

Phylogenomic analysis expands the known repertoire of single-stranded DNA viruses in benthic zones of the South Indian Ocean

Oliver K.I. Bezuidt^{1,2,*} and Thulani P. Makhalanyane^{2,3,*}

¹DSI/NRF South African Research Chair in Marine Microbiomics, Department of Biochemistry, Genetics and Microbiology, microbiome@UP, University of Pretoria, Pretoria, 0028, South Africa

²Department of Microbiology, Faculty of Science, Stellenbosch University, Stellenbosch 7600, South Africa

³Centre for Epidemic Response and Innovation, The School for Data Science and Computational Thinking, Stellenbosch University, Stellenbosch 7600, South Africa

*Corresponding authors: Oliver K.I. Bezuidt, DSI/NRF SARCHI Chair in Marine Microbiomics, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Hatfield, Pretoria 0028, South Africa. Email: bezuidt@gmail.com and Thulani P. Makhalanyane, Department of Microbiology, Stellenbosch University, van der Byl Street, JC Smuts Building, Office 310, Stellenbosch, 7600 South Africa. Email: tpm@sun.ac.za

Abstract

Single-stranded (ss) DNA viruses are ubiquitous and constitute some of the most diverse entities on Earth. Most studies have focused on ssDNA viruses from terrestrial environments resulting in a significant deficit in benthic ecosystems including aphotic zones of the South Indian Ocean (SIO). Here, we assess the diversity and phylogeny of ssDNA in deep waters of the SIO using a combination of established viral taxonomy tools and a Hidden Markov Model based approach. Replication initiator protein-associated (Rep) phylogenetic reconstruction and sequence similarity networks were used to show that the SIO hosts divergent and as yet unknown circular Rep-encoding ssDNA viruses. Several sequences appear to represent entirely novel families, expanding the repertoire of known ssDNA viruses. Results suggest that a small proportion of these viruses may be circular genetic elements, which may strongly influence the diversity of both eukaryotes and prokaryotes in the SIO. Taken together, our data show that the SIO harbours a diverse assortment of previously unknown ssDNA viruses. Due to their potential to infect a variety of hosts, these viruses may be crucial for marine nutrient recycling through their influence of the biological carbon pump.

Keywords: biogeochemical cycling, CRESS-DNA, single stranded DNA viruses, Rep and Capsid proteins, South Indian Ocean, viral diversity

Introduction

Disentangling the phylogenetically diverse assemblages of bacteria, fungi, phytoplankton, and viruses and their contributions to ecosystem services in the global ocean remains a major endeavour. There is strong evidence that these assemblages influence the biological carbon pump in marine environments [1, 2]. Among these assemblages, viruses are the most numerically abundant [3], and recent evidence has demonstrated their profound influence on prokaryotic lifestyles in the oceans [4–8]. These studies have shown that viruses play key roles in determining ecological patterns in marine ecosystems and mediating nutrient recycling through the transfer of genes between both eukaryotic and prokaryotic hosts [9–11]. Through lysogeny, viruses enhance the release of organic matter, promoting the recycling of nutrients such as dissolved organic carbon, through the microbial loop [12]. While there are increased insights regarding viral contributions, most studies have focused on limited geographic locations. As a result, comparatively less is known regarding the phylogeny and function of single- and double-stranded DNA viruses in large regions of the global ocean.

An accumulating body of research suggests that viruses are central drivers of metabolic processes through auxiliary

metabolic genes (AMGs) [13–15]. Previous studies have also shown that double stranded DNA (dsDNA) viruses mediate key metabolic processes, known to modulate microbial metabolic pathways during infection [16–19]. Most studies on viral diversity and functional contributions have been derived from studies conducted mostly in euphotic oceanic zones [11, 20–23]. This has resulted in a substantial knowledge gap regarding the evolutionary structure and function of viruses in benthic ecosystems. Current studies on benthic ecosystems have, however, provided several insights regarding the diverse dsDNA prokaryotic viruses and their encoded AMGs [24–28], and less on assemblages and potential functional contributions of ssDNA viruses.

More recently, uncharacterized groups of ssDNA viruses associated with vertebrates from terrestrial environments have become the subject of extensive research [29–34]. Although these studies have provided substantial insights regarding the diversity and versatility of ssDNA, the results have led to an overrepresentation of virus sequences from several habitats including terrestrial ecosystems and mammal derived samples [35–37]. Several studies have noted the lower proportion of studies on ssDNA viruses in marine ecosystems [36, 38, 39]. Given the importance of viruses as mediators of the biological carbon pump, the lack of studies on ssDNA viruses may limit our understanding regarding their

Received: 28 February 2024. Revised: 25 April 2024. Accepted: 29 April 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of the International Society for Microbial Ecology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

functional contributions. The diversity and functional contributions of ssDNA in the oceans may substantially contribute to marine nutrient recycling and the biological carbon pump.

Previous studies on viral communities in benthic zones have demonstrated that ssDNA viruses constitute dominant constituents of these environments [40, 41]. However, to the best of our knowledge, the evolutionary relationships, diversity, and distribution of ssDNA viruses in the SIO remains unexplored. We predict that the distinct environmental conditions and water masses in the SIO may select for phylogenetically diverse ssDNA viruses. Here, we explore the diversity and phylogeny of ssDNA from benthic zones of the SIO. In addition to using conventional approaches to search for putative viral contigs, we applied a hidden Markov model (HMM) workflow to expand the current repertoire of ssDNA viruses.

Materials and methods

Sample collection and molecular ecological analyses

Seawater samples were collected aboard the RV SA *Agulhas II* from the Crossroads transect in the South Indian and Southern Oceans as detailed previously [42, 43]. As part of the cruise, 39 samples (13 × 3) were collected across a 1000 km transect. At each site, three samples were collected from the deepest depth (Supplementary Table S1). A Sea-Bird SBE-911plus V2 CTD System (Sea-Bird Electronics, Inc., Bellevue, Washington, USA) was used to collect samples ~10 m above the seafloor. Depending on the conditions, the CTD was retrieved within 5 hours. Once on deck, 5 L of seawater was retrieved from each of the three Niskin bottles and subjected to a two-step filtration process to allow for the collection of particle-associated viruses (hereinafter referred to as viruses), potential symbionts including candidate phyla radiation, and free-living microbial biomass. Following filtration, all samples were immediately stored at -25°C until processing.

DNA was extracted from membrane filters using a phenol-chloroform method [44] with extractions performed in duplicate, with minor modifications. Specifically, the pH of the extraction buffer was adjusted from 8 to 9.5 to compensate for low DNA concentration. Using sterile forceps, both 0.45 and 0.2 μm membrane filters were cut in half and used as the samples in the protocol, and as a result the number of glass beads used in the protocol was lowered from 0.4–0.5 to 0.25 ml (0.10–0.11 mm diameter). Samples with the highest DNA concentration ($n=6$) were sent for sequencing at the Molecular Research DNA (MR DNA) sequencing facility (Shallowater, TX, USA). These six samples are pseudo replicates from same sampling sites as indicated in Supplementary Table S1 (CR11, CR12 for CR1; CR21, CR22 for CR9; and CR31, CR32 for CR10). Due to generally low DNA yields, whole metagenome amplification was performed using REPLI-g Midi kit (Qiagen, Hilden, Germany). Libraries were prepared using the Nextera DNA Sample preparation kit (Illumina, Inc., San Diego, CA) with a small insert size (<1 kbp). The final libraries were then pooled and sequenced as paired end reads for 300 cycles, using the Illumina HiSeq 2500 system (Illumina, Inc. San Diego, CA, USA).

Bioinformatics analysis

Shotgun metagenome analysis and processing

Raw metagenomic reads were inspected for quality, and the presence of sequencing adapters, using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc). Following this, the reads were processed using BBSplit version 38.00 to remove PhiX174 (CP004084.1) using default settings. The resultant PhiX174 free

reads were further trimmed using Trimmomatic version 0.36 [45]. The total sequencing coverage was estimated using Nonpareil version 3.301 [46] (Supplementary Fig. S1). We assembled the metagenomes using metaSPAdes version 3.14.1 as detailed previously [47] following default parameters. All contigs <1000 bp were removed and the remaining sequences were used for downstream analyses.

Bacterial taxonomic classification and functional annotation

To estimate the distribution and taxonomy of the taxa recovered from the metagenomes, contigs were aligned against a non-redundant (NR) protein database from the National Center for Biotechnology Information (NCBI) database [48]. We used DIAMOND BLASTx version 2.02.2 with the default settings [49]. The outputs were analysed using MEGAN version 6.20.19 [50] to estimate taxonomic distributions and determine overrepresented taxa across all metagenomes. Contigs for the top five most represented taxa were extracted and analysed using Kofamscan version 1.3.0 to determine their metabolic potential with the default settings [51].

Viral prediction and quality processing

To assess the distribution and classification, two approaches (detailed below) were used to predict viruses. For the first approach, we used standard viral prediction pipelines, whereas the second relied on the use of a HMM based approach [52].

Approach 1: All six assembled metagenomes were used to search for putative viral contigs using a combination of Virsorter version 2.2.3 [53] and VirFinder version 1.1 [54]. Contigs with sizes ≥ 1 kb, predicted using both Virsorter2 version 3 SOP ([dx.doi.org/10.17504/protocols.io.bwm5pc86](https://doi.org/10.17504/protocols.io.bwm5pc86)), as well as those predicted to have P -value <.01 by VirFinder, were retained and merged for downstream analyses. These merged contigs were clustered into viral operational taxonomic units (vOTUs) using CD-HIT version 4.8.1 [55], based on 95% sequence identity >80% of the shortest contig. The resultant vOTUs contig were processed using CheckV version 0.70 [56] to estimate overall quality and completeness (Supplementary Tables S2 and S3). Putative viral contigs, that were designated as complete, high, medium, and low were retained. These contigs were post-processed using the geNomad pipeline version 1.6.1 [57] with the default parameters. Contigs that were classified as affiliated with the class *Caudoviricetes* and order *Petitvirales* were assessed for the presence of auxiliary metabolic genes using VIBRANT version 1.2.1 [58], with the default specifications. The lifecycles associated with these putative viruses were determined using BACPHLIP [59] with the default settings. Lytic/virulent phages were identified with a minimum lifecycle prediction score of ≥ 0.8 . Following this, we inspected a set of putative ssDNA-specific viruses. Contigs (≥ 1 kb) predicted as ssDNA viruses by Virsorter version 2.2.3 were retained. These contigs were further clustered into viral OTUs using CD-HIT version 4.8.1 and further estimated for quality and completeness as detailed earlier. We retained contigs designated as complete, high, medium, and low quality. These putative virus predictions were validated using BLASTp [60] with e -value $1e-05$, against the NR protein database from the NCBI (accessed January 2023) [61].

Approach 2: To generate HMM-based profiles, we retrieved all *Cressdnaviricota* and *Phixviricota* protein sequences from GenBank [61], (<https://github.com/SAMicrobiomes/ssDNA>). To reduce redundancy, these sequences were clustered at 95% amino acid identity >90% of the shortest sequence, using CD-HIT version

4.8.1 [55]. The resultant sequences were compared, using all-vs-all BLASTp with e-value $1e-5$. The output sequences were further clustered, using Markov cluster (MCL) algorithm [62] with the inflation parameter set to 1.5. Clusters with proteins ≥ 10 annotated replication initiator (Rep) and major capsid (VP1) were aligned using MAFFT version 7.487 [63] with the —auto parameter. The alignments were used to create HMM profiles, using HMMER version 3.3.2 [52]. The profiles were searched against protein sequences, which were predicted using the -p meta function in Prodigal [64], from all metagenomes using HMMSCAN version 3.0 [65]. Metagenomic sequences, that shared similarity with conserved *Cressnaviricota* replication initiator (Rep) and *Phixviricota* major capsid (VP1) proteins, with HMMSCAN scores ≥ 50 were retrieved for downstream analysis. Viral contigs, predicted to harbour Rep and VP1 proteins, were validated using BLASTp [60] with e-value $1e-05$ against a non-redundant (NR) database acquired from the NCBI as well as HHpred [66] against the Protein Data Bank (PDB) and Protein Family Database (Pfam) database.

Phylogenetic analysis of single stranded DNA viruses

Complete Rep and VP1 protein sequences, from both our metagenomes and GenBank viral datasets, were used to reconstruct phylogenies based on conserved amino acid sequences. Rep proteins, predicted from our metagenomes, were compared with those from ssDNA viruses and plasmid sequences from previous studies [67]. These protein sequences were aligned using MAFFT-linsi and trimmed using Trimal with gap threshold 0.15 [68]. Phylogenetic trees were computed using FastTree with options -spr 4 -mlacc 2 -slow -lg [69].

Results and discussion

Diverse putative hosts dominated by prokaryotes

Viruses are crucial for the biological carbon pump and regulate microbial community structure and abundance, determining the genetic diversity and evolution of their hosts [70–74]. In turn, there is some evidence that marine viruses co-exist with potential hosts, which include phylogenetically diverse eukaryotes and prokaryotes [22, 71, 75–77]. For instance, previous studies have shown that eukaryotic hosts include copepods [78] and protists [79]. However, the diversity of potential hosts in understudied environments such as the SIO remains unknown.

Our dataset provides an overview of the taxonomy of bacterial, archaeal, and eukaryotic hosts in benthic zones of the SIO. The data suggest that bacteria and archaea may be the dominant hosts, compared with eukaryotes which constituted a minor fraction of our sequences (Fig. 1A). This finding is consistent with previous studies, which have demonstrated that prokaryotes far outnumber eukaryotes in all ecosystems [4, 80–83]. Taxonomic classifications suggests that three bacterial (*Proteobacteria*, *Candidatus Marinimicrobia*, and *Bacteroidota*) and one archaeal (*Thaumarchaeota*) phyla were the most overrepresented in the benthic SIO (Fig. 1B). Other studies on the diversity and distribution of marine microbiota in aphotic zones have reported similar findings [84–86]. Some of these taxa constitute ecologically rare taxa, which may represent active free-living bathypelagic microbiota [87] and may disproportionately contribute to the sequestration of key nutrients.

Although sequences affiliated with Eukaryota were present at comparatively low abundances, several taxa including Chlorophyta, Streptophyta, Foraminifera, Ascomycota, and members of the ecologically diverse SAR supergroup were

found (Supplementary Fig. S2). These taxa may have originated from deep-sea floor unicellular and multicellular species. We argue that some of these species may host eukaryotic ssDNA viruses found in our dataset. Our results are consistent with previous studies, which reported generally low abundances of eukaryotes in other benthic environments such as the Black Sea [88]. Based on their relatively high abundances, we predict that prokaryotes may be the primary hosts of marine viruses, which ultimately mediate important functional processes in the SIO.

Widespread functional capacity in diverse prokaryotic hosts

To reduce the knowledge deficit regarding contributions to ecosystem services, we determined functional capacity and putative ecological contributions of bacteria and archaea in the SIO. Metabolic analysis revealed a suite of complete pathways linked to carbon degradation, nitrogen, methanogenesis, and sulphur recycling (Fig. 2A). In addition to providing the first such data for this region, the results are consistent with recent findings showing that deep oceans possess remarkably diverse capacity for functional processes [89–92]. Linking these functional genes to microorganisms suggests that some taxa may exclusively drive specific metabolic processes. For instance, *Proteobacteria* appear to be the only taxa with metabolic capacity for methanogenesis, suggesting that they may contribute to oxidizing methane in the deep environments [93, 94]. This result suggests that these numerically dominant taxa may mediate key ecosystem processes in the SIO. Results from functional analysis suggests that some metabolic roles may be driven by a consortia rather than one numerically abundant microorganism. For instance, genes for sulphur metabolism were found in sequences affiliated with *Candidatus Marinimicrobia*, *Proteobacteria*, and *Thaumarchaeota*. Among these, ammonia oxidizing *Thaumarchaeota* appear to be the only taxa, with a complete pathway for dissimilatory sulphate reduction and sulphide oxidation. This result suggests that these *Thaumarchaeota* may play especially important roles in sulphur recycling [95, 96]. This observation also suggests that, in addition to other groups associated with marine and terrestrial environments (i.e. *Euryarchaeota*, *Crenarchaeota*, and *Aigarchaeota*) [97], *Thaumarchaeota* may augment the metabolic contributions of candidate sulphate-reducing archaea. Collectively, the detection of these genes provides some indication of the metabolic capacity in aphotic SIO waters, which may influence important ecological processes.

Evidence of extensive viral diversity and functional potential in the deep South Indian Ocean

Despite the importance of the SIO in modulating global climate and heat uptake in the ocean [1, 42, 98], we lack taxonomic insights regarding the functional roles of viruses. Using two of the most widely used pipelines [54, 99], we explored viral communities in deep SIO waters. Viral predictions were combined, checked for quality and completeness, and clustered into 3076 putative viral OTUs. In total, 1427 (46.4%) of these vOTUs were classified into known prokaryotic viral taxa. These viruses included members of the class *Caudoviricetes* (991), which included families *Straboviridae* ($n = 3$), *Kyanoviridae* ($n = 2$), *Autographiviridae* ($n = 1$), and *Herelleviridae* ($n = 1$). In addition, several sequences were affiliated with taxa from the order *Petitvirales* (436), including members of the *Microviridae* ($n = 406$) family.

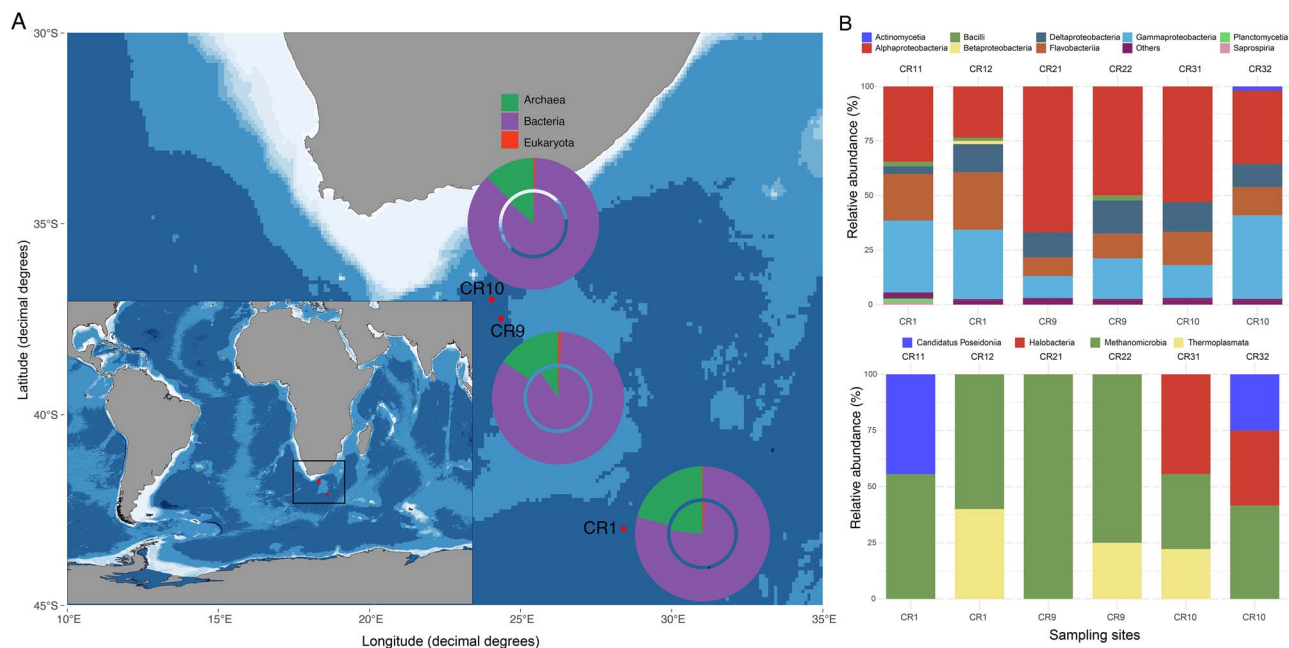


Figure 1. (A) A map showing the six sampling locations in the South Indian Ocean. The map also includes an overview of microbial diversity (pie charts) at each sampling location. (B) Bar plots showing the relative abundances of bacterial and archaeal classes, at each sampling site.

Only a small fraction of the total vOTUs appear to harbour AMGs. From 62 vOTUs, we identified 78 AMGs (Supplementary Table S4). These AMGs were mostly associated with functional traits related to the synthesis of amino acids, carbohydrates, secondary metabolites as well as terpenoides/polyketides metabolic pathways (Supplementary Table S5). Several of these AMGs related to amino acid metabolism included 2OG-Fe(II) oxygenases (arginine and proline metabolism). Previous studies suggest that these AMGs may modulate host nitrogen metabolism, stress response, and DNA repair mechanisms [100–102]. Some of the genes linked to these AMGs are involved in the metabolism of cysteine and methionine, linked to carbon, nitrogen, and sulphur utilization, and the reprogramming of cells to anabolic states [103]. The presence of viral AMGs, previously implicated in sulphur metabolisms (i.e. *dcm*, *cysH*, and *metK*) [104], correlates with the most complete pathways in our metagenomes. Genetic evidence suggests that viruses may contribute substantially to bacterial metabolic reprogramming in deep SIO waters [105]. Determining the lifecycles of Caudoviricetes found in the SIO suggests that these viruses may favour lytic ($n=948$) over lysogenic ($n=14$) cycles. This is consistent with previous reports of high prokaryotic mortality due to viral lysis in the dark ocean [106]. The study by Lara *et al.* [106] further suggests that a preference for viral lytic cycles may result in the production of dissolved organic carbon, supporting respiration in the dark ocean [107]. To estimate the diversity of our Caudoviricetes, we clustered all 991 representative contigs against viral RefSeq database (release 219) using 95% ANI across 85% of the sequence using a method detailed previously [56]. However, none of our contigs clustered with any of the viruses from the reference dataset, suggesting that the deep SIO hosts potentially novel dsDNA viruses.

To explore the diversity of eukaryotic DNA viruses in our dataset, a set of putative viral contigs highlighted as ssDNA by Virsorter 2.2.3 were retained [53]. These were clustered into 7627 viral OTUs. Of these, 6018 vOTUs were validly assigned

and predicted to be complete (24), high (1559), medium (2146), and low (2289) quality. Sequence similarity searches, using these assigned contigs, were used to estimate the diversity of ssDNA viruses. A total of 3440 contigs, with blast descriptions containing the terms *circo*, *CRESS*, *circular genetic*, *circular virus/DNA*, and which either possessed or lacked replication initiator proteins, were classified as rep-encoding ssDNA viruses. Among these, 150 contigs were classified as *Microviridae*, in addition to 810 other contigs associated with dsDNA phages. These numbers suggest a high proportion of false positive predictions (Supplementary Table S6).

Uncharacterized circular Rep-encoding ssDNA sequences dominate deep South Indian Ocean waters

Due to inherent methodological differences, current *in silico* approaches may yield contrasting results regarding viral divisions. While previous studies have revealed a diverse array of viruses [25, 108], sequence classification using standard approaches leads to false positive predictions of prokaryotic ssDNA viruses [109, 110]. A consequence of these false positive annotations may be the assignment of dsDNA viral contigs as ssDNA. To avoid this common issue, we used a Hidden Markov Model (HMM) based pipeline described previously [52] to investigate the diversity of Microviruses and Circular Rep-Encoding ssDNA (CRESS) viruses, in the SIO (Fig. 2B). Using this approach, we established HMM profiles for conserved Rep and VP1 proteins from Cressnaviricota and Phixviricota viral sequences retrieved from GenBank [111]. Based on these profiles, HMM searches across all metagenomes, revealed 3260 and 344 putative Rep and VP1 complete protein sequences, respectively. To reduce redundancy, these sequences were subsequently clustered (at $\geq 95\%$ amino acid identity, over 85% of the shortest sequence), resulting in 2307 (Rep) and 209 (VP1) representative protein sequences. The validation of these hallmark sequences, against a non-viral specific database such as NR, suggests that 2303 and 209 of Rep and VP1 sequences shared sequence similarities with both rep-encoding viruses and

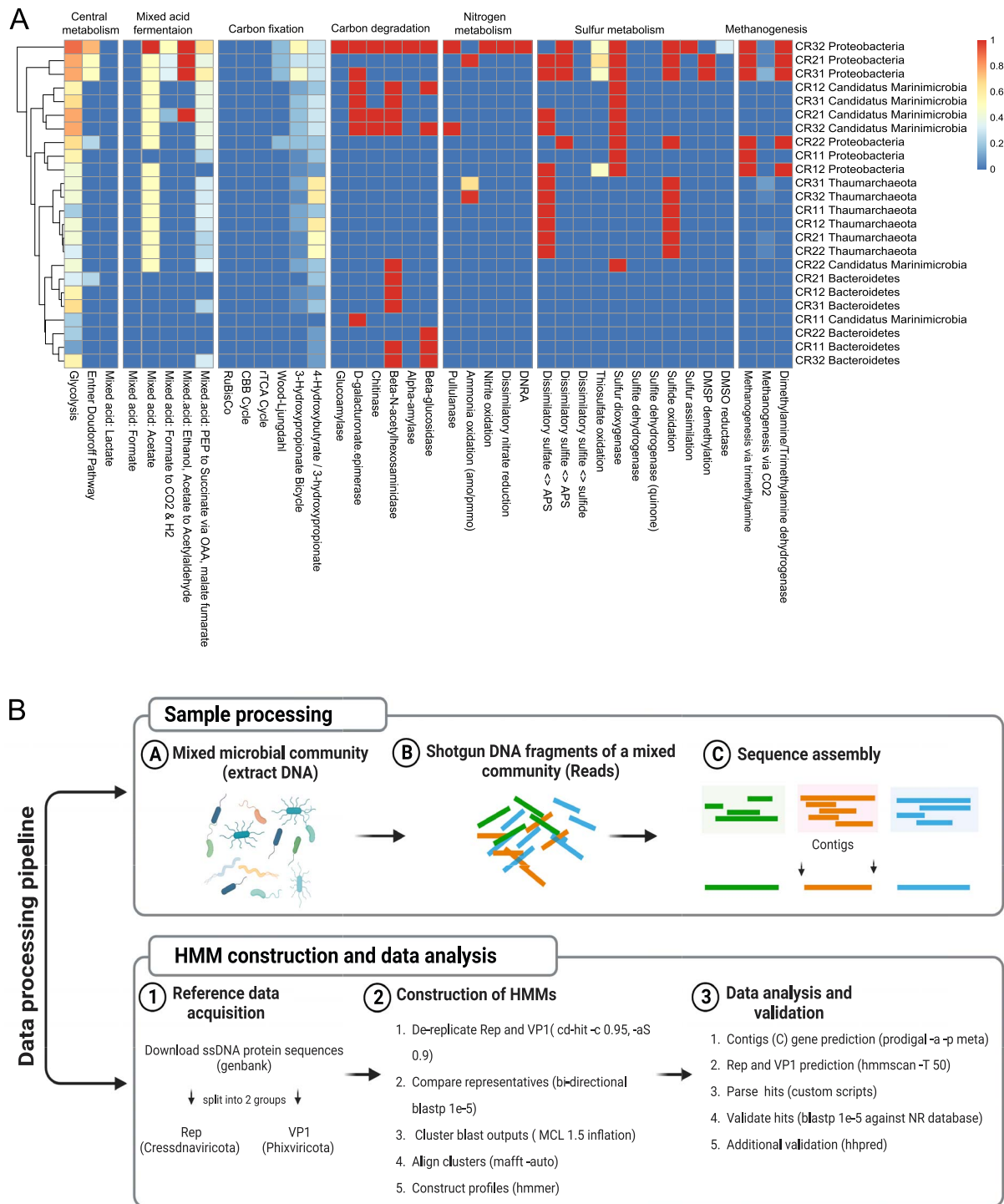


Figure 2. (A) Heatmap showing the metabolic potential of the four most overrepresented taxa, at each site. The plot shows the (B) overview of the methodological workflow followed in this study including sample processing and hidden Markov model (HMM) construction and data analysis.

Microviridae, respectively. The remaining four Rep sequences, without hits against NR database, were further validated using HHpred against PDB and pfam databases (Tables S7 and S8).

To reduce the knowledge deficit regarding these viruses, we estimated the diversity and phylogeny of CRESS sequences from the SIO. Phylogenetic analysis was used to compare 2307 Rep sequences with the 709 reported by Kazlauskas *et al.* [67] (Fig. 3). Our maximum likelihood reconstructions suggest that sequences

from the SIO may represent unclassified CRESS. These sequences were predominantly assigned with categorical Groups 1–5 as well as *Circoviridae*, which are known to infect a wide range of hosts including mammals and diatoms [112–114]. Consistent with previous studies [41, 115], these viruses were highly abundant in our samples. It is possible that these viruses may interact with phylogenetically diverse eukaryotic hosts. These hosts may include members of the division Chlorophyta and

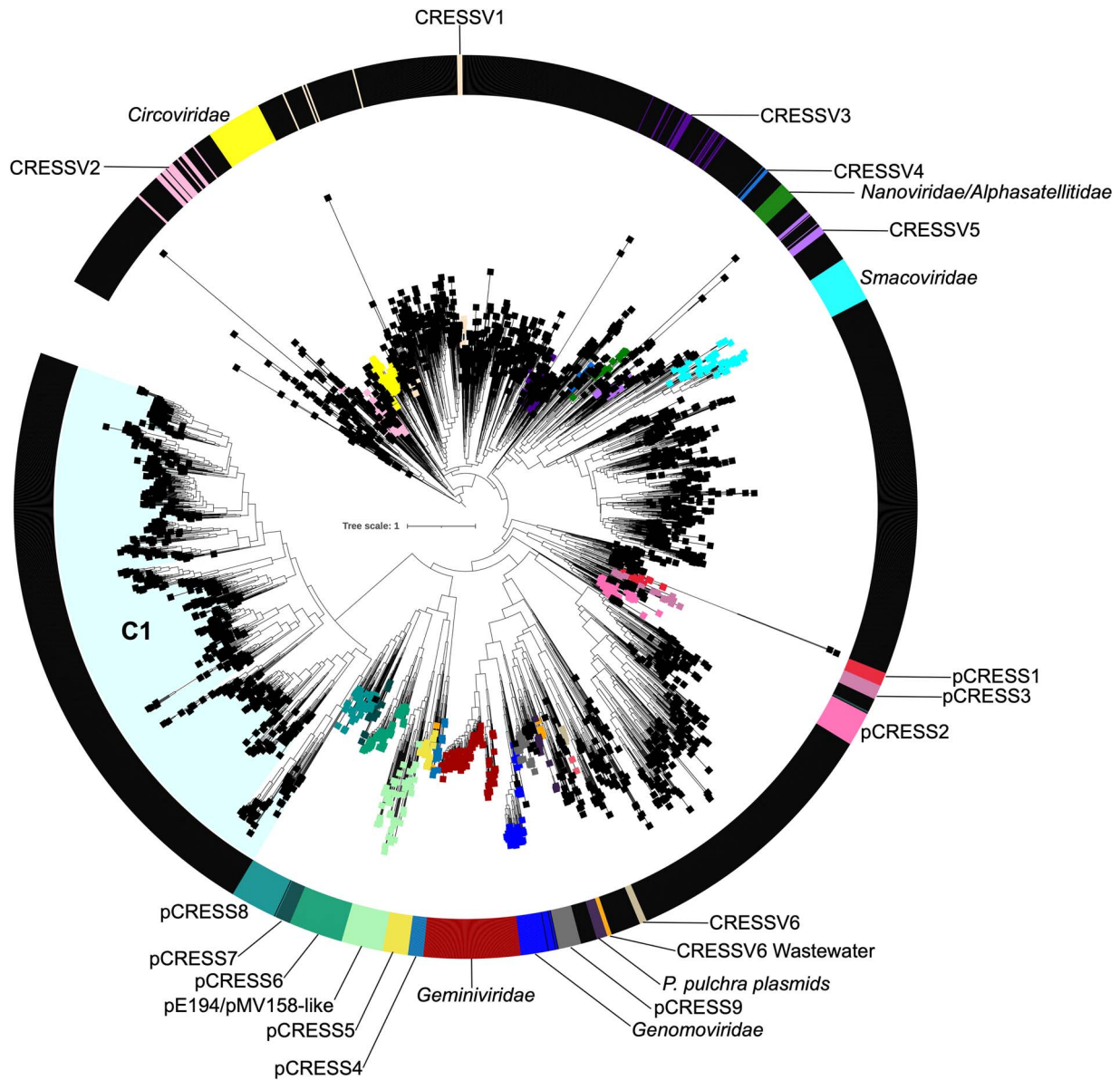


Figure 3. An unrooted maximum likelihood phylogenetic tree showing the diversity of Rep protein sequences. The black squares represent Rep proteins, from the South Indian Ocean, and CRESS DNA viruses. The other groups shown include rep proteins retrieved from a previous study by Kazlauskas et al. [67].

Streptophyta, Foraminifera, Ascomycota and the prevalent SAR supergroup (Supplementary Fig. S2). Our phylogenetic analysis revealed distinct unclassified clades, including C1 which appears to be comprised exclusively by viruses from the SIO (Fig. 3). These distinct clades expand the known diversity of marine viruses. In addition, sequences from the SIO include clades representing several as yet unidentified rep-encoding CRESS viruses. The data strongly suggests that these viruses may have a broader host range preference than previously thought. In addition, the may the environmental conditions in these deep SIO waters may select for distinct CRESS viruses adapted to these harsh conditions. The analyses further revealed that one sequence clustered with pCRESS8 and three grouped with Genomoviruses which are known plasmids and fungal pathogens, respectively [116, 117].

To explore the diversity of ssDNA viruses 2307 (Rep) and 209 (VP1) non-redundant proteins, recovered in this study, were compared with GenBank Rep and VP1 protein sequences. These

sequences were used to generate family-level clusters of sequence similarity networks (SSNs), as described previously by Krabberger et al. [118]. These SSNs included 87.1% Rep (2010) and 98.5% VP1 (206) of ssDNA protein sequences from the SIO dataset. Consistent with the phylogenetic analysis, SSN generated by comparisons of Cressdnaviricota Rep proteins showed that at least 89% (1792 of 2010) of the Rep sequences from our dataset, clustered with Circoviruses and unclassified CRESS viruses (Fig. 4). The SSNs also suggests that several sequences from our data represent distinct clusters, which were not similar to those deposited on GenBank. These divergent clusters (C2 and C3) overlapped with the distinct CRESS like viral C1 clade in the phylogenetic tree. In addition to confirming some overlap between the two methods, this result highlights the high diversity of ssDNA viruses in deep SIO waters (Fig. 4).

Contrary to CRESS signatures, HMM searches suggest that the numbers of Phixviricota-like VP1 protein sequences in the SIO

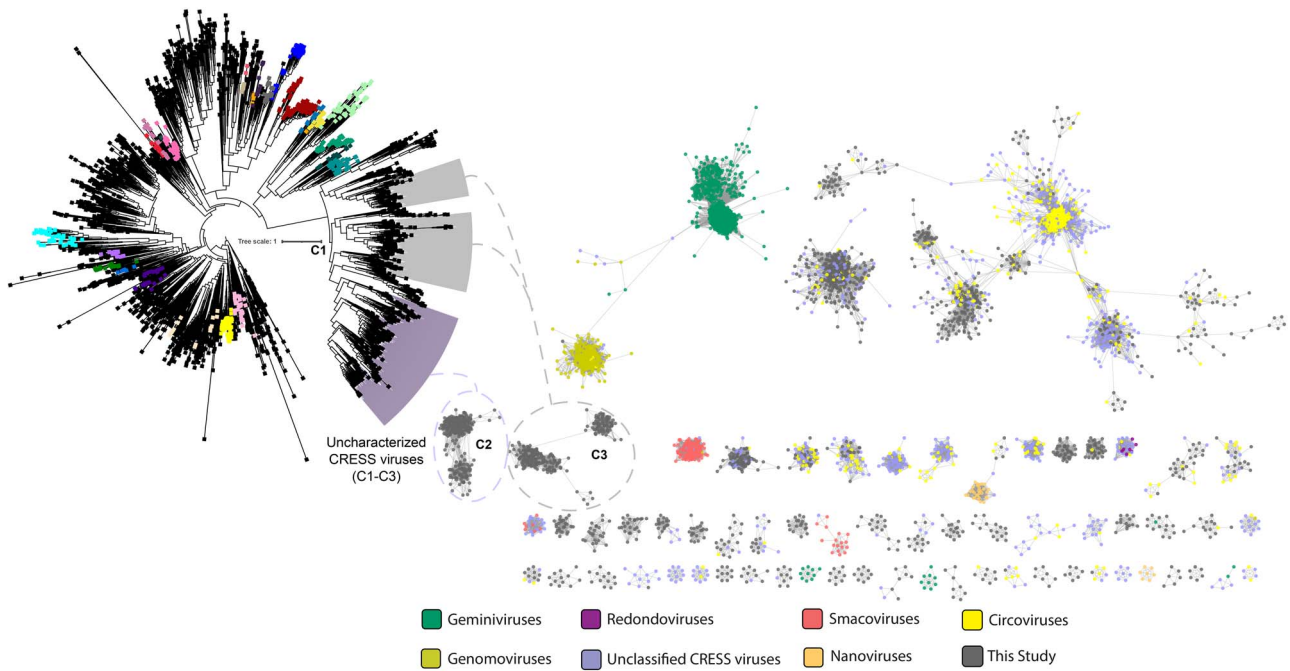


Figure 4. An unrooted maximum likelihood phylogenetic tree (A) and a sequence similarity network of rep protein sequences indicating some of the potentially novel CRESS viral families (B) from our South Indian Ocean data, annotated as C1 to C3.

may be low. Additionally, Phixviricota VP1-based SSN indicated that prokaryotic viral sequences from the SIO may be similar with known Microviruses (Supplementary Fig. S3). This result is somewhat surprising and is in contrast to previous studies, which have reported that Microviruses from benthic marine environments were highly abundant and diverse, compared with other viruses from the same family [119]. Based on our findings, we propose two possibilities. The first is that benthic Microviruses from the SIO may be less divergent due to the ecological selective pressures. Alternatively, the underrepresentation of these viruses in our data may be due to technical challenges. For instance, amplification biases during library preparation have been previously shown to obscure viral abundances in several studies [120, 121]. However, based on the remarkably low number of studies on the SIO, and the limited datasets, it is reasonable to conclude that these viruses may represent highly diverse and novel lineages. Validation of metagenomic assemblies supports this assertion and revealed that 2590 contigs from our dataset may be associated with circular genetic elements, which were previously described by Tisza *et al.* (2020) [34]. SSNs indicated that 12 and 4 clusters (of the total 161, with cluster sizes ≥ 5) were constituted by ssDNA viral Rep and Capsid associated proteins, respectively (Supplementary Fig. S4). This result suggest that the diversity of ssDNA viruses in the SIO may be as similar to levels previously reported in the North Atlantic Ocean [122] and other polar aquatic environments [123].

Taken together, the diverse ssDNA viruses identified in this study suggests that the relationships between virus and host populations may be dynamic or substantially driven by physico-chemical properties in the deep SIO [33]. However, the functional potential of eukaryotic viruses in the environment remains under-explored. Our results showing high abundances of Cressdnaviricota suggests that these viruses contribute significantly to the biological carbon pump in the SIO. Alternately, based on the phylogenetic evidence from our study it is reasonable to predict that viruses may have profound effects on ecological processes

across the water column, from the surface to the deepest depths. These effects may include recycling dissolved inorganic carbon in aphotic zones. However, studies deploying a marine snow catcher are required to confirm the role played by viruses in suspended and sinking particulate matter in the SIO. Given the importance of the SIO in global climate and nutrient circulation, these studies may shed light on the influence of these viruses on nutrient recycling of carbon and nitrogen.

Acknowledgements

We also thank the Centre for High Performance Computing (Cape Town, South Africa) and the Centre for Bioinformatics and Computational Biology, University of Pretoria, for providing computational resources. TPM acknowledges Stellenbosch University for supporting this study.

Supplementary material

Supplementary material is available at *ISME Communications* online.

Conflicts of interest

All authors declare that they have no competing interests.

Funding

We wish to thank the National Research Foundation of South Africa and the South African National Antarctic Program for supporting our work (UID 110717). We wish to acknowledge the Technology Innovation Agency for supporting the sequencing costs. The project also received funding from the European Union's Horizon 2020 research and innovation programme as part of the AtlantECO project under grant agreement No 862923.

Data availability

Raw sequence data linked to study have been deposited to the NCBI SRA under accession number PRJNA894371. Supplemental materials are available on the following link: <https://doi.org/10.6084/m9.figshare.24032697>.

References

- Castillo DJ, Dithugoe CD, Bezuidt OK et al. Microbial ecology of the Southern Ocean. *FEMS Microbiol Ecol* 2022;**98**:fiac123. <https://doi.org/10.1093/femsec/fiac123>
- Fuhrman JA, Cram JA, Needham DM. Marine microbial community dynamics and their ecological interpretation. *Nat Rev Microbiol* 2015;**13**:133–46. <https://doi.org/10.1038/nrmicro3417>
- Falkowski PG, Fenchel T, Delong EF. The microbial engines that drive Earth's biogeochemical cycles. *Science* 2008;**320**:1034–9. <https://doi.org/10.1126/science.1153213>
- Dominguez-Huerta G, Zayed AA, Wainaina JM et al. Diversity and ecological footprint of Global Ocean RNA viruses. *Science* 2022;**376**:1202–8. <https://doi.org/10.1126/science.abn6358>
- Gaia M, Meng L, Pelletier E et al. Mirusviruses link herpesviruses to giant viruses. *Nature* 2023;**616**:783–9. <https://doi.org/10.1038/s41586-023-05962-4>
- Maidanik I, Kirzner S, Pekarski I et al. Cyanophages from a less virulent clade dominate over their sister clade in global oceans. *The ISME Journal* 2022;**16**:2169–80. <https://doi.org/10.1038/s41396-022-01259-y>
- Wicaksono WA, Egamberdieva D, Cernava T et al. Viral community structure and potential functions in the dried-out Aral Sea basin change along a desiccation gradient. *mSystems*. 2023;**8**:e00994–22.
- Zayed AA, Wainaina JM, Dominguez-Huerta G et al. Cryptic and abundant marine viruses at the evolutionary origins of Earth's RNA virome. *Science* 2022;**376**:156–62. <https://doi.org/10.1126/science.abm5847>
- Ha AD, Moniruzzaman M, Aylward FO. Assessing the biogeography of marine giant viruses in four oceanic transects. *ISME Commun* 2023;**3**:43. <https://doi.org/10.1038/s43705-023-00252-6>
- Mayers KM, Kuhlisch C, Basso JT et al. Grazing on marine viruses and its biogeochemical implications. *MBio* 2023;**14**:e01921–1.
- Zimmerman AE, Howard-Varona C, Needham DM et al. Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat Rev Microbiol* 2020;**18**:21–34. <https://doi.org/10.1038/s41579-019-0270-x>
- Fuhrman JA. Marine viruses and their biogeochemical and ecological effects. *Nature* 1999;**399**:541–8. <https://doi.org/10.1038/21119>
- Luo X-Q, Wang P, Li J-L et al. Viral community-wide auxiliary metabolic genes differ by lifestyles, habitats, and hosts. *Microbiome* 2022b;**10**:1–18. <https://doi.org/10.1186/s40168-022-01384-y>
- Yi Y, Liu S, Hao Y et al. A systematic analysis of marine lysogens and proviruses. *Nat Commun* 2023;**14**:6013. <https://doi.org/10.1038/s41467-023-41699-4>
- Zhou K, Wong TY, Long L et al. Genomic and transcriptomic insights into complex virus–prokaryote interactions in marine biofilms. *ISME J* 2023;**17**:2303–12. <https://doi.org/10.1038/s41396-023-01546-2>
- Ahlgren NA, Fuchsmann CA, Rocap G et al. Discovery of several novel, widespread, and ecologically distinct marine Thaumarchaeota viruses that encode amoC nitrification genes. *ISME J* 2019;**13**:618–31. <https://doi.org/10.1038/s41396-018-0289-4>
- Hurwitz BL, Hallam SJ, Sullivan MB. Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol* 2013;**14**:R123. <https://doi.org/10.1186/gb-2013-14-11-r123>
- Roux S, Brum JR, Dutilh BE et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 2016a;**537**:689–93.
- Wittmers F, Needham DM, Hehenberger E et al. Genomes from uncultivated Pelagiphages reveal multiple phylogenetic clades exhibiting extensive auxiliary metabolic genes and cross-family multigene transfers. *Msystems* 2022;**e01522–1**.
- Brum JR, Ignacio-Espinoza JC, Roux S et al. Ocean plankton. Patterns and ecological drivers of ocean viral communities. *Science* 2015;**348**:1261498. <https://doi.org/10.1126/science.1261498>
- Chow C-ET, Suttle CA. Biogeography of viruses in the sea. *Annu Rev Virol* 2015;**2**:41–66. <https://doi.org/10.1146/annurev-virology-031413-085540>
- Luo E, Eppley JM, Romano AE et al. Double-stranded DNA viroplankton dynamics and reproductive strategies in the oligotrophic open ocean water column. *ISME J* 2020a;**14**:1304–15. <https://doi.org/10.1038/s41396-020-0604-8>
- Sussman JL, Lin D, Jiang J et al. Protein data Bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallogr D Biol Crystallogr* 1998;**54**:1078–84. <https://doi.org/10.1107/S0907444998009378>
- Gao C, Liang Y, Jiang Y et al. Viroplankton assemblages from challenger deep, the deepest place in the oceans. *iScience* 2022;**25**:104680. <https://doi.org/10.1016/j.isci.2022.104680>
- Li Z, Pan D, Wei G et al. Deep sea sediments associated with cold seeps are a subsurface reservoir of viral diversity. *ISME J* 2021;**15**:2366–78. <https://doi.org/10.1038/s41396-021-00932-y>
- Pratama AA, Bolduc B, Zayed AA et al. Expanding standards in viromics: in silico evaluation of dsDNA viral genome identification, classification, and auxiliary metabolic gene curation. *PeerJ* 2021;**9**:e11447. <https://doi.org/10.7717/peerj.11447>
- Rambo IM, Langwig MV, Leão P et al. Genomes of six viruses that infect Asgard archaea from deep-sea sediments. *Nat Microbiol* 2022;**7**:953–61. <https://doi.org/10.1038/s41564-022-01150-8>
- Suter EA, Pachiadaki MG, Montes E et al. Diverse nitrogen cycling pathways across a marine oxygen gradient indicate nitrogen loss coupled to chemoautotrophic activity. *Environ Microbiol* 2021;**23**:2747–64. <https://doi.org/10.1111/1462-2920.15187>
- Capozza P, Lanave G, Diakoudi G et al. Diversity of CRESS DNA viruses in Squamates recapitulates hosts dietary and environmental sources of exposure. *Microbiology Spectrum* 2022;**10**:e0078022. <https://doi.org/10.1128/spectrum.00780-22>
- Hillary LS, Adriaenssens EM, Jones DL et al. RNA-viromics reveals diverse communities of soil RNA viruses with the potential to affect grassland ecosystems across multiple trophic levels. *ISME Commun* 2022;**2**:1–10.
- Jansson JK, Wu R. Soil viral diversity, ecology and climate change. *Nat Rev Microbiol* 2023;**21**:296–311. <https://doi.org/10.1038/s41579-022-00811-z>
- Porter AF, Shi M, Eden J-S et al. Diversity and evolution of novel invertebrate DNA viruses revealed by meta-transcriptomics. *Viruses* 2019;**11**:1092. <https://doi.org/10.3390/v11121092>
- Reavy B, Swanson MM, Cock PJ et al. Distinct circular single-stranded DNA viruses exist in different soil types. *Appl Environ Microbiol* 2015;**81**:3934–45. <https://doi.org/10.1128/AEM.03878-14>

34. Tisza MJ, Pastrana DV, Welch NL et al. Discovery of several thousand highly diverse circular DNA viruses. *elife* 2020;**9**:e51971. <https://doi.org/10.7554/eLife.51971>
35. Mahmoudabadi G, Phillips R. A comprehensive and quantitative exploration of thousands of viral genomes. *elife* 2018;**7**:2418–24. <https://doi.org/10.7554/eLife.31955>
36. Rosario K, Duffy S, Breitbart M. Diverse circovirus-like genome architectures revealed by environmental metagenomics. *J Gen Virol* 2009;**90**:2418–24. <https://doi.org/10.1099/vir.0.012955-0>
37. Shkoporov AN, Stockdale SR, Lavelle A et al. Viral biogeography of the mammalian gut and parenchymal organs. *Nat Microbiol* 2022;**7**:1301–11. <https://doi.org/10.1038/s41564-022-01178-w>
38. Angly FE, Felts B, Breitbart M et al. The marine viromes of four oceanic regions. *PLoS Biol* 2006;**4**:e368. <https://doi.org/10.1371/journal.pbio.0040368>
39. Labonte JM, Suttle CA. Previously unknown and highly divergent ssDNA viruses populate the oceans. *ISMEJ* 2013;**7**:2169–77. <https://doi.org/10.1038/ismej.2013.110>
40. Cheng R, Li X, Jiang L et al. Virus diversity and interactions with hosts in deep-sea hydrothermal vents. *Microbiome* 2022;**10**:235. <https://doi.org/10.1186/s40168-022-01441-6>
41. Yoshida M, Takaki Y, Eitoku M et al. Metagenomic analysis of viral communities in (had) pelagic sediments. *PLoS One* 2013;**8**:e57271. <https://doi.org/10.1371/journal.pone.0057271>
42. Phoma BS, Makhalanyane TP. Depth-dependent variables shape community structure and functionality in the Prince Edward islands. *Microb Ecol* 2021;**81**:396–409. <https://doi.org/10.1007/s00248-020-01589-4>
43. Phoma S, Vikram S, Jansson JK et al. Agulhas current properties shape microbial community diversity and potential functionality. *Sci Rep* 2018;**8**:10542. <https://doi.org/10.1038/s41598-018-28939-0>
44. Miller DN, Bryant JE, Madsen EL et al. Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl Environ Microbiol* 1999;**65**:4715–24. <https://doi.org/10.1128/AEM.65.11.4715-4724.1999>
45. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**:2114–20. <https://doi.org/10.1093/bioinformatics/btu170>
46. Rodriguez-R LM, Gunturu S, Tiedje JM et al. Nonpareil 3: fast estimation of metagenomic coverage and sequence diversity. *mSystems* 2018;**3**. <https://doi.org/10.1128/msystems.00039-18>
47. Nurk S, Meleshko D, Korobeynikov A et al. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 2017;**27**:824–34. <https://doi.org/10.1101/gr.213959.116>
48. Sayers EW, Bolton EE, Brister JR et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res* 2022a;**50**:D20–6. <https://doi.org/10.1093/nar/gkab1112>
49. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2015;**12**:59–60. <https://doi.org/10.1038/nmeth.3176>
50. Huson DH, Beier S, Flade I et al. MEGAN Community edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput Biol* 2016;**12**:e1004957. <https://doi.org/10.1371/journal.pcbi.1004957>
51. Aramaki T, Blanc-Mathieu R, Endo H et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 2020;**36**:2251–2. <https://doi.org/10.1093/bioinformatics/btz859>
52. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol* 2011;**7**:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>
53. Guo J, Bolduc B, Zayed AA et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* 2021;**9**:37. <https://doi.org/10.1186/s40168-020-00990-y>
54. Ren J, Ahlgren NA, Lu YY et al. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* 2017;**5**:69. <https://doi.org/10.1186/s40168-017-0283-5>
55. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006;**22**:1658–9. <https://doi.org/10.1093/bioinformatics/btl158>
56. Nayfach S, Camargo AP, Schulz F et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol* 2021a;**39**:578–85. <https://doi.org/10.1038/s41587-020-00774-7>
57. Camargo AP, Roux S, Schulz F et al. Identification of mobile genetic elements with geNomad. *Nature Biotechnology* 2023; 1–10.
58. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* 2020;**8**:90. <https://doi.org/10.1186/s40168-020-00867-0>
59. Hockenberry AJ, Wilke CO. BACPHLIP: predicting bacteriophage lifestyle from conserved protein domains. *PeerJ* 2021;**9**:e11396. <https://doi.org/10.7717/peerj.11396>
60. Altschul SF, Gish W, Miller W et al. Basic local alignment search tool. *J Mol Biol* 1990;**215**:403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
61. Sayers EW, Cavanaugh M, Clark K et al. GenBank. *Nucleic Acids Res* 2022b;**50**:D161–4. <https://doi.org/10.1093/nar/gkab1135>
62. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 2002;**30**:1575–84. <https://doi.org/10.1093/nar/30.7.1575>
63. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772–80. <https://doi.org/10.1093/molbev/mst010>
64. Hyatt D, Chen G-L, LoCascio PF et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010;**11**:119. <https://doi.org/10.1186/1471-2105-11-119>
65. Potter SC, Luciani A, Eddy SR et al. HMMER web server: 2018 update. *Nucleic Acids Res* 2018;**46**:W200–4. <https://doi.org/10.1093/nar/gky448>
66. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 2005;**33**:W244–8. <https://doi.org/10.1093/nar/gki408>
67. Kazlauskas D, Varsani A, Koonin EV et al. Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nat Commun* 2019;**10**:1–12.
68. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;**25**:1972–3. <https://doi.org/10.1093/bioinformatics/btp348>
69. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;**5**:e9490. <https://doi.org/10.1371/journal.pone.0009490>
70. Chevallereau A, Pons BJ, van Houte S et al. Interactions between bacterial and phage communities in natural environments. *Nat Rev Microbiol* 2022;**20**:49–62. <https://doi.org/10.1038/s41579-021-00602-y>
71. Knowles B, Bonachela JA, Behrenfeld MJ et al. Temperate infection in a virus–host system previously known for virulent dynamics. *Nat Commun* 2020;**11**:1–13.

72. Suttle CA. Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol* 2007;**5**:801–12. <https://doi.org/10.1038/nrmicro1750>
73. Weinbauer MG, Rassoulzadegan F. Are viruses driving microbial diversification and diversity? *Environ Microbiol* 2004;**6**:1–11. <https://doi.org/10.1046/j.1462-2920.2003.00539.x>
74. Zhang R, Li Y, Yan W et al. Viral control of biomass and diversity of bacterioplankton in the deep sea. *Communications biology* 2020;**3**:1–10.
75. Coutinho FH, Silveira CB, Gregoracci GB et al. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nat Commun* 2017;**8**:1–12.
76. Rohwer F, Thurber RV. Viruses manipulate the marine environment. *Nature* 2009;**459**:207–12. <https://doi.org/10.1038/nature08060>
77. Wang K, Wommack KE, Chen F. Abundance and distribution of *Synechococcus* spp. and cyanophages in the Chesapeake Bay. *Appl Environ Microbiol* 2011;**77**:7459–68. <https://doi.org/10.1128/AEM.00267-11>
78. Dunlap DS, Ng TF, Rosario K et al. Molecular and microscopic evidence of viruses in marine copepods. *Proc Natl Acad Sci USA* 2013;**110**:1375–80. <https://doi.org/10.1073/pnas.1216595110>
79. Yoon HS, Price DC, Stepanauskas R et al. Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* 2011;**332**:714–7. <https://doi.org/10.1126/science.1203163>
80. Brueckner J, Martin WF. Bacterial genes outnumber archaeal genes in eukaryotic genomes. *Genome biology and evolution* 2020;**12**:282–92. <https://doi.org/10.1093/gbe/evaa047>
81. Guerrero LD, Vikram S, Cowan DA. Evidence of microbial rhodopsins in Antarctic Dry Valley edaphic systems. *Environ Microbiol*, 2017;**19**:3755–67. <https://doi.org/10.1111/1462-2920.13877>
82. Schleifer KH. Microbial diversity: facts, problems and prospects. *Syst Appl Microbiol* 2004;**27**:3–9. <https://doi.org/10.1078/0723-2020-00245>
83. Vikram S, Guerrero LD, Makhalanyane TP et al. Metagenomic analysis provides insights into functional capacity in a hyper-arid desert soil niche community. *Environ Microbiol* 2016;**18**:1875–88. <https://doi.org/10.1111/1462-2920.13088>
84. Aylward FO, Santoro AE. Heterotrophic Thaumarchaea with small genomes are widespread in the Dark Ocean. *mSystems* 2020;**5**:e00415–20. <https://doi.org/10.1128/mSystems.00415-20>
85. Bertagnolli AD, Padilla CC, Glass JB et al. Metabolic potential and in situ activity of marine Marinimicrobia bacteria in an anoxic water column. *Environ Microbiol* 2017;**19**:4392–416. <https://doi.org/10.1111/1462-2920.13879>
86. Zhang L, Kang M, Xu J et al. Bacterial and archaeal communities in the deep-sea sediments of inactive hydrothermal vents in the Southwest India ridge. *Sci Rep* 2016;**6**:25982. <https://doi.org/10.1038/srep25982>
87. Sebastián M, Giner CR, Balagué V et al. The active free-living bathypelagic microbiome is largely dominated by rare surface taxa. *ISME Commun* 2024;**4**:ycae015.
88. Schippers A, Kock D, Hoft C et al. Quantification of microbial communities in subsurface marine sediments of the Black Sea and off Namibia. *Front Microbiol* 2012;**3**:16.
89. Acinas SG, Sánchez P, Salazar G et al. Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. *Communications biology* 2021;**4**:1–15.
90. De Corte D, Muck S, Tiroch J et al. Microbes mediating the sulfur cycle in the Atlantic Ocean and their link to chemolithoautotrophy. *Environ Microbiol* 2021;**23**:7152–67. <https://doi.org/10.1111/1462-2920.15759>
91. Paoli L, Ruscheweyh H-J, Forneris CC et al. Biosynthetic potential of the global ocean microbiome. *Nature* 2022;**607**:111–8. <https://doi.org/10.1038/s41586-022-04862-3>
92. Zhou Z, Tran PQ, Kieft K et al. Genome diversification in globally distributed novel marine Proteobacteria is linked to environmental adaptation. *The ISME journal* 2020;**14**:2060–77. <https://doi.org/10.1038/s41396-020-0669-4>
93. Guan Y, Hikmawan T, Antunes A et al. Diversity of methanogens and sulfate-reducing bacteria in the interfaces of five deep-sea anoxic brines of the Red Sea. *Res Microbiol* 2015;**166**:688–99. <https://doi.org/10.1016/j.resmic.2015.07.002>
94. Newberry CJ, Webster G, Cragg BA et al. Diversity of prokaryotes and methanogenesis in deep subsurface sediments from the Nankai trough, ocean drilling program leg 190. *Environ Microbiol* 2004;**6**:274–87. <https://doi.org/10.1111/j.1462-2920.2004.00568.x>
95. Kajale S, Jani K, Sharma A. Contribution of archaea and bacteria in sustaining climate change by oxidizing ammonia and sulfur in an Arctic Fjord. *Genomics* 2021;**113**:1272–6. <https://doi.org/10.1016/j.ygeno.2020.11.005>
96. Liu Y, Beer LL, Whitman WB. Sulfur metabolism in archaea reveals novel processes. *Environ Microbiol* 2012;**14**:2632–44. <https://doi.org/10.1111/j.1462-2920.2012.02783.x>
97. Anantharaman K, Hausmann B, Jungbluth SP et al. Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. *ISME J* 2018;**12**:1715–28. <https://doi.org/10.1038/s41396-018-0078-0>
98. Yang L, Murtugudde R, Zhou L et al. A potential link between the Southern Ocean warming and the South Indian Ocean heat balance. *Journal of Geophysical Research: Oceans* 2020;**125**:e2020JC016132. <https://doi.org/10.1029/2020JC016132>
99. Roux S, Enault F, Hurwitz BL et al. VirSorter: mining viral signal from microbial genomic data. *PeerJ* 2015;**3**:e985. <https://doi.org/10.7717/peerj.985>
100. Focardi A, Ostrowski M, Goossen K et al. Investigating the diversity of marine bacteriophage in contrasting water masses associated with the east Australian current (EAC). *System Viruses* 2020;**12**:317. <https://doi.org/10.3390/v12030317>
101. Roux S, Hawley AK, Torres Beltran M et al. Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *elife* 2014;**3**:e03125. <https://doi.org/10.7554/eLife.03125>
102. Sullivan MB, Huang KH, Ignacio-Espinoza JC et al. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol* 2010;**12**:3035–56. <https://doi.org/10.1111/j.1462-2920.2010.02280.x>
103. Walvekar AS, Srinivasan R, Gupta R et al. Methionine coordinates a hierarchically organized anabolic program enabling proliferation. *Mol Biol Cell* 2018;**29**:3183–200. <https://doi.org/10.1091/mbc.E18-08-0515>
104. Kieft K, Breister AM, Huss P et al. Virus-associated organosulfur metabolism in human and environmental systems. *Cell Rep* 2021;**36**:109471. <https://doi.org/10.1016/j.celrep.2021.109471>
105. Luo E, Leu AO, Eppley JM et al. Diversity and origins of bacterial and archaeal viruses on sinking particles reaching the abyssal ocean. *The ISME journal* 2022a;**16**:1627–35. <https://doi.org/10.1038/s41396-022-01202-1>

106. Lara E, Vaque D, Sa EL *et al.* Unveiling the role and life strategies of viruses from the surface to the dark ocean. *Sci Adv* 2017;**3**:e1602565. <https://doi.org/10.1126/sciadv.1602565>
107. Aristegui J, Duarte CM, Agusti S *et al.* Dissolved organic carbon support of respiration in the dark ocean. *Science* 2002;**298**:1967. <https://doi.org/10.1126/science.1076746>
108. Gregory AC, Zayed AA, Conceicao-Neto N *et al.* Marine DNA viral macro- and microdiversity from pole to pole. *Cell* 2019;**177**:1109–1123 e1114.
109. Bi L, Yu D-T, Han L-L *et al.* Unravelling the ecological complexity of soil viromes: challenges and opportunities. *Sci Total Environ* 2022;**812**:152217. <https://doi.org/10.1016/j.scitotenv.2021.152217>
110. Khot V, Strous M, Hawley AK. Computational approaches in viral ecology. *Computational and Structural Biotechnology Journal* 2020;**18**:1605–12. <https://doi.org/10.1016/j.csbj.2020.06.019>
111. Benson DA, Cavanaugh M, Clark K *et al.* GenBank. *Nucleic Acids Res* 2013;**41**:D36–42. <https://doi.org/10.1093/nar/gks1195>
112. Delwart E, Li L. Rapidly expanding genetic diversity and host range of the Circoviridae viral family and other rep encoding small circular ssDNA genomes. *Virus Res* 2012;**164**:114–21. <https://doi.org/10.1016/j.virusres.2011.11.021>
113. Khalifeh A, Custer JM, Kraberger S *et al.* Novel viruses belonging to the family Circoviridae identified in wild American wigeon samples. *Arch Virol* 2021;**166**:3437–41. <https://doi.org/10.1007/s00705-021-05236-2>
114. Shi F, Li Q, Zou Z *et al.* The changes of immune-related molecules within the ileal mucosa of piglets infected with porcine circovirus type 2. *J Vet Sci* 2020;**21**. <https://doi.org/10.4142/jvs.2020.21.e78>
115. Chrzastek K, Kraberger S, Schmidlin K *et al.* Diverse single-stranded DNA viruses identified in chicken buccal swabs. *Microorganisms* 2021;**9**:2602. <https://doi.org/10.3390/microorganisms9122602>
116. Hao F, Wu M, Li G. Characterization of a novel genomovirus in the phytopathogenic fungus *Botrytis cinerea*. *Virology* 2021;**553**: 111–6. <https://doi.org/10.1016/j.virol.2020.11.007>
117. Tarasova E, Khayat R. A structural perspective of reps from CRESS-DNA viruses and their bacterial plasmid homologues. *Viruses* 2021;**14**:37. <https://doi.org/10.3390/v14010037>
118. Kraberger S, Schmidlin K, Fontenele RS *et al.* Unravelling the single-stranded DNA Virome of the New Zealand blackfly. *Viruses* 2019;**11**:532. <https://doi.org/10.3390/v11060532>
119. Yoshida M, Mochizuki T, Urayama SI *et al.* Quantitative viral community DNA analysis reveals the dominance of single-stranded DNA viruses in offshore upper bathyal sediment from Tohoku. *Japan Front Microbiol* 2018;**9**:75. <https://doi.org/10.3389/fmicb.2018.00075>
120. Martinez-Hernandez F, Fornas O, Lluesma Gomez M *et al.* Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat Commun* 2017;**8**:15892. <https://doi.org/10.1038/ncomms15892>
121. Roux S, Solonenko NE, Dang VT *et al.* Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ* 2016b;**4**:e2777. <https://doi.org/10.7717/peerj.2777>
122. Tucker KP, Parsons R, Symonds EM *et al.* Diversity and distribution of single-stranded DNA phages in the North Atlantic Ocean. *The ISME journal* 2011;**5**:822–30. <https://doi.org/10.1038/ismej.2010.188>
123. Yau S, Seth-Pasricha M. Viruses of polar aquatic environments. *Viruses* 2019;**11**:189. <https://doi.org/10.3390/v11020189>