**ARTICLE**    OPEN

# Genomics discovery of giant fungal viruses from subsurface oceanic crustal fluids

Ananda S. Bhattacharjee[1,3 ✉], Frederik Schulz[2], Tanja Woyke [ID][2], Beth N. Orcutt [ID][1] and Joaquín Martínez Martínez [ID][1 ✉]

© The Author(s) 2023

The oceanic igneous crust is a vast reservoir for microbial life, dominated by diverse and active bacteria, archaea, and fungi. Archaeal and bacterial viruses were previously detected in oceanic crustal fluids at the Juan de Fuca Ridge (JdFR). Here we report the discovery of two eukaryotic Nucleocytoviricota genomes from the same crustal fluids by sorting and sequencing single virions. Both genomes have a tRNA$^{Tyr}$ gene with an intron (20 bps) at the canonical position between nucleotide 37 and 38, a common feature in eukaryotic and archaeal tRNA genes with short introns (<100 bps), and fungal genes acquired through horizontal gene transfer (HGT) events. The dominance of *Ascomycota* fungi as the main eukaryotes in crustal fluids and the evidence for HGT point to these fungi as the putative hosts, making these the first putative fungi-Nucleocytoviricota specific association. Our study suggests active host-viral dynamics for the only eukaryotic group found in the subsurface oceanic crust and raises important questions about the impact of viral infection on the productivity and biogeochemical cycling in this ecosystem.

## INTRODUCTION

The oceanic igneous crust harbors the largest hydrologically active aquifer on Earth, and it is a site of dynamic exchange of fluids, chemical species, and microorganisms with the oceanic basin [1]. Phylogenetically and functionally diverse bacterial, archaeal, viral, and fungal taxa have been described in the oceanic basement [1]. Viruses are active participants in the ecology and evolution of microbial life [2]; yet, little is known about their diversity and activity in the marine crustal deep biosphere, with the main exception of a study revealing a diverse virus community in subsurface crustal fluids collected from the eastern flank of the Juan de Fuca Ridge (JdFR) [3]. These fluids were collected in 2011, 2013, and 2014 from borehole observatories installed in 2010 in Integrated Ocean Drilling Program (IODP) Holes U1362A (47°45.6628′N, 127°45.6720′W) and U1362B (47°45.4997′N 127°45.7312′W) [4]. The fluid samples were collected from 436 and 279 m below the seafloor, respectively [3]. Transmission electron microscopy and sequencing of microbial metagenomes obtained from tens to thousands of liters of fluid indicated that archaeon-infecting viruses and bacteriophages comprised the bulk of the viral community [3]. In addition, based on taxonomic affiliation, a small fraction of Nucleocytoviricota (Nucleocytoplasmic large DNA virus, NCLDV) genes were found within assembled scaffolds, but those genes were interpreted as belonging to archaeal viruses [3].

Nucleocytoviricota comprise a monophyletic group of double-stranded DNA viruses (70 kbp to 2.50 Mbp) [5] that infect a broad spectrum of unicellular and multicellular eukaryotes [6, 7] in many different habitats [8]. While their diversity is enormous [9, 10], they remain largely underexplored and poorly understood. Recent advances in metagenomics and targeted flow cytometry sorting and sequencing methods have improved the ability to capture and understand the genetic makeup of NCLDVs in environmental samples [9–18]. In this study, we sorted and sequenced the genome of single virus-like particles, including two NCLDVs, from a one-milliliter sample of crustal fluid collected in 2011 from the JdFR Hole U1362B. Here we report the taxonomic affiliations, and genomic characterization of those NCLDVS, and provide evidence of their ecology and interactions with putative hosts.

## MATERIALS AND METHODS

### Sample site and sampling

The oceanic basement fluid sample for this study was collected in July 2011 during cruise AT18-07 aboard the R/V *Atlantis* using ROV *Jason* II (cruise report available at: http://www.darkenergybiosphere.org/research/juandefuca.html). On dive J2-569, hydrothermal fluid samples were obtained from the Integrated Ocean Drilling Program (IODP) borehole U1362B (47°45.4997′N 127°45.7312′W) [19], located at an ocean depth of ~2650 m on the eastern flank of Juan de Fuca Ridge (JdFR) in the northeastern Pacific Ocean [20–22]. The sample was collected using a custom syringe sampler fired in the effluent of a free-flow chimney connected to the ball valve of the CORK (Circulation Obviation Retrofit Kit) well-head. The fluids originated from 359 m below the seafloor (~240 m of ocean sediments and 117 m into the basaltic basement) [3]. Upon recovery, a one-milliliter fluid aliquot was amended with 5% glycerol and 1 × TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) (final concentrations) and stored at −80 °C until further processing for fluorescence-activated particle sorting.

### Virus single-amplified genome (vSAG) generation and sequencing

After thawing on ice, the sample was diluted 100-fold with sterile 0.2 µm-filtered 1×TE buffer and subsequently stained with SYBR Green I (Life

[1]Bigelow Laboratory for Ocean Sciences, 60 Bigelow Drive, East Boothbay, ME, USA. [2]US Department of Energy Joint Genome Institute, Berkeley, CA, USA. [3]Present address: Department of Environmental Sciences, University of California Riverside, Riverside, CA, USA. ✉email: ananda.bhattacharjee@usda.gov; jmartinez@bigelow.org

Technologies, CA, USA), which has greater sensitivity for dsDNA, as described elsewhere [23]. The stained virus-like particles (VLPs), bacteria, and archaea were visualized by flow cytometry (Fig. S1A). Individual VLPs were sorted into one-half of a 384-well plate and bacteria/archaea cells in the other half. The plate layout included no-sort control wells (Fig. S1B). Sorting was performed at the Center for Aquatic Cytometry, Bigelow Laboratory for Ocean Sciences, with a BD Influx sorter following previously described methods [13].

Each well on the plate received 0.4 M KOH (final concentration) and was incubated at room temperature for 10 min to lyse the particles and render the gDNA available for amplification using Multiple Displacement Amplification (MDA) [24] (Fig. S1B). The MDA products from the whole plate were screened by PCR with primers for both bacterial and archaeal 16S rRNA genes as previously described [25] to confirm whether they derived from sorted cells or carry over cellular contamination (Fig. S1B). This work was carried out at the Single Cell Genomics Center, Bigelow Laboratory for Ocean Sciences. The 16S rRNA gene amplicons were used for phylogenetic analysis (Fig. S2).

Of the 27 putative vSAGs (i.e., successful MDA and no PCR-detectable 16S rRNA genes), eight were randomly selected for sequencing (vSAG1-8.JdFR). Nextera XT (Illumina, CA, USA) DNA libraries with a target size of 300 bp paired-end were prepared and sequenced on a MiSeq Benchtop sequencer (Illumina, CA, USA) at the Joint Genome Institute. Libraries from two aliquots of the same MDA amplicon were independently prepared and sequenced for vSAG1-8.JdFR.

## Sequence reads processing, assembly, and annotation

The Illumina raw reads (300 bps PE) were quality-trimmed with Trimmomatic v0.32 [26] using the following parameters: phred33 LEADING:0 TRAILING:5 SLIDINGWINDOW:4:15 MINLEN:36. Low complexity reads were removed using bbduk.sh from the BBTools suite [27] with a threshold of 0.5 (using entropy). Repairing of disordered quality filtered (QF) paired-end files was performed using repair.sh from BBTools suite. The QF reads for vSAGs were assembled using SPAdes v3.1.11 [28, 29] using the following parameters: --careful --sc --phred-offset 33 for the vSAGs. The scaffolds of at least 1000 bp were kept for further analyses. Six scaffolds (2.3–15.3Kbp) with 56–99% nucleotide level similarity present in all eight libraries were cautiously considered contamination and removed from further analysis. The contamination sequences are of bacterial and archaeal origin and are provided as supplementary data (https://github.com/asbhattacharjee/NCLDV-from-Juan-de-Fuca-Ridge-flank/blob/main/pre_annotation_containmination_files/Pre_annotation_contamination%20sequence_list.xlsx).

The vSAGs' coding sequences (CDS) were predicted using GeneMarkS [30] and their phyletic affiliations were determined through annotation using blastp (version 2.8.0+) [31] against the NCBI-NR database with e-value cut-off $10^{-10}$ (February 2021). The analysis was supplemented by blastp (version 2.8.0+) [31] analysis against NCBI Viral RefSeq (February 2021) and the reference viral database (RVDB) [32]. This analysis identified six of the vSAGs as bacteriophages. The remaining two, vSAG1.JdFR and vSAG8.JdFR, were identified as eukaryotic viruses. Only the latter two, vSAG1.JdFR and vSAG8.JdFR were retained for further analysis as part of this study. The vSAGs were further investigated for possible contaminations. We identified eleven scaffolds (1-17 kbp) in vSAG1.JdFR and vSAG8.JdFR libraries that were 100% identical at the nucleotide level to MarineAlpha5_Bin5 genes and had no viral genes. Those scaffolds were conservatively marked as putative contamination and removed from further analyses. All putative contamination sequences are made available as supplementary data (https://github.com/asbhattacharjee/NCLDV-from-Juan-de-Fuca-Ridge-flank/tree/main/post_annotation_contamination_files). The source of the putative contamination was not clear and was not further investigated in this study, although chemical analysis of the water sample and post-MDA screening with 16S rRNA suggested no contamination prior to sequencing. Post removal of putative contamination, a tBlastx [33] based alignment (nucleotide) of the vSAG1.JdFR and vSAG8.JdFR draft genomes was performed with the EasyFig [34] tool.

## vSAGs phylogenetic analysis

To infer the NCLDV species tree, hmmsearch [35] (version 3.1b2, hmmer.org) was used to identify five core NCVOGs NCLDV proteins [36, 37] using specific models (PMID: 28386012) for (i) DNA polymerase elongation subunit family B (NCVOG0038); (ii) D5-like helicase-primase (NCVOG0023); (iii) packaging ATPase (NCVOG0249); (iv) Poxvirus Late Transcription Factor VLTF3-like (NCVOG0262); and (v) DNA or RNA helicases of superfamily II (NCVOG0076). Proteins were aligned with mafft

v7.294b [38], and positions with less than 10% of information were removed from the alignment with trimAl version1.4 [39]. Proteins were then concatenated, and a tree was built with IQ-Tree 1.6.6 [40] and FastTree [41] using the best-fit highest-scoring evolutionary model based on the ModelFinder [42] feature in IQ-Tree data: LG + F + R86. In addition to vSAG1.JdFR and vSAG8.JdFR, the tree included 117 public NCLDV genomes (https://github.com/asbhattacharjee/NCLDV-from-Juan-de-Fuca-Ridge-flank/tree/main/species_tree) which were pre-clustered at ANI 95% (alignment fraction >70%) with FastANI 1.0 [43].

We used blastp (version 2.8.0+) [31, 33] against the NCBI-NR database for the phylogenetic analysis of the vSAGs' translational component eukaryotic initiation factor 4E (eIf-4E), a common gene among Mesomimiviridae members. The ten top hits were extracted and aligned with mafft v7.294b [38] and the gaps in the alignments were removed with alignment with trimAl version1.4 [39]. The phylogenetic tree of the trimmed, concatenated proteins was constructed with PhyML [44], which includes automatic selection of the best-fit substitution model, LG + G + F (Fig. S4).

Orthofinder 1.03 [45] was used to infer clusters of orthologous genes (COGs) on a representative dataset of 215 NCLDV genomes for comparative analysis. The presence and absence of genes in COGs, genes shared between NCLDV lineages, vSAG1.JdFR and vSAG8.JdFR were computed to understand the evolutionary relatedness of novel JdFR giant viruses. A custom python script [18] that includes the packages matplotlib [46] and UpSet [47] was used to visualize the genes shared between (a) different NCLDVs; and (b) within vSAGs. The analysis files are provided as supplementary data (https://github.com/asbhattacharjee/NCLDV-from-Juan-de-Fuca-Ridge-flank/tree/main/orthogroups).

## tRNA prediction and phylogeny

The tRNA gene predictions and their structures for vSAG1.JdFR and vSAG8.JdFR and other mesomimiviruses (Table S1) were performed using tRNAscan-SE [48] (parameters: mixed tRNA model) and ARAGORN v1.2.38 [49] (parameters: Type tRNA and tmRNA; Allow introns, 0-3000 bases; linear sequence topology; both single and double-strand) with default search mode settings. The predicted NCLDV Tyrosine tRNAs were analyzed with blastn (version 2.8.0+) [31] against the NCBI-NR database to determine their phylogeny. The top hits (a total 37) were used for phylogenetic analysis. The tRNA sequences were aligned on MEGA X [50] with MUSCLE [51] and trimmed with trimmed with trimal version1.4 [39]. Phylogenetic position of the NCLDV tRNAs with introns at canonical positions was determined with a maximum likelihood tree, bootstrap value of 100.

## Horizontal gene transfer analysis

Top 100 blastn vs nr hits of the blastp [31, 33] against the NCBI-NR database were used as references for phylogenetic analysis of translation system proteins to identify potential horizontally transferred genes. The translation system proteins were aligned with mafft-linsi v7.294b [38] and positions with less than 10% of information were trimmed with trimal 1.4 [39] >90% gaps. The trees were then built with IQ-tree 1.6.6 [40] with the mixture model C40 + Rx (best fit model based on a model test [42]).

## Phyletic affiliations

The phyletic affiliations of the genes from Juan de Fuca Ridge metagenomes from the IODP boreholes U1362A and U1362B [3] published on Integrated Microbial Genomes and Microbiomes (IMG/M) [52, 53] were determined as described above for the vSAG1.JdFR and vSAG8.JdFR (Fig. S3A, B). The viral gene homolog family affiliations of borehole 1362B JdFR metagenome (IMG metagenome ID 3300002532) were categorized as NCLDVs, other dsDNA viruses and Unclassified dsDNA viruses. The genes considered for analysis were, (a) non duplicates, and (b) best blast hits of >30% identity match.

Genes from JdFR metagenomes affiliated to eukaryotes were sorted based on best blast hits, (a) 1362B borehole (IMG metagenome ID 3300002532), with >90% similarity, and (b) U1362A borehole (IMG metagenome IDs 3300002481) with, >60% but less than <90% similarity, as no hits had >90% similarity (Fig. S5A, B).

## Analysis of 18S rRNA, 28S rRNA, and internal transcribed spacer (ITS) sequences

The taxonomic classification of the 18S rRNA gene, 28S rRNA gene, and ITS sequences from the IODP borehole U1362B assembled metagenome [52, 53] was done by blastn search [31, 33] against the SILVA database [54] (Table S2).
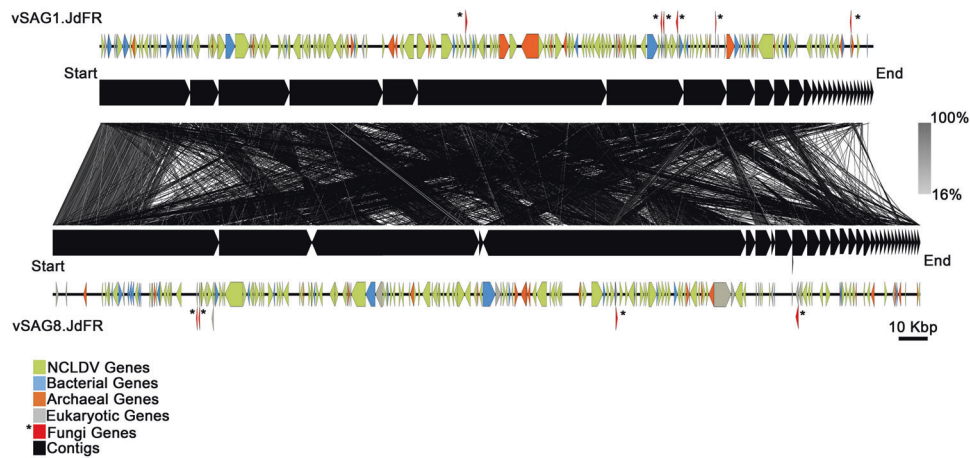
**Fig. 1 Nucleotide level alignment of vSAG1.JdFR and vSAG8.JdFR to each other.** The vSAG1.JdFR (28 scaffolds) and vSAG8.JdFR (33 scaffolds). Black tick marks denote the breaks between the concatenated (using 10 Ns) scaffolds. The lines between the vSAG1.JdFR and vSAG8.JdFR genomes represent the nucleotide level alignment based on tBlastx analysis (alignment ranging from 16 to 100%). The predicted genes were assigned taxonomic affiliation based on Blastp analysis. Genes affiliated to bacteria (blue), archaea (orange), NCLDV (green), eukaryotes (gray), fungal (red), and scaffolds (black) for vSAG1.JdFR and vSAG8.JdFR on the genome.

## RESULTS

Among the single-amplified genomes of particles sorted from the crustal fluid sample, we retrieved two related (average nucleotide identity (ANI) of 99.8%, alignment fraction of 70%) draft NCLDV genomes (vSAG1.JdFR and vSAG8.JdFR; Fig. 1, Tables S3 and S4). These NCLDV vSAGs were not represented in the metagenomes derived from samples collected at the same time from Holes U1362 A and U1362B [3]. However, our protein homology family affiliation analysis through the Integrated Microbial Genomes and Microbiomes (IMG/M) system [52] indicated that 3.31% of the unique viral genes and 25% of the total viral genes within the U1362B metagenome corresponded to NCLDVs (IMG ID 3300002532, Fig. S3).

### vSAG1.JdFR and vSAG8.JdFR belong to the mesomimiviridae family

Phylogenetic placement based on the concatenated alignment of five nearly universal NCLDV genes [55] showed vSAG1.JdFR and vSAG8.JdFR as sister lineages of *Phaeocystis globosa* viruses (PgVs) and *Chrysochromulina parva* viruses BQ2 (CpV BQ2) within the Mesomimiviridae family [56–58] (Fig. 2A). At the nucleotide level, vSAG1.JdFR and vSAG8.JdFR had ~65% and ~66% ANI with the PgVs and ~67 % and 69% ANI with CpV BQ2, respectively. Mapping the predicted genes from the relatively small vSAG1.JdFR and vSAG8.JdFR draft genomes (Fig. 2B) to the twenty established families of ancestral nucleocytoplasmic virus orthologous groups (NCVOGs [37]) showed that our vSAGs contain at least 11 and 13 ancestral NCVOGs, respectively (Fig. 2C, D; Table S5). The NCVOG presence/absence pattern for both vSAGs was most similar to other Mesomimiviridae viruses (Fig. 2D) and supported their evolutionary affinity. The vSAG1.JdFR and vSAG8.JdFR share 70 protein families (Fig. 2E) not identified in other Mesomimiviridae (as of *February 2021*), extending the Mesomimividae subfamily's gene repertoire known so far (supplementary data: https://github.com/asbhattacharjee/NCLDV-from-Juan-de-Fuca-Ridge-flank). On the other hand, the vSAGs lack 10 protein families present in previously known Mesomimiviridae members included in our study (Fig. 2E).

### vSAG1.JdFR and vSAG8.JdFR contain putative bacterial, archaeal, and eukaryotic genes

The partial vSAG1.JdFR and vSAG8.JdFR genomes harbor genes from bacteria (at least 32 and 43, respectively), archaea (at least 20 and 25, for vSAG1.JdFR and vSAG8.JdFR, respectively), and eukaryotes (at least 17 and 20, respectively), based on annotation using the NCBI-NR database (*as of February 2021*) (Fig. 1, and Tables S6, S7, and S8). However, as previously shown [9, 59], potential misclassifications of genes in databases, such as NCBI-NR, cannot be completely ruled out, particularly in the case of some of those previously classified as archaeal genes [9].

The bacterial and archaeal genes (Fig. 1, blue and orange, respectively) found on scaffolds with NCLDV genes are clustered in islands across the genomes. Taxonomic identification of the genes predicted as archaeal in vSAG1.JdFR and vSAG8.JdFR showed that twenty were affiliated to Euryarchaeota also present within crustal fluids at this site, but not bottom seawater (Table S6, [22]).

Among the eukaryotic genes, we identified fungal homolog genes in both vSAGs, which encode two posttranslational modification enzymes (ubiquitin-conjugating enzyme, vSAG1.JdFR Gene 190 and vSAG8.JdFR Gene 57; and iron-sulfur transporter atm1-mitochondrial, vSAG8.JdFR Gene 262), a peroxisome bio-proteinsis factor 10, variant 2 (vSAG1.JdFR gene 123 and vSAG8.JdFR gene 198), and a regulatory protein MIG1 (vSAG1.JdFR gene 195 and vSAG8.JdFR gene 62) (Table S9). Phylogenetic analysis of the posttranslational system proteins suggests horizontal gene transfer of genes of fungal origin (Fig. 3A, C).

Additionally, the JdFR vSAGs also contain translation-related genes. The vSAGs have a Tyrosine tRNA (tRNA$^{Tyr}$) gene, which we also found in the *Cafeteria roenbergensis* virus (CroV) (Mimiviridae), *Aureococcus anophagefferens* virus (AaV) (unclassified family-level lineage in the order Imitervirales), Mollivirus sibericum (Pandoraviridae), and *Paramecium bursaria Chlorella* virus (PBCV, *Phycodnaviridae*) (Table S1). Notably, the tRNA$^{Tyr}$ genes in vSAG1.JdFR, vSAG8.JdFR, AaV, and PBCV contain a 20 bps intron at the canonical position between nucleotides 37 and 38 of the precursor tRNA (Fig. 4A). Phylogenic analysis placed the vSAGs' tRNA$^{Tyr}$ gene (with introns) with eukaryotic tRNA$^{Tyr}$ genes (Fig. 4B). Furthermore, both vSAGs have the eukaryotic translation initiation factor 4E, which is commonly found among Mesomimiviridae and other giant viruses (Fig. S4) [60, 61].

### Fungi, predominantly Ascomycota, are the only confirmed eukaryotes in the crustal fluid samples

Our analysis of the previously published metagenome from Hole U1362B crustal fluids [3, 22] indicated that *Ascomycota* fungi (based on >90% sequence protein homology match) were the dominant eukaryotes in the igneous crust at this borehole
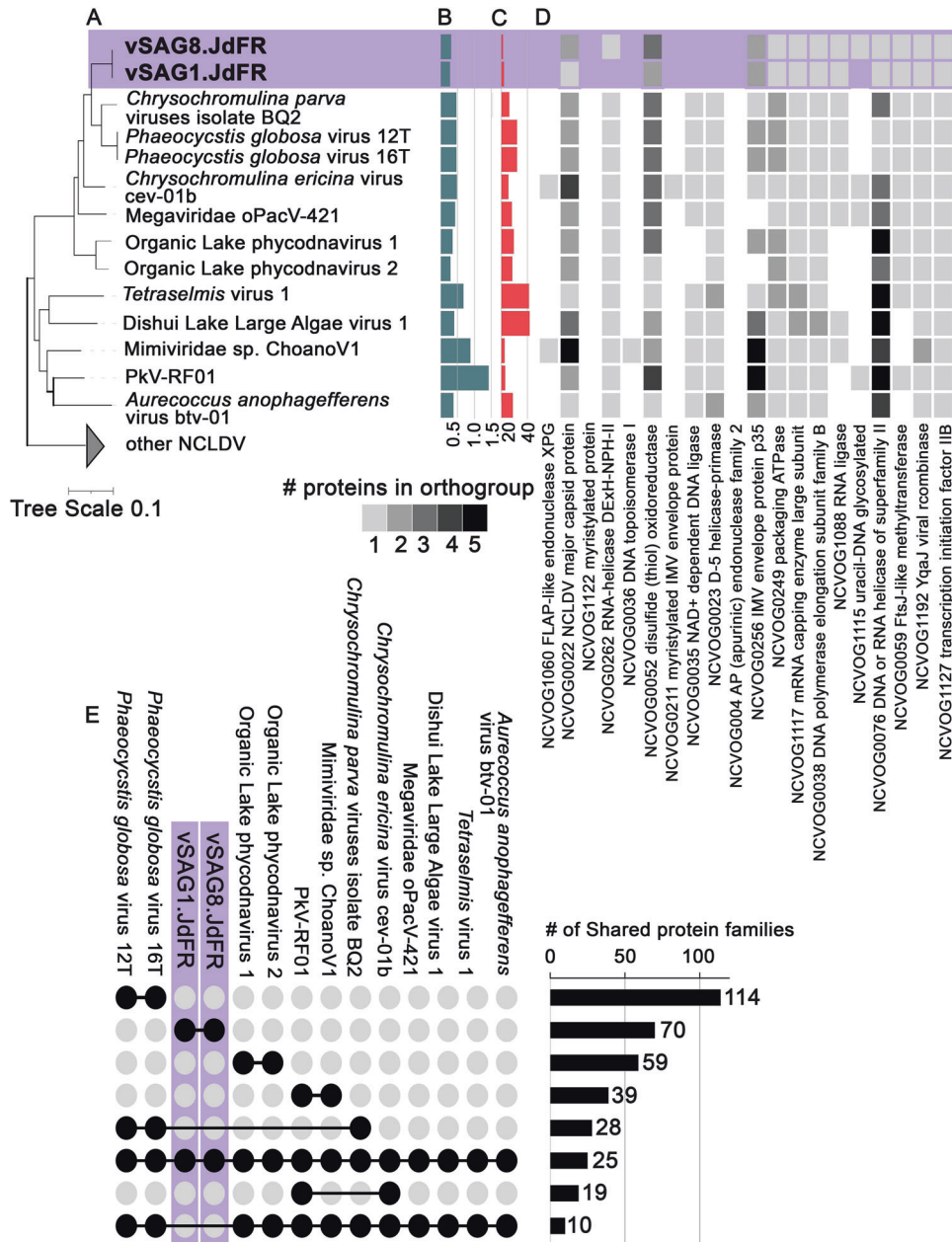
**Fig. 2  Phylogeny, genome characteristics, and genes identified in giant fungal viruses in oceanic crust. A** Phylogenetic position (maximum likelihood tree) of Juan de Fuca Ridge viral single-amplified genomes (vSAG1.JdFR and vSAG8.JdFR) based on a concatenated alignment of five core nucleocytoplasmic virus orthologous genes (NCVOGs) (1. DNA polymerase elongation subunit family B (NCVOG0038); 2. D5-like helicase-primase (NCVOG0023); 3. packaging ATPase (NCVOG0249); 4. Poxvirus Late Transcription Factor VLTF3-like (NCVOG0262), 5. DNA or RNA helicases of superfamily II (NCVOG0076)). The scale bar represents substitutions per site; (**B**) genome assembly sizes (Mbp); (**C**) number of NCVOGs identified in each genome; (**D**) heatmap of protein copy number of genes mapped to the twenty ancestral NCVOGs. Presence/absence and protein copy number patterns of these genes compared to closest related viruses in the phylogenetic tree are indicative of genome completeness; (**E**) number of protein families shared between selected members of the mesomimiviruses. Each set of compared genomes is displayed as solid circles connected by horizontal solid lines; the number of shared protein families and total number of distinct protein families in each lineage are shown as bars.

(Fig. S5A). *Ascomycota* were also present in the U1362A metagenome [3, 22] in relatively high abundance, although they did not comprise the majority of the identified eukaryotic genes, and protein homology matches were lower than 90% (Fig. S5B). Furthermore, *Ascomycota* taxa were the only eukaryotes that could be confirmed based on nearly complete 18S rRNA gene (Table S2), partially complete 28S rRNA gene (Table S10), and ITS sequences (Table S11) retrieved from the metagenomes. Short 18S rRNA fragments (Table S2) and other

eukaryotic gene sequences (Fig. S5A, B) in both metagenomes were affiliated to plants, algae, and animals, none of which are expected to be able to live under the environmental conditions of the igneous crust at this site. While it is possible that eukaryotic environmental DNA that reaches the igneous crust via recharge outcrops is preserved for some time, it is also possible that taxonomic affiliation was skewed by current information in public databases and by the limited information provided by the short sequence fragment's length.
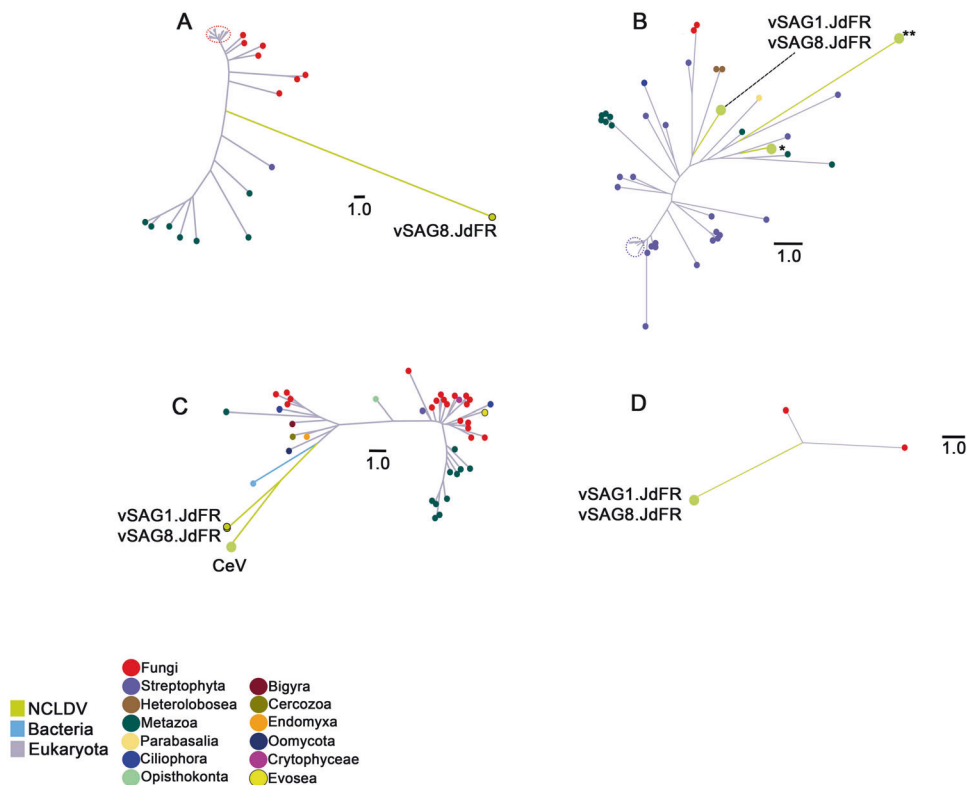
**Fig. 3 Horizontally transferred genes.** Phylogenetic analysis (unrooted trees) of vSAG genes (NCLDVs are green branches). **A** vSAG8.JdFR gene_262, iron-sulfur clusters transporter atm1, mitochondrial enzyme; (**B**) vSAG1.JdFR gene_123 and vSAG8.JdFR gene_198, peroxisome_bioproteinsis_factor_10, variant 2. *Marseillevirus LCMAC201 from deep sea sediments. **Hypothetical protein of Megaviridae environmental sample from marine metagenome; (**C**) vSAG1.JdFR gene_173 and vSAG8.JdFR gene_40, ubiquitin-conjugating enzyme E2-24 kDa-like isoform X2. CeV - *Chrysochromulina ericina* virus, CeV-01b (YP_009173309.1); (**D**) vSAG1.JdFR gene_195 and vSAG8.JdFR gene_62, regulatory protein MIG1.

## DISCUSSION

The two eukaryotic NCLDV draft genomes (vSAG1.JdFR and vSAG8.JdFR) we retrieved from an oceanic crustal fluid sample collected from the IODP borehole U1362B at the Juan de Fuca Ridge were missed in a previously published study of the viral diversity in this environment from samples collected in parallel to our sample from the same borehole and from the nearby borehole U1362A [3]. There were notable methodological differences between how our sample and the samples for the Nigro et al. study [3] were collected and processed. Nigro and colleagues performed metagenomic analyses from large volume (70–124 l) crustal fluid samples collected from Teflon-lined umbilicals and filtered in situ onto 0.22 µm pore-size polyethersulfone filter cartridges. In contrast, we performed single-virus genomics on particles sorted from one milliliter of water collected with a syringe sampler from the freely-venting opening of the borehole observatory. However, geochemical analyses showed consistency between the 64 °C anoxic crustal fluid samples of both studies (Table S12 and references [3, 20, 22]), supporting that the NCLDV particles originated from the crustal fluids, not from possible bottom seawater contamination during sampling. Important, yet possibly numerically rare, NCLDVs could have been missed due to relatively low metagenome sequencing depth, while our targeted method was better suited to detect relatively large virions with large dsDNA genomes.

While the presence of putative bacterial, archaeal, and eukaryotic genes in our vSAGs is not a novel feature for NCLDVs, it provides evidence about their origin from the oceanic basement and points toward fungi as their likely natural hosts. In addition to a core of 10–30 genes acquired mainly by vertical inheritance from ancient ancestors [62], NCLDVs contain multiple genes laterally

exchanged between viruses and hosts (current and ancestral), other viruses, or other cellular sources. In either case, the origin of such genes appears to be linked to the host's ecological niches, with the host providing the grounds for interaction and gene exchange between its viruses, prey, and symbionts [62]. The idea that NCLDVs exist in the crustal fluids with in situ hosts that interact with prey or symbiotic bacteria and archaea is further supported by the fact that, as for other NCLDVs, the bacterial and archaeal genes in our vSAGs are clustered in islands suggesting concomitant inheritance [63]. Also, most of their archaeal genes are affiliated to Euryarchaeota present in Juan de Fuca crustal fluids, but not bottom seawater [22].

The presence of translation-related genes, a universal feature for cellular life, in the Juan de Fuca Ridge NCLDVs is also not novel. For example, in this study we found that our vSAGs, CroV, AaV, PBCV, and Mollivirus siberium have a $tRNA^{Tyr}$ gene. Translational related genes are common in NCDLV, possibly piecemeal acquired from their eukaryotic hosts [18, 64]. Introns at the canonical position between nucleotides 37 and 38 of the precursor tRNA are common for the majority of bacterial, plastid, eukaryotic, and archaeal tRNA introns [65]. However, the short length (20 bps) of our vSAGs' introns, which is within the range of tRNA introns in eukaryotes and archaea (6–133 bps [65]), and the phylogenic placement of the vSAGs' $tRNA^{Tyr}$ genes (Fig. 4) suggest a eukaryotic origin.

All the known NCLDVs have eukaryotic hosts, and the evidence we presented above also support that is the case for the Juan de Fuca NCLDVs. We pose that Ascomycota fungi, the dominant (maybe the only) eukaryotes found in the oceanic crust ([66–70], this study), are the likely natural hosts for these viruses, making these viruses the exception to the rule, since all the fungal viruses recognized by
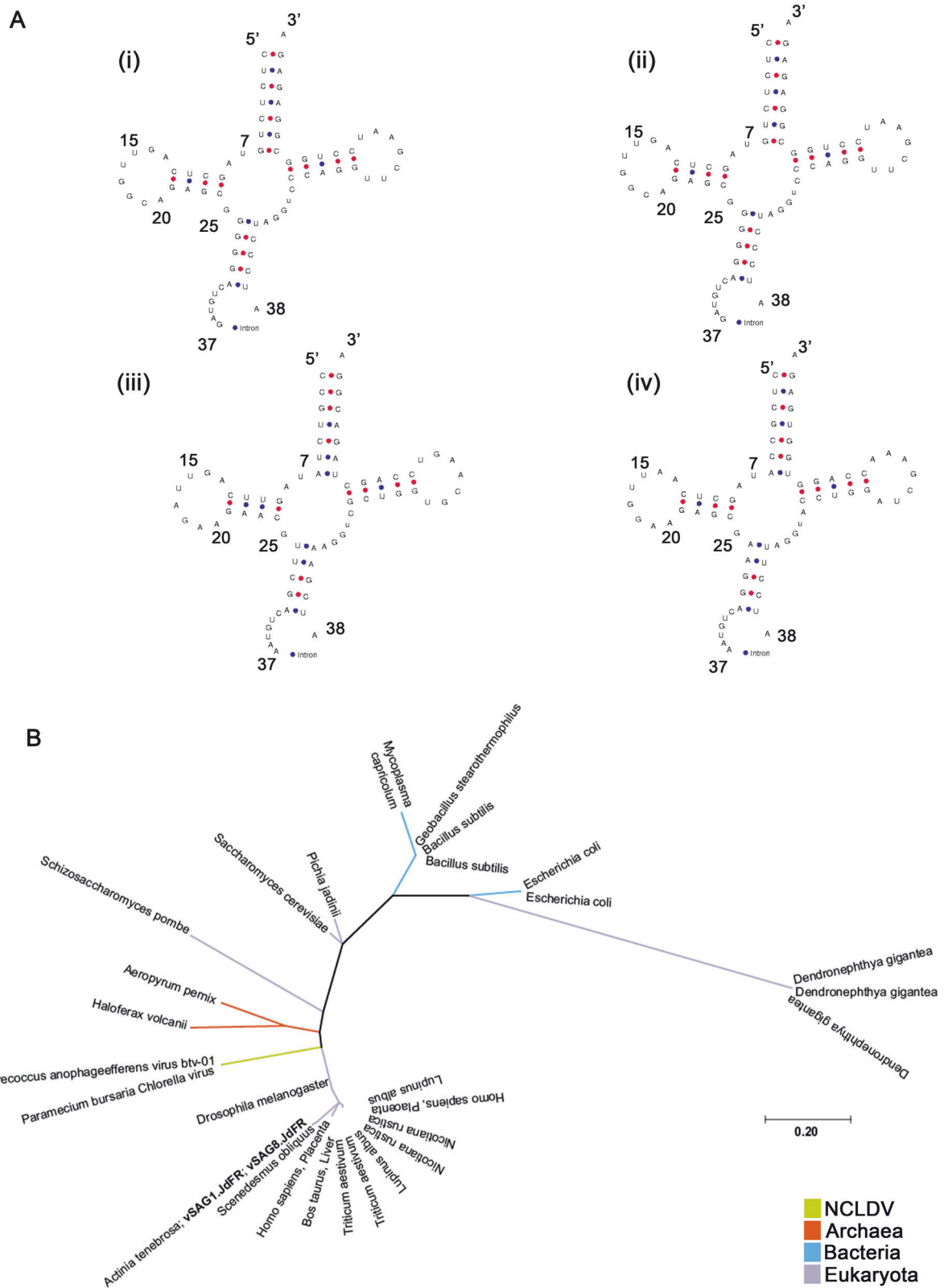
Fig. 4 **tRNATyr structure and phylogeny. A** Predicted tyrosine tRNA (tRNATyr) gene structure with intron at the canonical position between nucleotides 37 and 38 of the precursor tRNA **i.** vSAG1.JdFR; **ii.** vSAG8.JdFR; **iii.** *Aureococcus anophagefferens* virus (AaV); **iv.** *Paramecium bursaria Chlorella* virus (PBCV). **B** Phylogenetic position (maximum likelihood tree) of vSAGs based on tRNA[Tyr] gene.

the International Committee on Taxonomy of Viruses have RNA genomes, with the exception of ssDNA viruses in the family *Genomoviridae* and two retrotransposons [71]. Those types of fungal virus particles may exist within the oceanic crust; however, our sorting approach was suboptimal to detect them because of inherent limitations of flow cytometry to detect very small particles and the low sensitivity of the fluorescence dye SYBR Green I for RNA and ssDNA. Evidence supporting the hypothesis that the Juan de Fuca NCLDVs infect fungi includes co-occurrence of the viruses with the putative Ascomycota hosts, and the observation that vSAG1JdFR and vSAG8JdFR carry fungal genes co-localized in scaffolds with NCLDV genes. In addition, homolog fungal genes (over 70% sequence alignment) in our vSAGs encode two posttranslational modification enzymes (ubiquitin conjugating enzyme, and iron-sulfur transporter atm1-mitochondrial), a peroxisome bioproteinsis factor 10, variant 2, and a regulatory protein MIG1. Phylogenetic placement of those genes suggests a fungal origin, possibly via horizontal gene transfer (HGT), for the two posttranslational modification genes. Prior HGT network analyses also suggested fungi as putative hosts for Nucleocytoviricota klosneuviruses [9]. Additionally, the presence of core NCLDV genes D5-like primase-helicase in the fungus *Allomyces macrogynus* (*Blastocladiomycota* phylum), and B-family DNA polymerase in *Rhizophagus irregularis* (*Mucoromycota* phylum) and *Gonopodaya prolifera* (*Chytridiomycota* phylum) further supports the idea that NCLDVs may infect members of at least some ancestral fungal lineages [57]. Additional work is needed to expand the genomic catalog of NCLDV-host model systems, with a particular focus on fungi.

Fossil records and metagenomic data suggest that active anaerobic fungi may play important ecological and biogeochemical roles in the oceanic crust, where they might have originated as far as 2.4 Ga [66–70, 72, 73]. Recently, a live fungal strain from the genus *Exophiala* (phylum *Ascomycota*) was isolated from the igneous crust at North Pond on the western flank of the Mid-Atlantic Ridge [68, 74]. A study from the lower oceanic crust at Atlantis Bank Gabbro Massif, Indian Ocean revealed over 50% of fungal amplicon sequence variants were from the phylum *Ascomycota* [68]. Our discovery, facilitated by dsDNA-targeted single-virus genomics, of the presumed NCLDVs of fungi adds support to increasing evidence that fungi are active and ecologically important in the oceanic crust habitat [66–70, 72, 73]. Our findings of bacterial and archaeal genes in the vSAGs also support the hypothesis of a symbiotic-like association among fungi and chemoautotrophic archaea and bacteria in subseafloor basalts, where the prokaryotes may use the mycelia for growth while providing a source of carbohydrates or other nutrients to the fungi, as suggested elsewhere [70, 75]. Additionally, $H_2$ produced by the fungi in the anoxic environment could serve as an energy source for $H_2$-dependent chemoautotrophs such as sulfate-reducing bacteria [66, 69]. Such close physical relationships may allow lateral gene transfer between possible symbionts and the NCLDVs during infection of a putative fungal host. Finally, our study raises important questions regarding the extent of the impact that NCLDV infection of fungi has on productivity and biogeochemical cycling in the marine biosphere. Typically, mycoviruses cause persistent infections transmitted intracellularly that do not lyse the host cell [76]. However, not all such infections are asymptomatic. Some mycoviruses induce phenotypic changes that positively or negatively impact host growth, sporulation and virulence [77–80]. However, fungal associations with co-occurring microbes and co-infections [81] by more than one virus type complicate discerning the magnitude and nature of the interactions, i.e., synergistic, antagonistic, or mutualistic.

## DATA AVAILABILITY
The sequence data for vSAG1JdFR and vSAG8JdFR are available through the National Center for Biotechnology Information (NCBI) GenBank Sequence Read Archive under sample accession numbers SRX3120357 and SRX 3120352, respectively, under BioProject PRJNA398661. Annotations of vSAG1JdFR and vSAG8JdFR have been submitted to GenBank, NCBI and assigned accession numbers OP765507, and OP765584, respectively. This information is also linked at the Biological and Chemical Oceanography Data Management Office (BCO-DMO) database under dataset ID 717763. Sequence analysis scripts and data can be accessed on github: (https://github.com/asbhattacharjee/NCLDV-from-Juan-de-Fuca-Ridge-flank/blob/main/tRNA_analysis/tRNA_phylogenetic_tree.nwk).

## REFERENCES
1. Orcutt B, D'Angelo T, Jungbluth SP, Huber JA, Sylvan JB. Microbial life in oceanic crust. OSF Preprints, 2020; https://doi.org/10.31219/osf.io/2wxe6.
2. Koonin EV. On the origin of cells and viruses: primordial virus world scenario. Ann NY Acad Sci. 2009;1178:47–64.
3. Nigro OD, Jungbluth SP, Lin HT, Hsieh CC, Miranda JA, Schvarcz CR, et al. Viruses in the oceanic basement. MBio. 2017;8:1–15.
4. Wheat CG, Jannasch HW, Kastner M, Hulme S, Cowen J, Edwards KJ, et al. Fluid sampling from oceanic borehole observatories: design and methods for CORK activities (1990–2010). 2011. In Fisher AT, Tsuji T, Petronotis K, and the Expedition 327 Scientists, Proc. IODP, 327: Tokyo (Integrated Ocean Drilling Program Management International, Inc.). https://doi.org/10.2204/iodp.proc.327.109.2011.
5. Koonin EV, Yutin N Evolution of the large nucleocytoplasmic DNA viruses of eukaryotes and convergent origins of viral gigantism. *Advances in Virus Research*. 2019. Elsevier, pp 167–202.
6. Koonin EV, Senkevich TG, Dolja VV. The ancient Virus World and evolution of cells. Biol Direct. 2006;1:29.
7. Shinn GL, Bullard BL. Ultrastructure of Meelsvirus: A nuclear virus of arrow worms (phylum Chaetognatha) producing giant "tailed" virions. PLoS One. 2018;13:e0203282.
8. Wilhelm SW, Bird JT, Bonifer KS, Calfee BC, Chen T, Coy SR, et al. A student's guide to giant viruses infecting small eukaryotes: From Acanthamoeba to Zooxanthellae. Viruses. 2017;9:46–63.
9. Schulz F, Roux S, Paez-Espino D, Jungbluth S, Walsh DA, Denef VJ, et al. Giant virus diversity and host interactions through global metagenomics. Nature. 2020;578:432–6.
10. Moniruzzaman M, Martinez-Gutierrez CA, Weinheimer AR, Aylward FO. Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. Nat Commun. 2020;11:1–11.
11. Martínez Martínez J, Swan BK, Wilson WH. Marine viruses, a genetic reservoir revealed by targeted viromics. ISME J. 2014;8:1079–88.
12. Khalil JYB, Robert S, Reteno DG, Andreani J, Raoult D, La Scola B. High-throughput isolation of giant viruses in liquid medium using automated flow cytometry and fluorescence staining. Front Microbiol. 2016;7:1–9.
13. Wilson WH, Gilg IC, Moniruzzaman M, Field EK, Koren S, Lecleir GR, et al. Genomic exploration of individual giant ocean viruses. ISME J. 2017;11:1736–45.
14. Roux S, Chan LK, Egan R, Malmstrom RR, McMahon KD, Sullivan MB. Ecogenomics of virophages and their giant virus hosts assessed through time series metagenomics. Nat Commun. 2017;8:1–12.
15. Schulz F, Alteio L, Goudeau D, Blanchard J, Woyke T, Ryan EM, et al. Hidden diversity of soil giant viruses. Nat Commun. 2018;9:1–9.
16. Bäckström D, Yutin N, Jørgensen SL, Dharamshi J, Homa F, Zaremba-Niedwiedzka K, et al. Virus genomes from deep sea sediments expand the ocean megavirome and support independent origins of viral gigantism. mBio. 2019;10:e02497-18.
17. Martínez JM, Martinez-Hernandez F, Martinez-Garcia M. Single-virus genomics and beyond. Nat Rev Microbiol. 2020;18:705–16.
18. Schulz F, Yutin N, Ivanova NN, Ortega DR, Lee TK, Vierheilig J, et al. Giant viruses with an expanded complement of translation system components. Science (80-). 2017;356:82 LP–85.
19. Fisher A, Wheat CG, Becker K, Cowen J, Orcutt BN, Hulme SM, et al. Design, deployment, and status of borehole observatory systems used for single-hole and cross-hole experiments, IODP Expedition 327, eastern flank of Juan de Fuca Ridge. Proc Integr Ocean Drill Program; Juan Fuca Ridge flank Hydrogeol Exped 327 riserless Drill Platf from to Victoria, Br Columbia (Canada); Sites U1362, U1301, 1027, U1363; 5 July-5 Sept 2010. 2011;327:38.
20. Jungbluth SP, Grote J, Lin HT, Cowen JP, Rappé MS. Microbial diversity within basement fluids of the sediment-buried Juan de Fuca Ridge flank. ISME J. 2013;7:161–72.
21. Jungbluth SP, Bowers RM, Lin HT, Cowen JP, Rappé MS. Novel microbial assemblages inhabiting crustal fluids within mid-ocean ridge flank subsurface basalt. ISME J. 2016;10:2033–47.
22. Jungbluth SP, Amend JP, Rappé MS. Metagenome sequencing and 98 microbial genomes from Juan de Fuca Ridge flank subsurface fluids. Sci data. 2017;4:1–11.
23. Brussaard CPD. Optimization of procedures for counting viruses by flow cytometry. Appl Environ Microbiol. 2004;70:1506–13.

24. Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, et al. Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci. 2002;99:5261–6.

25. Labonté JM, Swan BK, Poulos B, Luo H, Koren S, Hallam SJ, et al. Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. ISME J. 2015;9:2386–99.

26. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.

27. Bushnell B, Rood J. BBTools. *Dep Energy Jt Genome Institute, Walnut Creek, CA* 2014.

28. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.

29. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol. 2013;20:714–37.

30. Besemer J, Lomsadze A, Borodovsky M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 2001;29:2607–18.

31. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

32. Goodacre N, Aljanahi A, Nandakumar S, Mikailov M, Khan AS, Delwart E, et al. A reference viral database (RVDB) to enhance bioinformatics analysis of high-throughput sequencing for novel virus detection. mSphere. 2018;3:e00069–18.

33. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 2004;32:W20–W25.

34. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinformatics. 2011;27:1009–10.

35. Johnson LS, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics. 2010;11:431.

36. Colson P, De Lamballerie X, Yutin N, Asgari S, Bigot Y, Bideshi DK, et al. "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. Arch Virol. 2013;158:2517–21.

37. Yutin N, Wolf YI, Raoult D, Koonin EV. Eukaryotic large nucleo-cytoplasmic DNA viruses: Clusters of orthologous genes and reconstruction of viral genome evolution. Virol J. 2009;6:1–13.

38. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.

39. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3.

40. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2014;32:268–74.

41. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009;26:1641–50.

42. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.

43. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9:1–8.

44. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst Biol. 2010;59:307–21.

45. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16:157.

46. Barrett P, Hunter J, Miller JT, Hsu J-C, Greenfield P matplotlib--A Portable Python Plotting Package. *Astron. data Anal. Softw. Syst. XIV.* 2005. p 91.

47. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: visualization of intersecting sets. IEEE Trans Vis Comput Graph. 2014;20:1983–92.

48. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955.

49. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004;32:11–16.

50. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35:1547.

51. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.

52. Chen IMA, Markowitz VM, Chu K, Palaniappan K, Szeto E, Pillay M, et al. IMG/M: Integrated genome and metagenome comparative data analysis system. Nucleic Acids Res. 2017;45:D507–D516.

53. Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, et al. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. 2012;40:D115–D122.

54. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2012;41:D590–D596.

55. Yutin N, Wolf YI, Koonin EV. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. Virology. 2014;466–467:38–52.

56. Mihara T, Koyano H, Hingamp P, Grimsley N, Goto S, Ogata H Taxon richness of "Megaviridae" exceeds those of bacteria and archaea in the ocean. Microbes Environ. 2018; ME17203.

57. Gallot-Lavallée L, Blanc G. A glimpse of nucleo-cytoplasmic large DNA virus biodiversity through the eukaryotic genomics window. Viruses. 2017;9:17.

58. Aylward FO, Moniruzzaman M, Ha AD, Koonin EV. A phylogenomic framework for charting the diversity and evolution of giant viruses. PLOS Biol. 2021;19:e3001430.

59. Andreani J, Verneau J, Raoult D, Levasseur A, La, Scola B. Deciphering viral presences: two novel partial giant viruses detected in marine metagenome and in a mine drainage metagenome. Virol J. 2018;15:66.

60. Koonin EV, Yutin N. Multiple evolutionary origins of giant viruses. F1000Research. 2018;7:1840.

61. Abrahao JS, Araujo R, Colson P, La, Scola B. The analysis of translation-related gene set boosts debates around origin and evolution of mimiviruses. PLoS Genet. 2017;13:e1006532.

62. Filée J, Chandler M. Gene exchange and the origin of giant viruses. Intervirology. 2010;53:354–61.

63. Koonin EV, Yutin N. Nucleo-cytoplasmic large DNA viruses (NCLDV) of eukaryotes. eLS. 2012. https://doi.org/10.1002/9780470015902.a0023268.

64. Abergel C, Rudinger-Thirion J, Giegé R, Claverie J-M. Virus-encoded aminoacyl-tRNA synthetases: structural and functional characterization of mimivirus TyrRS and MetRS. J Virol. 2007;81:12406–17.

65. Yoshihisa T. Handling tRNA introns, archaeal way and eukaryotic way. Front Genet. 2014;5:213.

66. Ivarsson M, Schnürer A, Bengtson S, Neubeck A. Anaerobic fungi: a potential source of biological $H_2$ in the oceanic crust. Front Microbiol. 2016;7:1–8.

67. Ivarsson M, Bengtson S, Neubeck A. The igneous oceanic crust e Earth' s largest fungal habitat? Fungal Ecol. 2016;20:249–55.

68. Quemener M, Mara P, Schubotz F, Beaudoin D, Li W, Pachiadaki M, et al. Meta-omics highlights the diversity, activity and adaptations of fungi in deep oceanic crust. Environ Microbiol. 2020;22:3950–67.

69. Drake H, Ivarsson M. The role of anaerobic fungi in fundamental biogeochemical cycles in the deep biosphere. Fungal Biol Rev. 2017;32:20–25.

70. Bengtson S, Rasmussen B, Ivarsson M, Muhling J, Broman C, Marone F, et al. Fungus-like mycelial fossils in 2.4-billion-year-old vesicular basalt. Nat Publ Gr. 2017;1:1–6.

71. Suzuki N. An introduction to fungal viruses. In Encyclopedia of Virology. Bamford DH, Zuckerman M. editors. 431–42. Oxford: Academic Press; 2021.

72. Ivarsson M, Broman C, Holmström SJM, Ahlbom M, Lindblom S, Holm NG. Putative fossilized fungi from the lithified volcaniclastic apron of Gran Canaria, Spain. Astrobiology. 2011;11:633–50.

73. Bengtson S, Ivarsson M, Astolfo A, Belivanova V, Broman C, Marone F, et al. Deep-biosphere consortium of fungi and prokaryotes in Eocene subseafloor basalts. Geobiology. 2014;12:489–96.

74. Hirayama H, Abe M, Miyazaki J, Sakai S, Nagano Y, Takai K Data report: cultivation of microorganisms from basaltic rock and sediment cores from the North Pond on the western flank of the Mid-Atlantic Ridge, IODP Expedition 336 1. 2015; 336.

75. Ivarsson M, Bengtson S, Skogby H, Lazor P, Broman C A Fungal-Prokaryotic Consortium at the Basalt-Zeolite Interface in Subseafloor A Fungal-Prokaryotic Consortium at the Basalt-Zeolite Interface in Subseafloor Igneous Crust. 2015.

76. Khan HA, Telengech P, Kondo H, Bhatti MF, Suzuki N. Mycovirus Hunting Revealed the Presence of Diverse Viruses in a Single Isolate of the Phytopathogenic Fungus Diplodia seriata From Pakistan. Front Cell Infect Microbiol. 2022;12:913619.

77. García-Pedrajas MD, Cañizares MC, Sarmiento-Villamil JL, Jacquat AG, Dambolena JS. Mycoviruses in biological control: From basic research to field implementation. Phytopathology. 2019;109:1828–39.

78. Ghabrial SA, Castón JR, Jiang D, Nibert ML, Suzuki N. 50-plus years of fungal viruses. Virology. 2015;479:356–68.

79. Okada R, Ichinose S, Takeshita K, Urayama S, Fukuhara T, Komatsu K, et al. Molecular characterization of a novel mycovirus in Alternaria alternata manifesting two-sided effects: Down-regulation of host growth and up-regulation of host plant pathogenicity. Virology. 2018;519:23–32.

80. Shah UA, Kotta-Loizou I, Fitt BDL, Coutts RHA. Mycovirus-induced hypervirulence of Leptosphaeria biglobosa enhances systemic acquired resistance to Leptosphaeria maculans in Brassica napus. Mol Plant-Microbe Interact. 2020;33:98–107.

81. Hillman BI, Annisa A, Suzuki N. Viruses of plant-interacting fungi. Adv Virus Res. 2018;100:99–116.

## COMPETING INTERESTS
The authors declare no competing interests.

## ADDITIONAL INFORMATION
**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s43705-022-00210-8.

**Correspondence** and requests for materials should be addressed to Ananda S. Bhattacharjee or Joaquín Martínez Martínez.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.