A genomic resource for exploring bacterial-viral dynamics in seagrass ecosystems

Cassandra L. Ettinger[1,#], Jason E. Stajich[1,2]

[1]Department of Microbiology and Plant Pathology, University of California, Riverside, CA, United States

[2]Institute for Integrative Genome Biology, University of California, Riverside, Riverside, CA, United States

[#]**Corresponding Author:** Cassandra L. Ettinger, cassande@ucr.edu

Keywords: *Zostera marina*, eelgrass, metagenomics, viral discovery, viral diversity, phage, bacteria, carbon cycling, metagenome-assembled genomes

Abstract:

Background: Seagrasses are globally distributed marine flowering plants that play foundational roles in coastal environments as ecosystem engineers. While research efforts have explored various aspects of seagrass-associated microbial communities, including describing the diversity of bacteria, fungi and microbial eukaryotes, little is known about viral diversity in these communities.

Results: To begin to address this, we leveraged metagenomic sequencing data to generate a catalog of bacterial metagenome-assembled genomes (MAGs) and phage genomes from the leaves of the seagrass, *Zostera marina*. We expanded the robustness of this viral catalog by incorporating publicly available metagenomic data from seagrass ecosystems. The final MAG set represents 85 high-quality draft and 62 medium-quality draft bacterial genomes. While the viral catalog represents 354 medium-quality, high-quality, and complete viral genomes. Predicted auxiliary metabolic genes in the final viral catalog had putative annotations largely related to carbon utilization, suggesting a possible role for phage in carbon cycling in seagrass ecosystems.

Conclusions: These genomic resources provide initial insight into bacterial-viral interactions in seagrass meadows and are a foundation on which to further explore these critical interkingdom interactions. These catalogs highlight a possible role for viruses in carbon cycling in seagrass beds which may have important implications for blue carbon management and climate change mitigation.

Background:

Seagrasses are submerged marine flowering plants that have critical roles as foundation species in coastal ecosystems worldwide. They provide essential ecosystem services such as stabilizing the seafloor, filtering pollutants, supporting fisheries, and driving biogeochemical cycles [1–3]. One of their most significant contributions is their ability to sequester carbon in both their tissues and surrounding sediments (i.e., blue carbon) [2,4]. Despite their ecological importance, seagrass ecosystems are under increasing threat from pollution, climate change, and coastal development. Preserving these ecosystems is critical not only for their carbon sequestration potential but also for maintaining the biodiversity and ecological services they provide to coastal communities.

In recent years, there has been growing recognition of the importance of microorganisms in maintaining the health of plants [5–7]. Studies have begun to describe the composition and structure of the bacterial community associated with seagrasses, particularly *Zostera marina*, the dominant seagrass in the Northern hemisphere [8–14]. These bacterial communities are thought to have important roles in facilitating nitrogen and sulfur cycling to benefit seagrass growth and survival [8,9,11,12,15–17].

While research has largely focused on the bacterial component of these microbial communities, advances in sequencing and bioinformatics now allow us to explore the role of viruses in these systems. Phages, in particular, have been shown to influence global ocean and soil biogeochemical cycles by modulating host population dynamics through auxiliary metabolic genes (AMGs), which can directly alter bacterial metabolism to increase overall host fitness [18–21]. However, the diversity and ecological roles of phage in seagrass ecosystems remains largely unexplored.

In this study, we generated a catalog of viruses from seagrass-associated metagenomic samples, and then investigated host-viral dynamics of viral operational taxonomic units (vOTUs) and bacterial metagenome-assembled genomes (MAGs) generated from *Z. marina* leaf samples from Bodega Bay, CA. Specifically our objectives were to: (i) create a catalog of vOTUs from *Z. marina* and other seagrass species using publicly available metagenomic sequencing data, (ii) assemble a catalog of bacterial MAGs from *Z. marina* leaf tissue, and (iii) explore bacterial-phage dynamics with a focus on exploring AMGs involved in nitrogen and sulfur metabolisms.

Methods:

Sequence generation

We extracted DNA from epiphytic washes from *Z. marina* leaves as part of previous work focused on characterizing the mycobiome using high-throughput sequencing of the ITS2 region [22]. We chose three DNA extracts from that work here for deep metagenomic sequencing with the goal of obtaining high quality metagenome-assembled genomes. We provided DNA to the UC Davis Genome Center DNA Technologies Core for sequencing and library preparation. DNA libraries were sequenced on an Illumina HiSeq4000 to generate 150 bp paired-end reads.

Metagenomic processing

We trimmed sequence reads using bbDuk v. 37.68 [23] with the following parameters: qtrim=rl trimq=10 maq=10. We then mapped against and removed any reads from the metagenomes matching the available genome for *Z. marina* v. 3.1 [24] using bowtie2 v. 2.4.5 [25] and samtools v. 1.11 [26]. We co-assembled the remaining reads from all three metagenomic samples using MEGAHIT v. 1.2.9 [27].

We identified and assessed metagenome-assembled genomes (MAGs) using the anvi'o v. 7.2 workflow [28]. First, we used bowtie2 v. 2.4.5 [25] and samtools v. 1.11 [26] to obtain read coverage for each metagenomic sample against the assembly. Then we used "anvi-gen-contigs-database" to generate a database from the co-assembly and to predict open-reading frames using Prodigal v. 2.6.3 [29]. This command also identifies single-copy bacterial [30], archaeal [31], and protista [32] genes using HMMER v. 3.2.1 [33] and ribosomal RNA genes using barrnap [34]. We predicted taxonomic assignments for each gene call using Kaiju v. 1.8.2 [35] with the NCBI BLAST non redundant protein database nr including fungi and microbial eukaryotes v. v. 2020-05-25. For each individual metagenomic sample, we then used "anvi-profile" to construct an anvi'o profile for contigs >1 kbp with the "–cluster-contigs" option. Next we ran several automatic binning algorithms including MetaBAT2 v. 2.15, MaxBin v. 2.2.1, BinSanity v.0.5.4 and CONCOCT v. 1.1.0 [36–39] to generate preliminary sets of bacterial MAGs. We provided the resulting set from each algorithm to DAStool v. 1.1.2 [40] to generate a single optimal MAG set. MAGs from this set were then manually assessed for contamination and refined using "anvi-refine". After manual refinement, MAGs were further de-contaminated using MAGpurify v. 2.1.2 [41].

We assessed MAG completeness and contamination using CheckM v. 1.2.1 [42] and CheckM2 v. 1.0.1 [43]. In this work we report all identified MAGs with >80% completion and <10% contamination. We refer to MAGs as high quality if they were >90% complete with <5% contamination, and medium quality if they were >50% complete with <10% contamination per established guidelines [44]. To obtain a putative taxonomy for each MAG, we used GTDB-Tk v. 2.2.6 [45], which uses a combination of average nucleotide identity and phylogenetic placement in the context of the Genome Taxonomy Database to taxonomically identify MAGs.

Viral identification

In order to have a robust dataset to compare our recovered virus catalog in the context of other studies, we downloaded publicly available data from NCBI GenBank from nine studies [11,12,17,46–51] representing 65 metagenomic and 18 metatranscriptomic samples collected from seagrass environments (Table S1).

We filter and trimmed public metagenomic data using bbDuk v. 37.68 [23] with the following parameters: ktrim=r k=23 mink=11 hdist=1 tpe tbo qtrim=rl trimq=10 maq=10. Then for each study, samples were co-assembled using MEGAHIT v. 1.2.9. We used the co-assemblies for each study, as well as the co-assembly from this study, when identifying viral sequences using a workflow similar to Guo et al. [52]

Briefly, we identified viral sequences from metagenomic co-assemblies using VirSorter2 v. 2.2.3 [53], a tool that uses multiple random forest classifiers to predict whether a sequence contains a DNA or RNA virus, with the following parameters: --min-length 5000, --min-score 0.5, --include-groups dsDNAphage,RNA,ssDNA,lavidaviridae. We then ran CheckV v0.8.1 [54] on the VirSorter2 predicted viral sequences using the “end_to_end” workflow. To be conservative in our analyses, we removed viral sequences with a CheckV quality score of “not-determined” and “low-quality” prior to downstream analysis. We then ran VirSorter2 again on the viral sequences from the CheckV workflow with the --prep-for-dramv option. We used DRAM-v v. 1.2.2 [55] to “annotate” viral sequences and then “distill” annotations into predicted auxiliary metabolic genes (AMGs) for phage.

Virus clustering and analysis

We clustered viral sequences ≥ 10 kbp in length into 95% similarity viral operational taxonomic units (vOTUs) using dRep v. 3.2.2 [56]. We used Prodigal v. 2.6.3 [29] to predict open reading

frames in vOTUs using the -p meta option. We then provided the predicted proteins from the phage vOTUs to VContact2 v. 0.9.19, as well as predicted proteins from the INPHARED August 2023 viral reference database, to generate viral clusters (VCs) based on viral gene-sharing networks [57,58]. We further used geNomad v. 1.7.4 to assign taxonomy to phage vOTUs [59]. We used iPHoP v. 1.3.2 [60], which integrates across multiple methods in a machine learning framework to assign host taxonomy at the genus level, to predict host-virus linkages using a combination of the final bacterial MAG collection from this study and iPHoP's reference host database.

We mapped reads from each metagenome to vOTUs using bowtie2 --sensitive with a minid=0.90 to quantify vOTU relative abundance [61]. We then used SAMtools and BEDTools genomecov to obtain coverage estimates for each vOTU across each individual metagenomic sample [26,62]. We used coverM in contig mode to parse bam files and calculate the trimmed pileup coverage (tpmean) of vOTUs which displayed ≥ 75% coverage over the length of the viral sequence. Thresholds for analysis of vOTUs were based on community guidelines for length (i.e. ≥ 10 kbp), similarity (i.e. ≥ 95% similarity), and detection (i.e. ≥ 75% of the viral genome length covered ≥ 1x by reads at ≥ 90% average nucleotide identity) [63,64]. The viral relative abundance (tpmean), CheckV quality, geNomad taxonomy, iPHoP host-prediction, MAG diversity, and DRAM-v annotation results were analyzed and visualized in R v. 4.3.0 [65] using the tidyverse v. 2.0.0 and phyloseq v. 1.44.0 [66,67].

To compare viral diversity between metagenomic samples (i.e. beta diversity), we calculated the Hellinger distance, the Euclidean distance of Hellinger transformed relative abundance (tpmean) data. We performed Hellinger transformations using the transform function in the microbiome v. 1.22.0 package [68], calculated the Hellinger distance using the ordinate function in phyloseq, and then visualized these distances using principal-coordinate analysis (PCoA).

Results and Discussion

Refined viral catalog from seagrass ecosystems

In total, we recovered 17,145 predicted viral genomic fragments which we clustered at 95% average nucleotide identity into 3,633 vOTUs. To ensure a high quality viral catalog, we filtered this dataset further using community thresholds for length and quality [63,64]. The refined viral collection represents 354 viral sequences comprising 351 double-stranded DNA phage, two lavidaviridae, and one RNA phage based on VirSorter2 random forest classification, with each sequence representing a unique vOTU (Table S2). Of these sequences, nine are integrated prophage (Figure 1A). The final catalog includes 44 complete, 28 high-quality and 282 medium-quality draft viral genomic fragments (Figure 1B).

To explore the taxonomy of the viral sequences reported here, we used two complementary approaches: VContact2 [57], a cluster based method, and geNomad [59], an alignment based tool. We ran VContact2 [57] with INPHARED [58] reference genomes to cluster phage vOTUs into 136 VCs, which represent genus-level groupings based on gene-sharing networks. Of these, 35 VCs represented clusters that contained reference genomes, while the other 101 VCs were unique, potentially representing novel phage genera (Figure S1A). However, 148 vOTUs could not be assigned confidently to any VC.

Taxonomic classification of viruses was further refined using geNomad [59]. Most vOTUs (98.58% of vOTUs) were assigned to the *Caudoviricetes* class of tailed double-stranded DNA bacteriophage (Figure S1B). However, viral taxonomy is currently in revision [69], and the majority of sequences could not be confidently placed into finer taxonomic ranks: 331 vOTUs were unclassified at the order level, and 352 at the family level. While geNomad provided

higher-level taxonomic predictions for 351 of the 354 vOTUs, more comprehensive classification awaits future updates to formal viral taxonomy definitions.

The recovery of phage genomes from metagenomes varied across experimental studies (Figure 1C). Of the 354 viral sequences in the final catalog, 50 were recovered from the new metagenomic data in this study, with the remainder derived from publicly available metagenomes. Studies that contributed more viral sequences to the final catalog generally employed deeper sequencing, with average depths ranging from 9.7 to 28.2 Gb per sample, compared to studies with few recovered viral sequences (1.2 to 9.5 Gb per sample).

When we examined viral community composition across metagenomes, clustering by scientific study was evident (Figure 1D). Notably, the viral communities in the metagenomes generated in this study formed a distinct cluster. However, technical differences (e.g., sequencing depth) across datasets prevented us from drawing broader conclusions regarding viral diversity based on seagrass species or tissue type.

Despite the use of deep metagenomic sequencing and the integration of public metagenomes, we recovered a relatively small, refined catalog of viral sequences. This highlights the limitations of relying solely on short-read metagenomic sequencing to capture viral diversity, at least in seagrass ecosystems. Viromics approaches (e.g., [70]) may provide more comprehensive insights into these viral communities, as even with improved bioinformatics pipelines, true viral diversity and abundance remains elusive.

MAG collection reflects abundant bacterial groups on Z. marina leaves

We assembled 147 total bacterial MAGs including 85 high-quality draft MAGs (>90% completion, <5% contamination), and 62 medium-quality MAGs (>80% completion, <10%

contamination) (Table S3). These MAGs largely belong to the following taxonomic classes (Figure 2A): Alphaproteobacteria (43), Gammaproteobacteria (35), Bacteroidia (29), and Plactomycetia (16). Within these classes, the most frequently recovered orders were: Rhodobacterales (28), Pirellulales (14), Flavobacteriales (12), Chitinophagales (11) and Pseudomonadales (11). Notably, using GTDB-Tk 30 MAGs could not be assigned to any known genus, six were unclassified at the family level, and one lacked an assignment at the order level, highlighting potential evolutionary novelty compared to the current GTDB database. These findings suggest that seagrass ecosystems may harbor novel bacterial species that remain uncharacterized.

The taxonomic distribution of the recovered MAGs is broadly consistent with previous DNA-based surveys of *Zostera*-associated bacteria. These studies also identified Alphaproteobacteria, Gammaproteobacteria, and Bacteroidia as dominant classes, along with orders such as Rhodobacterales and Flavobacteriales [9–13]. Further a global study of *Z. marina* reported that Planctomycetia was enriched on leaves relative to surrounding water [8]. Together, these results suggest that this MAG collection complements previous DNA-based surveys, providing an opportunity for exploring deeper insight into the genomics and functional potential of abundant bacterial taxa within *Z. marina* ecosystems.

Recovered phage largely predicted to infect most frequent MAG phyla

We used iPHoP [60] to predict bacteria-virus linkages but were only able to predict bacterial hosts for 18 vOTUs (Figure 2B). Of these, hosts generally were distributed across the most frequently recovered MAG classes (i.e., Alphaproteobacteria, Gammaproteobacteria, and Bacteroidia). While the MAG collection was added to the iPHoP database to enable prediction of direct links between MAGs and viruses, only one MAG, SGMAG-05 (a Saprospiraceae bacterium in the class Bacteroidia) was predicted to be host to a virus from the high-quality

catalog. This phage, vOTU566, was identified as a prophage belonging to the Caudoviricetes. The remaining host-virus predictions were derived from iPHoP's default genome library rather than the MAG collection. The limited number of host-virus links identified between the MAG and viral collections suggests that much more work remains to uncover these interactions within seagrass ecosystems. Future studies should endeavor to sequence the bacterial community alongside deeper virome sequencing to better capture viral diversity. Additionally, employing physical linking techniques such as Hi-C may help characterize novel host-viral interactions.

Auxiliary metabolic genes point to phage role in carbon cycling

We used DRAM-v [55] to explore predicted AMG functions in the viral collection (Figure 2C). Over half of putative AMGs were unannotated (57.47%), similar to other recent studies of phage from the environment [71,72]. Given the importance of nitrogen and sulfur cycling in seagrass ecosystems, we searched the annotated AMGs for putative functions related to these processes. However no annotated AMGs had predictions related to nitrogen fixation or sulfur cycling. This absence could be biological (e.g., such genes may be more likely to be present in root or rhizosphere samples vs. leaves) or could reflect the limitations of recovering viruses from metagenomic datasets. Virome sequencing will likely be necessary to confirm whether viruses play a role in these nutrient cycles.

In contrast, we identified a variety of predicted AMGs with putative annotations of functions related to carbon utilization, particularly carbohydrate-active enzymes (CAZymes) involved in organic carbon cycling, sugar processing and plant degradation (Figure 2D). Seagrass beds are known as hot spots for carbon sequestration and key contributors to blue carbon storage [2,4]. Additionally previous work in the Mediterranean seagrass *Posidonia oceanica* has shown that rhizosphere sediments are enriched in sugars [46]. These findings may suggest a possible role

for viruses in carbon cycling in seagrass beds, which may have important implications for blue carbon management and climate change mitigation.

Conclusion

This study provides a valuable community resource and catalog of refined seagrass-associated bacterial and viral genomes, serving as a genomic foundation for future research. These comprehensive collections likely capture the most abundant groups associated with the leaves of *Z. marina*. While we found no evidence of AMGs related to nitrogen or sulfur metabolism, we instead report the presence of AMGs putatively linked to carbon utilization, particularly a diverse array of CAZymes that may be involved in organic carbon cycling. These results suggest that viruses may play an important role in carbon sequestration and blue carbon storage within seagrass ecosystems, which may have significant implications for climate change mitigation. Moving forward, research should prioritize investigating the role of phages through viromic techniques, which would offer deeper insights into viral ecology and phage contributions to carbon cycling in seagrass habitats.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The raw metagenomic sequencing data was deposited at GenBank under accession no. PRJNA1140276. MAGs with > 90% completion estimates were deposited on NCBI under this BioProject at GenBank under accession no. PRJNA1140276, while MAGs with completion estimates between 80 and 90% were deposited on Zenodo (DOI: 10.5281/zenodo.14225973). Viral genomes were also deposited on Zenodo (DOI: 10.5281/zenodo.14226038). All code used in this work has been deposited on Github (casett/Seagrass_MAG_and_Virus_Catalog) and archived in Zenodo (DOI: 10.5281/zenodo.14226514).

Competing Interests

The authors declare no competing interests.

Funding

The sequencing data in this work was generated by grants from the UC Davis H. A. Lewin Family Fellowship and the UC Davis Center for Population Biology to CLE. CLE was supported by the National Science Foundation (NSF) under a NSF Ocean Sciences Postdoctoral Fellowship (Award No. 2205744). JES is a CIFAR Fellow in the program Fungal Kingdom: Threats and Opportunities and was partially supported by NSF awards EF-2125066 and IOS-2134912. Computations were performed using the computer clusters and data storage

resources of the UC Riverside HPCC, which were funded by grants from NSF (MRI-2215705, MRI-1429826) and NIH (1S10OD016290-01A1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

CLE conceived and designed the experiments, performed sampling, analyzed the data, prepared figures and/or tables, wrote and reviewed drafts of the paper. JES reviewed drafts of the paper.

Acknowledgements

We would like to thank Jonathan A. Eisen for feedback on this manuscript. We would like to thank Katherine Dynarski (ORCID: 0000-0001-5101-9666) and Sonia Ghose (ORCID: 0000-0001-5667-6876) for their help with sample collection. We would also like to thank John J. Stachowicz for use of his scientific sampling permit, California Department of Fish and Wildlife Scientific Collecting Permit # SC 4874.

References

1. Hemminga MA, Duarte CM. Seagrass Ecology. Cambridge University Press; 2000. doi:10.1017/CBO9780511525551

2. Fourqurean JW, Duarte CM, Kennedy H, Marbà N, Holmer M, Mateo MA, et al. Seagrass ecosystems as a globally significant carbon stock. Nat Geosci. 2012;5: 505–509. doi:10.1038/ngeo1477

3. Orth RJ, Carruthers TJB, Dennison WC, Duarte CM, Fourqurean JW, Heck KL, et al. A Global Crisis for Seagrass Ecosystems. BioScience. 2006;56: 987. doi:10.1641/0006-3568(2006)56[987:agcfse]2.0.co;2

4. Mcleod E, Chmura GL, Bouillon S, Salm R, Björk M, Duarte CM, et al. A blueprint for blue carbon: toward an improved understanding of the role of vegetated coastal habitats in sequestering CO2. Front Ecol Environ. 2011;9: 552–560. doi:10.1890/110004

5. Trivedi P, Batista BD, Bazany KE, Singh BK. Plant–microbiome interactions under a changing world: responses, consequences and perspectives. New Phytol. 2022;234: 1951–1959. doi:10.1111/nph.18016

6. Vandenkoornhuyse P, Quaiser A, Duhamel M, Le Van A, Dufresne A. The importance of the microbiome of the plant holobiont. New Phytol. 2015;206: 1196–1206. doi:10.1111/nph.13312

7. Compant S, Samad A, Faist H, Sessitsch A. A review on the plant microbiome: Ecology, functions, and emerging trends in microbial application. J Advert Res. 2019;19: 29–37. doi:10.1016/j.jare.2019.03.004

8. Fahimipour AK, Kardish MR, Lang JM, Green JL, Eisen JA, Stachowicz JJ. Global-Scale Structure of the Eelgrass Microbiome. Appl Environ Microbiol. 2017;83. doi:10.1128/AEM.03391-16

9. Ettinger CL, Voerman SE, Lang JM, Stachowicz JJ, Eisen JA. Microbial communities in sediment from *Zostera marina* patches, but not the leaf or root microbiomes, vary in relation to distance from patch edge. PeerJ. 2017;5: e3246. doi:10.7717/peerj.3246

10. Cúcio C, Engelen AH, Costa R, Muyzer G. Rhizosphere Microbiomes of European Seagrasses Are Selected by the Plant, But Are Not Species Specific. Front Microbiol. 2016;7: 440. doi:10.3389/fmicb.2016.00440

11. Cúcio C, Overmars L, Engelen AH, Muyzer G. Metagenomic Analysis Shows the Presence of Bacteria Related to Free-Living Forms of Sulfur-Oxidizing Chemolithoautotrophic Symbionts in the Rhizosphere of the Seagrass *Zostera marina*. Front Mar Sci. 2018;5: 343793. doi:10.3389/fmars.2018.00171

12. Crump BC, Wojahn JM, Tomas F, Mueller RS. Metatranscriptomics and Amplicon Sequencing Reveal Mutualisms in Seagrass Microbiomes. Front Microbiol. 2018;9: 332425. doi:10.3389/fmicb.2018.00388

13. Hurtado-McCormick V, Kahlke T, Petrou K, Jeffries T, Ralph PJ, Seymour JR. Regional and Microenvironmental Scale Characterization of the Seagrass Microbiome. Front

Microbiol. 2019;10: 1011. doi:10.3389/fmicb.2019.01011

14. Ettinger CL, Williams SL, Abbott JM, Stachowicz JJ, Eisen JA. Microbiome succession during ammonification in eelgrass bed sediments. PeerJ. 2017;5: e3674. doi:10.7717/peerj.3674

15. Sun F, Zhang X, Zhang Q, Liu F, Zhang J, Gong J. Seagrass (*Zostera marina*) Colonization Promotes the Accumulation of Diazotrophic Bacteria and Alters the Relative Abundances of Specific Bacterial Lineages Involved in Benthic Carbon and Sulfur Cycling. Appl Environ Microbiol. 2015;81. doi:10.1128/AEM.01382-15

16. Wang L, Tomas F, Mueller RS. Nutrient enrichment increases size of *Zostera marina* shoots and enriches for sulfur and nitrogen cycling bacteria in root-associated microbiomes. FEMS Microbiol Ecol. 2020;96. doi:10.1093/femsec/fiaa129

17. Mohr W, Lehnen N, Ahmerkamp S, Marchant HK, Graf JS, Tschitschko B, et al. Terrestrial-type nitrogen-fixing symbiosis between seagrass and a marine bacterium. Nature. 2021;600: 105–109. doi:10.1038/s41586-021-04063-4

18. Kristensen DM, Mushegian AR, Dolja VV, Koonin EV. New dimensions of the virus world discovered through metagenomics. Trends Microbiol. 2010;18. doi:10.1016/j.tim.2009.11.003

19. Jin M, Guo X, Zhang R, Qu W, Gao B, Zeng R. Diversities and potential biogeochemical impacts of mangrove soil viruses. Microbiome. 2019;7. doi:10.1186/s40168-019-0675-9

20. Hurwitz BL, Westveld AH, Brum JR, Sullivan MB. Modeling ecological drivers in marine viral communities using comparative metagenomics and network analyses. Proc Natl Acad Sci U S A. 2014;111. doi:10.1073/pnas.1319778111

21. Breitbart MYA, Thompson LR, Suttle CA, Sullivan MB. Exploring the vast diversity of marine viruses. Oceanography. 2007;20: 135–139. Available: https://www.jstor.org/stable/24860053

22. Ettinger CL, Eisen JA. Characterization of the Mycobiome of the Seagrass, *Zostera marina*, Reveals Putative Associations With Marine Chytrids. Front Microbiol. 2019;10: 2476. doi:10.3389/fmicb.2019.02476

23. Bushnell B. BBTools. 2022. Available: https://sourceforge.net/projects/bbmap/

24. Ma X, Olsen JL, Reusch TBH, Procaccini G, Kudrna D, Williams M, et al. Improved chromosome-level genome assembly and annotation of the seagrass, *Zostera marina* (eelgrass). F1000Res. 2021;10. doi:10.12688/f1000research.38156.1

25. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology. 2009;10: R25. doi:10.1186/gb-2009-10-3-r25

26. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078–2079. doi:10.1093/bioinformatics/btp352

27. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution

for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31: 1674–1676. doi:10.1093/bioinformatics/btv033

28. Eren AM, Murat Eren A, Esen ÖC, Quince C, Vineis JH, Morrison HG, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ. 2015;3: e1319. doi:10.7717/peerj.1319

29. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11: 119. doi:10.1186/1471-2105-11-119

30. Lee MD. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics. 2019;35: 4162–4164. doi:10.1093/bioinformatics/btz188

31. Lee MD. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics. 2019;35: 4162–4164. doi:10.1093/bioinformatics/btz188

32. Delmont T. Assessing the completion of eukaryotic bins with anvi'o. 2018. Available: http://merenlab.org/delmont-euk-scgs

33. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011;7: e1002195. doi:10.1371/journal.pcbi.1002195

34. Seemann T. barrnap: BAsic Rapid Ribosomal RNA Predictor. Available: https://github.com/tseemann/barrnap

35. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun. 2016;7: 11257. doi:10.1038/ncomms11257

36. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014;11: 1144–1146. doi:10.1038/nmeth.3103

37. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7: e7359. doi:10.7717/peerj.7359

38. Graham ED, Heidelberg JF, Tully BJ. BinSanity: unsupervised clustering of environmental microbial assemblies using coverage and affinity propagation. PeerJ. 2017;5: e3035. doi:10.7717/peerj.3035

39. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome. 2014;2: 1–18. doi:10.1186/2049-2618-2-26

40. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of genomes from metagenomes via a dereplication, aggregation, and scoring strategy. Nature Microbiology. 2018;3: 836–843. doi:10.1101/107789

41. Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC. New insights from uncultivated genomes of the global human gut microbiome. Nature. 2019;568: 505–510. doi:10.1038/s41586-019-1058-x

42. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25: 1043–1055. doi:10.1101/gr.186072.114

43. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nat Methods. 2023;20: 1203–1212. doi:10.1038/s41592-023-01940-w

44. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2017;35: 725–731. doi:10.1038/nbt.3893

45. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2019. doi:10.1093/bioinformatics/btz848

46. Sogin EM, Michellod D, Gruber-Vodicka HR, Bourceau P, Geier B, Meier DV, et al. Sugars dominate the seagrass rhizosphere. Nature Ecology & Evolution. 2022;6: 866–877. doi:10.1038/s41559-022-01740-z

47. Rubio-Portillo E, Martin-Cuadrado A-B, Ramos-Esplá AÁ, Antón J. Metagenomics Unveils *Posidonia oceanica* “Banquettes” as a Potential Source of Novel Bioactive Compounds and Carbohydrate Active Enzymes (CAZymes). mSystems. 2021. doi:10.1128/msystems.00866-21

48. Miranda K, Weigel BL, Fogarty EC, Veseli IA, Giblin AE, Murat Eren A, et al. The Diversity and Functional Capacity of Microbes Associated with Coastal Macrophytes. mSystems. 2022. doi:10.1128/msystems.00592-22

49. Schorn S, Ahmerkamp S, Bullock E, Weber M, Lott C, Liebeke M, et al. Diverse methylotrophic methanogenic archaea cause high methane emissions from seagrass meadows. Proceedings of the National Academy of Sciences. 2022;119: e2106628119. doi:10.1073/pnas.2106628119

50. Fraser MW, Martin BC, Wong HL, Burns BP, Kendrick GA. Sulfide intrusion in a habitat forming seagrass can be predicted from relative abundance of sulfur cycling genes in sediments. Sci Total Environ. 2023;864: 161144. doi:10.1016/j.scitotenv.2022.161144

51. Fraser MW, Gleeson DB, Grierson PF, Laverock B, Kendrick GA. Metagenomic Evidence of Microbial Community Responsiveness to Phosphorus and Salinity Gradients in Seagrass Sediments. Front Microbiol. 2018;9: 378381. doi:10.3389/fmicb.2018.01703

52. Guo J, Vik D, Pratama AA, Roux S, Sullivan M. Viral sequence identification SOP with VirSorter2. protocols.io. 2021. Available: https://www.protocols.io/view/viral-sequence-identification-sop-with-virsorter2-btv8nn9w

53. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome. 2021;9: 37. doi:10.1186/s40168-020-00990-y

54. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV

assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol. 2021;39: 578–585. doi:10.1038/s41587-020-00774-7

55. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Res. 2020;48: 8883–8900. doi:10.1093/nar/gkaa621

56. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017;11: 2864–2868. doi:10.1038/ismej.2017.126

57. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat Biotechnol. 2019;37: 632–639. doi:10.1038/s41587-019-0100-8

58. Cook R, Brown N, Redgwell T, Rihtman B, Barnes M, Clokie M, et al. INfrastructure for a PHAge REference Database: Identification of Large-Scale Biases in the Current Collection of Cultured Phage Genomes. PHAGE. 2021;2: 214–223. doi:10.1089/phage.2021.0007

59. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, et al. Identification of mobile genetic elements with geNomad. Nat Biotechnol. 2024;42: 1303–1312. doi:10.1038/s41587-023-01953-y

60. Roux S, Camargo AP, Coutinho FH, Dabdoub SM, Dutilh BE, Nayfach S, et al. iPHoP: An integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. PLoS Biol. 2023;21: e3002083. doi:10.1371/journal.pbio.3002083

61. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012;9. doi:10.1038/nmeth.1923

62. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26: 841–842. doi:10.1093/bioinformatics/btq033

63. Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, et al. Minimum Information about an Uncultivated Virus Genome (MIUViG). Nat Biotechnol. 2018;37: 29–37. doi:10.1038/nbt.4306

64. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. Benchmarking viromics: an evaluation of metagenome-enabled estimates of viral community composition and diversity. PeerJ. 2017;5: e3817. doi:10.7717/peerj.3817

65. R Core Team. R: A Language and Environment for Statistical Computing. 2021. Available: https://www.r-project.org/

66. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One. 2013;8: e61217. doi:10.1371/journal.pone.0061217

67. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the Tidyverse. Journal of Open Source Software. 2019;4: 1686. doi:10.21105/joss.01686

68. Lahti L, Shetty S. Tools for microbiome analysis in R - Microbiome package. 2017.

Available: http://microbiome.github.com/microbiome

69. Turner D, Kropinski AM, Adriaenssens EM. A Roadmap for Genome-Based Phage Taxonomy. Viruses. 2021;13. doi:10.3390/v13030506

70. Santos-Medellin C, Zinke LA, Ter Horst AM, Gelardi DL, Parikh SJ, Emerson JB. Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural soil viral communities. ISME J. 2021;15: 1956–1970. doi:10.1038/s41396-021-00897-y

71. Ettinger CL, Saunders M, Selbmann L, Delgado-Baquerizo M, Donati C, Albanese D, et al. Highly diverse and unknown viruses may enhance Antarctic endoliths' adaptability. Microbiome. 2023;11: 103. doi:10.1186/s40168-023-01554-6

72. An L, Liu X, Wang J, Xu J, Chen X, Liu X, et al. Global diversity and ecological functions of viruses inhabiting oil reservoirs. Nat Commun. 2024;15: 6789. doi:10.1038/s41467-024-51101-6

Figure 1. Putative viral sequences identified from seagrass ecosystems. (A) The proviral status of predicted viral genomes is shown as a bar graph, with the number above the bar representing the total number of sequences. (B) CheckV quality metrics for viral sequences are shown as a bar graph with bars colored by quality and the number above the bar representing the total number of sequences in each bar. (C) Bar chart depicting the number of viral sequences identified in the metagenomic co-assembly from each study, with the number to the right of the bar representing the total number of sequences in each bar. (D) Principal-coordinate analysis (PCoA) visualization of Hellinger distances of relative abundance of viral communities across metagenomic samples. Samples are colored by study, and have shapes based on sample type.

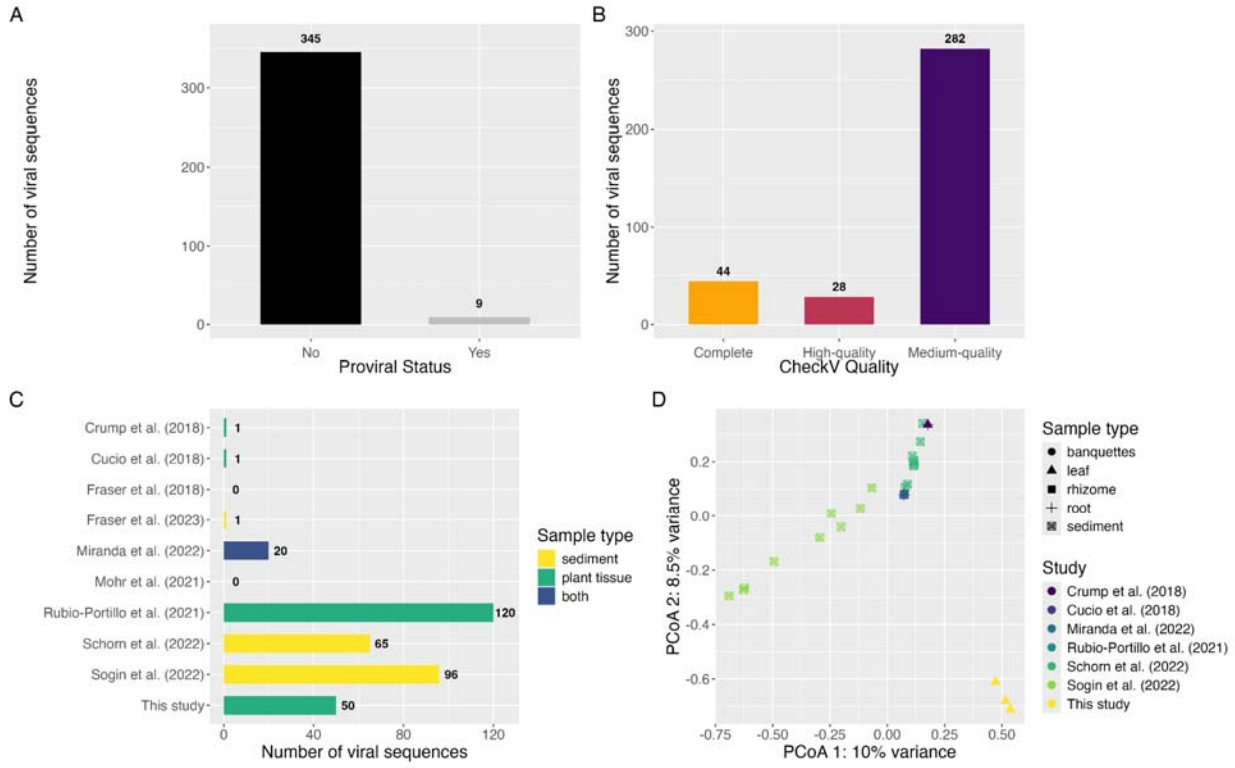
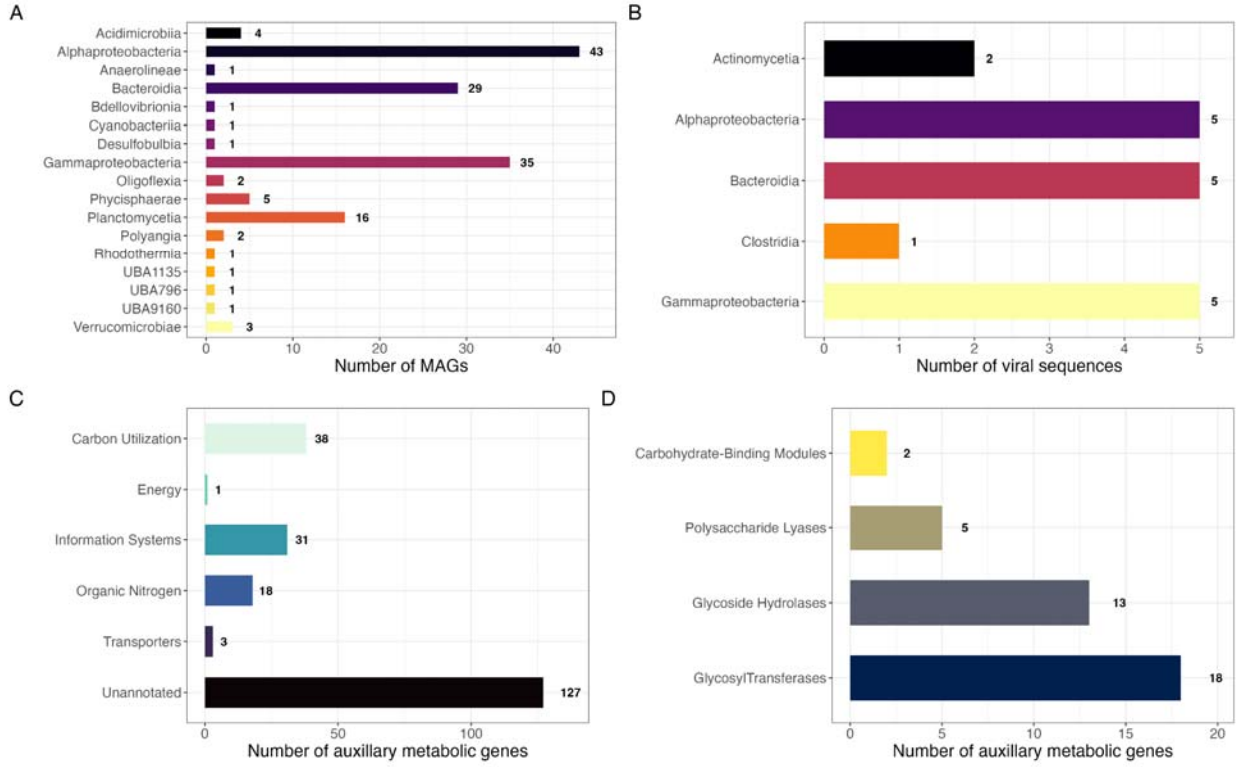
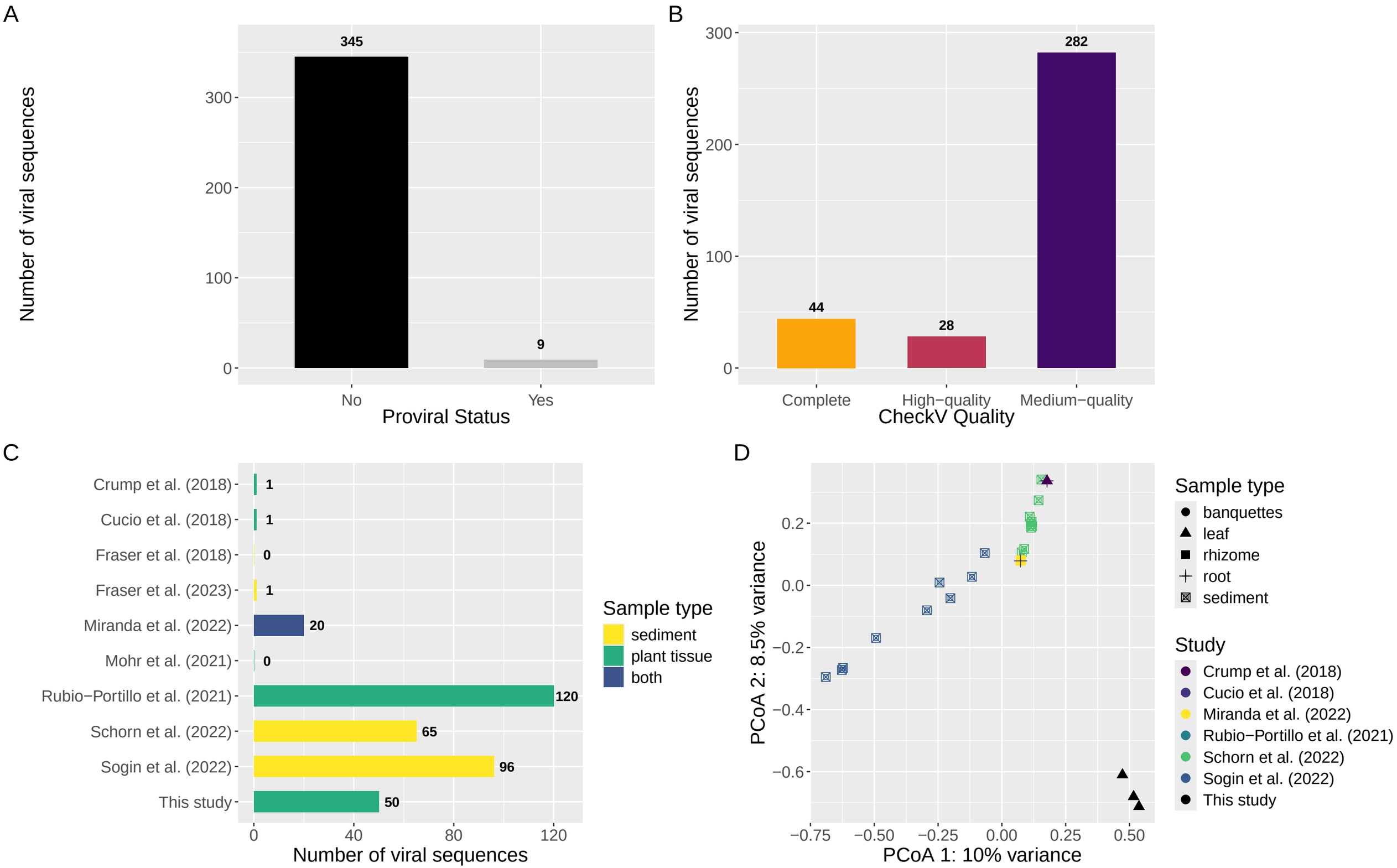
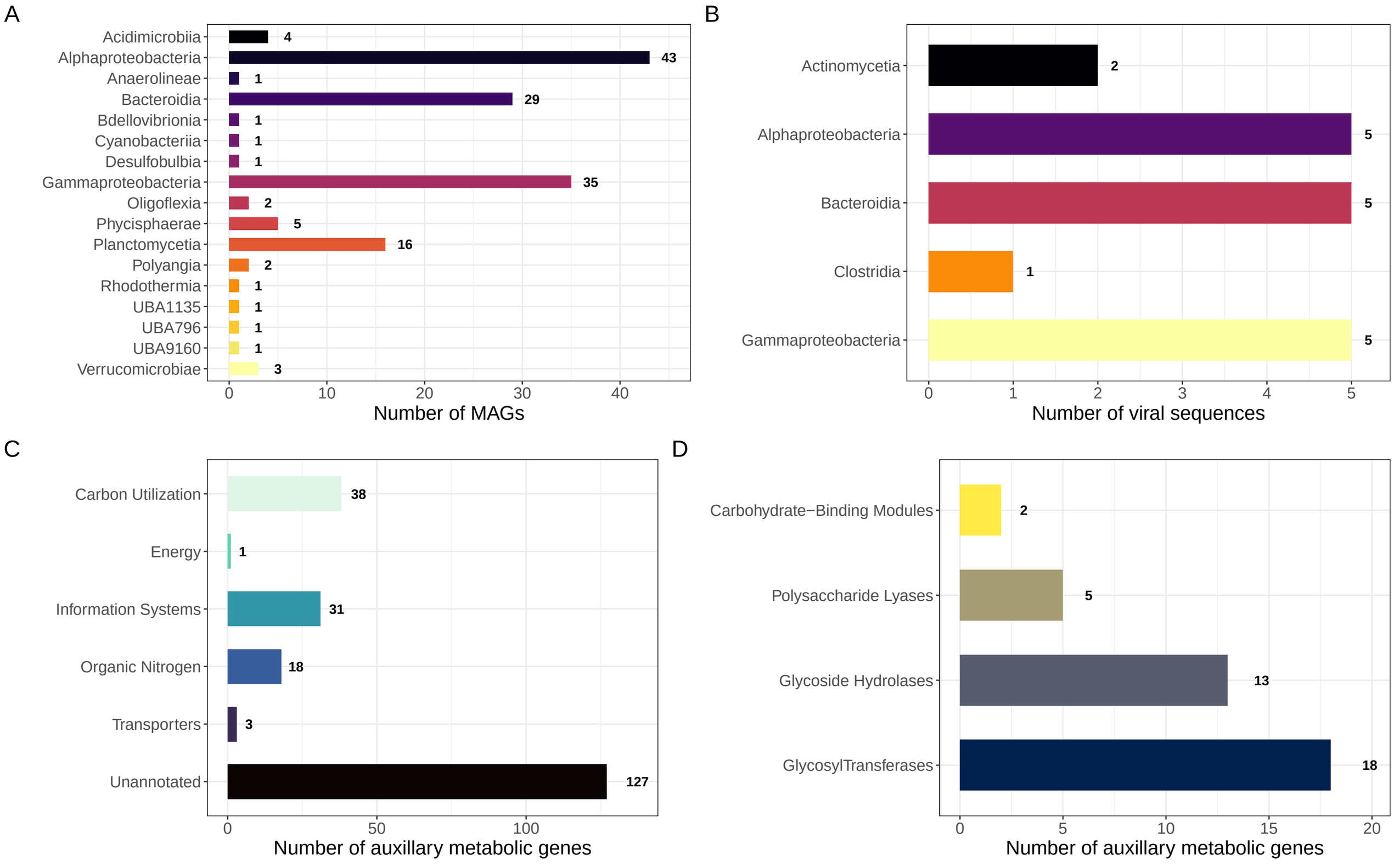


Figure 2. Viral-Bacterial interactions likely related to carbon metabolism. (A) Bar chart depicting the number of metagenome-assembled genomes (MAGs) identified in this study, colored by taxonomic class, and with the number to the right of the bar representing the total number of MAGs in each bar. (B) Bar chart depicting the viral host predictions, colored by host taxonomic class, and with the number to the right of the bar representing the total number of viral sequences in each bar. (C) Bar chart showing the number of predicted phage auxiliary metabolic genes (AMGs) summarized by DRAM-v distilled metabolic categories, with the number to the right of the bar representing the total number of AMGs in each bar. (D) Bar chart showing the number of predicted phage AMGs in different CAZyme families with-in the carbon utilization category, with the number to the right of the bar representing the total number of AMGs in each bar.




A
Number of viral sequences
300
200
100
0
345
9
No
Yes
Proviral Status
B
Number of viral sequences
300
200
100
0
44
28
282
Complete
High−quality
Medium−quality
CheckV Quality
C
Crump et al. (2018)
Cucio et al. (2018)
Fraser et al. (2018)
Fraser et al. (2023)
Miranda et al. (2022)
Mohr et al. (2021)
Rubio−Portillo et al. (2021)
Schorn et al. (2022)
Sogin et al. (2022)
This study
1
1
0
1
20
0
120
65
96
50
0
40
80
120
Number of viral sequences
Sample type
sediment
plant tissue
both
D
PCoA 2: 8.5% variance
0.2
0.0
−0.2
−0.4
−0.6
−0.75
−0.50
−0.25
0.00
0.25
0.50
PCoA 1: 10% variance
Sample type
banquettes
leaf
rhizome
root
sediment
Study
Crump et al. (2018)
Cucio et al. (2018)
Miranda et al. (2022)
Rubio−Portillo et al. (2021)
Schorn et al. (2022)
Sogin et al. (2022)
This study


A
Acidimicrobiia
Alphaproteobacteria
Anaerolineae
Bacteroidia
Bdellovibrionia
Cyanobacteriia
Desulfobulbia
Gammaproteobacteria
Oligoflexia
Phycisphaerae
Planctomycetia
Polyangia
Rhodothermia
UBA1135
UBA796
UBA9160
Verrucomicrobiae
4
43
1
29
1
1
1
35
2
5
16
2
1
1
1
1
3
0
10
20
30
40
Number of MAGs
B
Actinomycetia
Alphaproteobacteria
Bacteroidia
Clostridia
Gammaproteobacteria
2
5
5
1
5
0
1
2
3
4
5
Number of viral sequences
C
Carbon Utilization
Energy
Information Systems
Organic Nitrogen
Transporters
Unannotated
38
1
31
18
3
127
0
50
100
Number of auxillary metabolic genes
D
Carbohydrate−Binding Modules
Polysaccharide Lyases
Glycoside Hydrolases
GlycosylTransferases
2
5
13
18
0
5
10
15
20
Number of auxillary metabolic genes