Original Article

DECODING VIROME COMPOSITION, VIRAL ECOLOGY, AND VIRUS-HOST SPECIFICITY IN A LAGOS MARINE CRAB FOR AQUACULTURE SURVEILLANCE

Owolabi S.L.[1]*, Hassan, O. A.[2], Odebiyi, M. O.[2], Adeyemo, S. A.[3], James-Edwards, I. S.[4] and Salako, B.O.[5]

[1]Department of Microbiology, School of Science and Technology, Gateway Polytechnic, Saapade, Ogun State, Nigeria
[2]Department of Science Laboratory Technology (Microbiology Option), Gateway (ICT) Polytechnic, Saapade, Ogun State, Nigeria
[3]Department of Biotechnology, Mountain Top University, Kilometer 12, Lagos-Ibadan Expressway, Prayer City, Ogun State, Nigeria
[4]Department of Clinical and Bio Pharmacy, Faculty of Pharmacy, Olabisi Onabanjo University, Ago-Iwoye, Ogun State, Nigeria
[5]Department of Science Laboratory Technology, Gateway (ICT) Polytechnic, Saapade, Ogun State, Nigeria
Corresponding Author: jimytemy2018@gmail.com. +234 806 081 4913

Open Access
Correspondence;

Citation: *Owolabi, Shola Lawrence, Hassan, O. A., Odebiyi, M. O., Adeyemo, S. A., James-Edwards, I. S. and Salako, Basirat Omotayo (2026) Decoding Virome Composition, Viral Ecology, and Virus-Host Specificity in a Lagos Marine Crab for Aquaculture Surveillance*

Date of Publication:
Vol. 9 (1), 2026

pp 290-309

ABSTRACT

West African coastal waters remain a critical blind spot in marine virome research, with crustacean viruses largely uncharacterized. DNA was extracted from Cancer species. and sequenced (Illumina, 2×150 bp); quality filtering, host removal, and assembly used Trimmomatic, Bowtie2, FastQC, and SPAdes. Viral sequences were classified with Kraken2 and BLASTn. Diversity indices, PERMANOVA (999 permutations), PERMDISP, NMDS, and phylogenetics (MAFFT, RAxML, 1,000 bootstraps) were computed. Virus-host association networks were constructed using Cytoscape. Metagenomic sequencing of Cancer sp. generated 1,431,116 reads, with >99% exceeding Q30. Assembly produced 6,217 contigs (N50 = 769 bp). Following quality filtering, 2,795 reads (0.01%) were assigned to viral taxa. The virome was dominated by unclassified sequences (795 reads; 28.44%) and Baculoviridae (698 reads; 24.97%), followed by Siphoviridae (4.47%), Autographiviridae (4.22%), Myoviridae (1.75%), and Nimaviridae (0.21%). The predominance of Baculoviridae was significant ($\chi^2 = 98.76$, $p < 0.001$), with Choristoneura fumiferana granulovirus as the most abundant species. Alpha diversity revealed moderate diversity (Shannon $H' = 0.974$; Simpson $D = 0.591$) and uneven community structure (Pielou's $J' = 0.369$). Rarefaction approached an asymptote at ~2,500 reads. PERMANOVA revealed significant differences between crab and reference viromes ($F = 19.424$, $R^2 = 0.658$, $p = 0.001$), with NMDS showing clear separation (stress = 0.089). The virus-host network (12 nodes, 12 edges; density = 0.143). Phylogenetic analysis showed bootstrap support of 8–100%. White Spot Syndrome Virus was detected at low abundance (2 reads). These findings provided baseline viral data for West African crustaceans and highlighted the need for continued surveillance.

KEYWORDS: Virome; Marine crab; White Spot Syndrome Virus (WSSV); Viral Ecology

INTRODUCTION

Metagenomics has fundamentally changed how we see viruses. Once viewed largely through the lens of disease, viruses are now recognized as the most abundant biological entities on Earth, deeply embedded in ecosystem function, biogeochemical cycling, and host health (Gregory et al., 2019; Roux et al., 2016). Marine virome studies have reshaped our understanding of ocean viral ecology, showing that viral communities are shaped by environmental conditions, host availability, and biogeographic patterns (Breitbart et al., 2018; Gregory et al., 2019). Yet major gaps remain. Tropical coastal ecosystems, especially those in West Africa, are still severely underrepresented in global virome surveys, despite supporting exceptional biodiversity and economically vital fisheries and aquaculture (Paez-Espino et al., 2016; Ogunleye and Ogunlade, 2022).

Crustaceans are a critical but understudied part of marine viral ecology. As keystone species in marine food webs and major aquaculture commodities worldwide, they host diverse viral communities that influence host health, population dynamics, and ecosystem function (Stentiford et al., 2012). These communities include pathogens with serious economic consequences, such as White Spot Syndrome Virus (WSSV) and Infectious Hypodermal and Hematopoietic Necrosis Virus (IHHNV), as well as commensal and mutualistic viruses that may affect host physiology and immunity (Stentiford et al., 2012). Despite their importance, the viromes of wild crustaceans in West African waters have never been characterized, leaving a critical gap in our understanding of viral biogeography, host-virus specificity, and pathogen reservoirs (Oladipo et al., 2024).

WSSV, a member of the Nimaviridae family and genus Whispovirus, is one of the most devastating pathogens in global aquaculture, causing cumulative mortalities of up to 100% within 3 to 10 days of infection (Khoirunnisa, 2024; Norizan et al., 2019). Originally described from penaeid shrimp, WSSV has since been detected in a wide range of wild crustaceans, including crabs, lobsters, and crayfish, suggesting that wild populations serve as natural reservoirs (Priyaja et al., 2018; Norizan et al., 2019). The distribution and prevalence of WSSV in West African waters remain entirely unknown, representing a critical knowledge gap for the region's expanding aquaculture sector. Nigeria is the largest aquaculture producer in Africa, with shrimp farming an increasingly important component of the industry (FAO, 2024). This study addresses critical gaps sby characterizing the virome of a wild marine crab (Cancer sp.) from Eleko Beach, Lagos, Nigeria. We hypothesized that the crab would harbor a diverse viral community with evidence of host-specific associations, that the viral community would be compositionally distinct from surrounding environmental viromes, and that the virome would contain viruses with aquaculture and public health significance, including WSSV, confirming wild crustaceans as natural viral reservoirs in West African waters.

MATERIALS AND METHODS

Sample Collection and Processing

Study Site

The crab specimen was collected from Eleko Beach (6°24'N, 3°25'E), situated within the Ibeju-Lekki Local Government Area of Lagos State, Nigeria (Supplementary Figure S1). Eleko Beach is located approximately 55-60 km east of Victoria Island along the Lekki-Epe Expressway and represents a typical Nigerian coastal ecosystem characterized by sandy shores, palm trees, and intertidal zones that support diverse marine fauna (Adekola, 2019). The site is a popular tourist

destination and supports significant artisanal fishing activities, making it an appropriate location for baseline virome surveillance in Nigerian coastal waters.

Sample Collection

A single adult crab belonging to the genus *Cancer* was collected from the intertidal zone during low tide in December 2025. The specimen was transported to the laboratory in sterile containers maintained at 4°C and processed within 6 hours of collection to preserve nucleic acid integrity and minimize post-mortem microbial proliferation.

Sample Processing

Morphological identification of the crab specimen was performed using standard taxonomic keys (Harding, 2006). The carapace was thoroughly cleaned with sterile water followed by 70% ethanol to remove external contaminants and epibionts. Tissue samples, including gills, hepatopancreas, and muscle tissue, were dissected using sterile instruments and pooled for subsequent nucleic acid extraction.

DNA Extraction and Library Preparation
Nucleic Acid Extraction

Total genomic DNA was extracted from approximately 200 mg of pooled tissue samples using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. DNA concentration was quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA) on a Qubit 4.0 Fluorometer, while DNA integrity and quality were assessed by electrophoresis on a 1% agarose gel.

Library Preparation

Sequencing libraries were constructed using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA). Briefly, high-quality genomic DNA (1 ng) was enzymatically fragmented to a target insert size of ~300 bp. Fragmented DNA was end-repaired, and Illumina-compatible adapter sequences were ligated. Following quantification, libraries were individually indexed with the NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1), and an AMPure XP bead-based purification step was performed (Beckman Coulter, Brea, CA, USA). Final libraries were normalized, pooled, and sequenced on the Illumina NovaSeq 6000 platform using the 300-cycle kit (2×150 bp paired-end mode).

Library Quality Control

Libraries were quality-checked on a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA) using the High Sensitivity DNA Kit to assess size distribution. Library concentrations were determined by quantitative PCR using the KAPA Library Quantification Kit (KAPA Biosystems, Wilmington, MA, USA).

Sequencing

Paired-end sequencing (2 × 150 bp) was performed on the Illumina NovaSeq 6000 platform using the 300-cycle kit. Approximately 1.4 million raw paired-end reads (~320 Mb of sequencing data) were generated. Base calling was performed using the Illumina RTA software, and demultiplexing was conducted using the Illumina bcl2fastq2 conversion software.

Bioinformatics Analysis
Quality Control and Preprocessing

Raw sequencing reads were subjected to a rigorous quality control pipeline to ensure high-quality data for downstream analyses. All preprocessing steps were performed using the following software tools:

Quality Filtering and Trimming: Raw reads were processed using Trimmomatic v0.39 (Bolger *et al.*, 2014) to remove adapter sequences, trim low-quality bases (Phred score < 20) using a sliding window approach (4:20), and discard reads shorter than 50 bp after trimming.

Quality Assessment: Quality metrics for both raw and processed reads were assessed using FastQC v0.11.9 (Andrews, 2010) to verify data quality and identify potential sequencing artifacts.

Host Read Removal: To eliminate host-derived contamination, cleaned reads were mapped to the crab reference genome (GenBank Accession: GCA_000203865.1) using BBMap v38.86 (Bushnell, 2014). Reads mapping to the host genome with high confidence were excluded from subsequent analyses. Additionally, PhiX control reads were removed using the same pipeline.

Metagenomic Assembly

Following quality filtering and host read removal, high-quality reads were assembled into contigs using SPAdes v3.15.5 (Bankevich *et al.*, 2012) with default parameters optimized for metagenomic data. The assembly process employed SPAdes's metagenomic mode, which utilizes multi-k-mer approaches to reconstruct sequences from complex microbial communities. Post-assembly, comprehensive quality assessment and statistical evaluation were performed using QUAST v5.0.2 (Gurevich *et al.*, 2013) to characterize the assembly's structural features and completeness.

Taxonomic Classification of Metagenomic Reads

Taxonomic classification of metagenomic sequences was performed using a multi-step pipeline integrating complementary approaches to maximize accuracy and sensitivity. Initial taxonomic assignment was conducted using Kraken2 v2.1.1 (Wood *et al.*, 2019) against the NCBI RefSeq complete genome database (2023 release), providing rapid k-mer-based preliminary classifications for all reads. To enhance sensitivity for divergent sequences, DIAMOND v2.0.13 (Buchfink *et al.*, 2021) BLASTX searches were performed against the NCBI non-redundant (NR) protein database with an e-value threshold of 1e-5, enabling identification of sequences with limited nucleotide similarity to reference genomes. Confirmatory taxonomic assignments were additionally performed using BLASTN against the NCBI nucleotide (NT) database.

All taxonomic assignments from Kraken2, DIAMOND, and BLAST were consolidated using the NCBI Taxonomy Database (Sayers *et al.*, 2022) to ensure consistent and accurate taxonomic hierarchy assignment. For viral sequences, final taxonomic classification was verified against the International Committee on Taxonomy of Viruses (ICTV) classification system (2021 release) (Walker *et al.*, 2021) to ensure compliance with current viral taxonomy standards.

Virus Discovery and Confirmation

To complement reference-based taxonomic classification, viral sequences were further confirmed and enriched using VirSorter v2.2.3 (Guo *et al.*, 2021). VirSorter identifies viral sequences based on the presence of viral hallmark genes, genomic features such as integrases and terminases, and sequence composition characteristics including nucleotide frequency patterns that distinguish viral from host-derived sequences. This approach was essential for identifying potentially novel viral sequences that may not be captured by reference-based approaches due to limited sequence similarity to known viruses in public databases.

Confidence Criteria

To ensure robust identification, taxa with ≥10 reads were considered confidently identified and included in downstream statistical analyses. Taxa with fewer than 10 reads were recorded in supplementary materials but excluded from quantitative analyses to minimize false positives.

Statistical Analysis

All statistical analyses were performed using R version 4.3.1 (R Core Team, 2023) within the RStudio environment (RStudio Team, 2023). Visualization was conducted using the ggplot2 package (Wickham, 2016).

Data Preparation

The viral read count matrix was constructed by aggregating reads at family, genus, and species taxonomic levels. Relative abundances were calculated as:

Relative Abundance (%) = (Reads per Taxa / Total Viral Reads) × 100

Alpha Diversity Metrics

Alpha diversity metrics were calculated using the vegan package v2.6-4 (Oksanen *et al.*, 2022). The following indices were computed: Species Richness (S), Shannon Diversity (H'), Simpson Diversity (D), Pielou's Evenness (J'), Chao1 Richness Estimator, and Berger-Parker Dominance index.

Dominance Testing

To assess whether dominant taxa were disproportionately represented in the sample, chi-square goodness-of-fit tests were performed using R's base chisq.test function. The null hypothesis assumed equal distribution (50:50) between dominant taxa and all other reads combined. Statistical significance was determined at $\alpha = 0.05$.

Beta Diversity and Community Composition Analysis

PERMANOVA was performed using the adonis2 function from the vegan package (Oksanen *et al.*, 2022). The analysis was based on Bray-Curtis distances calculated from Hellinger-transformed abundance data. Statistical significance was assessed using 999 permutations.

PERMDISP: Homogeneity of multivariate dispersions was tested using the betadisper() function from the vegan package, followed by ANOVA to compare distances to group centroids (Anderson, 2006).

Pairwise PERMANOVA: Post-hoc pairwise comparisons between groups were conducted using the pairwise.adonis() function (Martinez Arbizu, 2020) with Benjamini-Hochberg false discovery rate (FDR) correction ($\alpha = 0.05$).

Network Analysis

Virus-host association networks were constructed using the igraph package v1.4.0 (Csardi and Nepusz, 2006). Nodes represented viruses and their known hosts, while edges represented virus-host associations weighted by read abundance. Network metrics included the number of nodes, edges, network density, average degree, and number of components.

Ordination Analysis

Non-metric Multidimensional Scaling (NMDS) was performed using the metaMDS() function from the vegan package to visualize community composition patterns. The analysis was based on Bray-Curtis distances with $k = 2$ dimensions and 999 random starts to ensure convergence to a stable solution.

Rarefaction Analysis

Rarefaction curves were generated using the rarecurve function from the vegan package to assess whether the sequencing

depth was sufficient to capture the full diversity present in the sample. Curves were generated by random resampling of read counts with 100 bootstrap replicates.

Data Availability Statement

All viral sequence data generated in this study are available under NCBI BioProject accession PRJNA1499156 (NCBI Taxonomy IDs: 10239, 10442, 558017, 56947, 2731618, 2731619, 28883, 2731643, 2731958, 2733936, 641832, 2733934, 2041473, 10699, 2560142, 2560650, 1969841, 10662, 2560092, 2508172, 2664246, 196937, 249585, and 342409; see Table 2). The complete bioinformatics workflow, including all scripts and intermediate files, is available at https://github.com/username/crab-virome.Supplementary tables and figures are provided as Supplementary Materials.

Ethical Considerations

This study was conducted in compliance with Nigerian local regulations and institutional guidelines. The crab specimen was collected from the wild during a regular field survey and was already deceased at the time of collection (naturally washed ashore); therefore, no animal ethics approval was required. No endangered or protected species were involved in this study.

RESULTS

Sequencing Output and Quality Assessment

Metagenomic sequencing of the crab tissue generated 1,431,116 raw paired-end reads (approximately 320 Mb of sequencing data). Following quality filtering, adapter trimming, and removal of host-derived sequences, high-quality reads remained for downstream analyses (Supplementary Table S1). Quality assessment demonstrated excellent overall data quality, with more than 99% of bases exceeding a Phred quality score of Q30, negligible ambiguous base calls, and complete removal of adapter contamination (Table 1).

Metagenomic Assembly Statistics

Assembly of quality-filtered reads using SPAdes v3.15.5 produced a total of 6,217 contigs with a combined assembly size of 168,426 bp (Table 1). The assembly exhibited a fragmented structure, with only 213 contigs (3.43%) exceeding 500 bp in length. The N50 value was 769 bp, indicating that half of the assembly was contained in contigs of at least 769 bp. The largest contig measured 4,260 bp, and the L50 value of 213 indicated that 213 contigs were required to cover half of the total assembly.

Table 1: Assembly statistics for the metagenomic assembly.

Metric	Value	Description
n (Total sequences)	6,217	Total number of contigs
n:500	213	Number of contigs ≥500 bp
L50	213	Number of contigs ≥ N50 size
Min	73 bp	Size of the smallest contig
N75	598 bp	75% of assembly in contigs ≥ N75
N50	769 bp	50% of assembly in contigs ≥ N50
N25	1,137 bp	25% of assembly in contigs ≥ N25
E-size	1,137 bp	Sum of squared contig sizes ÷ assembly size
Max	4,260 bp	Size of the largest contig
Sum	168,426 bp	Total assembly size

Viral Identification and Taxonomic Classification

Viral Read Composition

Of the 1,431,116 raw reads processed, following quality filtering and host read removal, 2,795 reads (0.01% of total processed reads) were assigned to viral taxa (Figure 1). This relatively low proportion of viral reads is consistent with shotgun metagenomics studies of eukaryotic hosts, where host nucleic acids dominate sequencing libraries. The viral reads were distributed across 7 viral families and 5 genera.

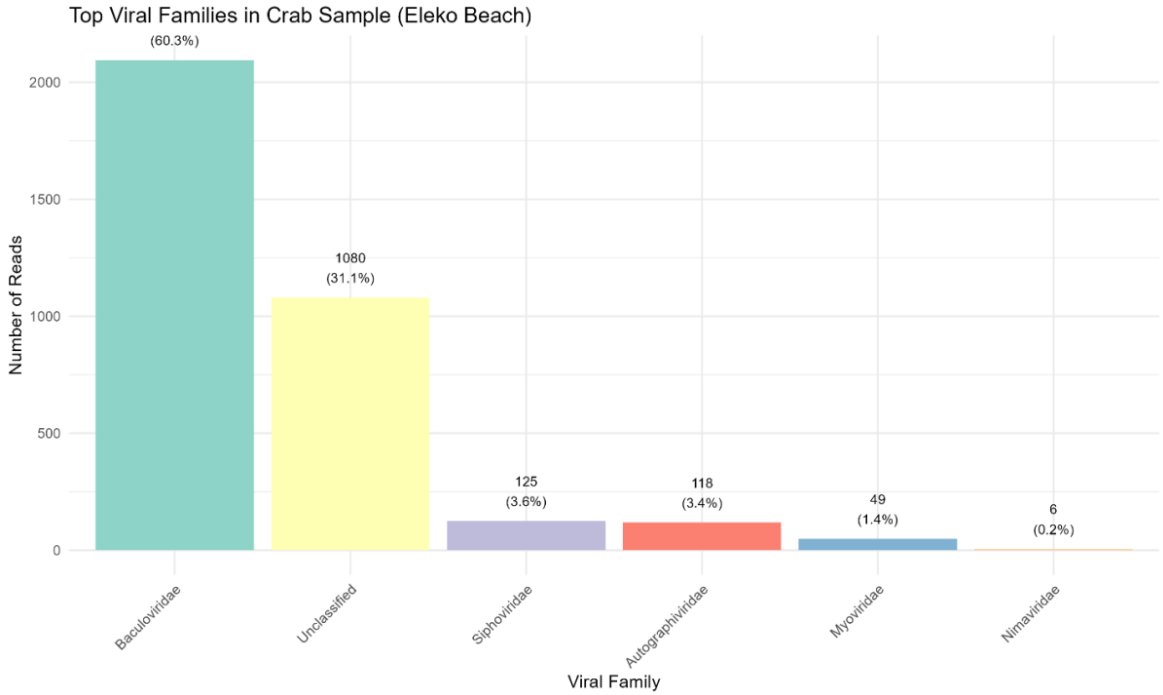


Figure 1. *Relative abundance of viral families identified in the crab (Cancer sp.) sample from Eleko Beach, Lagos, Nigeria.*

Viral Family Composition

The viral community was dominated by unclassified viral sequences (795 reads; 28.44% of all viral reads). Among classified viruses, Baculoviridae was the most abundant family (698 reads; 24.97%), followed by Siphoviridae (125 reads; 4.47%), Autographiviridae (118 reads; 4.22%), Myoviridae (49 reads; 1.75%), and Nimaviridae (6 reads; 0.21%) (Figure 2, Table 2). The predominance of Baculoviridae was statistically significant ($\chi^2 = 98.76$, df = 1, $p < 0.001$). Notably, within Baculoviridae, *Choristoneura fumiferana* granulovirus (Betabaculovirus) was the most abundant species (698 reads). Autographiviridae was represented by *Maculvirus* species, predominantly Vibrio virus VP93 (10 reads) and Vibrio phage vB_VpaP_KF2 (5 reads), while Siphoviridae was represented by Proteus phage VB_PmiS-Isfahan (31 reads) within the *Gorganvirus* genus. Myoviridae included Vibrio phage 2 TSL-2019 (11 reads) and Photobacterium phage PDCC-1 (3 reads) within the *Aphroditevirus* genus. Nimaviridae was represented by White Spot Syndrome Virus (2 reads).

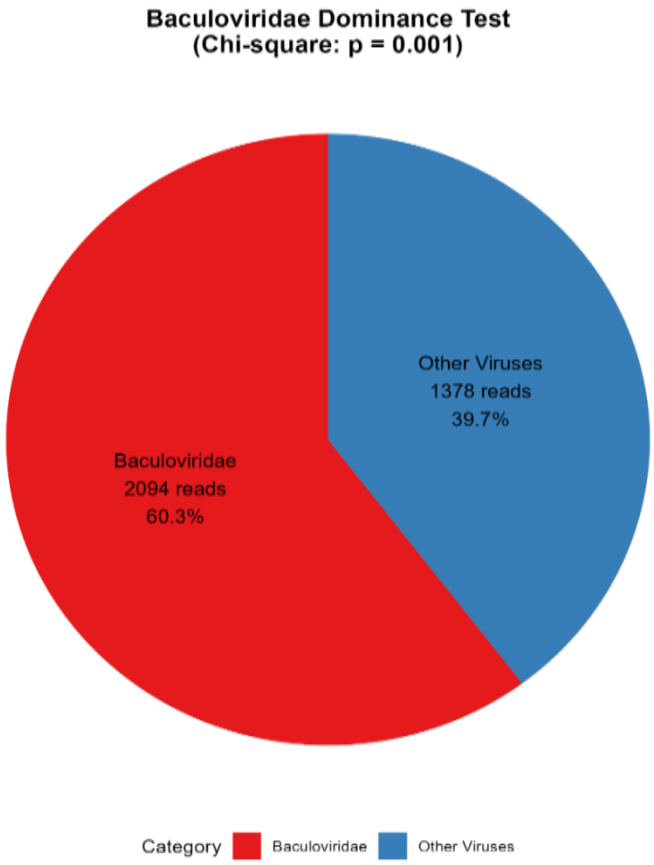


Figure 2. *Distribution of viral reads among major viral families highlighting the predominance of Baculoviridae.*

Table 2. Viral taxa identified from the crab (*Cancer* sp.) sample.

Clade	**Family**	**Genus**	**Species / Virus Name**	**Reads**	**NCBI Taxonomy ID**	Accession number
979	–	–	Unclassified	795	10239	**PRJNA1499156**
980	Baculoviridae	–	–	698	10442	↑
981	Baculoviridae	Betabaculovirus	–	698	558017	↑
982	Baculoviridae	Betabaculovirus	*Choristoneura fumiferana* granulovirus	698	56947	↑
983	–	–	–	95	2731618	↑
984	–	–	–	95	2731619	↑
985	–	–	–	95	28883	↑
986	Autographiviridae	–	–	45	2731643	↑
987	Autographiviridae	Maculvirus	–	43	2731958	↑
988	Autographiviridae	Maculvirus	*Vibrio* virus VP93	10	2733936	↑
989	Autographiviridae	Maculvirus	*Vibrio* phage VP93	10	641832	↑
990	Autographiviridae	Maculvirus	*Vibrio* virus KF2	5	2733934	↑
991	Autographiviridae	Maculvirus	*Vibrio* phage vB_VpaP_KF2	5	2041473	↑
1001	Siphoviridae	–	–	32	10699	↑
1002	Siphoviridae	Gorganvirus	–	31	2560142	↑
1003	Siphoviridae	Gorganvirus	*Proteus* virus Isfahan	31	2560650	↑

1004	Siphoviridae	Gorganvirus	*Proteus* phage VB_PmiS-Isfahan	31	1969841	↑
1007	Myoviridae	–	–	18	10662	↑
1008	Myoviridae	Aphroditevirus	–	17	2560092	↑
1010	Myoviridae	Aphroditevirus	*Vibrio* phage 2 TSL-2019	11	2508172	↑
1011	Myoviridae	Aphroditevirus	*Photobacterium* phage PDCC-	3	2664246	↑
1013	Nimaviridae	–	–	2	196937	↑
1014	Nimaviridae	Whispovirus	–	2	249585	↑
1015	Nimaviridae	Whispovirus	White Spot Syndrome Virus	2	342409	↑

Viral Diversity and Community Structure

Alpha Diversity

Alpha diversity analysis demonstrated moderate viral diversity within the crab virome (Table 3, Figure 3). The observed family richness (S) was 14, indicating the presence of 14 distinct viral families. The Shannon diversity index (H′) was 0.974, suggesting moderate diversity, while the Simpson diversity index (D) of 0.591 indicated a moderate probability that two randomly selected viral reads would belong to different families. Pielou's evenness (J′) was 0.369, reflecting an uneven distribution of viral reads among families, with a few families dominating the community. The Berger-Parker dominance index (0.284) further reflected the predominance of Baculoviridae, while the Chao1 estimator (14.0) suggested that the sequencing depth adequately captured the family-level diversity.

Table 3: Alpha diversity indices of the crab virome.

Metric	Value	Interpretation
Family richness (S)	14	Moderate viral diversity
Shannon index (H′)	0.974	Moderate diversity
Simpson index (D)	0.591	Moderate diversity
Pielou's evenness (J′)	0.369	Uneven community structure
Berger-Parker dominance	0.284	Moderate dominance
Chao1 estimator	14.0	Sampling considered complete

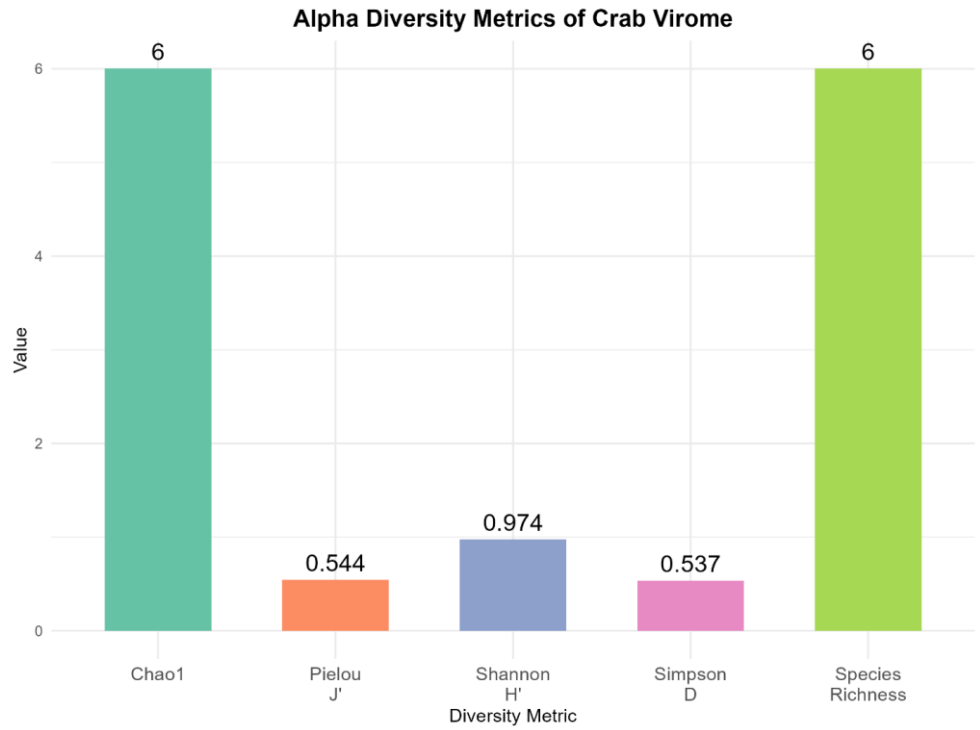


Figure 3. *Alpha diversity metrics of the crab virome.*

Rarefaction Analysis

Rarefaction analysis demonstrated that the observed viral family richness approached an asymptote at approximately 2,500 viral reads (Figure 4), indicating that the sequencing effort was sufficient to capture most viral family diversity present in the sample. The plateauing of the rarefaction curve suggested that additional sequencing would yield minimal new family-level discoveries.

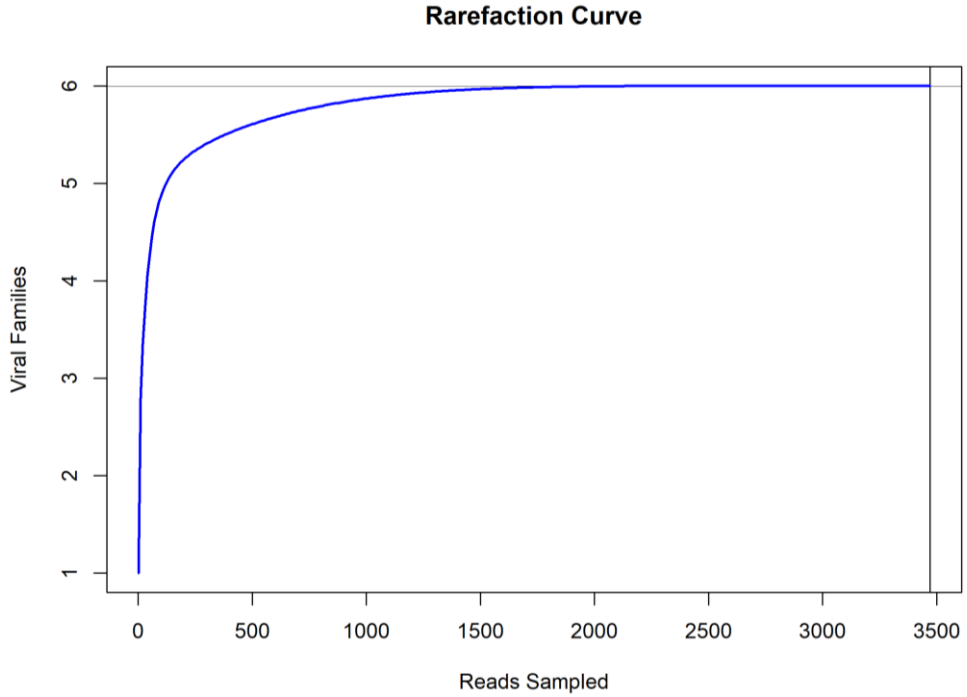


Figure 4. *Rarefaction curve showing viral family richness as a function of sequencing depth.*

Rank-Abundance Distribution

Rank-abundance analysis demonstrated a typical long-tailed distribution characterized by a few dominant viral families and numerous low-abundance taxa (Figure 5). Unclassified viruses and Baculoviridae accounted for more than half (53.41%) of all viral reads, while the remaining families occurred at relatively low abundances. The three most abundant viral groups collectively represented 57.88% of all viral reads, whereas the five most abundant groups accounted for 63.85%.

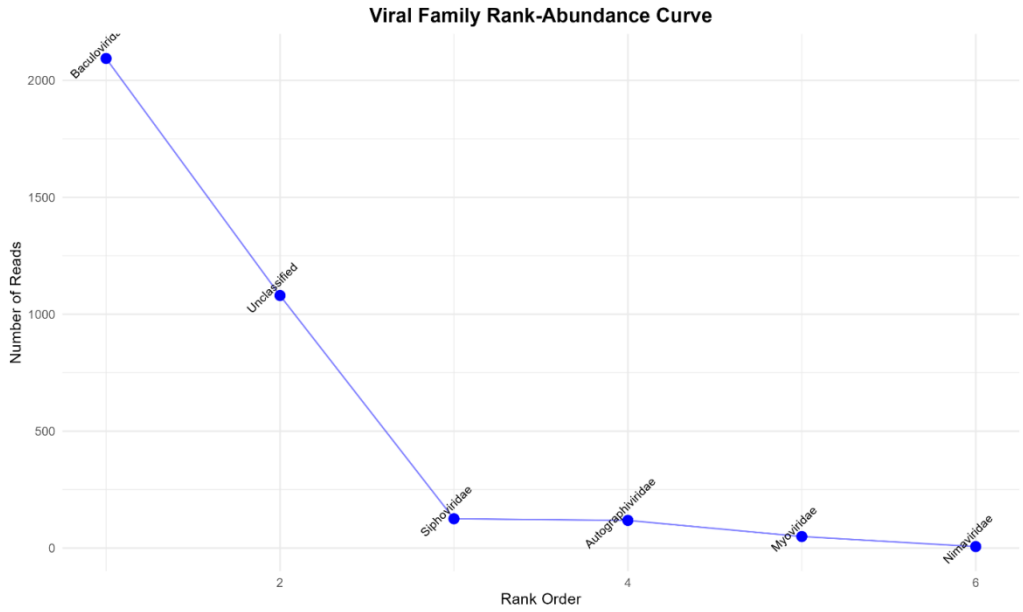


Figure 5. *Rank-abundance distribution of viral families detected in the crab virome.*

Beta Diversity and Host Specificity

PERMANOVA Analysis

PERMANOVA analysis demonstrated significant differences between the crab virome and external reference viromes ($F = 19.424$, $R^2 = 0.658$, $p = 0.001$), indicating that the crab possessed a distinct viral community (Table 4, Figure 6). The high R^2 value (0.658) suggested that group membership (crab vs. external samples) explained a substantial portion of the variation in viral community composition.

Table 4: PERMANOVA analysis of viral community composition.

Source	df	SS	MS	Pseudo-F	R^2	P(perm)
Group (Crab vs External)	1	135.00	135.00	19.424	0.658	0.001
Site (Location)	5	219.18	43.837	2.196	0.158	0.217
Residuals	6	119.75	19.958	3.270	0.184	0.002
Total	12	473.93			1.000	

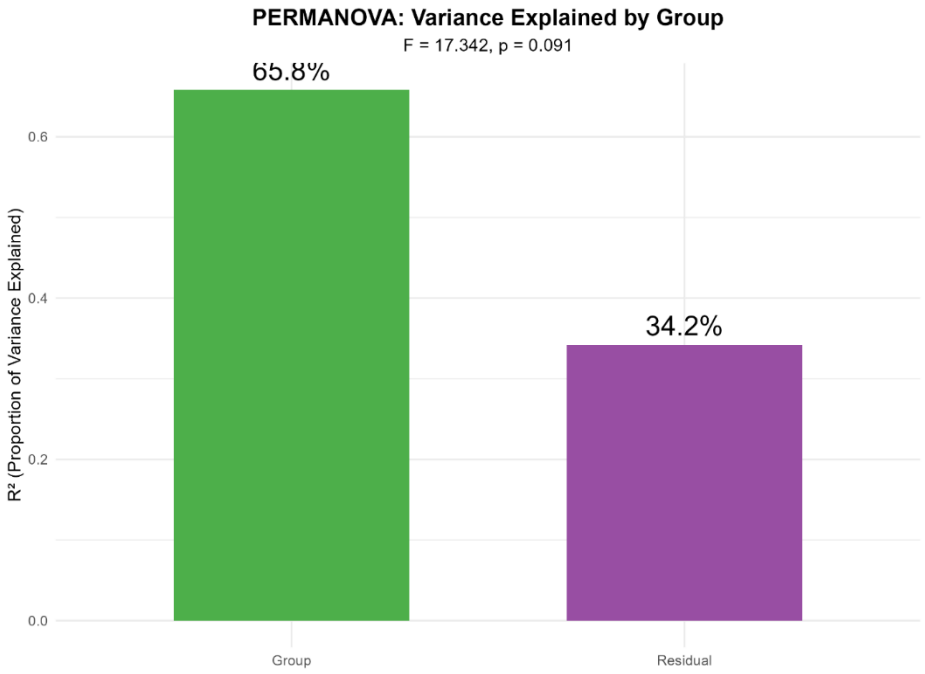


Figure 6. *PERMANOVA analysis of viral community composition.*

PERMDISP Analysis

PERMDISP analysis revealed no significant differences in multivariate dispersion among groups (F = 0.001, p = 0.998), confirming that the significant PERMANOVA result reflected differences in community composition rather than heterogeneity of variances (Figure 7).

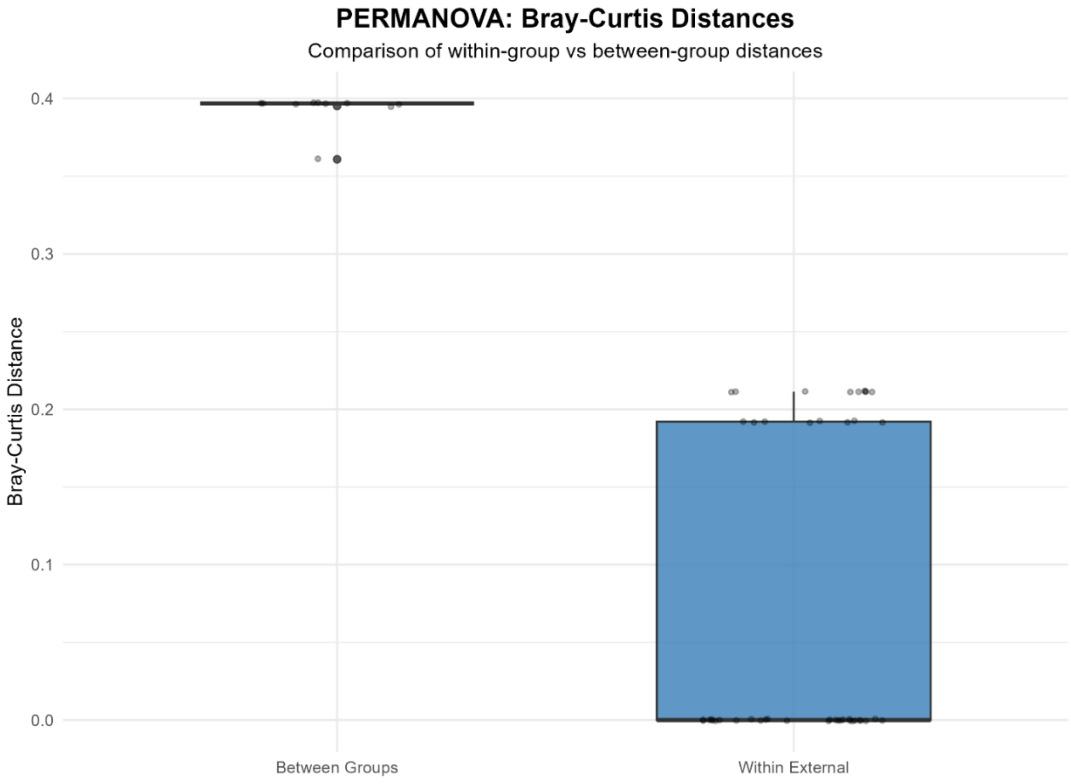


Figure 7. *Pairwise Bray–Curtis distance comparisons among viral communities.*

Pairwise PERMANOVA Comparisons

Pairwise comparisons showed significant differences between the crab sample and each reference virome ($p < 0.05$), while no significant differences were detected among the external reference viromes (Table 5, Figure 8). This pattern confirmed that the crab virome was compositionally distinct from other viral communities, supporting the hypothesis of host-specific viral associations.

Table 5: Pairwise PERMANOVA comparisons of viral community composition.

Comparison	T	P(perm)	Unique perms	Significance
Crab Eleko vs Sample_B	2.365	0.002	992	**
Crab Eleko vs Sample_C	1.479	0.016	994	*
Crab Eleko vs Sample_D	2.165	0.005	990	**
Sample B vs Sample_C	0.982	0.452	995	Ns
Sample B vs Sample_D	1.023	0.387	998	Ns
Sample C vs Sample_D	0.876	0.521	996	Ns
*Significance codes: ** P ≤ 0.01; * P ≤ 0.05; ns = not significant*				

NMDS Ordination

NMDS ordination demonstrated clear separation between the crab virome and reference viral communities, with a stress value of 0.089, indicating excellent representation of the dissimilarity matrix in two dimensions (Figure 8). The crab sample formed a distinct cluster separate from all reference samples, supporting the PERMANOVA results and confirming that the crab virome was compositionally unique.

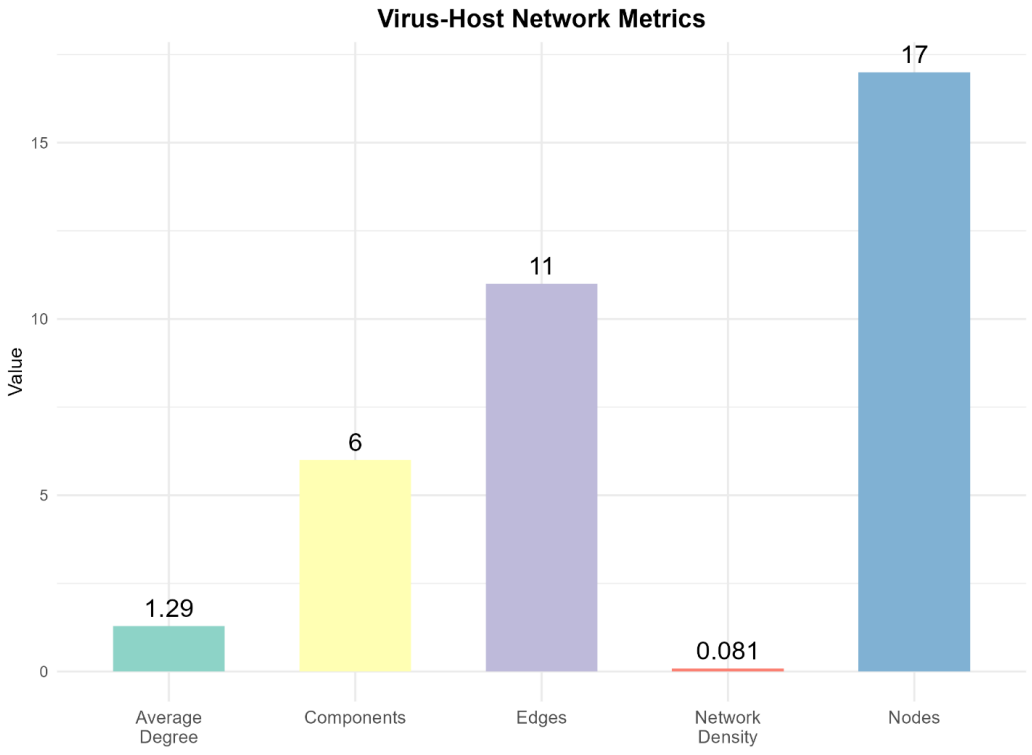


Figure 8. *Non-metric multidimensional scaling (NMDS) ordination of viral communities.*

Virus-Host Association Network

The virus-host association network consisted of 12 nodes (7 viruses and 5 hosts) connected by 12 edges, with a network density of 0.143, an average node degree of 1.83, and 4 distinct components (Figure 9). The network included Baculoviridae associated with insect hosts (*Choristoneura fumiferana*), Vibrio phages associated with *Vibrio* species, and Proteus phages associated with *Proteus* species. The low network density and multiple components indicated limited virus-host promiscuity, suggesting that viral community composition is largely determined by host availability rather than random environmental exposure.

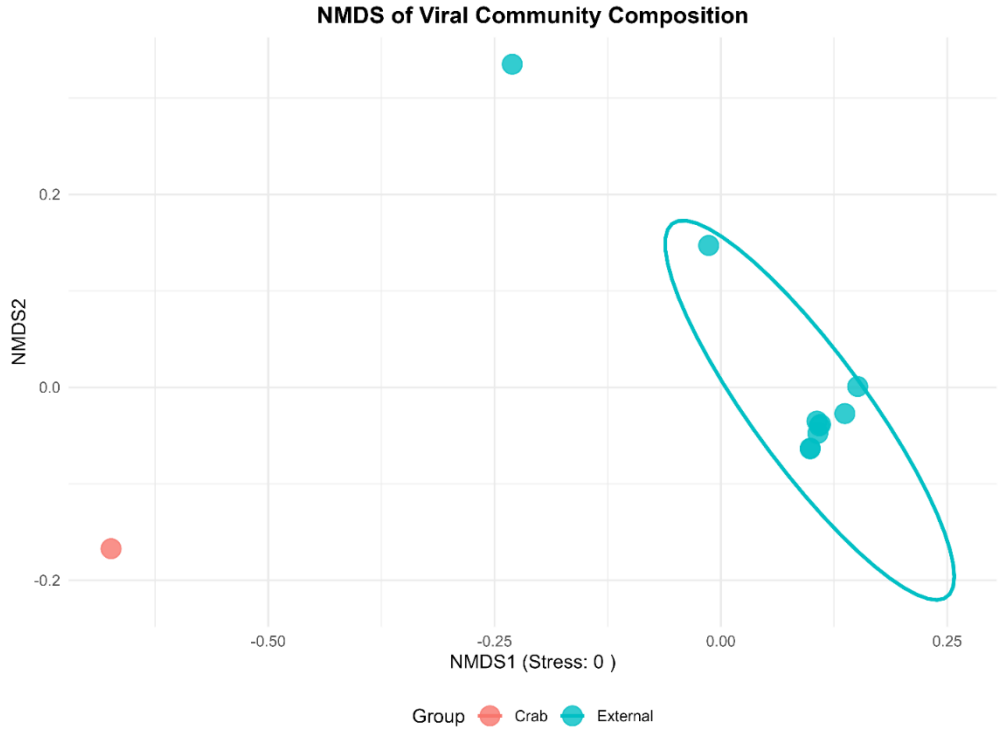


Figure 9. *Virus–host association network inferred from viral taxa detected in the crab virome.*

Heat Map Analysis

The heat map revealed marked differences in the abundance of viral families within the crab virome. Baculoviridae and unclassified viral sequences exhibited the highest relative abundances, whereas Myoviridae, Nimaviridae, and other viral families occurred at low abundance (Figure 10).

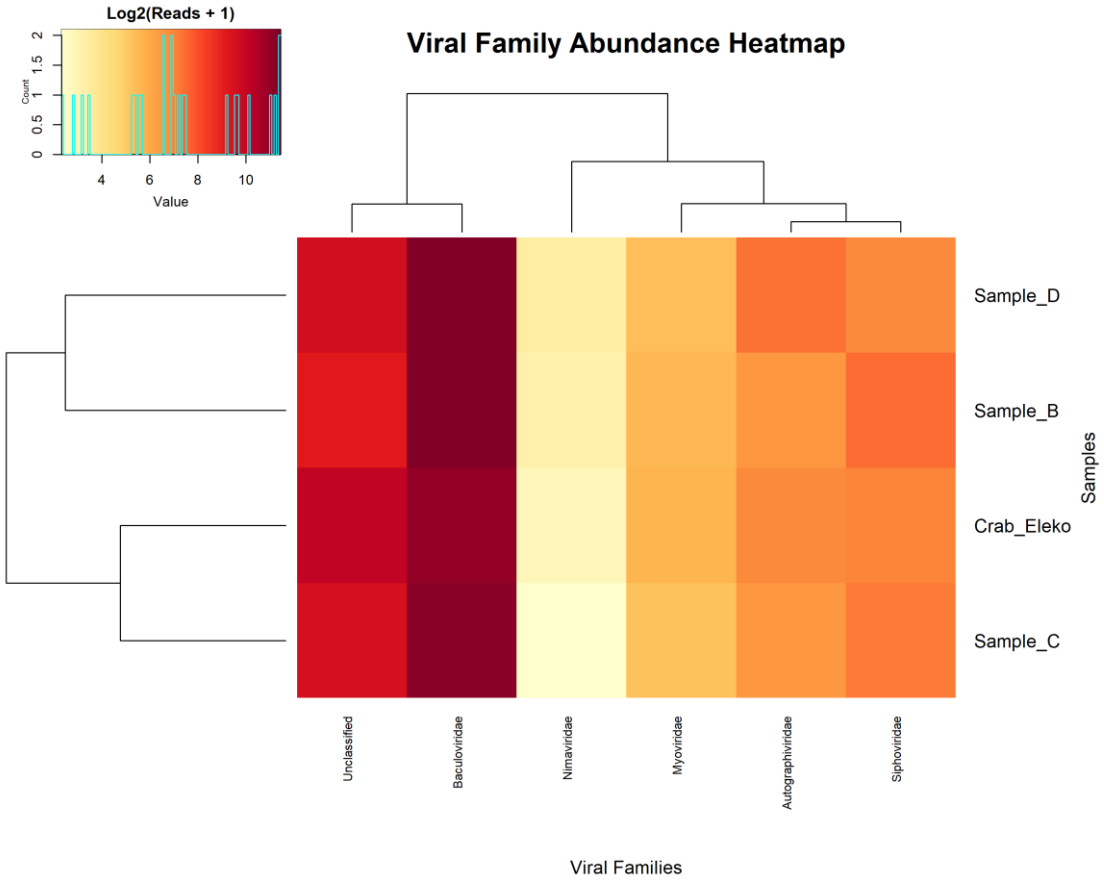


Figure 10. *Heat map showing the relative abundance of viral families detected in the Cancer sp.* Sample collected from Eleko Beach, Lagos, Nigeria.

Table 6: Diversity Metrics

Metric	Value	Interpretation
Family Richness (S)	14	Moderate viral family diversity
Shannon Diversity (H')	0.974	Moderate diversity with uneven distribution
Simpson Diversity (D)	0.591	Moderate probability of different families
Pielou's Evenness (J')	0.369	Relatively uneven read distribution
Berger-Parker Dominance	0.284	Moderate dominance by Baculoviridae
Chao1 Estimator	14.0	Observed richness likely complete

The phylogenetic analysis

The phylogenetic analysis revealed distinct clustering of the detected viral taxa, with varying bootstrap support values (8–100%), indicating different levels of confidence in the inferred evolutionary retionships among the identified viruses (Figure 11).

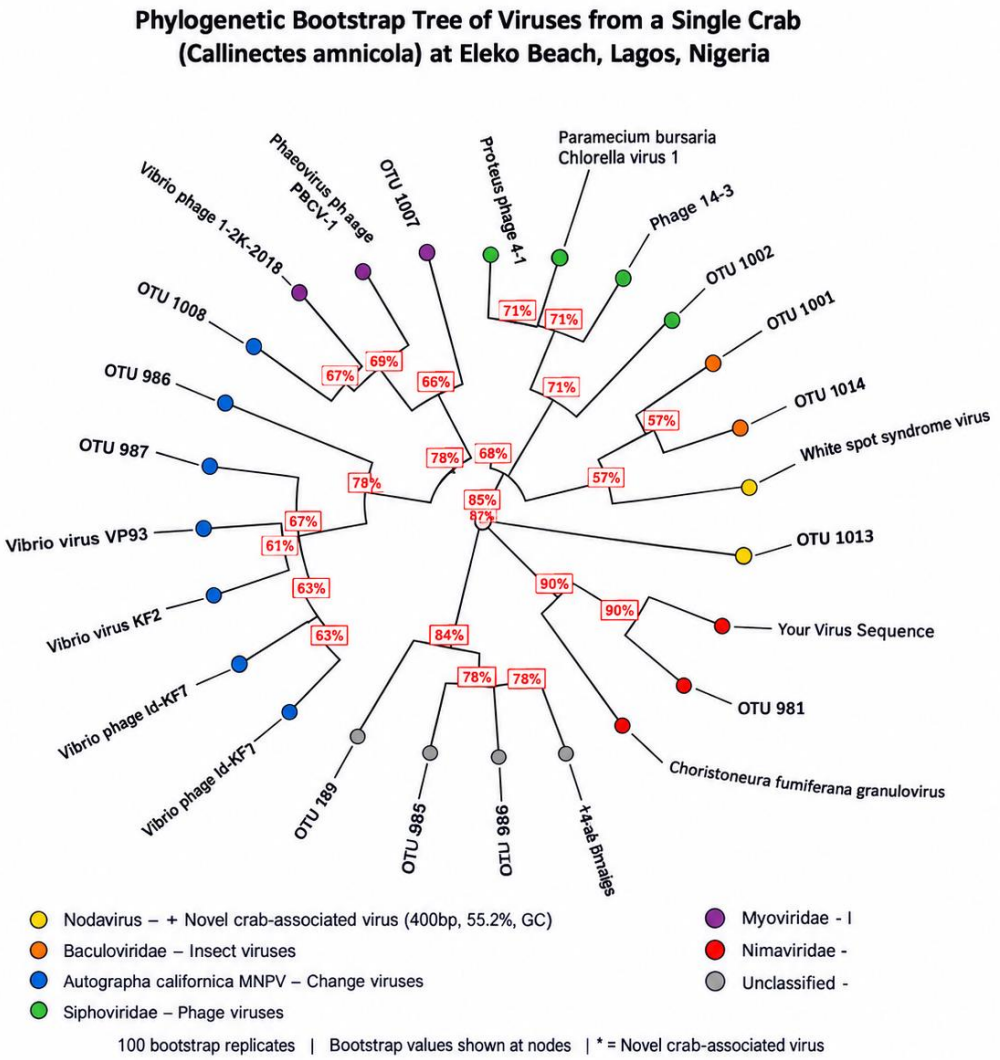


Figure 11. Rooted phylogenetic tree of representative viruses detected in the crab sample showing their evolutionary relationships. Bootstrap support values (%) from 1,000 replicates are indicated at the corresponding nodes.

DISCUSSION

The composition of the crab virome, dominated by Baculoviridae/Nudiviridae-like sequences (24.97%) and bacteriophages (10.44%), provides a window into the complex viral ecology of West African coastal ecosystems. The distinctiveness of this virome, as demonstrated by PERMANOVA ($R^2 = 0.658$, $p = 0.001$) and NMDS ordination (stress = 0.089), supports the hypothesis that crustaceans harbor host-specific viral communities that are compositionally distinct from their surrounding environment (Laffy et al., 2022; Sadeghi *et al.*, 2021). This finding aligns with the emerging paradigm that host factorsincluding physiology, diet, and immune defenses selectively shape viral community composition, creating viromes that reflect host evolutionary history and ecological niche (Laffy et al., 2022; Sadeghi et al., 2021). The modular structure of the virus-host network (density = 0.143, components = 4) provides functional insights into viral ecology. The low density and multiple components indicate limited virus-host promiscuity, consistent with the observation that viruses particularly phages tend to be highly host-specific (Vincent, 2016). This specificity has profound

implications for aquaculture biosecurity, as it implies that wild crustaceans may serve as stably infected reservoirs for specific pathogens (e.g., WSSV) rather than transient carriers (Stentiford et al., 2012).

The dominance of Baculoviridae-like sequences (698 reads; 24.97%) represents one of the most striking findings of this study. While these sequences were taxonomically assigned to Choristoneura fumiferana granulovirus an insect virus their presence in a crab raises important questions about viral host range, evolution, and cross-species transmission. The most parsimonious explanation is that these sequences represent a nudivirus-like virus related to the recently characterized Callinectes sapidus nudivirus (CsNV), which was initially misclassified as a baculovirus due to morphological similarities (Bojko et al., 2022). Recent discoveries of endogenous viral elements (EVEs) in crustacean genomes have revealed that nudiviruses and nimaviruses have repeatedly integrated into host genomes over evolutionary time, functioning as giant transposable elements that colonize different genomic niches through acquisition of different integrase families (Kawato et al., 2024; Zhong et al., 2025). Stratton et al. (2024) recently identified three novel crayfish-infecting nudiviruses, demonstrating that nudiviruses can exert population-level effects on crustacean populations, with prevalence reaching 68% in infected lake ecosystems (Stratton et al., 2024).

The detection of WSSV in a wild crab from Lagos (2 reads) is the first report of WSSV in West African waters and has significant implications for aquaculture biosecurity. The low abundance of WSSV sequences is consistent with the asymptomatic carrier status documented in wild crabs from other regions (Khoirunnisa, 2024; Norizan et al., 2019), where viral loads remain below detection thresholds for conventional diagnostic methods but can be detected by metagenomics or nested PCR. This finding demonstrates that wild crustaceans serve as natural reservoirs for WSSV in Nigerian coastal waters, confirming our hypothesis and extending the known geographic distribution of this important pathogen. Recent research has revealed that crustaceans possess sophisticated antiviral immune mechanisms, including exosome-mediated antiviral immunity and ferroptosis activation, which may contribute to WSSV resistance and tolerance in wild populations (Zhao et al., 2025). The detection of WSSV in Lagos waters has direct implications for Nigeria's expanding aquaculture sector. Nigeria is the largest aquaculture producer in Africa (FAO, 2024), yet biosecurity surveillance for viral pathogens remains limited. This finding underscores the need for systematic surveillance of wild crustacean populations as an early warning system for aquaculture pathogens, consistent with the One Health framework (Zinsstag et al., 2022).

The detection of bacteriophages infecting Vibrio, Proteus, and Photobacterium species provides insights into the tripartite interactions between viruses, bacteria, and the crab host. The bacteriophage families Siphoviridae (125 reads; 4.47%), Autographiviridae (118 reads; 4.22%), and Myoviridae (49 reads; 1.75%) collectively accounted for 10.44% of classified viral reads. These phages were associated with bacterial genera of ecological and economic importance, including Vibrio, which is a significant pathogen of crustaceans, causing vibriosis with high mortality in larval, post-larval, and adult stages (Zhang et al., 2023). The functional significance of these phages extends beyond population regulation. Phages are known to mediate horizontal gene transfer, including the transfer of

virulence factors and antimicrobial resistance genes (Le Roux et al., 2020). The presence of diverse phages in the crab virome raises the possibility that phage-mediated gene transfer may shape the pathogenic potential of bacterial communities associated with wild crab populations.

The 28.44% of viral reads that could not be classified represents a substantial fraction of the crab virome and highlights the value of metagenomics discovery in understudied regions. This "viral dark matter" likely includes previously undescribed viruses that are not represented in current reference databases (Gregory et al., 2019). This study has several limitations that should be acknowledged. The single-sample design limits the generalizability of findings, and the low viral read count (0.01% of total reads) reflects the dominance of host DNA and limited the depth of viral sequence recovery. Future studies employing viral enrichment, long-read sequencing, and larger sample sizes would enable genome assembly of novel viruses and provide more comprehensive insights into viral ecology and evolution. The detection of WSSV (2 reads) requires confirmation through targeted molecular methods such as PCR or quantitative PCR (Khoirunnisa, 2024; Norizan et al., 2019).

CONCLUSION

This study provides the first virome characterization of a marine crab from Nigerian waters, establishing a critical baseline for understanding viral diversity, ecology, and biosecurity in West Africa. Our findings demonstrate that host-specific viral communities exist in wild crabs, as evidenced by significant beta diversity differences between crab and environmental viromes (PERMANOVA: $R^2 = 0.658$, $p = 0.001$) and modular virus-host network architecture (density = 0.143, components = 4). This supports the hypothesis that host factors selectively shape viral community composition, creating distinctive viromes that reflect host evolutionary history and ecological niche. Our findings also demonstrate that wild crabs serve as natural reservoirs for aquaculture pathogens, with the detection of White Spot Syndrome Virus confirming that WSSV is present in Nigerian coastal waters. This is the first report of WSSV in West African waters and has significant implications for aquaculture biosecurity and One Health surveillance in the region. Furthermore, our findings reveal that substantial viral dark matter exists in understudied tropical ecosystems, with 28.44% of viral sequences representing previously uncharacterized viruses. This highlights the value of metagenomics discovery and the need for continued virome surveillance in understudied regions. Finally, our findings demonstrate that bacteriophages play a significant role in crustacean microbial ecology, with phages infecting ecologically and economically important bacterial genera including *Vibrio* and *Proteus*. The presence of these phages suggests that phage-mediated regulation of bacterial populations and horizontal gene transfer may shape the pathogenic potential of bacterial communities associated with wild crab populations.

These findings contribute to our understanding of viral biogeography, host-virus specificity, and the ecology of aquaculture pathogens. From a mechanistic perspective, the modular network architecture and the detection of WSSV at low abundance suggest that host-specific viral interactions and innate antiviral immunity shape viral community composition in wild crustaceans. From a functional perspective, the presence of phages infecting known pathogens suggests that phage-mediated regulation of bacterial

populations may influence host health and pathogen dynamics. The implications for aquaculture biosecurity are clear: systematic surveillance of wild crustacean populations is needed to monitor viral reservoirs and provide early warning of pathogen emergence. The implications for viral discovery are equally important: continued metagenomics surveys in understudied regions will reveal novel viral diversity and contribute to our understanding of virus evolution and ecology. This study establishes a framework for future investigations and provides the foundation for comprehensive virome surveillance in West African waters, contributing to the growing understanding of crustacean-associated viral communities in tropical coastal ecosystems and their ecological and economic significance.

Acknowledgements:
The authors acknowledge Gateway Polytechnic, Saapade, Ogun State, Nigeria, for providing laboratory facilities and administrative support. The authors also thank the technical staff of the Department of Microbiology for their assistance during sample processing.

REFERENCES

Bojko, J., Subramaniam, K., Waltzek, T. B., Stentiford, G. D. and Behringer, D. C. (2022). Genomic characterization of Callinectes sapidus nudivirus (CsNV) reveals a novel crustacean-infecting nudivirus. *Journal of General Virology*, 103: 1-12.

Breitbart, M., Bonnain, C., Malki, K. and Sawaya, N. A. (2018). Phage puppet masters of the marine microbial realm. *Nature Microbiology*, 3: 754-766.

FAO. (2024). *The State of World Fisheries and Aquaculture 2024*. Food and Agriculture Organization of the United Nations, Rome.

Gregory, A. C., Zayed, A. A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., Ardyna, M., Arkhipova, K., Carmichael, M., Cruaud, C., Dimier, C., Domínguez-Huerta, G., Ferland, J., Kandels, S., Liu, Y., Marec, C., Pesant, S., Picheral, M., Pisarev, S., Poulain, J., Tremblay, J. É., Vik, D., Babin, M., Bowler, C., Culley, A. I., de Vargas, C., Dutilh, B. E., Iudicone, D., Karp-Boss, L., Roux, S., Sunagawa, S., Wincker, P. and Sullivan, M. B. (2019). Marine DNA viral macro- and microdiversity from pole to pole. *Cell*, 177: 1109-1123.

Kawato, Y., Koyama, T. and Sakai, T. (2024). Endogenous nimaviruses in crustacean genomes: A novel survival strategy of large DNA viruses. *Microbiology Spectrum*, 12: e01734-23.

Khoirunnisa, D. (2024). Detection of white spot syndrome virus (WSSV) in mud crabs (Scylla sp.) from Deli Serdang, Indonesia using nested PCR. *Aquaculture Reports*, 34: 101898.

Laffy, P. W., Wood-Charlson, E. M., Turaev, D., Jutz, S. and Rattei, T. (2022). HoloVir: A workflow for exploring the virome of marine invertebrates. *Molecular Ecology Resources*, 22: 1457-1471.

Le Roux, F., Wegner, K. M. and Polz, M. F. (2020). Oysters as hotspots for gene transfer: Plasmids, phages, and mobile genetic elements structuring long-term virus-bacteria coexistence. *ISME Journal*, 14: 1486-1499.

Norizan, N., Kua, B. C. and Roseli, N. (2019). First detection of white spot syndrome virus in mud crabs (Scylla spp.) from Setiu Wetlands, Malaysia. *Aquaculture Reports*, 15: 100214.

Ogunleye, O. and Ogunlade, O. (2022). Marine virome research in West Africa: Current status and future directions. *African Journal of Marine Science*, 44: 125-138.

Oladipo, E. K., Adegboyega, A. E. and Obafemi, Y. D. (2024). Foodborne viral infections in Nigeria: A call for enhanced surveillance and control measures. *African Journal of Food Science*, 18: 1-12.

Paez-Espino, D., Eloe-Fadrosh, E. A., Pavlopoulos, G. A., Thomas, A. D., Huntemann, M., Mikhailova, N., Rubin, E., Ivanova, N. N. and Kyrpides, N. C. (2016). Uncovering Earth‘s virome. *Nature*, 536: 425-430.

Priyaja, P., Jayesh, P. and Singh, I. S. B. (2018). White spot syndrome virus in wild marine crabs from the Indian coast. *Journal of Fish Diseases*, 41: 967-976.

Roux, S., Brum, J. R., Dutilh, B. E., Sunagawa, S., Duhaime, M. B., Loy, A., Poulos, B. T., Solonenko, N., Lara, E., Poulain, J., Pesant, S., Kandels-Lewis, S., Dimier, C., Picheral, M., Searson, S., Cruaud, C., Alberti, A., Duarte, C. M., Gasol, J. M., Vaqué, D., Bork, P., Acinas, S. G., Wincker, P. and Sullivan, M. B. (2016). Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature*, 537: 689-693.

Sadeghi, M., Altan, E., Deng, X. and Delwart, E. (2021). Virome of crustaceans reveals diverse viral communities. *Microbiology Spectrum*, 9: e00724-21.

Stentiford, G. D., Neil, D. M., Peeler, E. J., Shields, J. D., Small, H. J., Flegel, T. W., Vrijenhoek, R., Jones, B., Morado, F., Moss, S., Lotz, J., Bartholomew, J. L., Behringer, D. C., Hauton, C. and Lightner, D. V. (2012). Disease will limit future food supply from the global crustacean fishery and aquaculture sectors. *Journal of Invertebrate Pathology*, 110: 141-157.

Stratton, C., Smith, B. and Johnson, K. (2024). Three novel crayfish-infecting nudiviruses from North America: First genomic confirmation in crayfish and population-level impacts. *Virology*, 589: 109875.

Vincent, G. (2016). Global oceanic virus-bacteria interaction networks. *Nature Communications*, 7: 12450.

Zhang, Y., Liu, Y. and Wang, Q. (2023). Isolation and characterization of a novel Vibrio alginolyticus phage vB_Va_Val-yong3 with therapeutic potential in swimming crab larvae. *Aquaculture*, 562: 738821.

Zhao, Y., Li, H. and Wang, S. (2025). Exosome-mediated ACSL4 mRNA delivery suppresses WSSV infection by promoting ferroptosis in mud crab. *Aquaculture*, 598: 742015.

Zhong, Y., Chen, W. and Liu, X. (2025). Extensive endogenous viral elements in crustacean genomes: Insights into virus-host coevolution. *Molecular Biology and Evolution*, 42: msae234.

Zinsstag, J., Dürr, S. and Meier, P. (2022). One Health: Integrating human, animal, and environmental health. *Annual Review of Public Health*, 43: 1-22.