NCBI accession metadata for Eukaryotic viruses encoding ribosomal protein eL40 from samples collected on KM1419 and KM1108 from Mar 2011 to Sep 2014

Website: https://www.bco-dmo.org/dataset/949101
Data Type: Cruise Results
Version: 1
Version Date: 2025-01-22

Project
» Giant viruses in the open ocean: Is large size adaptive where cells are scarce? (GVs NPSG)

Contributors	Affiliation	Role
Edwards, Kyle F.	University of Hawai'i at Mānoa	Principal Investigator
Steward, Grieg	University of Hawai'i at Mānoa	Principal Investigator
Schvarcz, Christopher	University of Hawai'i at Mānoa	Scientist
Thomy, Julie	University of Hawai'i at Mānoa	Scientist
McBeain, Kelsey	University of Hawai'i at Mānoa	Technician
Mickle, Audrey	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
This dataset contains sample collection metadata, as well as GenBank accessions and relevant Bioproject numbers for FloV-SA2 samples collected on KM1419 and KM1108 at Station ALOHA from Mar 2011 to Sep 2014. This study analyzes the genome of FloV-SA2 (phylum Nucleocytoviricota), a cultured marine virus isolated from open ocean seawater in the Pacific Ocean using a marine microalga strain (UHM3020) in the genus Florenciella (class Dictyochophyceae) as a host. The analysis highlights unique features of the genome, including the encoding of a ribosomal protein (eL40) and a group II viral rhodopsin. The research explores the affiliations and possible origins of these genes, supported by metagenomic and metatranscriptomic data indicating the presence and expression of eL40 in other giant viruses. This study expands the understanding of the metabolic versatility of eukaryoviruses and proposes new mechanisms by which these viruses can manipulate host resources and energy.

Table of Contents

Coverage

Location: Pacific Ocean North waters, Station ALOHA, 22°45' N, 158°00' W depth 25m and 45m
Spatial Extent: **N**:22.90065 **E**:-157.886274 **S**:21.246098 **W**:-159.107376
Temporal Extent: 2011-03-27 - 2014-09-15

Methods & Sampling

Note: The detailed protocols are described in Thomy et al., 2024

For eukaryote isolation, seawater sampling was carried out on March 02, 2011 from the oligotrophic open-ocean site at Station ALOHA, in the North Pacific Subtropical Gyre at a depth of 45 meters. Seawater samples were enriched with Keller (K) medium and unialgal cultures were then isolated by serial dilution to extinction. *Florenciella* sp. strain UHM3020 was further identified by small subunit ribosomal RNA gene (18S rRNA gene) sequencing. DNA was extracted from the pellets using the MasterPure Complete DNA and RNA Purification Kit (Epicentre). *Florenciella* 18S rRNA was amplified by PCR then cloned and extracted using the Zyppy Plasmid Miniprep Kit (Zymo Research). Near-full-length 18S rRNA gene for *Florenciella* was sequenced using Sanger method.

For virus isolation, seawater sampling was carried out on September 15, 2014 from the same site as described previously for the host isolation at a depth of 25 meters. Forty liters of seawater was filtered through 0.8 µm pore size filters to remove larger cells while minimizing losses of large viruses. Viral particles in the filtrate were concentrated by tangential flow filtration (TFF; 30 kDA molecular weight cut-off). The concentrate was amended with nutrients to match K medium and then used to challenge a culture of a healthy *Florenciella* culture isolated previously from the same water. Viruses in the filtrate were concentrated by TFF (30 kDa) to 300 mL volume, further concentrated to 0.5 mL by centrifugal ultrafiltration (30 kDa) and then purified in a CsCl buoyant density gradient. DNA was extracted from the virus peak in the gradient using Masterpure Complete DNA and RNA Purification Kit (LGC Biosearch Technologies). DNA was sequenced using Illumina (NextSeq System) and PacBio methods.

The FloV-SA2 complete genome sequence was deposited in GenBank with accession number PP542043 (see related datasets) as well as the ubiquitin-60S ribosomal protein eL40 gene sequence encoded in the *Florenciella* sp. host genome PP665604 (see related datasets). The gene annotations were published as Supplementary Table 1 in Thomy et al., 2024.

Data Processing Description

The FloV-SA2 genome was assembled from PacBio sequencing reads using Canu v1.0 and polished using a combination of pbalign v0.2.0.141024 and Quiver v2.0.0.

Initial gene prediction was conducted with Prokka v1.14.5. Functional annotations were performed using a BLASTp search using Diamond (v2.1.4) against:

*NCBI Refseq databases (O'Leary et al., 2015)

*InterProScan (Jones et al., 2014)

BCO-DMO Processing Description

-imported "Data-samples-info-V2-2024-10-15.xlsx" into the BCO-DMO data system
-split column containing lat and lon
-converted lon to negative to represent decimal degrees
-converted date to YYYY-mm-dd
-renamed fields to conform with BCO-DMO naming conventions
-exported file as "949101_v1_eukaryotic_viruses_flov-sa2.csv"

[table of contents | back to top]

Data Files

File
949101_v1_eukaryotic_viruses_flov-sa2.csv (Comma Separated Values (.csv), 297 bytes) MD5:c178b2d4462302a940490d28dd0494ad Primary data file for dataset ID 949101, version 1

[table of contents | back to top]

Related Publications

Buchfink, B., Xie, C., & Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. Nature Methods, 12(1), 59–60. https://doi.org/10.1038/nmeth.3176
Methods

Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., Turner, S. W., & Korlach, J. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nature Methods, 10(6), 563–569. https://doi.org/10.1038/nmeth.2474
Methods

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetijew, M., Yong, S-Y., Lopez, R., and Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics, 30(9), 1236–1240. doi:10.1093/bioinformatics/btu031
Methods

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Research, 27(5), 722–736. https://doi.org/10.1101/gr.215087.116
Methods

Lawrence, J. E., & Steward, G. F. (2010). Purification of viruses by centrifugation. Manual of Aquatic Viral Ecology, 166–181. https://doi.org/10.4319/mave.2010.978-0-9845591-0-7.166
Methods

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., … Pruitt, K. D. (2015). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Research, 44(D1), D733–D745. https://doi.org/10.1093/nar/gkv1189
Methods

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics, 30(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153
Software

Thomy, J., Schvarcz, C. R., McBeain, K. A., Edwards, K. F., & Steward, G. F. (2024). Eukaryotic viruses encode the ribosomal protein eL40. Npj Viruses, 2(1). https://doi.org/10.1038/s44298-024-00060-2
Methods

[table of contents | back to top]

Related Datasets

References

Thomy,J., Schvarcz,C.R., McBeain,K.A., Edwards,K.F. and Steward,G.F. (2024). Florenciella sp. strain UHM3020 ubiquitin-60S ribosomal protein eL40 gene, complete cds. GenBank accession number PP665604 [GenBank]. https://www.ncbi.nlm.nih.gov/nuccore/PP665604

Thomy,J., Schvarcz,C.R., McBeain,K.A., Edwards,K.F. and Steward,G.F. (2024). Florenciella sp. virus SA2 isolate FloV-SA2, complete genome. GenBank accession number PP542043. [GenBank]. https://www.ncbi.nlm.nih.gov/nuccore/PP542043

University of Hawaii. Florenciella sp. UHM3020, Florenciella sp.strain UHM3020. 2024/10. In: BioProject [Internet]. Bethesda, MD: National Library of Medicine (US), National Center for Biotechnology Information; 2011-. Available from: http://www.ncbi.nlm.nih.gov/bioproject/PRJNA1169929. NCBI:BioProject: PRJNA1169929.

University of Hawaii. Florenciella sp. virus SA2 isolate genome sequencing. 2024/10. In: BioProject [Internet]. Bethesda, MD: National Library of Medicine (US), National Center for Biotechnology Information; 2011-. Available from: http://www.ncbi.nlm.nih.gov/bioproject/PRJNA1169927. NCBI:BioProject: PRJNA1169927.

[table of contents | back to top]

Parameters

Parameter	Description	Units
Taxa_name	Taxa name of genome of FloV-SA2 (phylum Nucleocytoviricota) and Florenciella sp.strain UHM3020	unitless
Source_of_sample	Source of sample used for derived genome	unitless
Genbank_accession	Genbank assession number associated with FloV-SA2 and Florenciella sp.strain UHM3020	unitless
Bioproject_number	Bioproject number associated with FloV-SA2 and Florenciella sp.strain UHM3020	unitless
Latitude	Latitude for sample collection in decimal degrees, postive values are North	decimal degrees
Longitude	Coordinates for sample collection in decimal degrees, postive values are East	decimal degrees
Date_Isolated	Date genome was isolated from Pacific ocean at Station Aloha in Marine Viral Ecology Laboratories	unitless

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	PacBio P6 C4
Generic Instrument Name	Automated DNA Sequencer
Dataset-specific Description	DNA was sequenced using Illumina (NextSeq System) and PacBio methods.
Generic Instrument Description	A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences.

Dataset-specific Instrument Name	Illumina (NextSeq System)
Generic Instrument Name	Automated DNA Sequencer
Dataset-specific Description	DNA was sequenced using Illumina (NextSeq System) and PacBio methods.
Generic Instrument Description	A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences.

Dataset-specific Instrument Name	centrifugal ultrafiltration
Generic Instrument Name	Centrifuge
Dataset-specific Description	Viruses in the filtrate were concentrated by TFF (30 kDa) to 300 mL volume, further concentrated to 0.5 mL by centrifugal ultrafiltration (30 kDa) and then purified in a CsCl buoyant density gradient.
Generic Instrument Description	A machine with a rapidly rotating container that applies centrifugal force to its contents, typically to separate fluids of different densities (e.g., cream from milk) or liquids from solids.

Dataset-specific Instrument Name	PCR
Generic Instrument Name	Thermal Cycler
Dataset-specific Description	Florenciella 18S rRNA was amplified by PCR then cloned and extracted using the Zyppy Plasmid Miniprep Kit (Zymo Research).
Generic Instrument Description	A thermal cycler or "thermocycler" is a general term for a type of laboratory apparatus, commonly used for performing polymerase chain reaction (PCR), that is capable of repeatedly altering and maintaining specific temperatures for defined periods of time. The device has a thermal block with holes where tubes with the PCR reaction mixtures can be inserted. The cycler then raises and lowers the temperature of the block in discrete, pre-programmed steps. They can also be used to facilitate other temperature-sensitive reactions, including restriction enzyme digestion or rapid diagnostics. (adapted from http://serc.carleton.edu/microbelife/research_methods/genomics/pcr.html)

[table of contents | back to top]

Deployments

KM1419

Website	https://www.bco-dmo.org/deployment/949113
Platform	R/V Kilo Moana
Start Date	2014-09-13
End Date	2014-09-17
Description	Project: Hawaii Ocean Timeseries (HOT), Cruise 265

KM1108

Website	https://www.bco-dmo.org/deployment/731928
Platform	R/V Kilo Moana
Start Date	2011-02-27
End Date	2011-03-03
Description	Project: Hawaii Ocean Timeseries (HOT), Cruise 230

[table of contents | back to top]

Project Information

Giant viruses in the open ocean: Is large size adaptive where cells are scarce? (GVs NPSG)

Coverage: North Pacific

NSF Award Abstract:
Viruses can infect all forms of life. Viruses are highly diverse, and one aspect of diversity is size: genomes of viruses vary more than a thousandfold in length, and the size of viral particles varies nearly a millionfold. The discovery of “giant” viruses was astounding because they can be physically larger and code for more genes than many free-living microorganisms. There is growing evidence that giant viruses are widespread and diverse in the ocean, but much about their ecology remains unknown. What critical ecological tradeoffs vary with virus size, allowing small and large viruses to coexist? Do these tradeoffs cause the distribution of virus sizes to vary across habitats? This project aims to answer these questions for viruses that infect phytoplankton, the microscopic plants that are the foundation of ocean productivity. This research can also influence a diverse array of scientific fields because virus size varies greatly in other ecosystems and host-associated microbiomes. The fundamental constraints on size may be broadly similar across systems, but the processes driving virus size have not been thoroughly investigated in any of them. This project supports the training of a postdoctoral researcher, two graduate students, and undergraduate students in integrative science that includes field, laboratory, and modeling components. National Science Foundation-supported Research Experience for Undergraduates and Tribal Colleges and Universities programs at UH Manoa that serve Pacific Islanders and other underrepresented groups are used for recruiting students. In addition, science outreach at public events in Hawai’i includes an interactive game to communicate ideas about giant viruses and their role in the ocean.

Large viruses may have four advantages over smaller viruses: i) ability to infect a greater diversity of host genotypes, ii) better control of host metabolism, iii) large enough size to enter host cells by ingestion, and iv) greater persistence in the extracellular environment. These advantages may compensate for the advantages held by smaller viruses: higher contact rates with their hosts and greater offspring number per infection. The advantages of large size may be more consequential in oligotrophic habitats, where the microbial eukaryote community is primarily small phagotrophic flagellates (mixotrophs and heterotrophs), at low population densities, with resource-limited growth. The project goals are: (1) To test whether giant viruses indeed dominate in the oligotrophic ocean compared to a productive coastal location, as suggested by initial observations of this research team; (2) To test the above four hypotheses about the advantages of large size by conducting laboratory experiments with diverse viral isolates, and (3) To use an eco-evolutionary model of eukaryotic microbes and their viruses to explain observed size patterns.

This award reflects NSF's statutory mission and has been deemed worthy of support through evaluation using the Foundation's intellectual merit and broader impacts review criteria.

[table of contents | back to top]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-2129697
Simons Foundation (Simons)	566853
NSF Office of Integrative Activities (NSF OIA)	OIA-1736030

[table of contents | back to top]