Supplementary Material

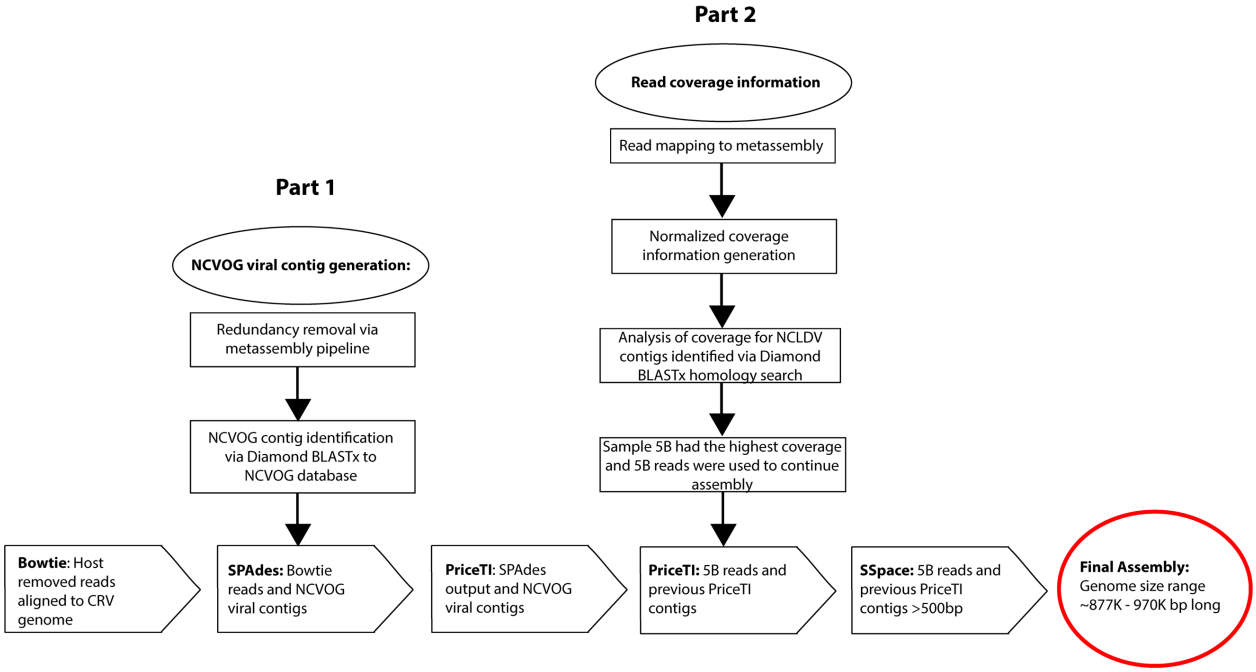
1 Supplementary Methods

Bacterial Negative Control
The negative control was a simulated bacterial metagenome (NC_016603.1) created via the Grinder program (parameters: -cf 50 -rd 150) (Angly et al. 2012). This simulated metagenome was run through Prodigal, and aligned against the Viral RefSeq Protein database using diamond BLASTx. The histogram of resulting e-values was analyzed, and the e-value at the 95% confidence interval ($p <= 0.05$) was $1.7e^{-13}$.

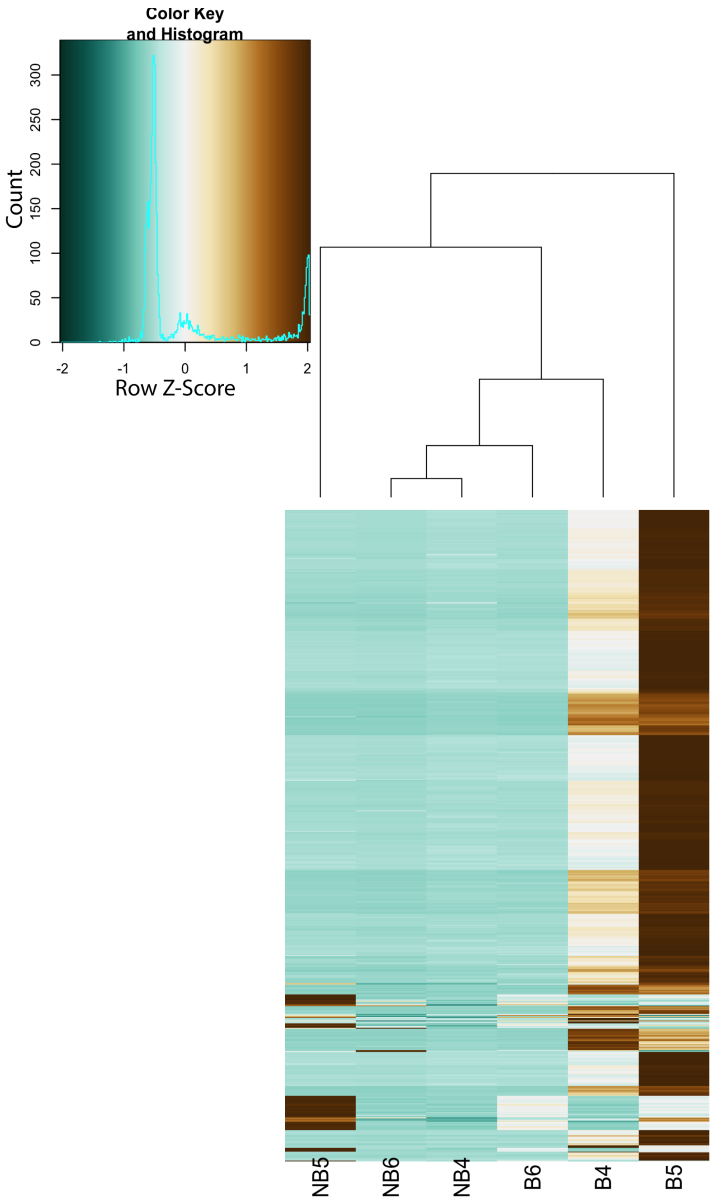
NCLDV genome recruitment average coverage calculations
tBLASTx results were converted to sam files via blast2sam.pl, which were then converted to sorted bam files (samtools v1.9). The coverage, or number of similarities, at each NCLDV genome position (i.e. base), was calculated with bedtools (v2.25.0) genomecov command (-d parameter). The average coverage was calculated by summing the coverage at each base and dividing by the total length of the NCLDV reference genome.

2 Supplementary Figures and Tables

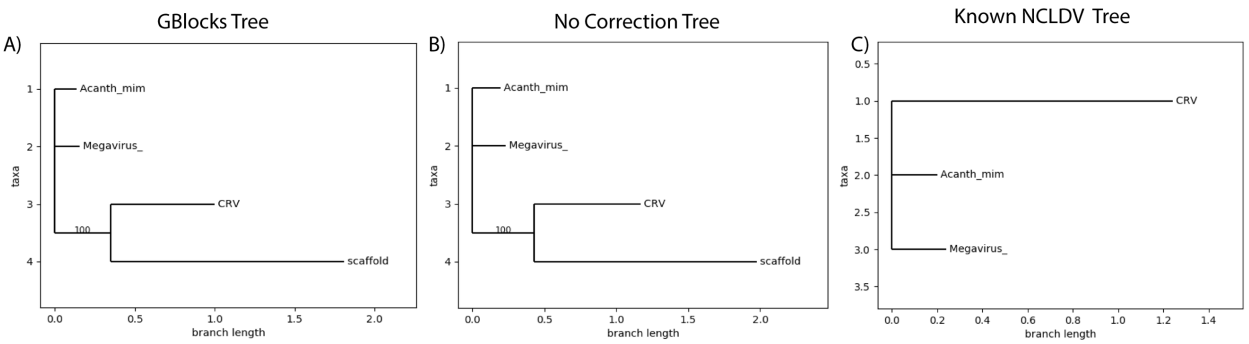
2.1 Supplementary Figures



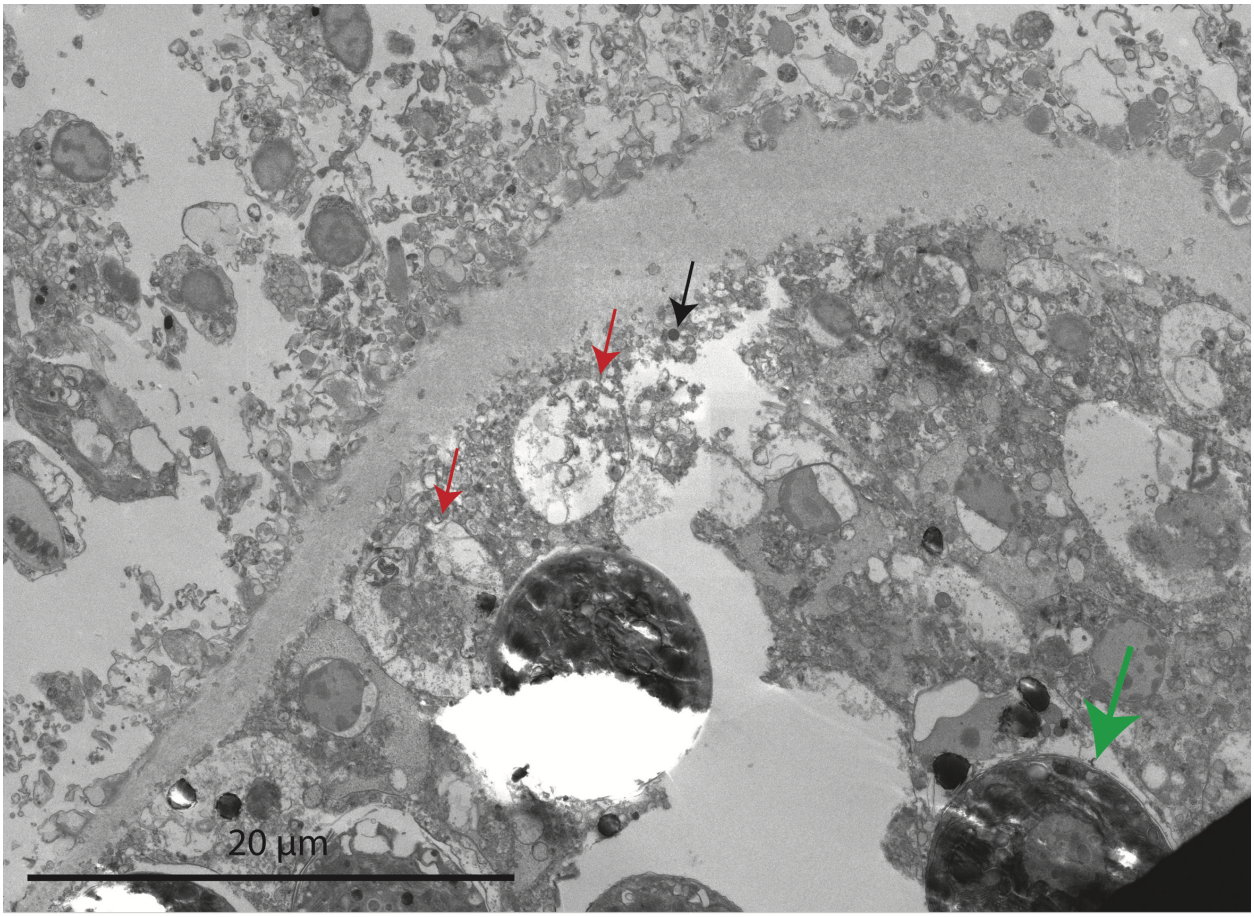
Supplemental Figure 1. Two-part assembly workflow used to generate a coral giant virus genome. Part 1 utilizes reads mapped to the CroV genome as well as meta-assembly contigs with similarities to NCVOG. Part 2 utilizes a single informative read library from sample 5B to continue assembly and for a final SSPACE elongation.



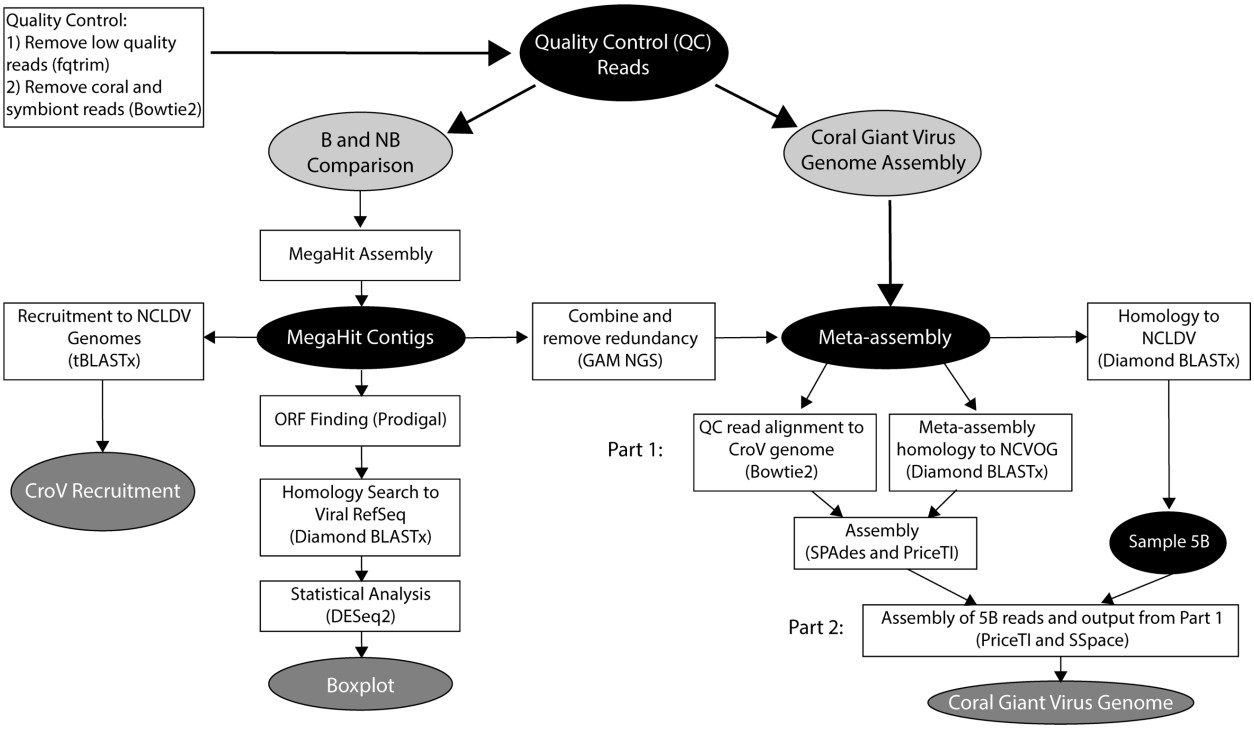
Supplemental Figure 2. Meta-assembly contigs were searched against the NCVOG database to identify contigs with significant homology to NCLDV (i.e. NCLDV contigs). Reads from each sample library (columns) were recruited to these NCLDV contigs (rows). Darker brown color (higher Z-Score) indicates higher read recruitment to NCLDV contigs, lighter blue/green color (lower Z-Score) indicates lower read recruitment to NCLDV contigs. Sample 5B read library was chosen as the most informative library for generating the final assembly because of high read recruitment to NCLDV contigs from the meta-assembly.



Supplemental Figure 3. Phylogenetic trees for our coral giant virus draft genome (noted as 'scaffold') against three genes from *Acanthamoeba polyphaga Mimivirus* (NC_014649.1, noted as 'Acanth_mim')), *Cafeteria roenbergensis virus* BV-PW1 (NC_014637.1, noted as 'CRV'), and *Megavirus chiliensis* (JN258408.1, noted as 'Megavirus_'; Supplemental Table 2 provides NCBI accession numbers of genes used for each NCLDV). A) GBlocks corrected tree, B) Tree without GBlocks correction, C) Phylogenetic tree for the 3 known NCLDV genes. Based on these results we could not resolve the phylogenetic placement of the coral giant virus genome based on known NCLDV genes.

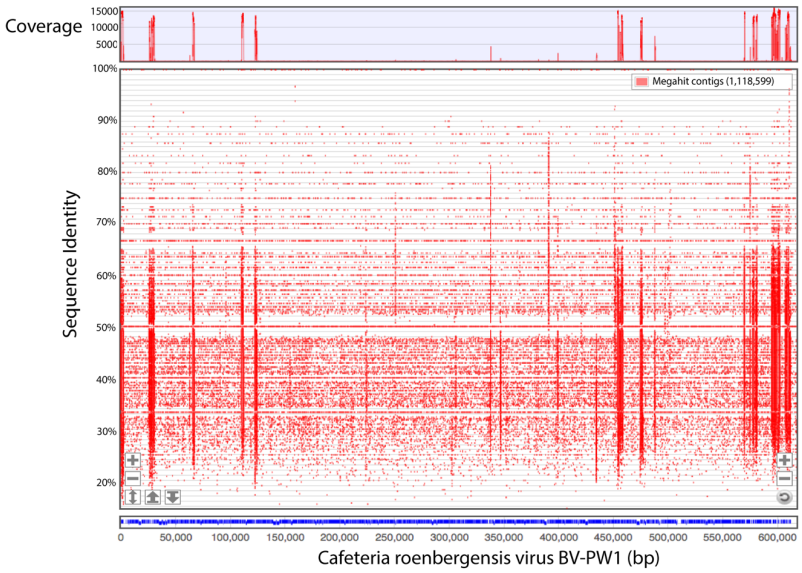


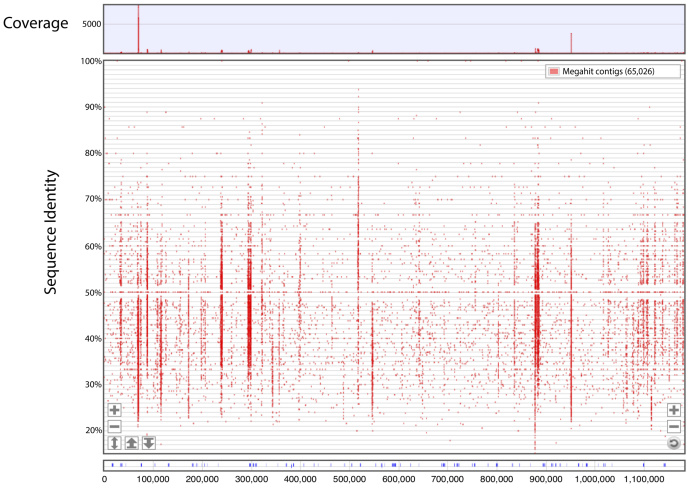
Supplemental Figure 4. Bleached coral EM showing coral cell sloughing (top left) in the epidermis, coral cells (red arrows), and Symbiodiniaceae (green arrow) within the coral gastroderm. Megavirus-like VLP with hair-like projections indicated by black arrow (close-up in main text Fig. 3B) appears outside of coral cells within the gastroderm.



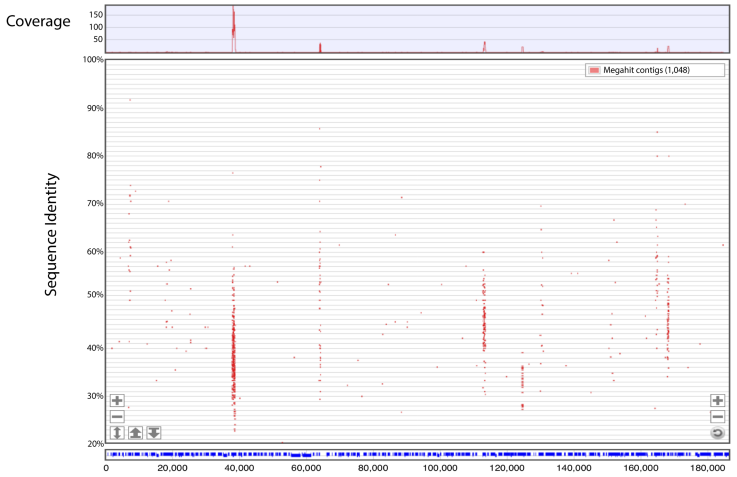
	Contig Statistics		
Assembly	N50	Average Length	# of Contigs
NCVOG virus contigs	2,452	1,942	1,079
SPAdes	2,771	2,163	929
PriceTI	3,522	2,992	491
PriceTI	5,103	4,479	196
SSpace	5,103	4,572	192

Supplemental Figure 5. Full workflow for B and NB comparisons and coral giant virus assembly. Coral giant virus assembly flow-chart and contig statistics table. The main steps of the assembly workflow as well as the intermediate steps that referenced NCVOGs and specified the most informative samples to continue assembly are shown on the right side of the figure. The table shows various assembly statistics for each main step of the coral giant virus assembly workflow.

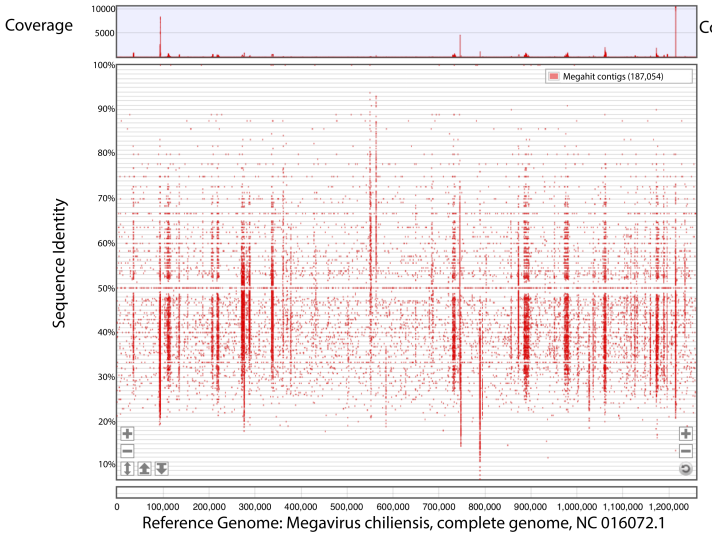




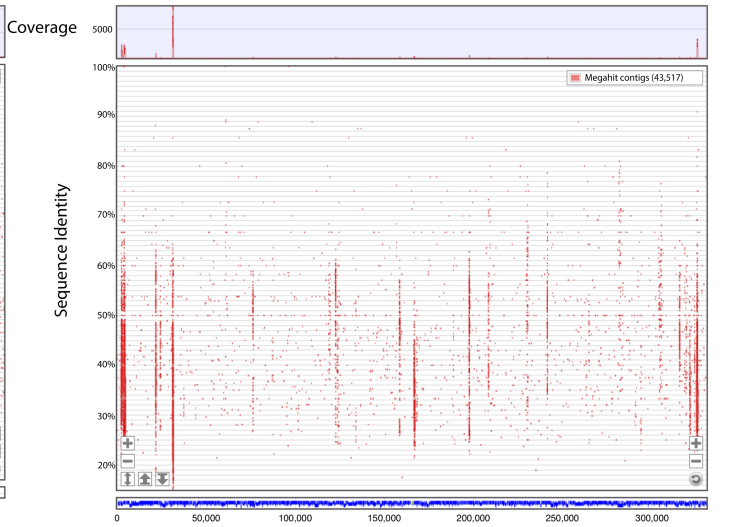
Reference Genome: Acanthamoeba polyphaga mimivirus, complete genome, NC 014649.1



Reference Genome: Lymphocystis disease virus - isolate China, complete genome, NC 005902.1

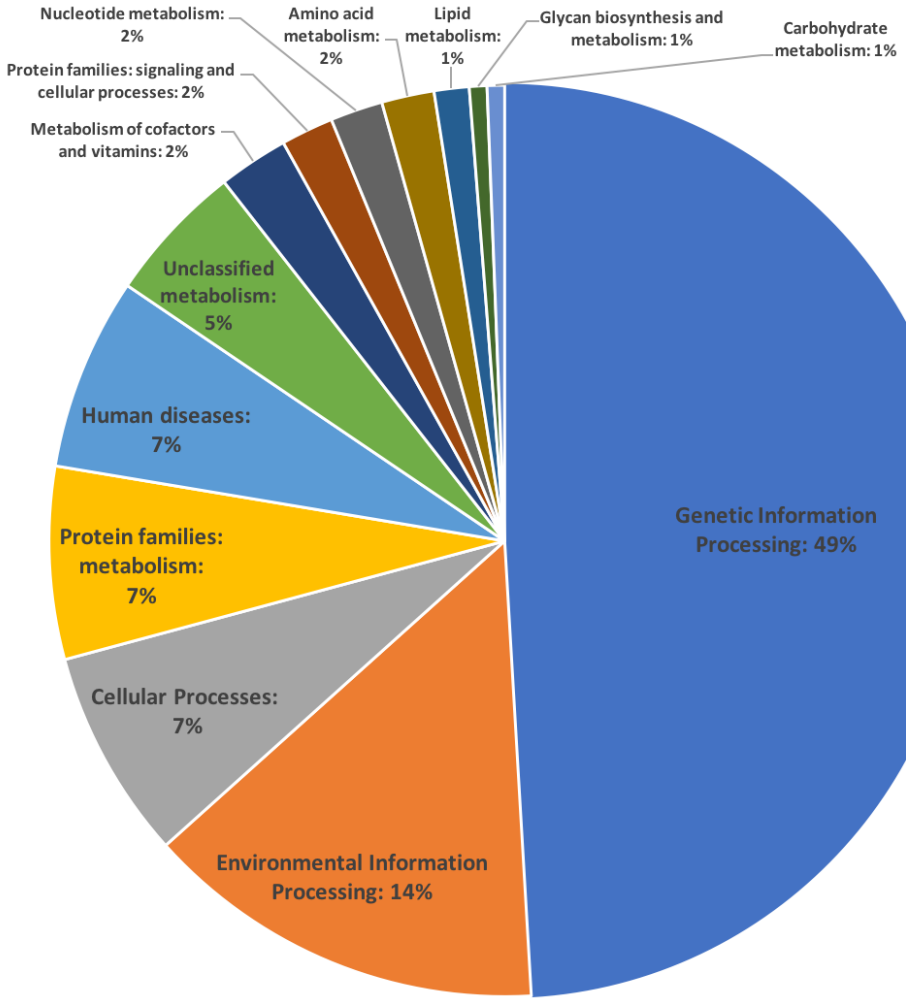


Reference Genome: Megavirus chiliensis, complete genome, NC 016072.1



Reference Genome: Paramecium bursaria Chlorella virus 1, complete genome, NC 000852.5

Supplemental Figure 6. Each recruitment panel shows MegaHit contig recruitments to the analyzed NCLDV: *Cafeteria roenbergensis virus* (CroV) BV-PW1 genome (NC 014837.1, length: 617,453bp), *Acanthamoeba polyphaga mimivirus* (NC_014649.1, length: 1,181,549), *Lymphocystis disease virus* (NC_005902.1, length: 208,501), *Megavirus chiliensis* (NC_016072.1, length: 1,259,197), and *Paramecium bursaria Chlorella virus* (NC_000852.5, length: 330,611). NCLDV genomes are noted on the x-axis and the number of similarities are indicated in the legend on the top right of the sequence identity plot. Peaks in each top coverage plot indicate recruitment coverage over genome regions. For all NCLDV analyzed, most recruitment occurred below 70% sequence identity. CroV had the most recruitments out of all NCLDV analyzed (see Table 1 in main text). Distinct bands across the 50% and 65% sequence identities may have resulted from artefacts within the recruitment program algorithm. Recruitment plot from MGAViewer.



Supplemental Figure 7. From the 652 ORFs generated from the final assembly, 161 (24.7%) had functional annotations. Chart shows the percentage of annotations to each functional category. Search was conducted via GhostKOALA against the KEGG Genes genus_prokaryotes + family_eukaryotes + viruses database.

2.2 Supplementary Tables

Supplemental Table 1. Changes in read numbers at each step of the contig assembly workflow.

	Workflow Step			
Sample	# reads generated after sequencing	# reads post quality trimming	# reads post host read removal	# MegaHit contigs generated
4B	71,989,873	63,555,592	60,719,967	575,514
5B	56,952,950	56,718,192	56,058,962	866,365
6B	73,749,506	69,518,044	67,823,249	728,465
4NB	77,183,161	72,775,920	35,242,651	418,864
5NB	84,334,865	83,406,191	82,488,153	471,607
6NB	60,730,815	52,035,997	36,517,743	415,158

Supplemental Table 2. NCLDV genes and reference numbers used for phylogenetic analysis (Supplemental Figure 3 for phylogenetic trees).

	Gene and NCBI Reference Number		
NCDLV	ribonucleoside-diphosphate reductase	DnaJ/Hsp40	DNA topoisomerase 2
Acanthamoeba polyphaga Mimivirus (NC_014649.1)	YP_003986815.1	YP_003986758.1	YP_003986988.1
Cafeteria roenbergensis virus BV-PW1 (NC_014637.1)	YP_003970085.1	YP_003969759.1	YP_003970179.1
Megavirus chiliensis (JN258408.1)	YP_004894645.1	YP_004894410.1	YP_004894454.1

Supplemental Table 3. Viral contig similarities (noted as 'hits') to viral families. Adjusted p-values from DeSeq2 comparison (RStudio, Wald test and parametric fit type) on B and NB samples.

Viral Family	B Relative % Hits	NB Relative % Hits	Adjusted p-value
Unclassified phage	0.81	3.01	0.0029
Baculoviridae	4.51	1.47	0.0166
Myoviridae	6.13	18.68	0.0166
Polydnaviridae	2.14	0.53	0.0166
Retroviridae	3.91	1.36	0.0166
Herpesviridae	1.29	0.11	0.0473
Podoviridae	0.38	1.40	0.0473
Siphoviridae	3.11	6.53	0.0473
Poxviridae	9.77	3.39	0.0587
Pandoraviridae	3.07	1.42	0.1661
Unclassified virophage	2.44	3.37	0.2463
Marseilleviridae	2.90	1.68	0.3986
Microviridae	8.63	18.07	0.8833
Adenoviridae	0.04	0.12	0.9913
Alloherpesviridae	0.09	0.08	0.9913
Ascoviridae	0.48	0.04	0.9913
Asfarviridae	0.06	0.11	0.9913
Bicaudaviridae	0.03	0.02	0.9913
Bidnaviridae	0.33	0.50	0.9913
Caulimoviridae	11.37	6.72	0.9913
Hytrosaviridae	0.19	0.19	0.9913
Iridoviridae	10.34	7.73	0.9913
Lavidaviridae	0.03	0.00	0.9913
Mimiviridae	8.81	7.02	0.9913
Nimaviridae	0.05	0.00	0.9913
Nudiviridae	0.43	0.21	0.9913
Phycodnaviridae	14.63	12.85	0.9913
Pithoviridae	0.40	0.23	0.9913
Unclassified archaeal virus	0.00	0.02	0.9913
Unclassified eukaryotic virus	3.52	3.10	0.9913
Unclassified virus	0.09	0.05	0.9913

Supplemental Table 4. Final assembly ORF search against the JGI IMG VR all proteins database, producing 124 annotations analyzed at 3 categories: 1) Viral cluster, 2) Host, 3) Ecosystem. Annotations, number of annotations, and % out of 124 are shown per category. Number of annotations to iVG reference genome families are included in dotted-line table section.

Annotation Category	Annotation: number of annotations (% out of 124)	
Viral cluster	Environmental: 53 (42.7%)	iVG and prophages*: 33 (26.6%)
	Host-associated: 21 (16.9%)	Engineered: 16 (12.9%)
	Freshwater: 1 (0.81%)	
Host	NA: 64 (51.6%)	Bacteria: 36 (29.0%)
	Eukaryote: 24 (19.3%)	
Ecosystem	iVG*: 33 (26.6%)	Aquatic Marine: 28 (22.5%)
	Aquatic Freshwater: 18 (14.5%)	Bioremediation: 12 (9.6%)
	Green Algae: 10 (8.1%)	Plants: 4 (3.2%)
	Wastewater: 4 (3.2%)	Aquatic non-Marine: 3 (2.4%)
	Terrestrial: 3 (2.4%)	Digestive System: 2 (1.6%)
	Plants Rhizoplane: 2 (1.6%)	Fungi: 1 (0.81%)
	Human: 1 (0.81%)	Marine: 1 (0.81%)
	NA: 1 (0.81%)	Red algae: 1 (0.81%)
*iVG Annotations	Viral Family	Viral Group
	Iridoviridae: 10	dsDNA
	Poxviridae: 10	dsDNA
	Mimiviridae: 4	dsDNA
	Pandoraviridae: 4	dsDNA
	Retroviridae: 2	RNA-RT
	Marseilleviridae: 1	dsDNA
	Unclassified, Mollivirus sibericum: 1	dsDNA
	Undefined: 1	Undefined

References

Angly, Florent, Dana Willner, Forest Rohwer, Hugenholtz Philip, and Gene Tyson. 2012. "Grinder: A Versatile Amplicon and Shotgun Sequence Simulator." *Nucleic Acids Research* 40 (March): e94. https://doi.org/10.1093/nar/gks251.