

Supplementary Table Information

Marine fungi harbor, and are co-diversified with, remnants of giant endogenous viral elements

By: Syrena K. Whitner^{1†}, Anthony S. Amend¹, and Alexander I. Culley¹.

¹Pacific Biosciences Research Center, University of Hawaii at Mānoa, Honolulu, HI, USA

[†]Corresponding author: Syrena Whitner (email: swhitner@hawaii.edu)

Supplementary Table 1. Descriptions of HMM marker sets used in analyses.

Supplementary Table 2. Description of fungal homologous genes used in the LLM analysis.

Supplementary Table 3. Reference NCLDV and fungal homologs used in the LLM analysis of GVOG0461 / TopoII.

Supplementary Table 4. Reference NCLDV and fungal homologs used in the LLM analysis of GVOG0013 / SFII.

Supplementary Table 5. Description of Myconoviridae isolates used in this study to generate the phylogeny in Figure 1. Accessed from Myers & Schulz. 2024. Giant viruses in early-diverging fungi: Myconoviridae [Data set]. University of Michigan - Deep Blue. <https://doi.org/10.7302/p5ay-ng61>

Supplementary Table 6. Hierarchical taxonomy based on a benchmarked phylogeny of 1,382 Nucleocytoviricota genomes from ICTV used to generate the phylogeny in Figure 1. Accessed from <https://faylward.github.io/GVDB/> (Aylward et al., 2021)

Supplementary Table 7. Description of Giant Virus Orthologous Groups (GVOGs) used to annotate viral-like genes in this study. There are a set of 8,863 protein families found in Nucleocytoviricota. Accessed from <https://faylward.github.io/GVDB/> (Aylward et al., 2021)

Supplementary Table 8. Viral Region statistics (generated via ViralRecall) including isolate ID, contig_id (originating from fungal genome assembly, containing viral region), contig length, and per contig coverage.

Supplementary Table 9. Genome Statistics for combined viral regions including number of viral regions per isolate, N50 of viral regions, and GC content of viral regions.

Supplementary Table 10. Genome statistics for host fungal genome assemblies. Columns B-F generated by QUAST, column G calculation detailed in methods.

Supplementary Table 11. BUSCO genome-mode completeness assessment of the 18 marine fungal isolates analyzed in this study. Genome assemblies were evaluated against the fungi_odb10 lineage dataset (n = 758 BUSCO groups) using BUSCO v6.0.0 in genome mode. Columns report the percentage and raw count of complete (C), single-copy (S), duplicated (D), fragmented (F), and missing (M) BUSCOs, as well as the percentage and count of complete BUSCOs containing internal stop codons.

Supplementary Table 12. Detailed description of software used during data analysis.