

PNAS

www.pnas.org

Supplementary Information for

Seasonal and Diel Patterns of Abundance and Activity of Viruses in the Red Sea

Gur Hevroni, José Flores-Uribe, Oded Béjà, and Alon Philosofo

Gur Hevroni

Email: hevronig@gmail.com

Alon Philosofo

Email: aphilosofo@gmail.com

This PDF file **includes**:

- Figures S1 to S9
- Tables S1 to S3:
Available here as PDF. Excel files can be download from
<https://doi.org/10.17605/osf.io/b74mt>
- Table S4 to S5:
Excel files can be download from <https://doi.org/10.17605/osf.io/b74mt>
- SI References

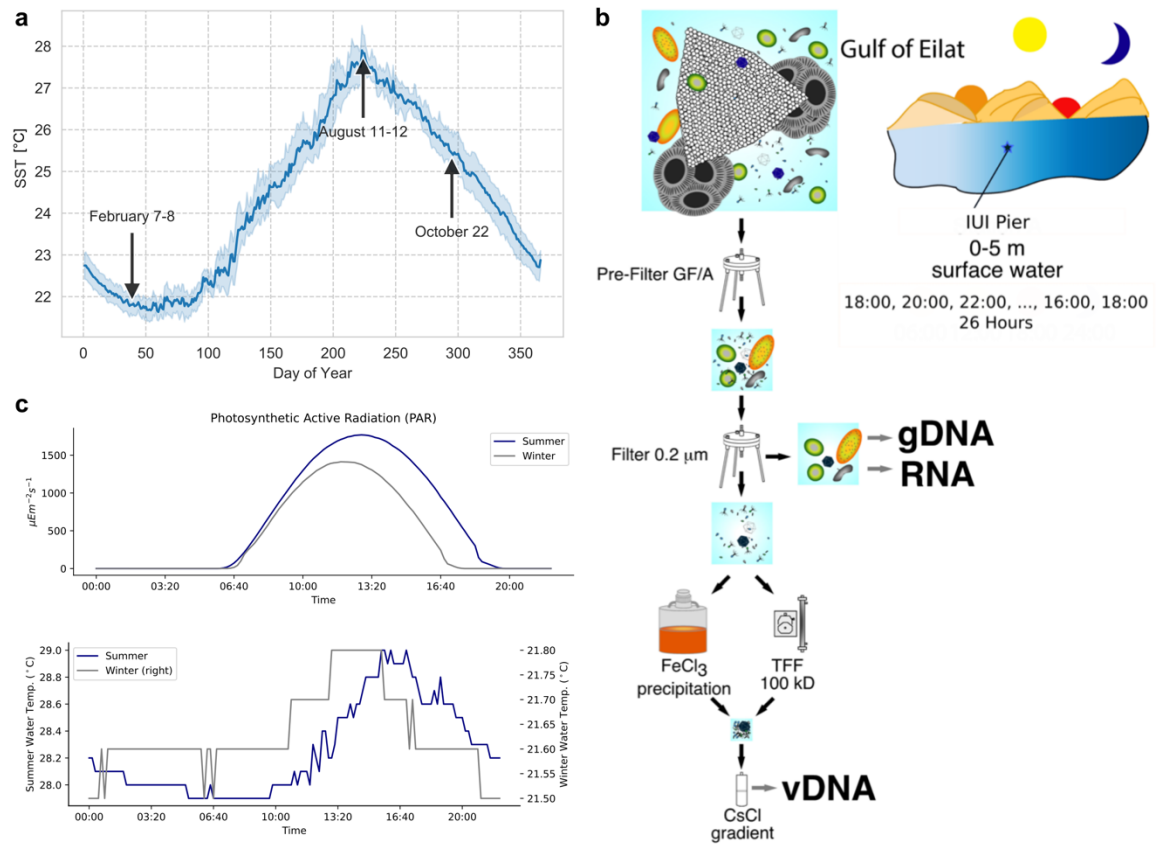


Fig. S1. Environmental parameters and experimental design. **(a)** Sea surface temperature (SST) measured at the IUI Pier from 2011-2019. x-axis: Day of Year, y-axis: mean SST with 95% confidence bands (1000 bootstraps). Arrows indicate day-of-year of samples mentioned in this study. **(b)** Sampling scheme. Sea water sampled every two hours, filtered and fractionated based on microorganismal size and nucleic acid content. **(c)** Photosynthetic Active Radiation (PAR) and water temperature recorded at the IUI pier during days of sampling in August 2015 (blue) and February 2016 (grey).

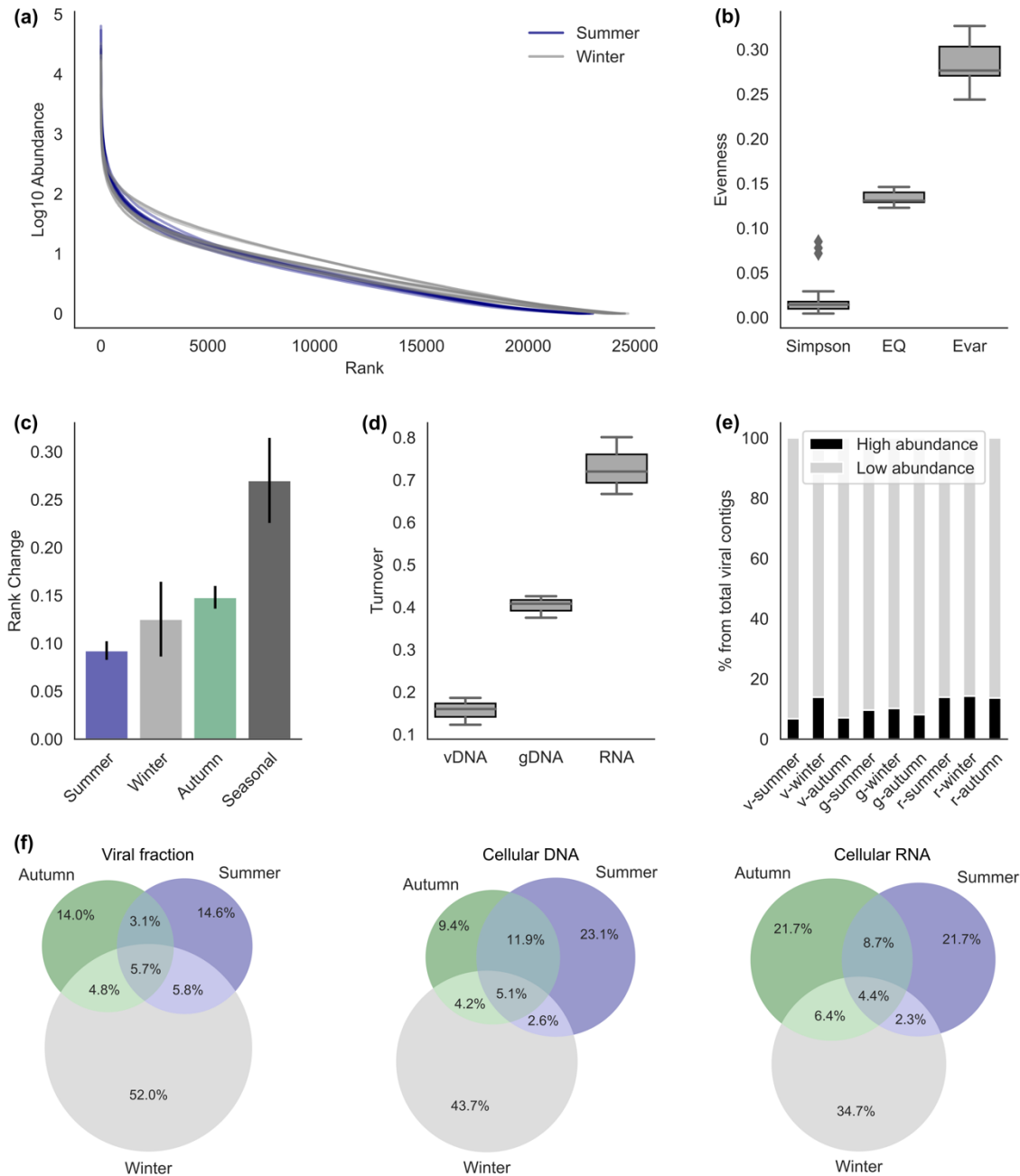


Fig. S2. Viral community structure. **(a)** Rank abundance curves of the viral contigs in the diel summer (blue) and winter (grey) samples. **(b)** *Simpsons*, *E_Q* and *E_{var}* evenness measures of viral contigs abundance in the diel summer and winter samples. **(c)** Scaled rank change, measuring how much the abundance of viral contigs changes over time relative to each other in the diel samples per season (summer, blue; winter, light grey; autumn, green), or between seasons (dark grey). The error bars represent the standard deviation of rank changes. **(d)** Viral contigs turnover between the three seasons, measured as the proportion of contigs either gained or lost relative to the total number of contigs observed across seasons. **(e)** The proportion of contigs that account for 80% of the viral abundance by sample (high abundance) from the total number of viral contigs. **(f)** Percentage of high abundance viral contigs shared and unshared between seasons in the viral fraction (n=6863), cellular DNA fraction (n=4944) and cellular RNA fraction (n=2544).

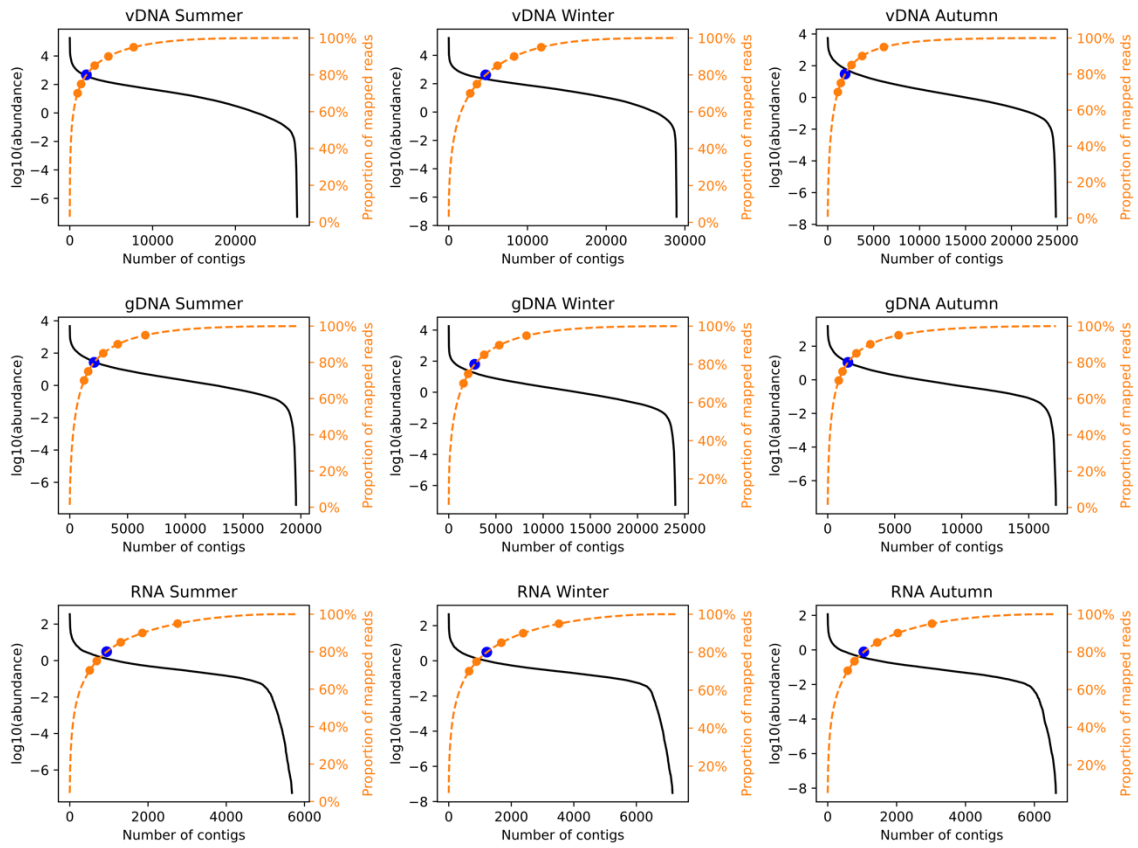


Fig. S3. Rank abundance curves of viral contigs in all samples (black) overlaid with cumulative percentage curve (orange). Cumulative percentage is calculated by dividing the cumulative sum of abundances by the sum abundance of each sample and multiplying by 100. Dots on the orange curve represent a range of cumulative percentage cutoffs (70-90%), where the blue dots represent the 80% cutoff used in this study.

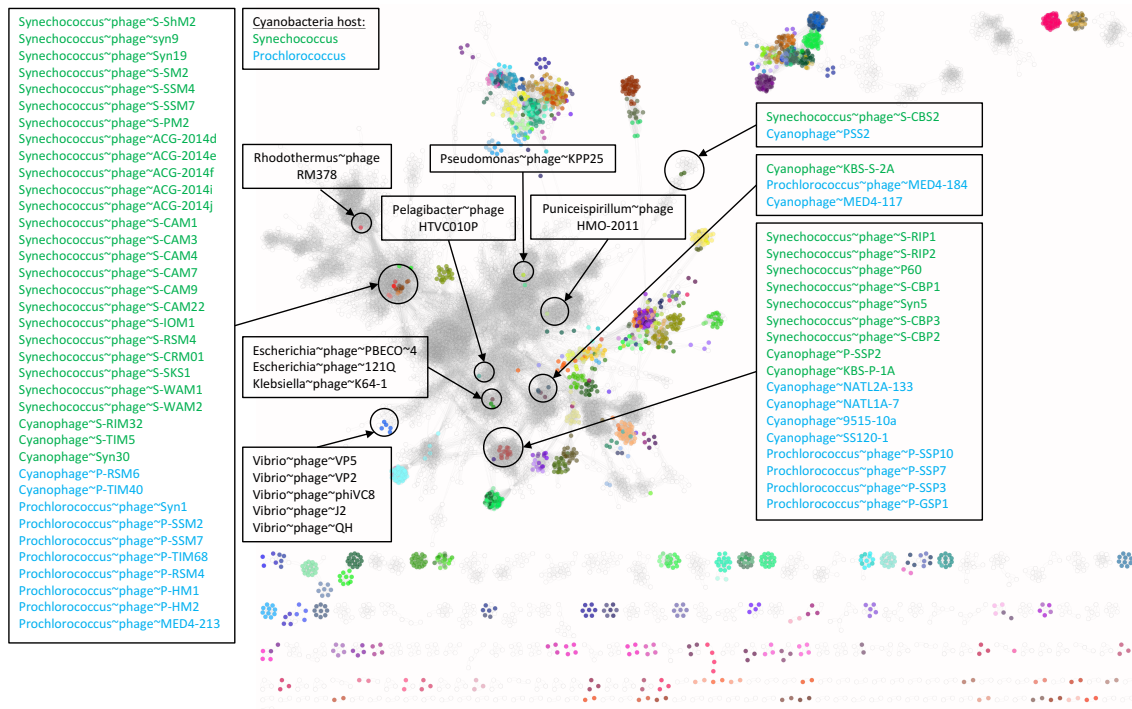


Fig. S4. Gene sharing network of Refseq viral genomes and Red Sea predicted viral contigs (2,239 and 7,047 genomes, respectively). Grey nodes represent Red Sea viral contigs. Colored nodes represent RefSeq viral genomes. Labeled clusters represent taxonomic annotation based on Refseq viral genomes that appear in a specific cluster.

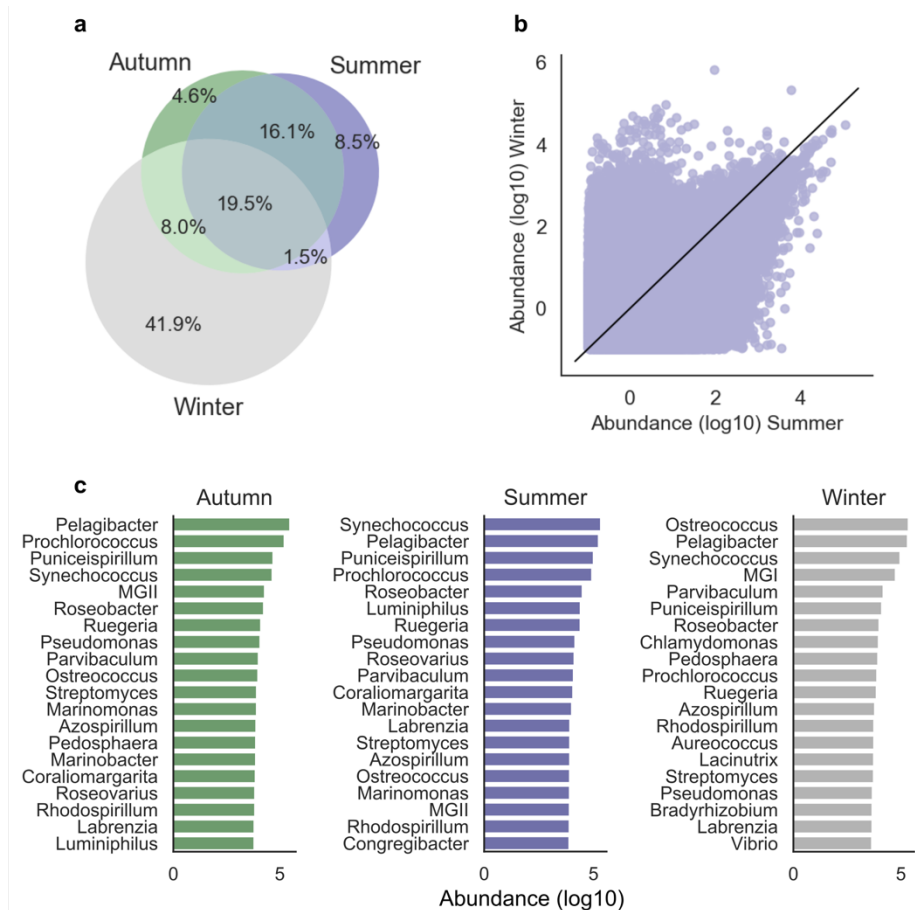


Fig. S5. Microbial community seasonal abundance. **(a)** Percentage of microbial contigs shared between seasons with count ≥ 1 RPKM in the metagenome samples. **(b)** Abundance of microbial contigs in the summer vs. winter cellular DNA samples. A 1:1 line is shown for reference. **(c)** Most abundant microbial genera by season: autumn (green), summer (blue) and winter (grey).

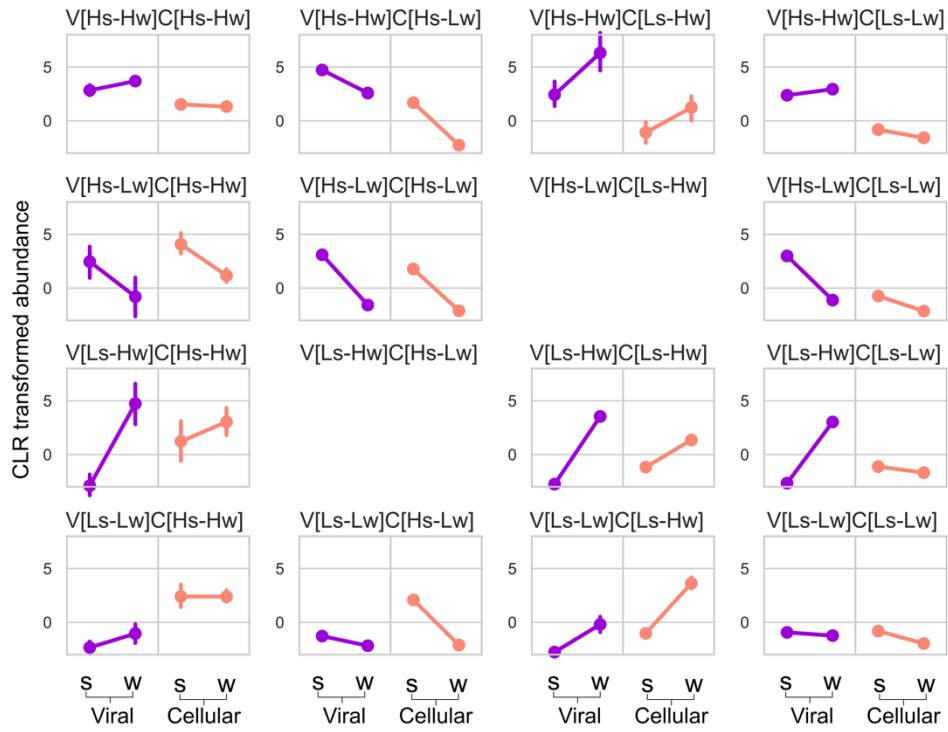


Fig. S6. The centered log ratio (CLR) transformed abundance (RPKM) of the rank-states patterns described in the paper. The CLR-based results show that these patterns are also detectable using methods that are insensitive to the biases associated with relative abundance measures in metagenomic data (1).

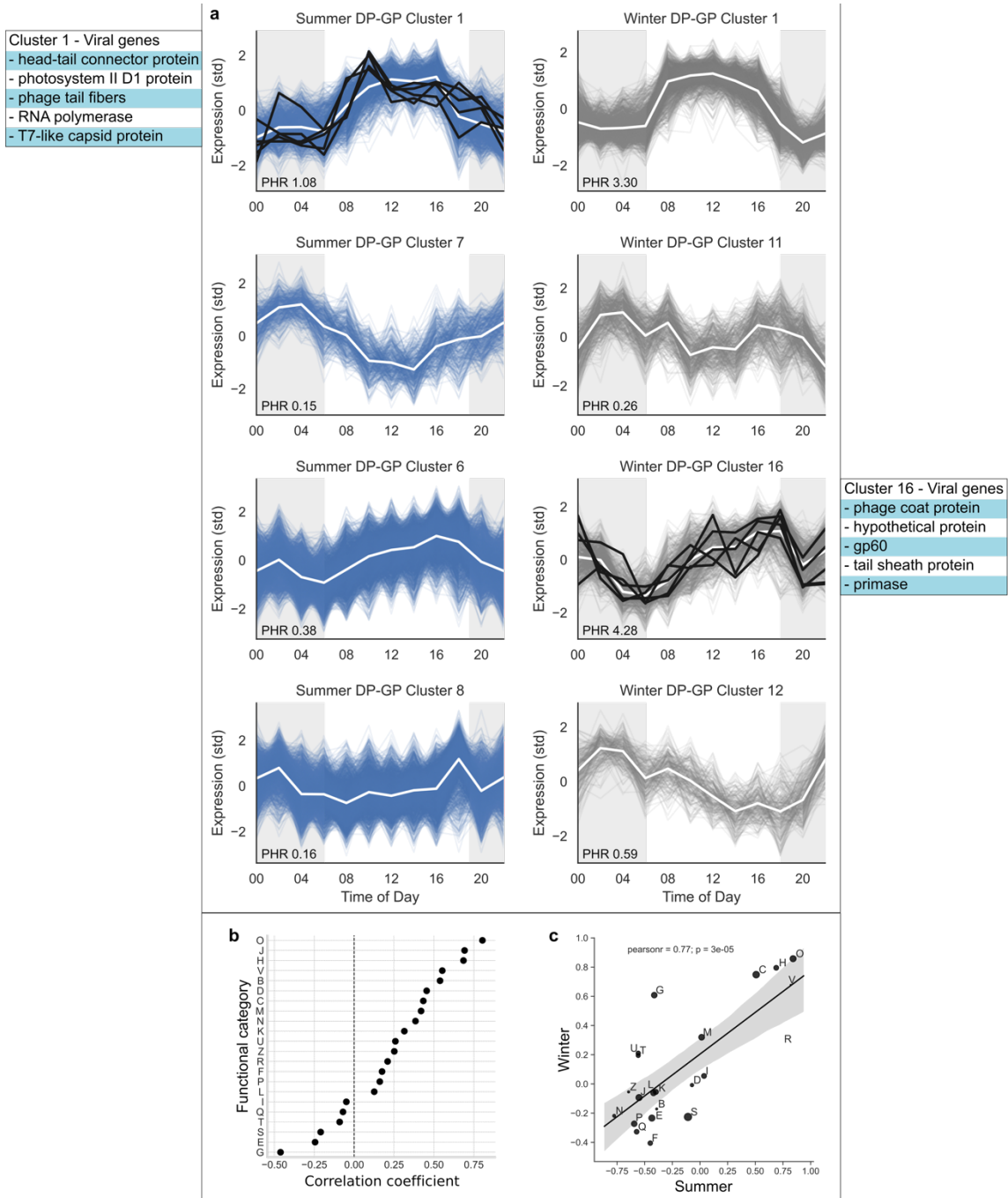


Fig. S7. Microbial community expression patterns. **(a)** Selected gene clusters of similar expression patterns in the summer (blue) and winter (grey) metatranscriptomic samples. The genes were clustered using Gaussian mixture models coupled with Dirichlet process (DP-GP)(2). The black lines in clusters 1 (summer) and 16 (winter) represent viral genes that were clustered together with the microbial genes. x-axis: time of day, y-axis: standard deviations from the mean, confidence bands (dashed lines) represent two standard deviations from the mean, horizontal bars: light (yellow) and dark (black) hours. **(b)** Correlation of the mean expression by functional category between seasons. x-axis: Spearman R correlation coefficient, y-axis: COG functional categories **(c)** Correlation of functional categories with diel light intensity. x-axis: Correlation coefficient of summer functional groups, y-axis: Correlation coefficient of winter functional groups. The fitted linear correlation line is shown for reference.

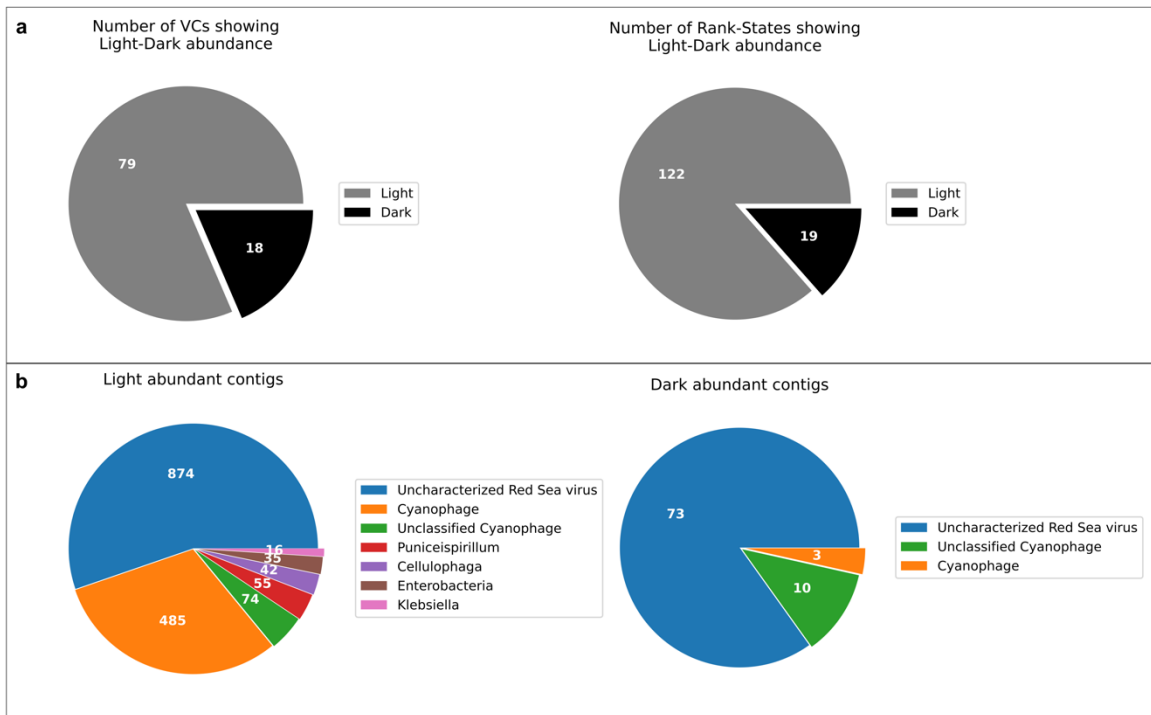


Fig. S8. Light-dependent virus-host interaction **(a)** Number of viral clusters (minimum of 10 ORFs in cluster and minimum counts > 0 RPKM by ORF in either the light or dark samples), with at least one rank-state that displayed differential light-dark abundance (Mann-Whitney test < 0.05). **(b)** Number of contigs from 16 viral clusters and 9 rank-states (grouped by predicted host genus) with significant light-dark differential abundance (Mann-Whitney test < 0.05).

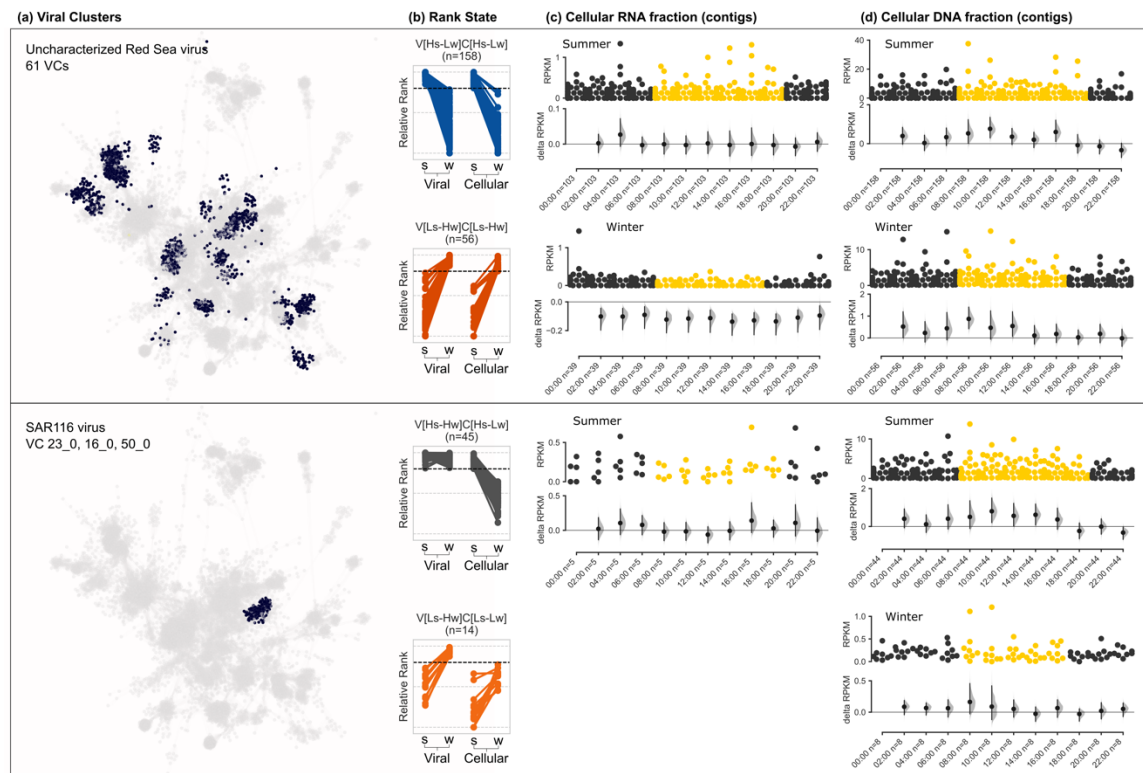


Fig. S9. Intracellular diel abundance patterns of uncharacterized Red Sea virus VCs (top) and SAR116 virus VCs (bottom). **(a)** Superclusters of selected VCs that displayed a differential light-dark signal (Mann-Whitney U test, $p < 0.05$). **(b)** VCs contigs that were classified into two rank-states with strong seasonal abundance pattern (V[Hs-Lw]C[Hs-Lw] and V[LS-Hw]C[LS-Hw]). **(c-d)** Diel distribution in the cellular RNA (c) and DNA (d) fractions of viral contigs from specific VCs and rank-state from (b); x-axis, hours of the day when a sample was collected and the number of viral contigs expressed; upper panel, abundance distribution of viral contigs by time point (each point represents a contig); point color represent samples collected in the dark (black) and light (yellow); y-axis, contig abundance (RPKM); lower panel, estimation plot (3) displaying the effect size as a 95% confidence interval (1,000 bootstraps) of the mean differences between each time point compared against midnight as a reference group.

Table S1. Distribution of rank-states for the viral contigs in the Active group. Top: number of contigs by rank-state. Bottom: Percentage of contigs by rank-state from total.

# Contigs by rank-state				
RankState	C[Hs-Hw]	C[Hs-Lw]	C[Ls-Hw]	C[Ls-Lw]
V[Hs-Hw]	75	212	8	1049
V[Hs-Lw]	5	679	0	461
V[Ls-Hw]	8	0	239	2017
V[Ls-Lw]	17	259	39	196

% Contigs by rank-state				
RankState	C[Hs-Hw]	C[Hs-Lw]	C[Ls-Hw]	C[Ls-Lw]
V[Hs-Hw]	1.42%	4.03%	0.15%	19.93%
V[Hs-Lw]	0.09%	12.90%	0.00%	8.76%
V[Ls-Hw]	0.15%	0.00%	4.54%	38.32%
V[Ls-Lw]	0.32%	4.92%	0.74%	3.72%

Table S3. DP-GP Clusters (2) of the microbial community diel gene expression, with p-value of the Light-Dark differential abundance (Mann-Whitney) by cluster, the effect size and correlation. PHR represents the photoautotrophs to heterotrophs in each cluster.

Season	Cluster	p-value of light-dark differential expression (Mann-Whitney)	Effect size (%)	Rank-biserial correlation (light +1, dark -1)	Photoautotroph to Heterotroph Ratio (PHR)
Summer	1	0.0E+00	100.00	1.00	1.08
Summer	2	8.0E-205	22.76	-0.54	0.26
Summer	3	3.7E-259	77.80	0.56	0.44
Summer	4	7.5E-214	21.21	-0.58	0.15
Summer	5	2.4E-82	1.00	-0.98	0.00
Summer	6	0.0E+00	99.61	0.99	0.38
Summer	7	6.5E-120	0.00	-1.00	0.15
Summer	8	1.7E-93	31.17	-0.38	0.16
Summer	9	3.0E-80	10.88	-0.78	0.24
Summer	10	5.5E-227	1.87	-0.96	0.29
Summer	11	3.9E-06	40.08	-0.20	0.57
Summer	12	1.2E-68	0.00	-1.00	1.12
Summer	13	4.8E-54	0.57	-0.99	0.95
Summer	14	4.7E-05	59.34	0.19	0.81
Summer	15	7.5E-70	99.96	1.00	1.67
Summer	16	3.5E-01	51.55	0.03	0.00
Summer	17	1.1E-30	7.40	-0.85	1.00
Summer	18	1.7E-06	100.00	1.00	
Winter	1	0.0E+00	100.00	1.00	3.30
Winter	2	3.2E-05	45.34	-0.09	0.47
Winter	3	4.2E-151	8.88	-0.82	0.13
Winter	4	6.4E-25	57.62	0.15	0.20
Winter	5	0.0E+00	98.41	0.97	0.31
Winter	6	0.0E+00	94.90	0.90	0.63
Winter	7	1.8E-199	100.00	1.00	2.58
Winter	8	0.0E+00	0.20	-1.00	0.09
Winter	9	4.0E-117	20.46	-0.59	0.49
Winter	10	1.9E-01	46.96	-0.06	0.83
Winter	11	1.3E-258	0.27	-0.99	0.26
Winter	12	2.1E-243	0.08	-1.00	0.59
Winter	13	1.1E-264	99.97	1.00	3.61
Winter	14	8.0E-196	94.57	0.89	0.48
Winter	15	9.0E-142	100.00	1.00	1.75
Winter	16	3.4E-79	84.74	0.69	4.28
Winter	17	9.3E-103	99.79	1.00	3.36
Winter	18	2.8E-31	98.35	0.97	0.00
Winter	19	1.4E-01	53.19	0.06	4.57
Winter	20	1.8E-22	100.00	1.00	2.29
Winter	21	2.6E-01	53.02	0.06	0.00
Winter	22	1.9E-48	0.66	-0.99	0.26
Winter	23	1.2E-04	94.44	0.89	

Table S4. Active contigs abundance in the winter and summer samples. Relative abundance, rank, rank-percentile and assigned rank-state for the contigs in the Active viral contigs group. Excel file can be downloaded from <https://doi.org/10.17605/osf.io/b74mt>

Table S5. Active viral contigs group viral cluster assignment (VC). Excel file can be downloaded from <https://doi.org/10.17605/osf.io/b74mt>

SI References

1. MR McLaren, AD Willis, BJ Callahan, Consistent and correctable bias in metagenomic sequencing experiments. *Elife* 8, e46923 (2019).
2. IC McDowell, et al., Clustering gene expression time series data using an infinite Gaussian process mixture model. *PLOS Comput. Biol.* 14, e1005896 (2018).
3. J Ho, T Tumkaya, S Aryal, H Choi, A Claridge-Chang, Moving beyond p values: data analysis with estimation graphics. *Nat. Methods* 16, 565–566 (2019).