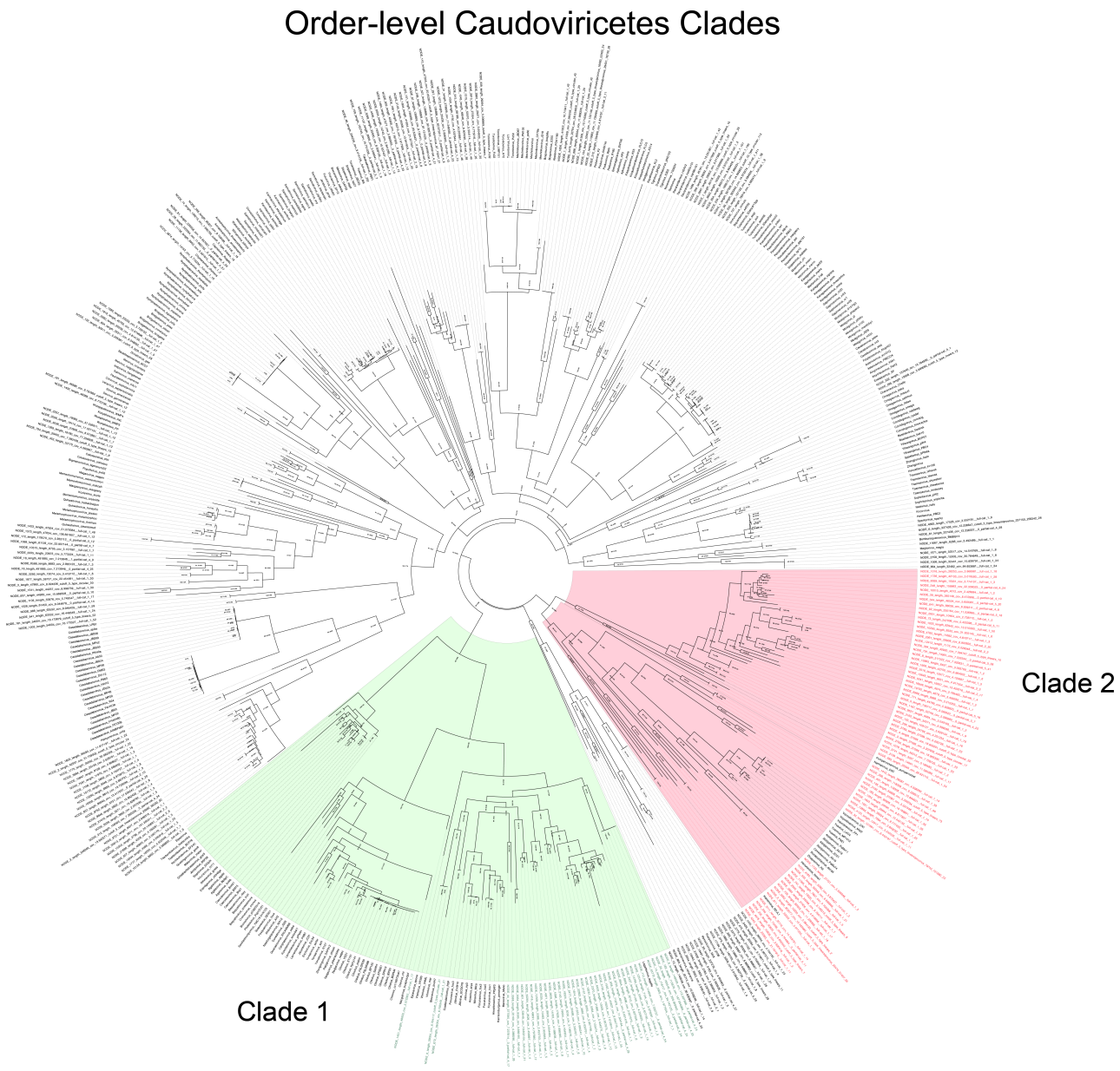


Figure 05. A) Counts of viral accessory genes in individual viral genomes. Columns represent viral metabolic gene counts per contig. The top row indicates vConTACT3 viral cluster (VC) assignments, with spacing grouping viruses within each VC; singletons are shown as individual VCs. Heatmap colors denote gene counts as shown in the legend. **B)** Gene diagrams of viral metabolic gene distribution across five recovered genomes from vOTU 158, enriched in putative AMGs. Genomes were inspected for flanking viral genes and AMG annotation coverage/identity. Arrows denote genes, and colors indicate AMG categories. The full curated AMG list is provided in Supplemental Material S8.

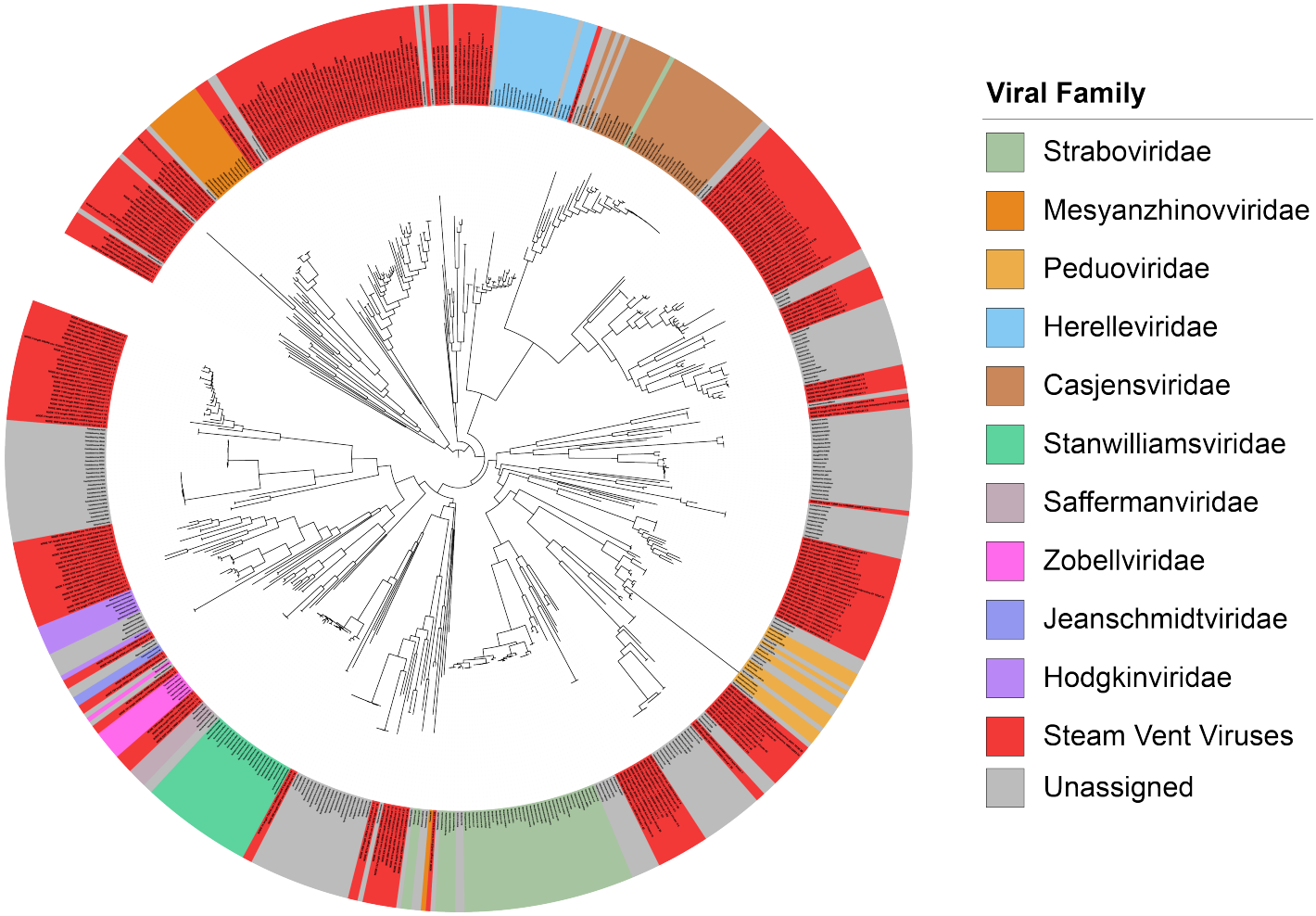
Supplemental Figures



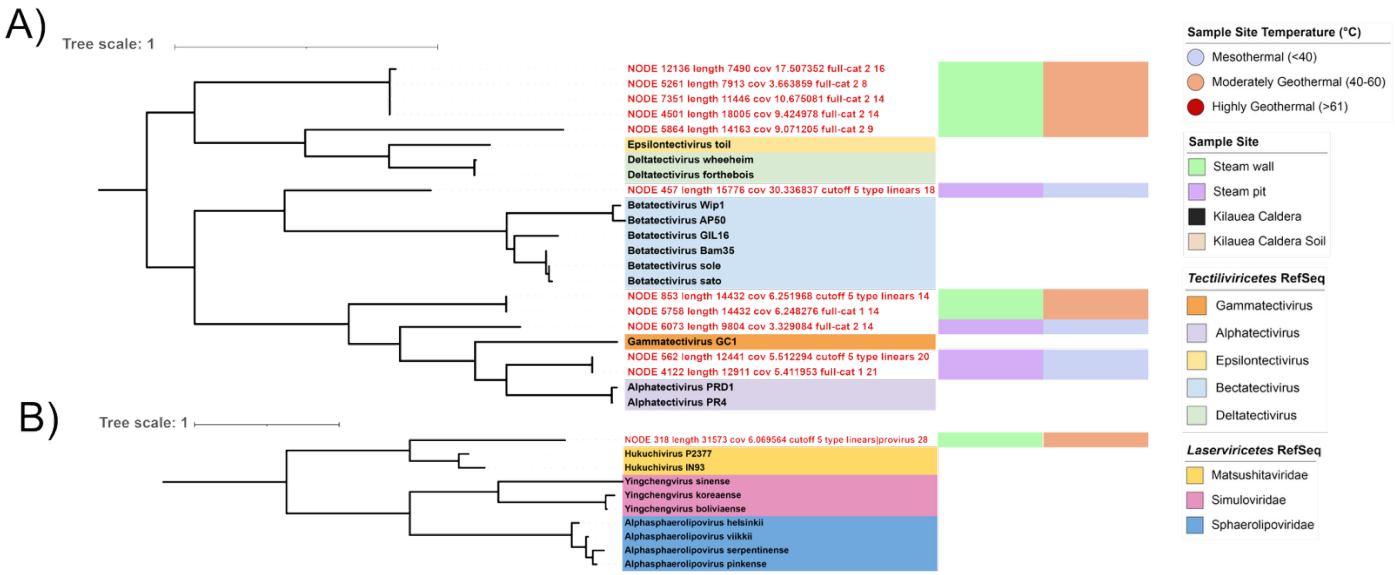
Supplemental Figure 01. Upset plot of vOTUs identified by each tool. Bars (with counts above) show the number of vOTUs shared across tool combinations indicated by connected, filled points below. The left panel shows total vOTUs identified per tool.



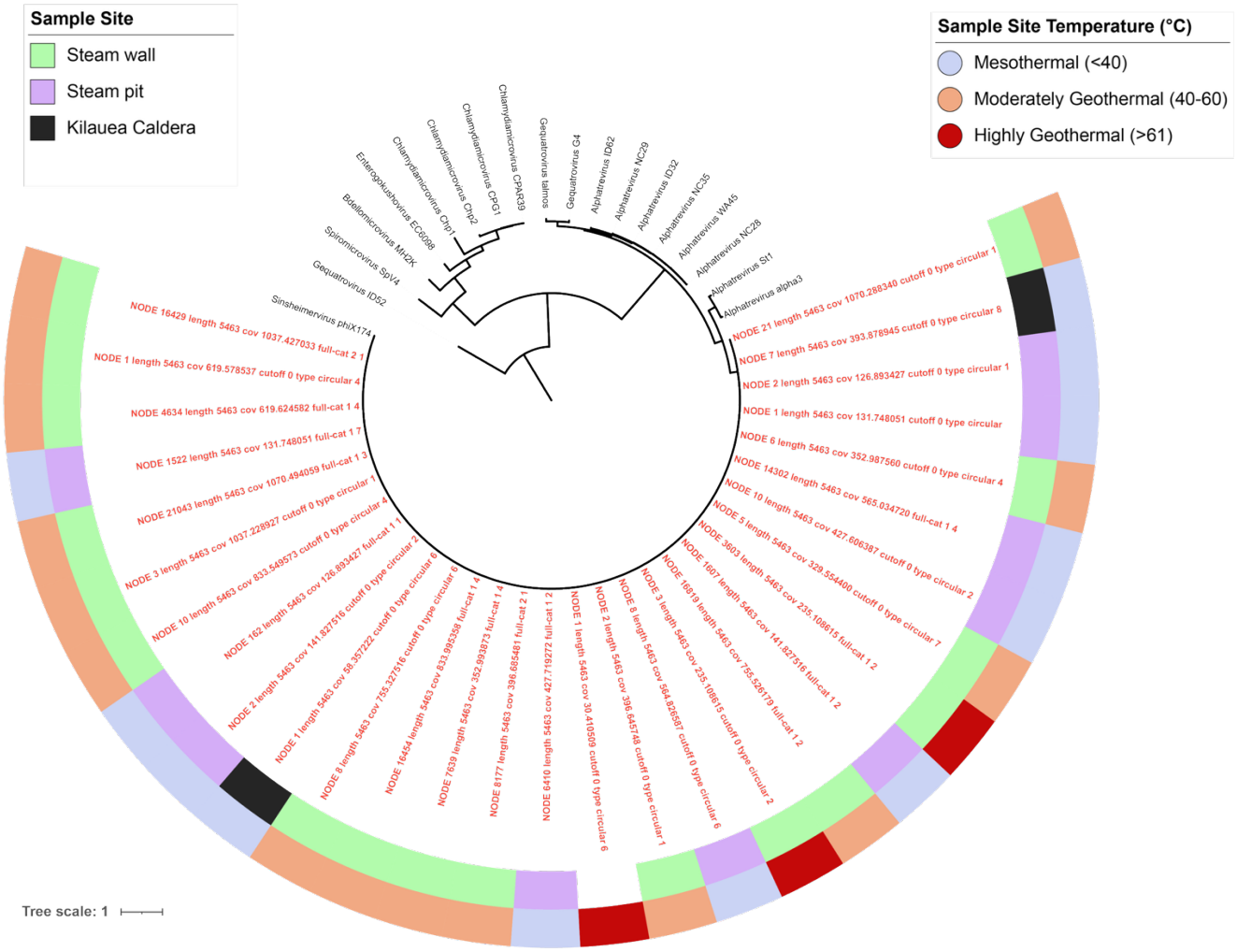
Supplemental Figure 02. Maximum likelihood tree of Caudoviricetes (TerL) amino acid sequences showing two previously undescribed order-level clades. Clade 1 (Kilaueavirales) is highlighted in green, with green taxon labels indicating fumarole-derived viruses and black taxon labels representing RefSeq viruses (first neighbors). Clade 2 (Pahoavirales) is highlighted in red, with red taxon labels indicating fumarole-derived viruses and black taxon labels representing RefSeq viruses. Node support values are shown and available in the IQ-TREE files on Zenodo.



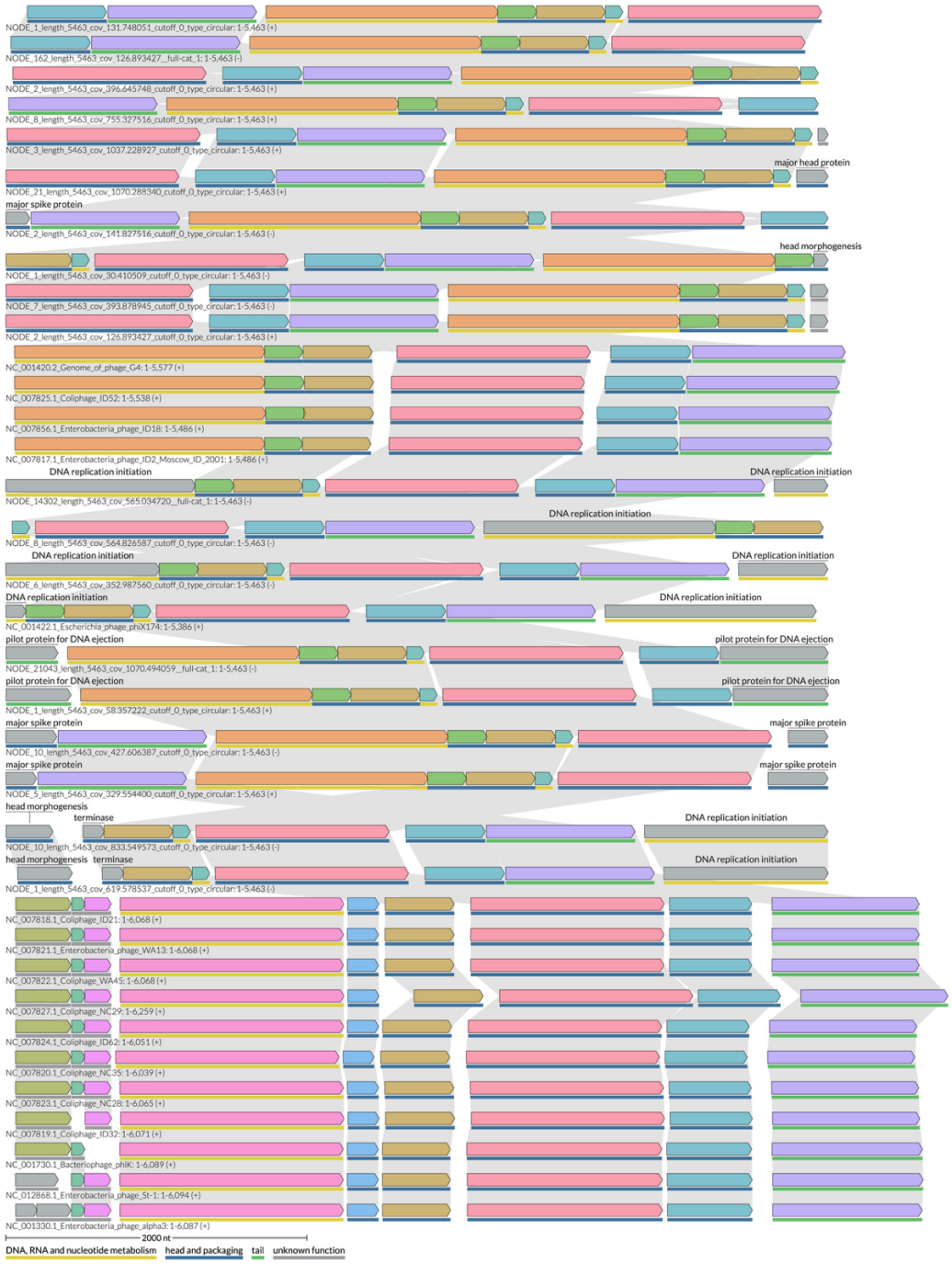
Supplemental Figure 03. Maximum likelihood tree of Caudoviricetes (TerL) amino acid sequences showing relationships between fumarole viruses and previously described viral families. Node label colors (outer ring) indicate viral family classification (many lacking order-level assignment); grey denotes unassigned viruses, and red denotes steam vent virus TerL. Clade 1 (Kilaueavirales) groups with RefSeq genomes from two families lacking prior order assignment—Casjensviridae and Herelleviridae—and includes 30 fumarole vOTUs and 17 unassigned RefSeq viruses. Clade 2 (Pahoavirales) includes 56 fumarole vOTUs and places Mesyanzhinovviridae as more recently diverged relative to some fumarole viruses. Additional fumarole viruses show close phylogenetic similarity to incertae sedis families, including Peduoviridae, Zobellviridae, and Saffermanviridae.



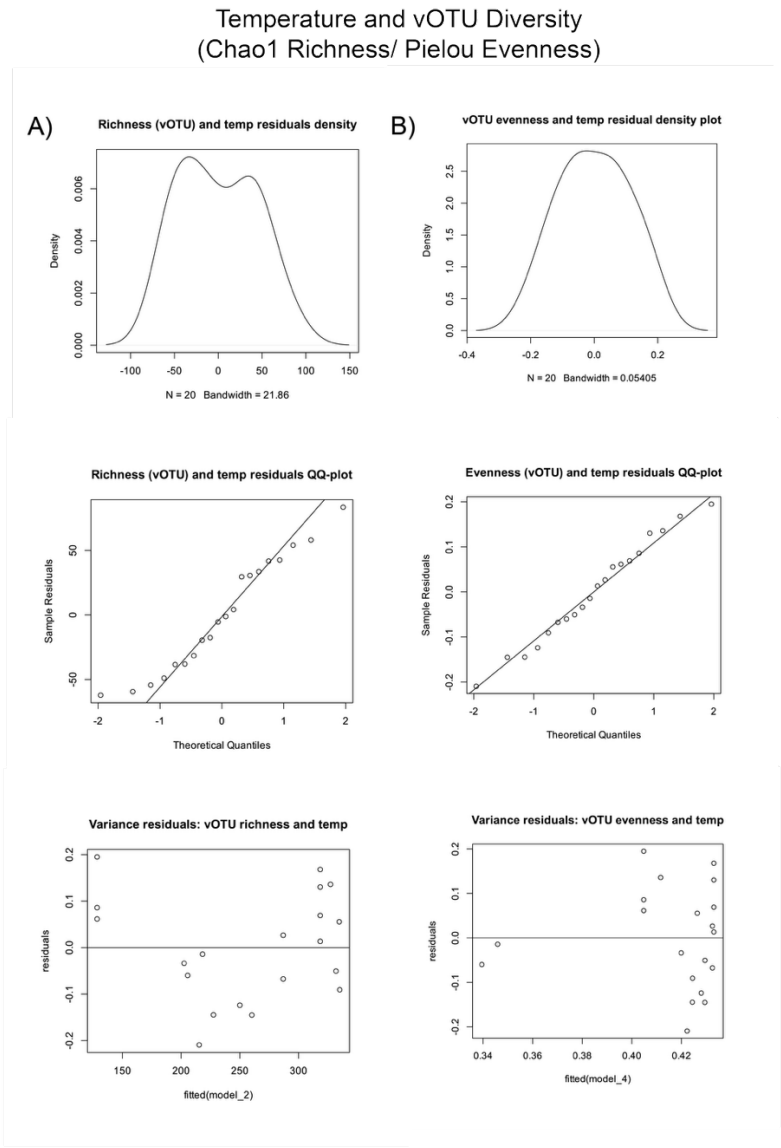
Supplemental Figure 04. A) Maximum likelihood class-level tree of Tectiviricetes packaging ATPase amino acid sequences (annotated with HHblits). Red labels indicate fumarole-derived viruses, while Tectiviricetes RefSeq viruses are highlighted with genus-specific colors as described in the figure legend. Alphatectivirus is the only genus within Tectiviricetes containing plasmid-dependent phages, and all five genera are included as references. **B)** Maximum likelihood class-level tree of Laserviricetes packaging ATPase amino acid sequences (annotated with HHblits). RefSeq viruses are colored by genus as described in the figure legend, and the red label denotes a fumarole-derived virus closely related to Hukuchivirus, a genus of thermophilic Thermus phages.



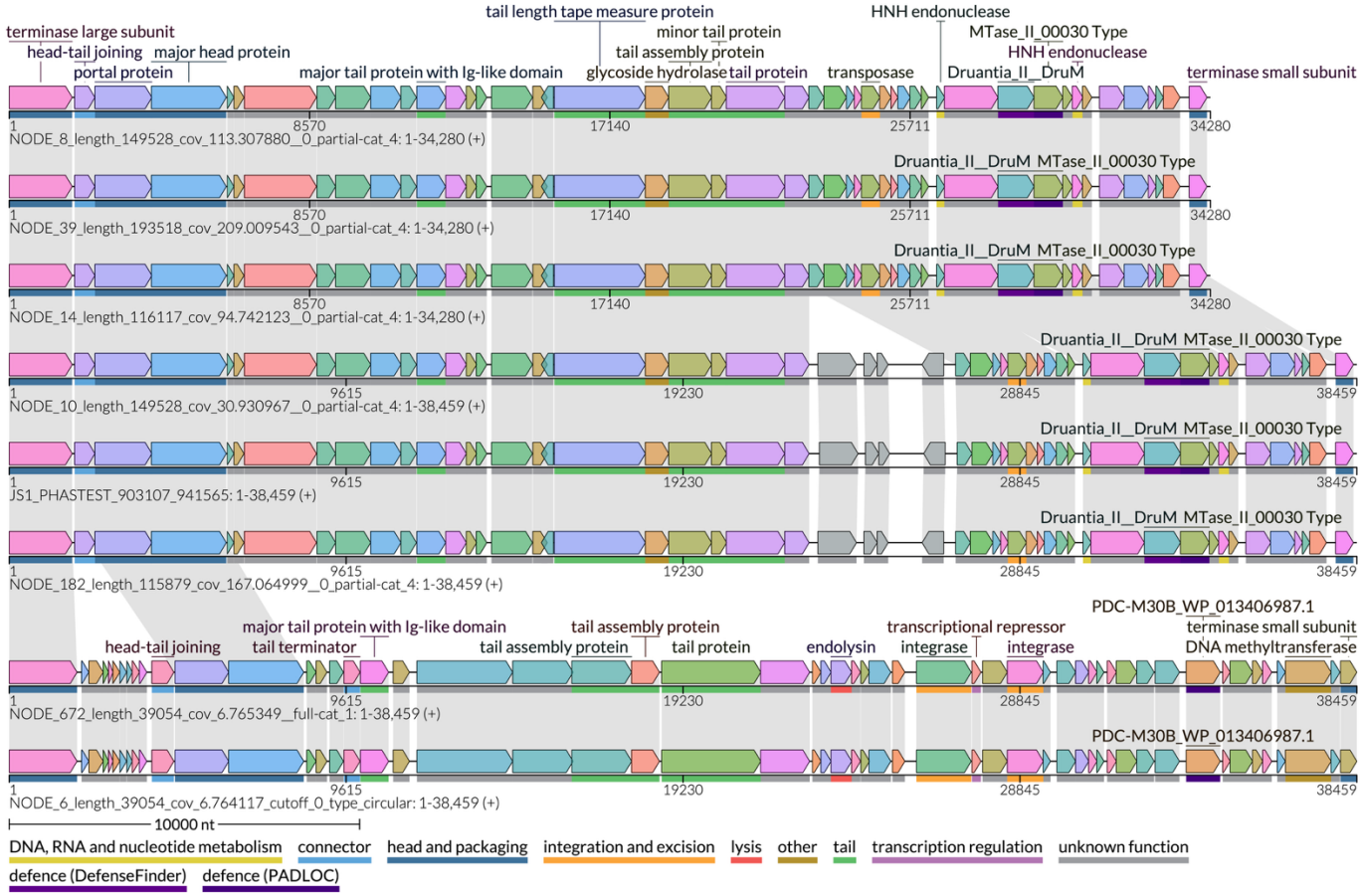
Supplemental Figure 05. Maximum likelihood class-level tree of Malgrandaviricetes major capsid protein (MCP) amino acid sequences. RefSeq viruses from a shared cluster in the gene-sharing network (whole genomes) were used as references for phylogenetic estimation. Red labels indicate fumarole viruses. The outer ring denotes the temperature range of the source sample, and the middle ring indicates the corresponding sampling site. Fumarole viruses form a poorly resolved clade with phiX174.



Supplemental Figure 06. Whole-genome comparative map of Microviridae viruses within the Bullavirinae (phiX-like) subfamily. Genome names are shown to the left below each gene map, with corresponding NCBI accessions when available. Genes are depicted as blocks, with shared colors indicating conserved proteins. Conserved and variable genes are identified using a pairwise proteome composition distance matrix (LoVis4u), and syntenic rearrangements are indicated by linkages between genomes. The bar beneath each genome shows functional annotations from LoVis4u based on local alignments (MMSeq2). Compared to RefSeq Bullavirinae (phiX-like viruses), fumarole Microviridae display greater syntenic variation and distinct genomic differences both from RefSeq genomes and among themselves.



Supplemental Figure 07. A) Diagnostic plots for the non-linear regression of temperature and vOTU richness (Chao1), including density, QQ, and variance plots of residuals. Residuals show a bimodal distribution in the density and QQ-plots, with constant variance. This supports the use of ordinary least squares (OLS) regression to assess the relationship between temperature and vOTU richness. B) Diagnostic plots for the non-linear regression of temperature and vOTU evenness (Pielou), including density, QQ, and variance plots of residuals. Residuals appear normally distributed with heteroscedasticity, indicating non-constant variance; thus, least squares regression is not sufficient to model the relationship between temperature and vOTU evenness.



Supplemental Figure 08. Comparative genomic map of three fumarole vOTUs within a vConTACT3 genus-level cluster. These vOTUs were detected in high abundance only at the steam pit site; two vOTUs show CRISPR-spacer matches to Gloeobacter kilaueensis, representing the first report of a virus infecting this genus. Genome names are shown to the left below each map, with NCBI accessions where available. Genes are depicted as blocks with shared colors indicating conserved proteins. Conserved and variable genes were identified using a pairwise proteome composition distance matrix (LoVis4u), with syntenic rearrangements indicated by linkages. The bar beneath each genome shows functional annotations from LoVis4u based on local alignments (MMSeq2).