# Identification and classification of the genomes of novel Microviruses in poultry slaughterhouse

Ke-Ming Xie [#, 1, 2], Ben-Fu Lin[#, 3], Peng Zhu[2, 4], Xin-Yu Sun[1], Chang Liu[2], Guang-Feng Liu[2], Xu-Dong Cao[5], Jing-Qi Pan[1], Sui-Ping Qiu[1], Xiao-Qi Yuan[1], Meng-Shi Liang[1], Jing-Zhe Jiang*[, 1, 2], Li-Hong Yuan*[, 1]

[1.]School of Life Sciences and Biopharmaceutics, Guangdong Pharmaceutical University, Guangzhou 510006, Guangdong, China

[2.]Key Laboratory of South China Sea Fishery Resources Exploitation & Utilization, Ministry of Agriculture and Rural Affairs, South China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou 510300, Guangdong, China

[3.]Huadu District Animal Health Supervision Institution, Guangzhou510800, Guangdong, China

[4.]College of Marine Ecology and Environment, Shanghai Ocean University, Shanghai, China

[5.]Department of Chemical and Biological Engineering, University of Ottawa, Ottawa, ON, Canada

***Corresponding author:**

Li-Hong Yuan: ylh@gdpu.edu.cn orcid: 0000-0002-8752-0572

Jing-Zhe Jiang: jingzhejiang@gmail.com orcid: 0000-0001-5260-7822

**[#]Equal contribution**

**Abstract:** Microviridae is a family of phages with circular ssDNA genomes and they are widely found in various environments and organisms. In this study, Virome techniques were employed to explore potential members of Microviridae in poultry slaughterhouse, leading to the identification of 98 novel and complete microvirus genomes. Using a similarity clustering network classification approach, these viruses were found to belong to at least 6 new subfamilies within Microviridae and 3 higher-level taxonomic units. Analysis of their genomes found that the genome size, GC content and genome structure of these new taxa showed evident regularities, validating the rationality of our classification method. Compared with the 19 families classified by previous researchers for microviruses dataset, our method can divide microviruses into about 45 more detailed clusters, which may serve as a new standard for classifying Microviridae members.

Furthermore, addressing the scarcity of host information for microviruses, this study significantly broadened their host range and discovered over 20 possible new hosts, including important pathogenic bacteria such as Helicobacter pylori and Vibrio cholerae, as well as different taxa demonstrated differential host specificity. The findings of this study effectively expand the diversity of the Microviridae, providing new insights for their classification and identification. Additionally, it offers a novel perspective for monitoring and controlling pathogenic microorganisms in poultry slaughterhouse environments.

**Keywords:** Poultry slaughterhouse; Microviruses; Genome; Clustering; Host

# 1 Introduction

China is a major player in livestock and poultry farming and consumption. According to the statistics, China's total meat consumption is nearly 100 million tons, accounting for 27% of the global total. In 2022, the domestic meat production reached 92.27 million tons, with poultry meat contributing 24.43 million tons, constituting 26.5% of the total global meat production(1).

Slaughterhouses play a crucial role as an essential pathway for livestock and poultry meat products to move from farms to consumers' tables. They are also the key points for the gathering and transmission of pathogenic microorganisms(2). Due to the high density and mobility of poultry when entering the market or slaughterhouses, poultry comes from diverse sources and has varying hygienic conditions, and may carry multiple pathogenic microorganisms(3). The slaughter process is prone to contaminating the environment and the personnel involved. Additionally, the waste generated during poultry slaughter and processing further provides favorable conditions for the proliferation of pathogenic microorganisms(4). The interaction between animals, the environment, and occupational personnel forms a closed-loop microbial transmission chain. Some pathogenic microorganisms can infect occupational personnel through direct contact, while others may have an indirect impact by contaminating the environment. Existing research indicates that the detection rate of certain pathogenic microorganisms, such as Campylobacter and Salmonella in the case of bacteria, avian influenza viruses in the case of viruses, is significantly higher among occupational personnel in comparison to the general population(5-9). Therefore, conducting

extensive microbial research at the interface of animals, the environment, and occupational personnel in poultry slaughterhouses is of significant importance.

Bacteriophages, a type of viruses that specifically infect bacteria, are the most abundant life forms on earth(10). It is estimated that there are as many as $10^{31}$ virus particles on earth(11, 12), representing a vast and largely untapped reservoir of biological resources. In the preliminary research conducted by our research group, we identified a significant number of pathogens from poultry slaughterhouse samples(2), along with a vast amount of novel bacteriophages (unpublished data). On the one hand, the abundant presence of pathogens in slaughterhouses creates favorable conditions for the survival of bacteriophages. Investigating the diversity, types, and hosts of bacteriophages in poultry slaughterhouses can enhance our understanding of the composition, transmission, and the interplay between pathogens and bacteriophages in such a unique environment. On the other hand, in poultry slaughterhouses, occupational personnel are at the core of operations, and exposures to pathogenic bacteria increase the risk of infections of this particular group of population, which is a major public health safety hazard. Therefore, it would be advantageous to fully explore and develop potential functional phage species based on the high diversity of phages in poultry slaughterhouses, we can effectively purify the environment, block the spread of pathogenic bacteria in poultry slaughterhouses to safeguard public health safety.

Members of the Microviridae are one of the most widely distributed single-stranded DNA viruses and their natural hosts include pathogenic bacteria such as Spiroplasma, Chlamydia, and Enterobacteria(13). Despite earlier limited attention to the Microviridae, recent research indicates their significant importance in the virosphere(14). At present, the only subfamilies in Microviridae recognized by ICTV are Bullavirinae and Gokushovirinae(15), which cannot fully reflect the diversity of viruses in this familyWhile more Microviridae subfamilies, such as Alpavirinae(16) and Pichovirinae(14) have recently been proposed, the number of classified groups and host information about Microviridae remain severely limited in the literature. This study takes the unique and biologically significant environment of a poultry slaughterhouse in Guangzhou in Guangdong Province in China and employs a multiple displacement amplification (MDA) method(17). Combined with metagenomics sequencing to obtain environmental virus sequencing data from the poultry slaughterhouse (DSV, Dataset of Slaughterhouse Virome) in Guangzhou. Within this dataset, we discovered a diverse set of novel viruses belonging to the Microviridae. A

detailed analysis of 98 nearly complete Microviridae genomes revealed their classification into at least six new subfamilies and three higher-level taxonomic units. Comparative analysis with publicly available viral databases demonstrated the high resolution of our classification. Additionally, over 20 potential hosts for microviruses were identified. This study expands our knowledge of the evolution, diversity, and host range of Microviridae, providing insights into the potential biosecurity and ecological significance of these microviruses in poultry slaughterhouses.

# 2 Materials and Methods

## 2.1 Sample Collection

Atotal of three types of samples were collected from a poultry slaughterhouse in a district of Guangzhou: animals, occupational personnel, and environmental samples. The environmental samples included air, soil, sludge, swabs from transportation vehicles, and swabs from the slaughterhouse workshop. The collection protocols were as follows: (1) Animal Samples: Sterile cotton swabs were inserted into the oral cavity and cloaca of chickens or ducks, rotated three times, and then removed. The swab's tail was discarded, and the swab was immersed in sterile 0.5% BSA-PBS buffer for preservation. Three chickens or ducks from each of the three spaces (caged area, pre-slaughter area, slaughter area) had their oral and cloacal swabs mixed to form one sample. (2) Occupational Personnel Nasal Swab Samples: To collect nasal swab samples from occupational personnel, a sterile cotton swab was gently inserted into the nasal pharynx of the participating volunteer. After a few seconds, the swab was gently rotated and removed. The swab's tail was discarded, and the swab was immersed in sterile 0.5% BSA-PBS buffer for preservation. Nasal swab samples from a single person with both nostrils were placed in an individual sample collection tube. Written informed consent was obtained from all participants. (3) Air Samples: BioSamplers KIT (225-9595, SKC, Eighty Four, PA) were installed at approximately 1.5 meters above floor at the ventilation points in the slaughter area, pre-slaughter area, and caged area (3 sampling points in total). Using 0.5% BSA-PBS buffer at a flow rate of 8 mL/h, sampling was conducted for 12 hours per day at 110V. Each day's collection was considered one air sample, and this process was repeated continuously for 3 days. The collected samples were stored in PBS buffer. (4) Soil Samples: Soil samples were collected using the quincunx sampling method at

various spaces(18), including the entrance of the poultry slaughterhouse, the slaughter workshop, and the pre-slaughter caged area. Each sample weighed 5-10g. (5) Sludge Samples: Sludge samples were collected at the four corners of the sewage discharge pool, with approximately 10 mL of sewage collected per sample. (6) Environmental Swab Samples: Sterile cotton swabs were used to collect environmental samples from the slaughter workshop, pre-slaughter area, caged area, and poultry transportation vehicles. Five swab samples were collected from each space or vehicle, discarding the swab tails and placing them in sterile 0.5% BSA-PBS buffer for preservation. After collection, all samples were stored at 4°C, transported to the laboratory in a cooler, and then stored long-term at -80°C in an ultra-low-temperature freezer. This study was approved by the Medical Ethics Committee of the School of Public Health, Sun Yat-sen University (Permit No. [2018] No. 001).

## 2.2 Sample Pool Preparation

In order to analyze the virus content and types in samples from different spaces (i.e. slaughter area, pre-slaughter area, caged area) and different types (i.e. air, animals, sludge.) within the slaughterhouse, we combined samples of the same type collected from the same space to prepare sample pools: (1) Combined oral and cloacal swabs from 20 ducks in each of the three spaces (caged area, pre-slaughter area, slaughter area) to create a pool (total of 60 ducks). Combined oral and cloacal swabs from 30 chickens in each of the three spaces to create a pool (total of 90 chickens). (2) Combined nasal swab samples from 20 frontline slaughterhouse workers into one pool. (3) Combined air samples collected continuously for 3 days from each sampling point, creating one pool per sampling point. (4) Combined soil samples collected from each space (mixed samples with four or more points) into one pool. (5) Combined sludge samples collected from each sewage discharge pool into one pool. (6) Combined swab samples from seven slaughterhouse process points in the workshop into one pool. (7) Combined swab samples collected from three poultry transportation vehicles into one pool. Sample pool information is provided in Supplementary Table S2.

Table 1. Sample information.

| Sample | Source | Sample | Sample Pool Names | Sequencing |
|---|---|---|---|---|

| Classification | | | Quantity | (Abbreviations) | Number |
|---|---|---|---|---|---|
| Animal | | Chicken | 90 | SC | C-SXAO |
| | | Duck | 60 | SD | D-SXAO |
| Human | | Occupational Personnel | 20 | SN | CD-SXO-S |
| Environment | Air | caged area | 3 | AC | CD-SXA-1 |
| | | pre-slaughter area | 3 | AW | CD-SXA-2 |
| | | slaughter area | 3 | AS | CD-SXA-3 |
| | | soil | 12 | Soil | CD-SXG |
| | | sludge | 4 | Sludge | CD-SXD |
| | swab | transport vehicle | 15 | ST | CD-SXt |
| | | slaughterhouse | 35 | SS | CD-SXS |
| | | workshop | | | |

## 2.3 Virus Enrichment, Nucleic Acid Extraction and Amplification

Virus-like particles (VLPs) were enriched separately based on the different properties of the samples. Approximately 0.4g of sludge and soil samples were taken, and each was added to about 2–5 volumes of sterile SB buffer (0.2 M NaCl, 50 mM Tris-HCl, 5 mM CaCl2, 5 mM MgCl2, pH 7.5). For air samples and swab samples, they were directly added to 2–5 volumes of sterile SB buffer and shaken to fully dissolve the virus particles. After three cycles of freeze-thawing, the particles were completely resuspended in 10 times the volume of pre-chilled SB buffer. All samples were centrifuged at 1,000, 3,000, 5,000, 8,000, 10,000, and 12,000 × g for 5 minutes at 4°C using a Sigma 3K30 centrifuge (Sigma Laborzentrifugen GmbH, Germany), and the supernatant was collected. Subsequently, the supernatant was filtered through 0.22 μm Millipore filters (Burlington, MA) to further remove any cell debris and organelles. The filtrate was transferred to 28% sucrose solution and ultra centrifuged at 300,000 × g for 2 hours in a Himac CP 100WX ultracentrifuge (Hitachi, Tokyo, Japan). The supernatant was discarded, and the pellet was re-suspended in 720 μl of water, 90 μl of 10 × DNase I Buffer, and 90 μl of DNase I (1 U/μl) (TAKARA, Japan). The suspension was thoroughly re-suspended, incubated at 37°C with shaking for 60 minutes, stored overnight at 4°C, and then transferred to a 2 ml centrifuge tube.

158 Total nucleic acids were extracted using the HP Virus DNA/RNA Kit (R6873; Omega
159 Bio-Tek, Norcross, USA), and carrier RNA was not used during the process to avoid potential
160 interference with sequencing results. The concentration of RNA was quantified using the Qubit™
161 dsDNA HS Assay Kit (Q32851) and Qubit™ RNA HS Assay Kit (Q32855) (Thermo Fisher
162 Scientific, Waltham, USA).
163 Virome research heavily relies on amplification, as the viral biomass in natural samples is
164 often very low. Due to variations in most amplification methods, quantitative studies of viral data
165 present challenges at present(19, 20). In the current study, uniform genome amplification (WGA)
166 and transcriptome amplification (WTA) were performed using the repi-g Cell WGA and WTA Kit
167 (150052, Qiagen, Hilden, Germany), based on the multiple displacement amplification (MDA)
168 method(17, 21-23).

169 ## 2.4 Library Construction and Sequencing

170 The amplified DNA was quantified using gel electrophoresis and Nanodrop 2000
171 spectrophotometer (Thermo Fisher Scientific, Waltham, MA). Ultrasonic random shearing
172 (Covaris M220) was performed to generate fragments with lengths ≤800 bp. Fragment ends were
173 repaired using T4 DNA Polymerase (M4211, Promega, Madison, Wisconsin), Klenow DNA
174 Polymerase (KP810250, Epicentre, Madison, Wisconsin), and T4 Polynucleotide Kinase (EK0031,
175 Thermo Fisher Scientific, Waltham, MA). Fragments in the range of 300-800 bp were collected
176 after electrophoresis. After amplification, the libraries were pooled, and paired-end sequencing of
177 150 bp, 250 bp, or 300 bp was performed on the Novaseq 6000, HiSeq X ten, and Miseq platforms
178 (Illumina, San Diego, California)(24-26).

179 ## 2.5 Sequence Filtering

180 All samples underwent metavirome sequencing, yielding approximately 700 million raw
181 sequence reads. The sequencing data were subjected to quality control and removal of low-quality
182 and adapter sequences using Fastp (version 0.20.0) (27). The reads were then assembled into
183 contigs using Megahit (version 1.2.9) (28, 29). The contigs were aligned and annotated against the
184 NCBI non-redundant protein database using Diamond (version 0.9.24.125) (30). Subsequently,

Megan6was employed for further classification of the annotated results(31). A total of 98 viral sequences (Contig ID are shown in Table S1) were identified as complete genomes and annotated as belonging to the Microviridae for further in-depth analysis.

## 2.6 Open Reading Frame (ORF) Prediction and Alignment

Cenote-Taker2 was used to predict open reading frames (ORFs) in the 98 viral genomes(32). The major capsid protein or capsid protein sequence (major capsid protein is preferred if available, otherwise capsid protein is chosen. These proteins are collectively referred to as "Cap") was selected from the predicted results of each viral sequence. NCBI BLASTP(33, 34) was used to compare ORF sequences with the NR database, with an Expect threshold (e-value) set to $10^{-5}$. For each ORF in the alignment results, the top ten protein sequences with their complete genomic sequences were downloaded based on identity. Duplicate sequences were removed from all downloaded sequences. SnapGene (www.snapgene.com, version 4.3.6) was utilized to open the Cenote-Taker2 output file for visualizing the genomic structure.

## 2.7 Clustering Analysis Based on Sequence Similarity

Cap sequences predicted for DSV microvirus were collected, along with top 10 ranked Cap sequences from the aforementioned BLASTP results, and introduced 20 Cap sequences from microviruses that have been definitively classified by the ICTV. A total of 577 sequences were aligned with each other using DIAMOND (version 0.9.14.115) to build a matrix of sequence similarities. A clustering network graph was constructed based on alignment scores using Gephi(35) (version 0.9.7). The nodes were colored on different sequence sources, hosts, or virus classification results. Furthermore, the Cap sequences used by Paul et al. (36) were integrated with the above data. The same method was employed to construct a similarity clustering network graph and color it, aiming to compare the network clustering resolution of our research method with that of Paul et al.

## 2.8 Host prediction

All complete genome sequences included in the analysis in section 2.7 were subjected to host prediction using hostG(37) (output results taking genus, genus_score > 0.7) and cherry(38) (output results taking Top_1_label, Score_1 > 0.7), analyzing the relationship between these viruses and their hosts, as well as the proportion of these hosts.

### 2.9 Phylogenetic Tree Based on Cap Sequences

Cap is a conserved gene of microviruses(39) with approximately 500 amino acids in length, and is commonly used as a phylogenetic marker for the classification of evolutionary branches or subfamilies within the Microviridae. Multiple sequence alignment was performed using MAFFT(40) ambiguous regions were removed using TrimAl(41) , and a maximum likelihood phylogenetic tree based on Cap sequences was constructed using IQtree(42) (version 2.1.4). ModelFinder(43) was set to MFP (for ModelFinder Plus), and 1000 ultrafast bootstrap replicates were used. The tree was visualized using iTOL(44) (version 6.5.2) (https://itol.embl.de).

### 2.10 Principles of Classification and Naming of Viral Sequences

According to the clustering in Figure 1 and cherry host prediction results, DSV-related viral sequences are named respectively. Taking cluster_1 as an example, if a sequence has host prediction results, it is named based on the host, such as the contig sequences CD-SXS-WGA-1-k141_397009 and CD-SXD-WGA-1-k141_230904 are named Bdellovibrio microvirus C1_1 and Bdellovibrio microvirus C1_2, respectively. Similarly, CD-SXG-WGA-1-k141_33139 and CD-SXG-WGA-1-k141_32996 are named Escherichia microvirus C1_1 and Escherichia microvirus C1_2. If the sequence has no host prediction results, contig sequences like CD-SXD-SXG-WGA-all--k141_113185 and CD-SXD-SXG-WGA-all--k141_328845 are named DSV microvirus C1_1 and DSV microvirus C1_2, and so forth. The original sequence ID and their corresponding names are listed in Table S1.

## 3 Results

### 3.1 Discovery of Novel Subfamilies of Microviridae

237 According to the ICTV standards, Microviridae includes two subfamilies (Bullavirinae and
238 Gokushovirinae) and seven described genera(15). Among them, the subfamily Bullavirinae has
239 three genera, comprising 14 species. The subfamily Gokushovirinae has four genera, consisting of
240 eight species. We selected 98 complete DNA viral genomes from the DSV dataset, annotated as
241 Microviridae, with a genome integrity exceeding 90%, for in-depth analysis. All DSV genomes
242 have lengths ranging from 4 - 6 kb, consistent with the genome size of Microviridae(13). As
243 predicted, these viruses all have Cap with lengths of 450-600 amino acids (AA). According to the
244 Cap similarity clustering network graph (Figure 1), the Microviridae sequences from DSV, along
245 with the related sequences aligned in NR and the Microviridae sequences from ICTV (totaling 577
246 sequences), roughly cluster into 9 clusters (cluster_1 to 8 and Bullavirinae). Among them, the 14
247 sequences of Bullavirinae (14 distinct species recognized by ICTV) formed a separate cluster
248 (light green). It is worth noting that cluster_1 (C1) and cluster_2 (C2) include an additional 6
249 ICTV sequences, all of which belong to Gokushovirinae. In C1, there are 5 ICTV sequences, 4 of
250 which belong to Chlamydiamicrovirus, and 1 belongs to Spiromicrovirus. In C2, 1 ICTV sequence
251 belongs to Bdellomicrovirus. According to this classification standard, the other unclassified
252 viruses in C1 and C2 should also belong to the Gokushovirinae. The remaining 6 clusters
253 (cluster_3 to 8) do not include ICTV sequences, suggesting that these clusters might represent
254 novel subfamilies within the Microviridae.

255 Based on the similarity of these 9 clusters, we can further categorize them into 5 major
256 families, tentatively referred to as Family_Red, Family_Blue, Family_Green, Family_Yellow, and
257 the independent Bullavirinae cluster. Among them, Family_Yellow includes 3 clusters, namely
258 cluster_6 (C6), cluster_7 (C7), and cluster_8 (C8). Family_Red includes C1, C2, and cluster_3
259 (C3), which contain 6 Gokushovirinae ICTV sequences. Therefore, for now, we equate
260 Family_Red with the Gokushovirinae. However, Family_Blue, Family_Green, and
261 Family_Yellow do not include any ICTVsequences, suggesting that they are newly discovered
262 taxonomic units in this study, at least on par with Gokushovirinae and Bullavirinae. Since
263 Microviridae is already classified at the family level, whether Family_Blue, Family_Green, and
264 Family_Yellow are proposed as new subfamilies or families requires further discussion.
265

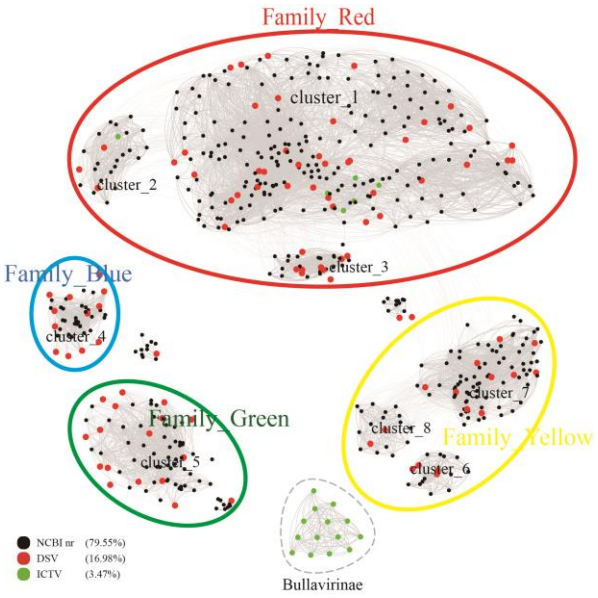



**Figure 1.** Similarity clustering network of Cap of microviruses from poultry slaughterhouses and related microvirus groups. The network includes identified microvirus Cap sequences from the DSV data (n=98), along with related Cap sequences from the NR data (n=459) and microvirus Cap sequences from the ICTV data (n=20). The similarity clustering network was constructed using Gephi (version 0.9.7) based on Diamond (version 0.9.14.115) alignment score. Gray connections represent Diamond Blastp score >480.

## 3.2 Expanding the Potential Hosts of Microviruses.

Host prediction was performed on the 98 newly discovered microvirus sequences from this study, along with their associated 459 NR sequences and 20 ICTV sequences, using hostG(37) and cherry(38) . In the hostG results, only NC_002643.1 from ICTV was accurately predicted to have a host (Bdellovibrio). However, in the cherry results, the majority of hosts were consistent with the ICTV results, indicating that the success rate and accuracy of cherry predictions were higher than hostG. According to the hostG results, the main hosts for DSV were Bdellovibrio and Chlamydia (Figure 2a, b). For NR sequences, the hosts were mainly distributed in the Bdellovibrio, Chlamydia, and Parabacteroides. Although these results align well with the current understanding of microvirus hosts, results from cherry suggest (Figure 2c, d) that the hosts of microviruses may be far more diverse than these three genera.

Although Bdellovibrio bacteriovorus and Escherichia coli are the predominant hosts in the

NR-derived virus hosts, respectively. Cherry also predicted hosts such as Caulobacter vibrioides and Bacillus halmapalus, indicating numerous microvirus hosts that have not been previously reported. Among the DSV-derived virus hosts, Bdellovibrio bacteriovorus still dominated, followed by Caulobacter vibrioides and Candidatus Pelagibacter ubique, representing novel hosts. In addition, NR data revealed the presence of human and animal pathogens such as Helicobacter pylori and Enterobacter cancerogenus. The DSV data also identified potential hosts including Vibrio cholerae and Pseudomonas aeruginosa. This suggests that the host range of microviruses within the Microviridae may be extensive, and that there are likely more potential hosts yet to be discovered.

From the perspective of sample types, the highest abundance of microviruses was observed in soil and sludge samples, corresponding to a higher diversity and quantity of their respective hosts (Figure 2e). Bdellovibrio bacteriovorus, as a typical host for microviruses, showed a higher proportion across various samples. Caulobacter vibrioides also exhibited high abundance in sludge, soil, and the slaughterhouse workshop (Figure 2e). While Escherichia coli, Gordonia terrae, and Pseudomonas aeruginosa were predicted only in soil samples, Vibrio cholerae was exclusively found in sludge samples. Other host bacteria were detected across different sample types. This indicates a close relationship between the detection of microviruses and the distribution of their host bacteria, displaying certain characteristics in various sample types.

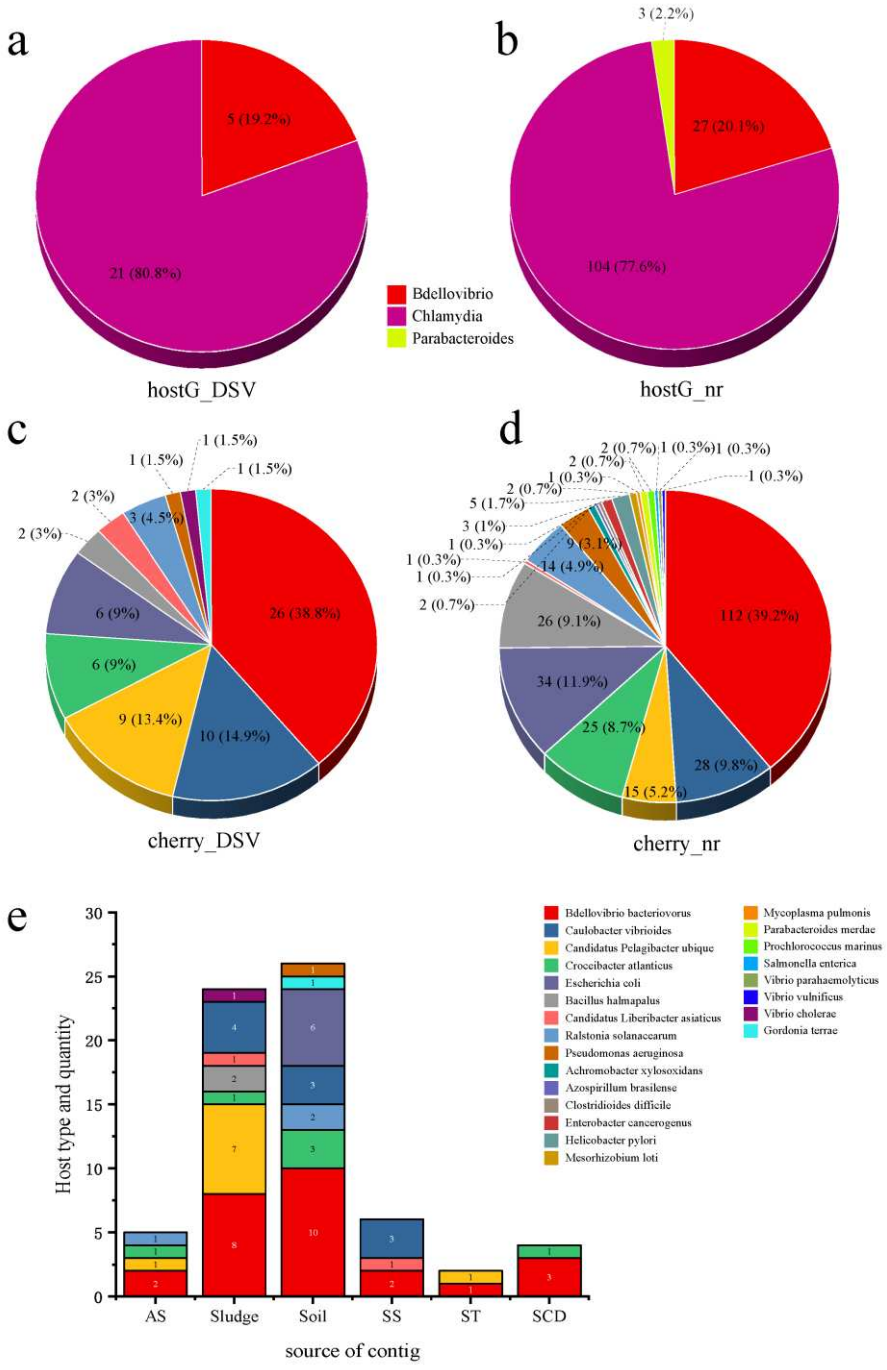



**Figure 2.** The host types and quantity statistics of microviruses from poultry slaughterhouses and related groups. (a) hostG(37) results of DSV sequences. (b) hostG results of NR sequences. (c) cherry(38) results of DSV sequences. (d) cherry results of NR sequences. Score > 0.7. (e) Host types and quantity predicted by cherry for corresponding DSV sequences. AS (All soil and sludge of slaughterhouse); Soil (Soil of slaughterhouse); Sludge (Sludge of slaughterhouse); SS (Swab of slaughterhouse workshop); ST (Swab of poultry transport vehicle); SCD (Oral and cloacal swabs of chickens and ducks).

## 3.3 Genome Length and GC Content

The genome sizes and GC content of viruses within the same family or genus are usually relatively consistent(45, 46). Based on the identification of 9 clusters in the previous sections, we further created boxplots illustrating their genome size and GC content (Figure 3). Both genome size and GC% exhibited high consistency within each of the 9 clusters, while significant differences were observed among different clusters. For instance, Bullavirinae showed distinct genome sizes and GC content compared to other groups. Individual scattered black dots outside the boxes in the figure represent sequences from the NR data. These results indicate that the genome characteristics of microviruses from different taxonomic groups exhibit good consistency and indirectly validate the reliability of our classification method based on the similarity clustering network graph.

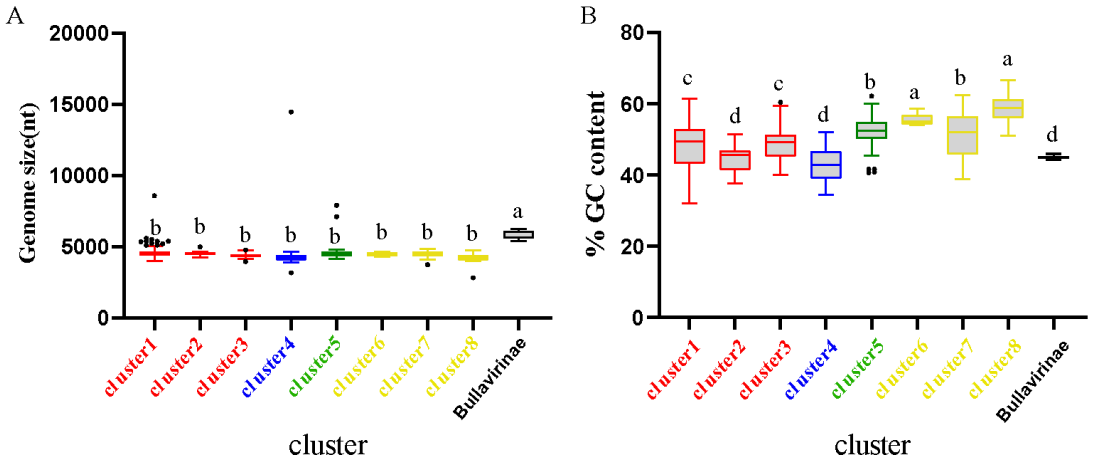



**Figure 3.** Genomic features of each cluster of microviruses from poultry slaughterhouses and related groups. (A) Distribution of microviruses genome sizes in each cluster from Result 3.1. (B) Distribution of microviruses genome %GC content in each cluster from Result 3.1. Red, blue, green, yellow, and black correspond to Family_Red, Family_Blue, Family_Green, Family_Yellow, and Bullavirinae, respectively. Turkey's test was used, where $P < 0.05$ indicates significant differences, and $P > 0.05$ indicates no significant differences. In the group where the maximum mean value is located, mark it with the letter "a." Then, compare this mean value with the mean values of other groups one by one. If there is no significant difference, label them with the same letter "a." Continue this process until encountering a mean value with a significant difference, then label it with the letter "b." Subsequently, use "b" as the standard for further comparisons. Repeat this process, labeling consecutive mean

values with the letters "b" until encountering a mean value with a significant difference, which is then labeled as the letter "c." This pattern continues for subsequent comparisons. The plot displays median values, 25th and 75th percentiles, 1.5 interquartile ranges, and outlier data points.

## 3.4 Phylogenetic Analysis Based on Cap Sequences

To better illustrate the diversity of DSV-related microviruses and their evolutionary origins, phylogenetic trees were constructed for each cluster based on the results in Figure 1. The sequences of DSV-related microviruses were classified and named according to the sample source and host type of the viruses (see Materials and Methods 2.10 for reference).

C1 is the cluster with the highest number of members among the eight clusters and exhibits the most diverse range of host sources (Figure 4. Displayed are partial positions of the phylogenetic tree. Some phylogenetic branches have been collapsed, the complete phylogenetic tree is detailed in Figure S1). Notably, in addition to typical hosts such as Escherichia coli and Bdellovibrio bacteriovorus, this cluster has hosts that were previously unreported, such as Caulobacter vibrioides, Pseudomonas aeruginosa, and Helicobacter pylori. Caulobacter vibrioides is a Gram-negative oligotrophic bacterium widely distributed in freshwater lakes and streams, serving as an important model organism for studying cell cycle regulation, asymmetric cell division, and cell differentiation(47). Pseudomonas aeruginosa is a common multidrug-resistant pathogen, characterized by its capsule, Gram-negative nature, and aerobic or facultatively anaerobic growth, causing diseases in plants and animals, including humans(48). Helicobacter pylori is a Gram-negative, flagellated spiral bacterium, classified as a class I carcinogen, responsible for approximately 89% of gastric cancer cases and associated with 5.5% of cancer cases worldwide(49-51). In general, hosts within the same clustering branch are relatively homogeneous. For example, in Figure 4, the hosts in the purple-colored block branch are primarily Bdellovibrio bacteriovorus, while the hosts in the deep blue-colored block branch are mainly Caulobacter vibrioides.

From the sample sources perspective, DSV viruses in this cluster mainly originate from soil (CD-SXG) and sludge (CD-SXD), with a few from swab samples taken in the slaughterhouse workshop environment (CD-SXS) (Supplementary Table S1). In comparison, the sources of NR

viruses are more diverse, including animal metagenomes, wastewater metagenomes, human metagenomes, and blackflies. As seen in the genomic structure diagram in Figure 4, members of C1 typically possess signature genes such as Major capsid protein or capsid protein(Cap), and Replication associated protein (Rep). Moreover, the genomes in this cluster often exhibit a sequential arrangement of Cap, Minor capsid protein (MinCP), Scaffold protein(SP), Rep, and DNA binding protein (DBP). However, a few viruses in this cluster have genome organization sequences that deviate from this pattern, such as DSV microvirus C1_4. Furthermore, Replication-associated protein was not predicted in DSV microvirus C1_1 and DSV microvirus C1_3. Coincidentally, these two sequences also lack host prediction results, likely suggesting the novelty of these viral genomes. Overall, sequences with closer phylogenetic relationships tend to exhibit more apparent consistency in host specificity and genomic structure.

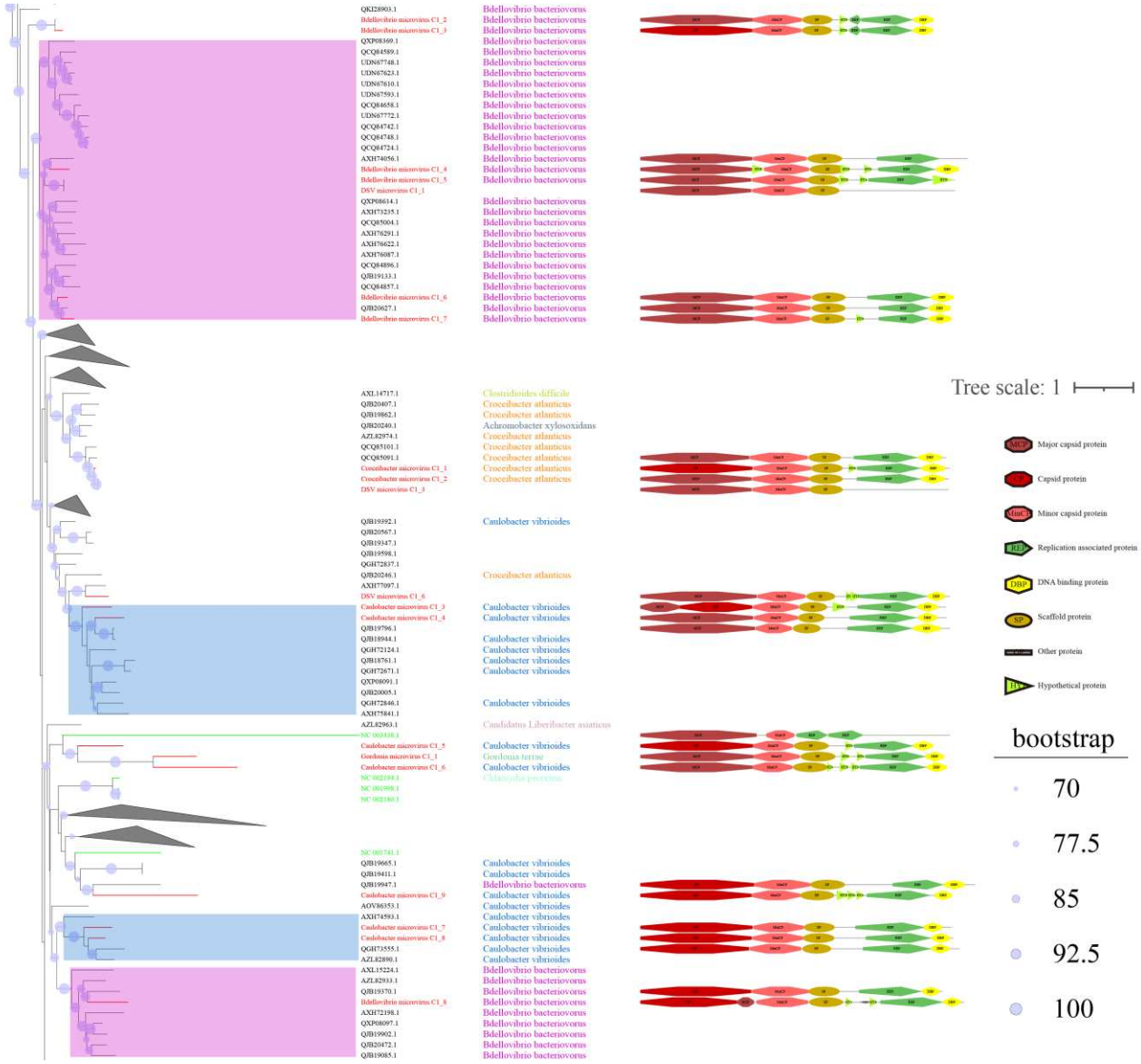


**Figure 4.** Phylogenetic tree, hosts, and genomic structure of cluster_1 microviruses from poultry slaughterhouse and related sources. The maximum likelihood phylogenetic tree was constructed based on the Cap sequences of Microviruses using IQtree (version 2.1.4). ModelFinder was set to MFP, and 1000 ultrafast bootstrap replicates were performed, displaying bootstrap values > 70. The red branches represent DSV microvirus sequences, green branches represent ICTV sequences, and black branches represent NR sequences. The third column shows host annotations predicted by Cherry, and the fourth column displays partial genomic structure diagrams.

C2 is a small viral cluster with a consistent host source, all being Bdellovibrio bacteriovorus, and a highly consistent genomic structure (Figure 5). The viruses in this cluster exhibit a sequential arrangement of Cap, MinCP, SP, DBP, and Rep, with Hypothetical protein (HYP) inserted on either side of DBP. Bdellovibrio microvirus C2_2 and Bdellovibrio microvirus C2_1 are closely related to AZL82867.1 and QJB19506.1, respectively. The genome of AZL82867.1 is derived from Honey bees, while QJB19506.1 originates from wastewater metagenome. This observation suggests the widespread presence of microviruses and their Bdellovibrio bacteriovorus hosts in various environmental settings. Only three sequences in C3 were predicted to have hosts (Figure S2), indicating that this group lacks sufficient host information and is a relatively novel group compared to C1 and C2. The predicted hosts for these three sequences are Azospirillum brasilense (A. brasilense) and Enterobacter cancerogenus (E. cancerogenus). A. brasilense is a microaerophilic nitrogen-fixing bacterium widely present in the rhizosphere worldwide, promoting plant growth(52, 53). E. cancerogenus is a significant pathogen commonly found in human clinical specimens such as blood and cerebrospinal fluid. It is not sensitive to penicillin and cephalosporin(54), and exploring bacteriophage targeting such multidrug-resistant pathogens is meaningful for developing phage therapy methods.

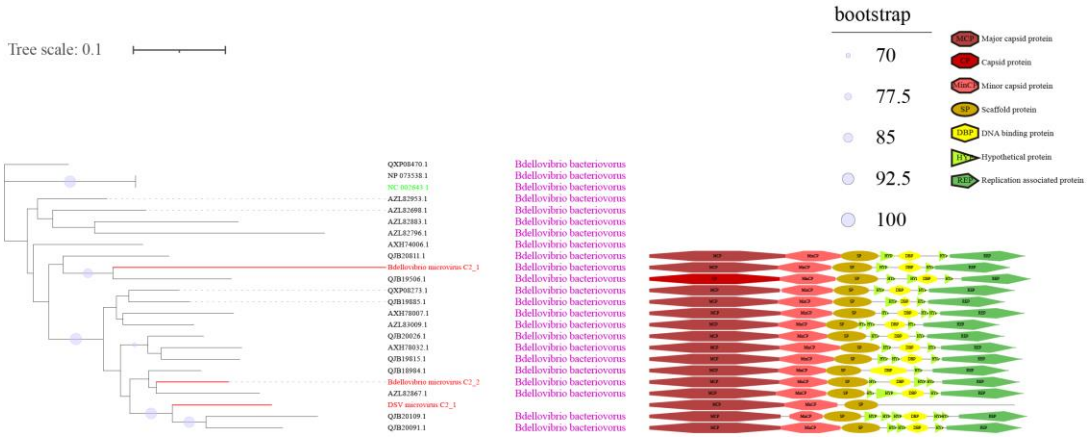



**Figure 5.** Phylogenetic tree, hosts, and genomic structure of cluster_2 microviruses from poultry slaughterhouse and related sources. The maximum likelihood phylogenetic tree was constructed based on the Cap sequences of Microviruses using IQtree (version 2.1.4). ModelFinder was set to MFP, and 1000 ultrafast bootstrap replicates were performed, displaying bootstrap values > 70. The red branches represent DSV microvirus sequences, green branches represent ICTV sequences, and black branches represent NR sequences. The third column shows host annotations predicted by Cherry, and the fourth column displays partial genomic structure diagrams.

Cluster_4 (C4) generally exhibits a relatively tidy genome structure (Figure S3). It is noteworthy that, despite being in the same cluster, there are significant differences in the host sources between DSV and NR viruses in C4. The majority of NR viruses are derived from wastewater metagenome samples, and their hosts are predominantly Bacillus halmapalus. Bacillus halmapalus, a halophilic bacterium, is a Gram-positive, alkaliphilic, alkalitolerant, facultative anaerobe. It is commonly isolated from soil, and its pathogenicity is not well understood(55). In DSV, only two viruses have Bacillus halmapalus as their host, and both are derived from sludge samples, aligning with the source of this bacterial species. Unlike NR, the primary hosts for DSV viruses are Candidatus Pelagibacter ubique. Except for Candidatus Pelagibacter microvirus C4_1, which originates from a swab of the transportation vehicle (CD-SXt), the rest are all from sludge samples (CD-SXD) (Table S1). Studies suggest that Candidatus Pelagibacter species may be among the most abundant bacteria globally and play a crucial role in the carbon cycle on Earth. Croceibacter atlanticus belongs to the phylum Bacteroidetes and is a species isolated from the Atlantic Ocean(56). Croceibacter microvirus C4_1 is the only virus in this cluster derived from swab of duck oral and cloaca (D-SXAO) (Table S1), and it is specifically associated with the host

Croceibacter atlanticus. This observation once again confirms the conclusion from Figure 2e that the detection of microviruses is closely related to the distribution of their host bacteria and the source of the samples.

Cluster_5 (C5), as shown in Figure S4. Although most NR sequences include Cap and Rep, DSV sequences, such as Bdellovibrio microvirus C5_2/4/7, Escherichia microvirus C5_1/2/13, only predict 2 ORFs: capsid protein and MinCP. We have not observed a correlation between this situation and sample sources, indicating a potentially higher novelty and lower conservation of genes in DSV sequences. C6 (Figure S5), similar to C3 (Figure S2), is a smaller cluster without predicted hosts, indicating a need for further research on this cluster. The genomic structure of the C7 sequences is primarily arranged in the order of Cap, MinCP, SP, Rep, and DBP (Figure S6). Croceibacter microvirus C7_1 and Croceibacter microvirus C7_2, two viruses within the same major branch, share Croceibacter atlanticus as their host (previously introduced in C4). This branch is the only one with predicted host results, while other branches lack host predictions. Therefore, C7 is also a potentially interesting virus cluster worthy of in-depth research. C8 (Figure S7) has two distinct hosts, with Ralstonia solanacearum being the dominant host and Achromobacter xylosoxidans as the second host. Ralstonia solanacearum is considered one of the most important plant pathogens due to its lethal nature, persistence, wide host range, and extensive geographical distribution(57). Achromobacter xylosoxidans belongs to the genus Achromobacter and is commonly found in moist environments, causing diseases such as bacteremia, pneumonia, pharyngitis, and urinary tract infections(58, 59).

## 3.5 Comparing DSV Viruses in Microvirus's Virosphere

To better understand the relationship between the identified microviruses in the poultry slaughterhouse and other reported microviruses in the Microviridae family, we expanded our focus beyond the 577 viral genomes highlighted in this study (Figure 1). To this end, an additional set of 4077 microvirus Cap sequences (utilizing 4,007 sequences for this study) studied by Paul et al.(36) were incorporated into our analysis for a more comprehensive clustering analysis. In the study by Paul et al., microviruses were broadly classified into 19 families, corresponding to the 19

color-coded clusters in Figure 6. Upon comparing the clustering results between Figure 1 and Figure 6, there is a good overall agreement between the two figures. Specifically, the four families identified in Figure 1 are concentrated within the purple cluster in Figure 6, representing the largest cluster in Family 3, as defined by Paul et al. DSV viral sequences are predominantly distributed within Gokushovirinae A(36), Shukshmavirinae(60) and Group D(61) of Family 3. Additionally, three scattered sequences are found in Pichovirinae(14), Gokushovirinae B and Gokushovirinae C(36), indicating that microviruses in the poultry slaughterhouse environment primarily belong to these groups. This result suggests that, although microviruses in the poultry slaughterhouse environment exhibit high diversity and novelty, they may still be relatively underrepresented in the entire microvirus virosphere. Family 3, possibly due to its close association with the human environment, is the largest group within the Microviridae. The expansion of other groups awaits further supplementation with samples from different sources and microbial hosts.

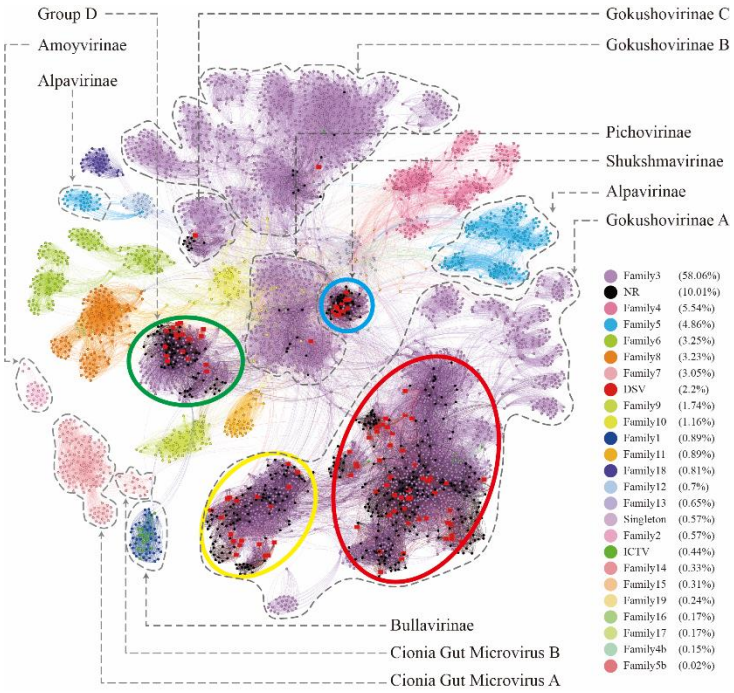



**Figure 6.** The microviruses collection with diverse taxa. Similarity clustering network was constructed using microviruses Cap sequences identified from DSV data (n=98), along with related Cap sequences from NR (n=459), ICTV microviruses Cap sequences (n=20), and an additional set of Cap sequences reported by Paul et al. (n=4007) (red dots represent DSV sequences, black dots represent NR sequences, and dark green dots represent ICTV sequences; other dots are colored based on the families defined by Paul et al.(36) ). Clusters corresponding

to those in Figure 1 are enclosed by ellipses of four different colors. Labels such as Pichovirinae, Shukshmavirinae, Group D, Alpavirinae, Gokushovirinae A/B/C correspond to subfamilies reported by previous studies(14, 16, 60, 61) and suggested classifications by Paul et al.(36) The similarity clustering network graph was created using Gephi (version 0.9.7) based on Diamond (version 0.9.14.115) alignment score, with gray edges indicating Diamond Blastp score >0.

On the other hand, according to the clustering results in Figure 6, the 19 major families delineated by Paul et al. can be further subdivided into approximately 45 smaller clusters. Particularly, within family 3 (the purple clusters), our clustering method can split it into as many as 20 smaller clusters. Specifically, Group D, Shukshmavirinae, Pichovirinae, and Gokushovirinae C each form independent cluster, while Gokushovirinae A and Gokushovirinae B can be further divided into multiple clusters at the same hierarchical level. These clusters are at least equivalent in status to Group D, Shukshmavirinae, and Pichovirinae. Additionally, families of other colors can also be further subdivided into smaller clusters. For example, family 5 identified as Alpavirinae(16) can be distinctly clustered into 4 different clusters in this study. This suggests that these smaller clusters may represent novel subfamilies or families that require further identification.

## 3.6 The Relationship between the Clusters of Microviruses and Their Host Sources

According to the cherry(38) results, the points in Figure 1 and Figure 6 are colored coded on host types in Figure 7. As shown in Figure 7a, clusters C2, C4, C5, a portion of C7, C8, and the Bullavirinae cluster exhibit clear host specificity, while the host colors in cluster C1 are highly mixed. Specifically, Bdellovibrio bacteriovorus is mainly the host for C2 and C5, the hosts for C4 are primarily Bacillus halmapalus and Candidatus Pelagibacter ubique, the main host for C8 is Ralstonia solanacearum, and only some sequences in C7 have host results, all of which are associated with Croceibacter atlanticus. Cluster C1 includes a significant number of Bdellovibrio bacteriovorus viruses, as well as viruses infecting Escherichia coli, Caulobacter vibrioides, Croceibacter atlanticus, and other bacteria. This may suggest that this group of viruses is more

prone to host jumping compared to other viruses group.

Similar to the results in Figure 7a, the host sources of Family_Red members in Figure 7b remain diverse, primarily involving Bdellovibrio bacteriovorus, Caulobacter vibrioides, Croceibacter atlanticus, Escherichia coli. As the number of members increases, Family_Green shows an expanded range of host types, mainly associated with Caulobacter vibrioides, Bdellovibrio bacteriovorus, and Candidatus Pelagibacter ubique. Family_Blue continues to be dominated by Bacillus halmapalus and Candidatus Hamiltonella defensa. Although the number of Family_Yellow members has increased significantly, a majority still lacks predicted hosts. Apart from the four main groups focused on in this study, the host types of most smaller groups are relatively singular, such as Bacillus halmapalus (family4), Escherichia coli (family5), Rhodobacter capsulatus (family7) (Figure 7b). These results suggest variations in host specificity among different viral groups. Moreover, the good correspondence between similarity clustering networks and host prediction results is evident.

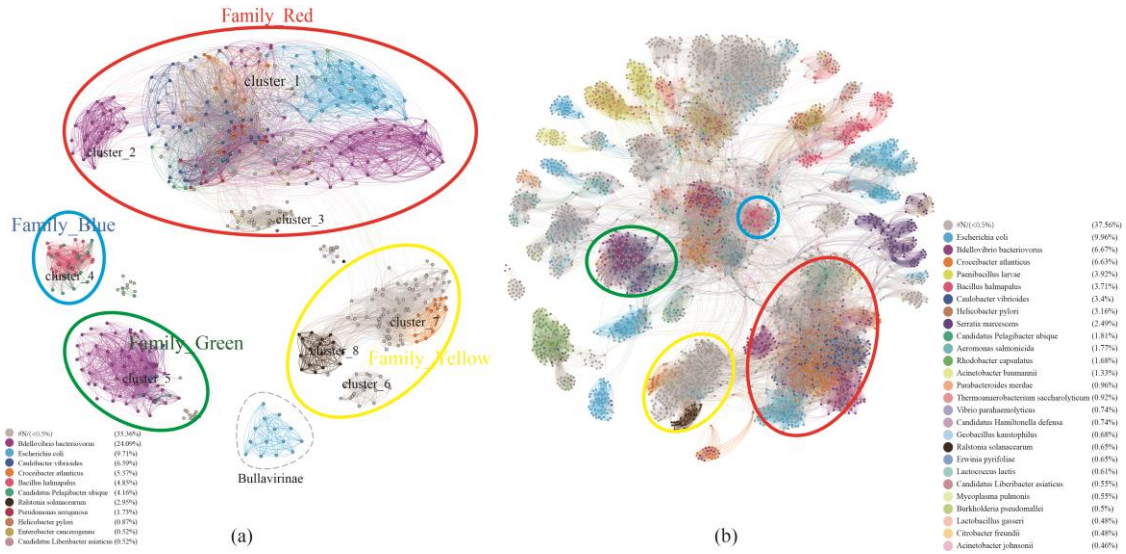



**Figure 7.** Host specificity of different clusters of Microviruses. (a) Similarity clustering network constructed using microviruses Cap sequences identified in DSV data (n=98), along with related Cap sequences from NR (n=459), and microviruses Cap sequences from ICTV (n=20), colored based on host types. (b) Based on the sequences in (a), an extended similarity clustering network was constructed by introducing Cap sequences reported by Paul et al. (n=4007), also colored according to host types. The similarity clustering network graph was created using Gephi (version 0.9.7) based on Diamond (version 0.9.14.115) alignment score, with gray edges indicating Diamond Blastp score >0.

# 4 Discussion

Viruses are the most abundant life forms on earth. It is estimated that there are as many as $10^{31}$ virus particles on earth(11, 12). However, The International Committee on Taxonomy of Viruses (ICTV) has officially recognized only around 12,000 known virus species (https://ictv.global/vmr, Version: VMR MSL38 v1). Viruses are often considered the "dark matter" of life sciences. Due to the challenges in cultivating many viruses, our understanding is limited to those that are easily cultivated and have significant impacts on humans or the economy. Advances in high-throughput sequencing and virome technologies have overcome the dependency on host cell cultures in traditional virology research, greatly enhancing the efficiency of discovering and identifying new viruses(62) . In recent years, virome technologies has been widely applied in various studies, including marine environments and research on vertebrates and invertebrates, leading to the identification of numerous novel viruses(63, 64) and significantly expanding our knowledge of the viral world.

Microviridae is one of the most common families of single-stranded DNA (ssDNA) viruses. Compared to double-stranded DNA (dsDNA) phages, the genomes of Microviridae are smaller, typically exhibiting higher safety by being less prone to carry virulence and resistance genes(36). Moreover, they are widely distributed across various ecosystems(65, 66), representing a relatively accessible and exploitable source of DNA resources. As of now, the ICTV recognizes only two subfamilies within Microviridae, namely Bullavirinae and Gokushovirinae. However, this classification does not fully capture the extensive diversity of newly reported microviruses(36). In recent years, numerous new taxonomic groups within the Microviridae family have been proposed. For instance, Krupovic et al. introduced a novel Microviridae subfamily named Alpavirinae, which was identified as prophage(16). Additionally, several newly proposed subfamilies of Microviridae include Pichovirinae from the human gut(14), Sukshmavirinae from termites(60), Group D from dragonflies(61), and Aravirinae and Stokavirinae from sphagnum- dominated peatlands(39). Paul et al. comprehensively analyzed the genomes of microviruses using their classification method, providing insights into the diversity, distribution, and host range of this viral

group(36). The proposed new classifications are clearly represented in the clustering network graph of this study (Figure 6), indicating a good validation across different research efforts.

As far as we know, this study represents a relatively comprehensive compilation of members of the Microviridae, providing an overview of the classification of Microviridae and holding significant importance for the identification, exploration, and expansion of Microviridae. However, due to the large number of potential new taxa, this study did not assign explicit taxonomic names to them, focusing instead on demonstrating relationships between clusters. We believe that as more members of Microviridae are discovered and identified, this family will continue to give rise to new taxa and may undergo redefinition. To address this situation, there is an urgent need for a universal and straightforward method for classification, such as utilizing numbers or letters to systematically name newly emerging taxa.

The evolutionary trajectory of dsDNA phages is primarily influenced by horizontal gene exchange, driving the diversity and adaptive evolution of this phage class. However, for ssDNA phages, the evolutionary patterns may fundamentally differ(67-69). For instance, in microviruses, gene recombination is not widespread, and the presence of Cap may limit the insertion of foreign DNA sequences(70) , potentially restricting gene transfer at the horizontal level. Despite these factors, microviruses exhibit high mutation rates in their genomes(71), suggesting that they may employ different evolutionary strategies to enhance adaptability. This adaptability is evident in the diverse clusters and extensive host range discovered in this study, as well as the diversity in hosts and genome structures found even within the same phylogenetic branch. Additionally, both this study and others(14) have observed differences in the genome structures of microviruses from various sample sources or types. This reflects the complexity and diversity of their evolution, showcasing their ability to adapt to different environments.

At present, the mainstream view suggests that the hosts of microviruses are primarily intracellular parasitic bacteria and Enterobacteria. For instance, the host of the Bullavirinae is Enterobacteria, and detailed studies have been conducted on representatives of this family, such as the phage ΦX174(69). Members of the the Gokushovirinae only infect Chlamydia, Bdellovibrio and Spiroplasma(13, 72). However, an increasing number of studies indicate that microviruses can infect a broader range of bacterial hosts, including Vibrio parahaemolyticus(73, 74) , Salmonella(75), Shigella flexneri(76) and other bacteria. To address the question of infecting hosts,

this study employed two new host prediction methods. The hostG utilizes shared protein clusters between viruses and prokaryotes to create a knowledge graph and trains a graph convolutional network for prediction(37). While it achieves high prediction accuracy, its results tend to be conservative and can only predict hosts at the genus level. Cherry is described as having the highest accuracy in identifying virus-prokaryote interactions, outperforming all existing methods at the species level, with an accuracy of 80%(38). The results from Cherry indicate that the hosts of the Microviridae exhibit extremely high diversity, including various pathogens such as Mycoplasma pulmonis, Helicobacter pylori, Vibrio cholerae, Clostridioides difficile, and Pseudomonas aeruginosa. Additionally, this study identified some plant-pathogenic bacteria, such as Ralstonia solanacearum (R. solanacearum) and Candidatus Liberibacter asiaticus (CLas). Bacterial wilt, caused by R. solanacearum, is economically significant as it can infect over 250 plant species, including potatoes, tomatoes, and tobacco, causing substantial yield losses in tropical and subtropical regions(77, 78). CLas is the pathogen responsible for citrus Huanglongbing (HLB, also known as citrus greening disease)(79), a highly destructive disease threatening global citrus production. There has been limited research on Microviridae infecting plant-pathogenic bacteria, and the findings of this study suggest that Microviridae also holds potential for applications in the control of bacterial diseases in plants.

# 5 Conclusion

This study employed virome techniques to thoroughly explore potential members of Microviridae in a poultry slaughterhouse, successfully identifying and analyzing 98 novel and complete microvirus genomes. Based on the similarity of Cap proteins, it was discovered that these genomes represent at least six new subfamilies within Microviridae, distinct from Bullavirinae and Gokushovirinae, as well as three higher-level classification units. These new taxa exhibit obvious regularities in genome size, GC content, and genome structure, further highlighting the rationality of the classification method used in this study. Additionally, based on the 19 families classified by previous researchers for all microviruses, our approach divides microviruses into about 45 more detailed clusters, which may serve as a new standard for classifying Microviridae members. The current information on microviruses' hosts remains limited,

and this study significantly expands their host range. In addition to typical hosts such as intracellular parasitic bacteria and Enterobacteria, we identified over 20 potential new hosts, including important pathogens like Helicobacter pylori and Vibrio cholerae. Moreover, we revealed distinct host specific differences among different taxa. These new findings will contribute to a deeper understanding of the interactions between Microviridae and their hosts.

## 6 Conflict of interest

The authors declare that they have no conflict of interest.

## 7 Author Contributions

**XKM**: Methodology, Validation, Formal analysis, Data Curation, Writing - Original Draft, and Visualization; **LBF:** Conceptualization, Methodology, Sample collection, Project administration, and Funding acquisition; **ZP, SXY, LC and LGF**: Data Curation, and Investigation; **CXD**: Writing - Review & Editing; **PJQ, QSP, YXQ and LMS**: Sample collection; **JJZ**: Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration, and Funding acquisition; **YLH**: Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision, Project administration, and Funding acquisition.
All authors read and approved the final manuscript.

## 8 Acknowledgments

This project was supported by the Natural Science Foundation of China (nos. 31872499 and 31972847) to Yuan LH and Jiang JZ; the Innovation Team Project of Guangdong Universities (No. 2022KCXTD017) to Yuan LH; the Central Public-Interest Scientific Institution Basal Research Fund, CAFS (nos. 2023TD44 and 2021SD05) to Jiang JZ. The funders had no role in the study design, data collection, and analysis, decision to publish, or manuscript preparation.

## 9 Data availability

The data set supporting the results of this article has been deposited in the National Center for Biotechnology Information (NCBI) under BioProject accession code PRJNA1053868. All viral

626 genomes obtained in this study were deposited in GenBank with the accession numbers:
627 OR998966-9063.
628

629 # References

630 1. Anonymous. 2023. Statistical Bulletin of the People's Republic of China on National Economic
631 and Social Development 2022. China Statistics:12-29. (in Chinese).
632 2. Benfu L, Jingqi P, Mengshi L, Suiping Q, Xiaoqi Y, Menghua W, Lihong Y. 2021. Serotyping of
633 Salmonella in Slaughterhouses in Guangzhou City of Guangdong Province and Analysis on Its
634 Antibiotic Resistance. "China Animal Health Inspection" Editorial Department 38:44-50. (in
635 Chinese).
636 3. Große-Peclum V, Siekmann L, Krischek C, Avramidis G, Ochs C, Viöl W, Plötz M. 2023. Using
637 TRIS-Buffered Plasma-Activated Water to Reduce Pathogenic Microorganisms on Poultry
638 Carcasses with Evaluation of Physicochemical and Sensory Parameters. Foods 12.
639 4. Wu JY, Lau EH, Yuan J, Lu ML, Xie CJ, Li KB, Ma XW, Chen JD, Liu YH, Cao L, Li MX, Di B, Liu YF,
640 Lu JY, Li TG, Xiao XC, Wang DH, Yang ZC, Lu JH. 2019. Transmission risk of avian influenza virus
641 along poultry supply chains in Guangdong, China. J Infect 79:43-48.
642 5. Rasschaert G, De Zutter L, Herman L, Heyndrickx M. 2020. Campylobacter contamination of
643 broilers: the role of transport and slaughterhouse. International Journal of Food Microbiology
644 322:108564.
645 6. García-Sánchez L, Melero B, Jaime I, Rossi M, Ortega I, Rovira J. 2019. Biofilm formation,
646 virulence and antimicrobial resistance of different Campylobacter jejuni isolates from a
647 poultry slaughterhouse. Food Microbiology 83:193-199.
648 7. Zeng H, De Reu K, Gabriël S, Mattheus W, De Zutter L, Rasschaert G. 2021. Salmonella
649 prevalence and persistence in industrialized poultry slaughterhouses. Poultry Science
650 100:100991.
651 8. Zhang Y, Dong J, Bo H, Dong L, Zou S, Li X, Shu Y, Wang D. 2019. Genetic and biological
652 characteristics of avian influenza virus subtype H1N8 in environments related to live poultry
653 markets in China. BMC Infect Dis 19:458.
654 9. Arzey GG, Kirkland PD, Arzey KE, Frost M, Maywood P, Conaty S, Hurt AC, Deng YM, Iannello P,
655 Barr I, Dwyer DE, Ratnamohan M, McPhie K, Selleck P. 2012. Influenza virus A (H10N7) in
656 chickens and poultry abattoir workers, Australia. Emerg Infect Dis 18:814-6.
657 10. Díaz-Muñoz SL, Koskella B. 2014. Bacteria-phage interactions in natural environments. Adv
658 Appl Microbiol 89:135-83.
659 11. Geoghegan JL, Holmes EC. 2017. Predicting virus emergence amid evolutionary noise. Open
660 Biol 7.
661 12. Mushegian AR. 2020. Are There 10(31) Virus Particles on Earth, or More, or Fewer? J
662 Bacteriol 202.
663 13. Anonymous. 2012. Family - Microviridae, p 385-393. *In* King AMQ, Adams MJ, Carstens EB,
664 Lefkowitz EJ (ed), Virus Taxonomy doi:https://doi.org/10.1016/B978-0-12-384684-6.00037-9.
665 Elsevier, San Diego.

14. Roux S, Krupovic M, Poulet A, Debroas D, Enault F. 2012. Evolution and diversity of the Microviridae viral family through a collection of 81 new complete genomes assembled from virome reads. PLoS One 7:e40418.
15. Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM, Alfenas-Zerbini P, Davison AJ, Dempsey DM, Dutilh BE, García ML, Harrach B, Harrison RL, Hendrickson RC, Junglen S, Knowles NJ, Krupovic M, Kuhn JH, Lambert AJ, Łobocka M, Nibert ML, Oksanen HM, Orton RJ, Robertson DL, Rubino L, Sabanadzovic S, Simmonds P, Smith DB, Suzuki N, Van Dooerslaer K, Vandamme A-M, Varsani A, Zerbini FM. 2021. Changes to virus taxonomy and to the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2021). Archives of Virology 166:2633-2648.
16. Krupovic M, Forterre P. 2011. Microviridae goes temperate: microvirus-related proviruses reside in the genomes of Bacteroidetes. PLoS One 6:e19893.
17. Lasken RS. 2009. Genomic DNA amplification by the multiple displacement amplification (MDA) method. Biochem Soc Trans 37:450-3.
18. Xiao C, You R, Zhu N, Mi X, Gao L, Zhou X, Zhou G. 2023. Variation of soil physicochemical properties of different vegetation restoration types on subtropical karst area in southern China. PLoS One 18:e0282620.
19. Hosono S, Faruqi AF, Dean FB, Du Y, Sun Z, Wu X, Du J, Kingsmore SF, Egholm M, Lasken RS. 2003. Unbiased whole-genome amplification directly from clinical samples. Genome Res 13:954-64.
20. Pan X, Durrett RE, Zhu H, Tanaka Y, Li Y, Zi X, Marjani SL, Euskirchen G, Ma C, LaMotte RH, Park I-H, Snyder MP, Mason CE, Weissman SM. 2013. Two methods for full-length RNA sequencing for low quantities of cells and single cells. 110:594-599.
21. Picher Á J, Budeus B, Wafzig O. 2016. TruePrime is a novel method for whole-genome amplification from single cells based on TthPrimPol. 7:13296.
22. Stepanauskas R, Fergusson EA, Brown J, Poulton NJ, Tupper B, Labonté JM, Becraft ED, Brown JM, Pachiadaki MG, Povilaitis T, Thompson BP, Mascena CJ, Bellows WK, Lubys A. 2017. Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. Nat Commun 8:84.
23. Tithi SS, Aylward FO, Jensen RV, Zhang L. 2018. FastViromeExplorer: a pipeline for virus and phage identification and abundance profiling in metagenomics data. PeerJ 6:e4227.
24. Modi A, Vai S, Caramelli D, Lari M. 2021. The Illumina Sequencing Protocol and the NovaSeq 6000 System. Methods Mol Biol 2242:15-42.
25. Patch AM, Nones K, Kazakoff SH, Newell F, Wood S, Leonard C, Holmes O, Xu Q, Addala V, Creaney J, Robinson BW, Fu S, Geng C, Li T, Zhang W, Liang X, Rao J, Wang J, Tian M, Zhao Y, Teng F, Gou H, Yang B, Jiang H, Mu F, Pearson JV, Waddell N. 2018. Germline and somatic variant identification using BGISEQ-500 and HiSeq X Ten whole genome sequencing. PLoS One 13:e0190264.
26. Ravi RK, Walton K, Khosroheidari M. 2018. MiSeq: A Next Generation Sequencing Platform for Genomic Analysis. Methods Mol Biol 1706:223-232.
27. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884-i890.
28. Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution

for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674-6.

29. Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, Yamashita H, Lam TW. 2016. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods 102:3-11.

30. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59-60.

31. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, Ruscheweyh HJ, Tappu R. 2016. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLoS Comput Biol 12:e1004957.

32. Tisza MJ, Belford AK, Domínguez-Huerta G, Bolduc B, Buck CB. 2020. Cenote-Taker 2 democratizes virus discovery and sequence annotation. Virus Evolution 7.

33. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403-10.

34. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-402.

35. Bastian M, Heymann S, Jacomy M. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. Proceedings of the International AAAI Conference on Web and Social Media 3:361-362.

36. Kirchberger PC, Martinez ZA, Ochman H. 2022. Organizing the Global Diversity of Microviruses. mBio 13:e0058822.

37. Shang J, Sun Y. 2021. Predicting the hosts of prokaryotic viruses using GCN-based semi-supervised learning. BMC Biology 19:250.

38. Shang J, Sun Y. 2022. CHERRY: a Computational metHod for accuratE pRediction of virus-pRokarYotic interactions using a graph encoder-decoder model. Brief Bioinform 23.

39. Quaiser A, Dufresne A, Ballaud F, Roux S, Zivanovic Y, Colombet J, Sime-Ngando T, Francez AJ. 2015. Diversity and comparative genomics of Microviridae in Sphagnum- dominated peatlands. Front Microbiol 6:375.

40. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772-80.

41. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972-3.

42. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol 37:1530-1534.

43. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587-589.

44. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44:W242-5.

45. Yuan W-G, Liu G-F, Shi Y-H, Xie K-M, Jiang J-Z, Yuan L-H. 2022. A discussion of RNA virus taxonomy based on the 2020 International Committee on Taxonomy of Viruses report. 13.

46. Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, Kuhn JH, Lavigne R, Brister JR, Varsani A, Amid C, Aziz RK, Bordenstein SR, Bork P, Breitbart M, Cochrane GR, Daly

RA, Desnues C, Duhaime MB, Emerson JB, Enault F, Fuhrman JA, Hingamp P, Hugenholtz P, Hurwitz BL, Ivanova NN, Labonté JM, Lee K-B, Malmstrom RR, Martinez-Garcia M, Mizrachi IK, Ogata H, Páez-Espino D, Petit M-A, Putonti C, Rattei T, Reyes A, Rodriguez-Valera F, Rosario K, Schriml L, Schulz F, Steward GF, Sullivan MB, Sunagawa S, Suttle CA, Temperton B, Tringe SG, Thurber RV, Webster NS, Whiteson KL, et al. 2019. Minimum Information about an Uncultivated Virus Genome (MIUViG). Nature Biotechnology 37:29-37.

47. Abraham W-R, Strömpl C, Meyer H, Lindholst S, Moore ERB, Christ R, Vancanneyt M, Tindall BJ, Bennasar A, Smit J, Tesar M. 1999. Phylogeny and polyphasic taxonomy of Caulobacter species. Proposal of Maricaulis gen. nov. with Maricaulis maris (Poindexter) comb. nov. as the type species, and emended description of the genera Brevundimonas and Caulobacter. 49:1053-1073.

48. Diggle SP, Whiteley M. 2020. Microbe Profile: Pseudomonas aeruginosa: opportunistic pathogen and lab rat. 166:30-33.

49. de Brito BB, da Silva FAF, Soares AS, Pereira VA, Santos MLC, Sampaio MM, Neves PHM, de Melo FF. 2019. Pathogenesis and clinical management of Helicobacter pylori gastric infection. World journal of gastroenterology 25:5578-5589.

50. Marghalani AM, Bin Salman TO, Faqeeh FJ, Asiri MK, Kabel AM. 2020. Gastric carcinoma: Insights into risk factors, methods of diagnosis, possible lines of management, and the role of primary care. Journal of Family Medicine and Primary Care 9.

51. Violeta Filip P, Cuciureanu D, Sorina Diaconu L, Maria Vladareanu A, Silvia Pop C. 2018. MALT lymphoma: epidemiology, clinical diagnosis and treatment. J Med Life 11:187-193.

52. Tarrand JJ, Krieg NR, Döbereiner J. 1978. A taxonomic study of the Spirillum lipoferum group, with descriptions of a new genus, Azospirillum gen. nov. and two species, Azospirillum lipoferum (Beijerinck) comb. nov. and Azospirillum brasilense sp. nov. Can J Microbiol 24:967-80.

53. Tien TM, Gaskins MH, Hubbell DH. 1979. Plant Growth Substances Produced by Azospirillum brasilense and Their Effect on the Growth of Pearl Millet (Pennisetum americanum L.). Appl Environ Microbiol 37:1016-24.

54. Farmer JJ, 3rd, Fanning GR, Davis BR, O'Hara CM, Riddle C, Hickman-Brenner FW, Asbury MA, Lowery VA, 3rd, Brenner DJ. 1985. Escherichia fergusonii and Enterobacter taylorae, two new species of Enterobacteriaceae isolated from clinical specimens. J Clin Microbiol 21:77-81.

55. Nielsen P, Fritze D, Priest FG. 1995. Phenetic diversity of alkaliphilic Bacillus strains: proposal for nine new species. 141:1745-1761.

56. Cho J-C, Giovannoni SJ. 2003. Croceibacter atlanticus gen. nov., sp. nov., A Novel Marine Bacterium in the Family Flavobacteriaceae. Systematic and Applied Microbiology 26:76-83.

57. Denny T. 2006. Plant pathogenic Ralstonia species, p 573-644 doi:10.1007/978-1-4020-4538-7_16.

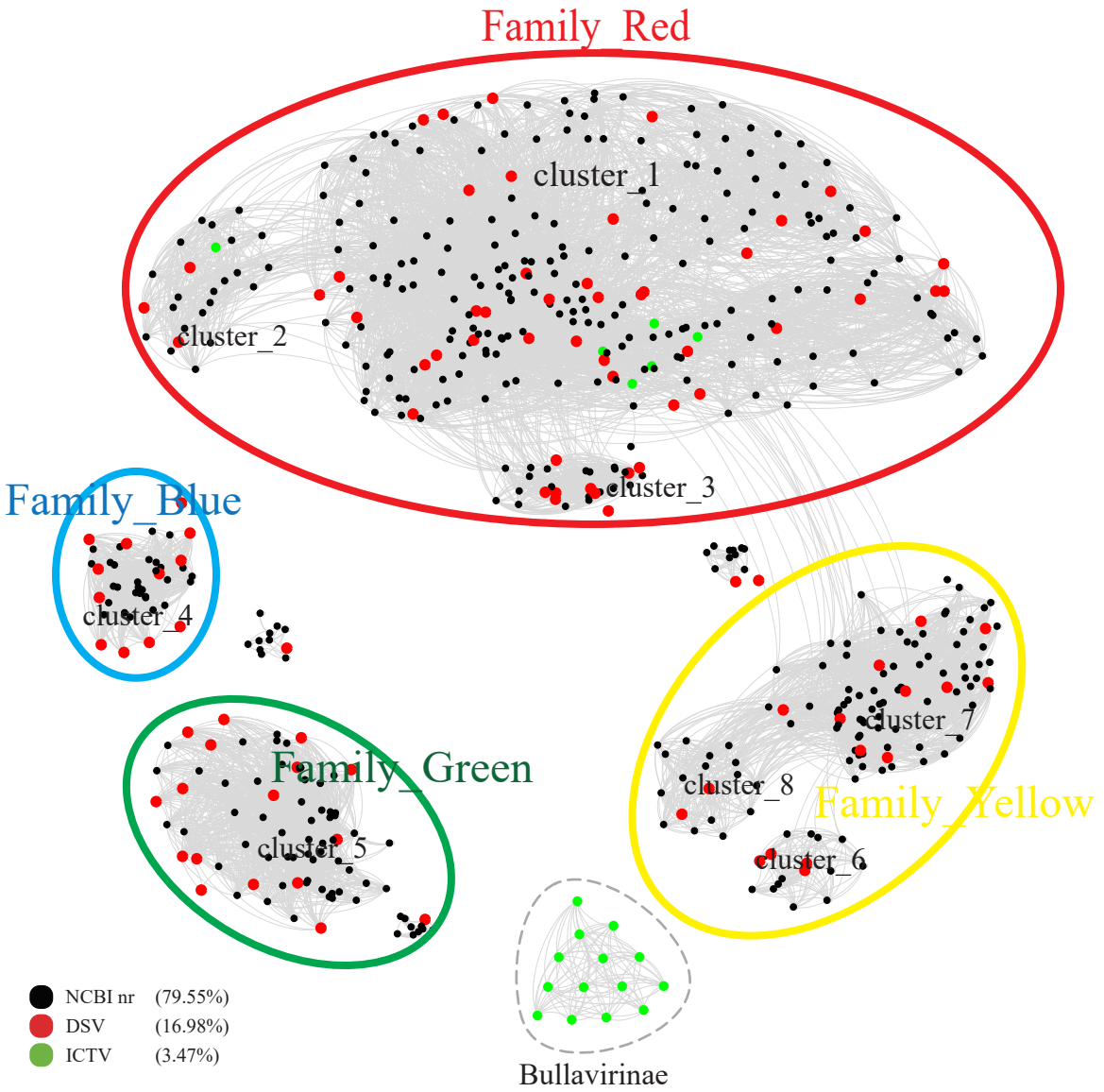
58. Igra-Siegman Y, Chmel H, Cobbs C. 1980. Clinical and laboratory characteristics of Achromobacter xylosoxidans infection. Journal of Clinical Microbiology 11:141-145.

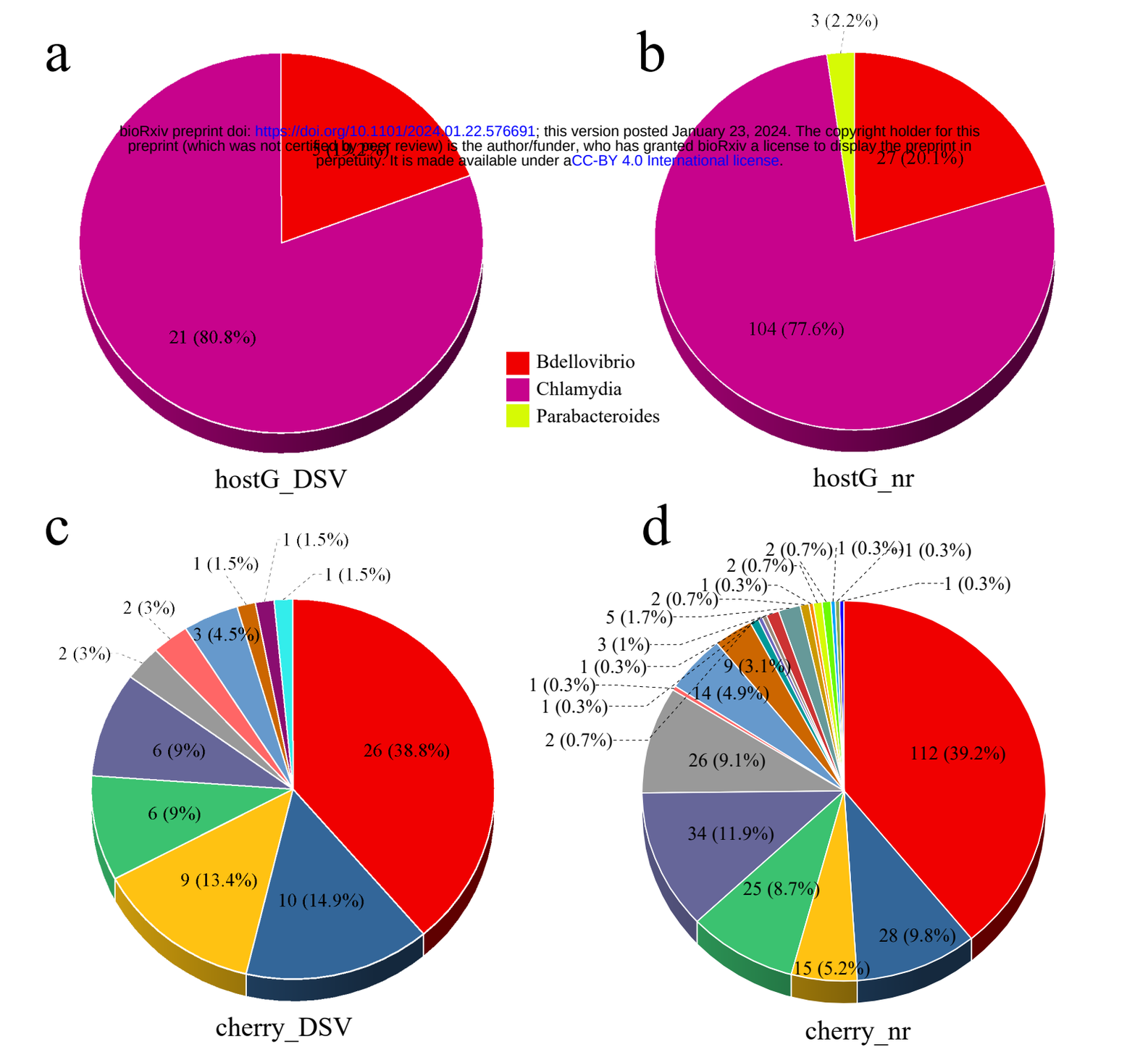
59. Duggan JM, Goldstein SJ, Chenoweth CE, Kauffman CA, Bradley SF. 1996. Achromobacter xylosoxidans Bacteremia: Report of Four Cases and Review of the Literature. Clinical Infectious Diseases 23:569-576.

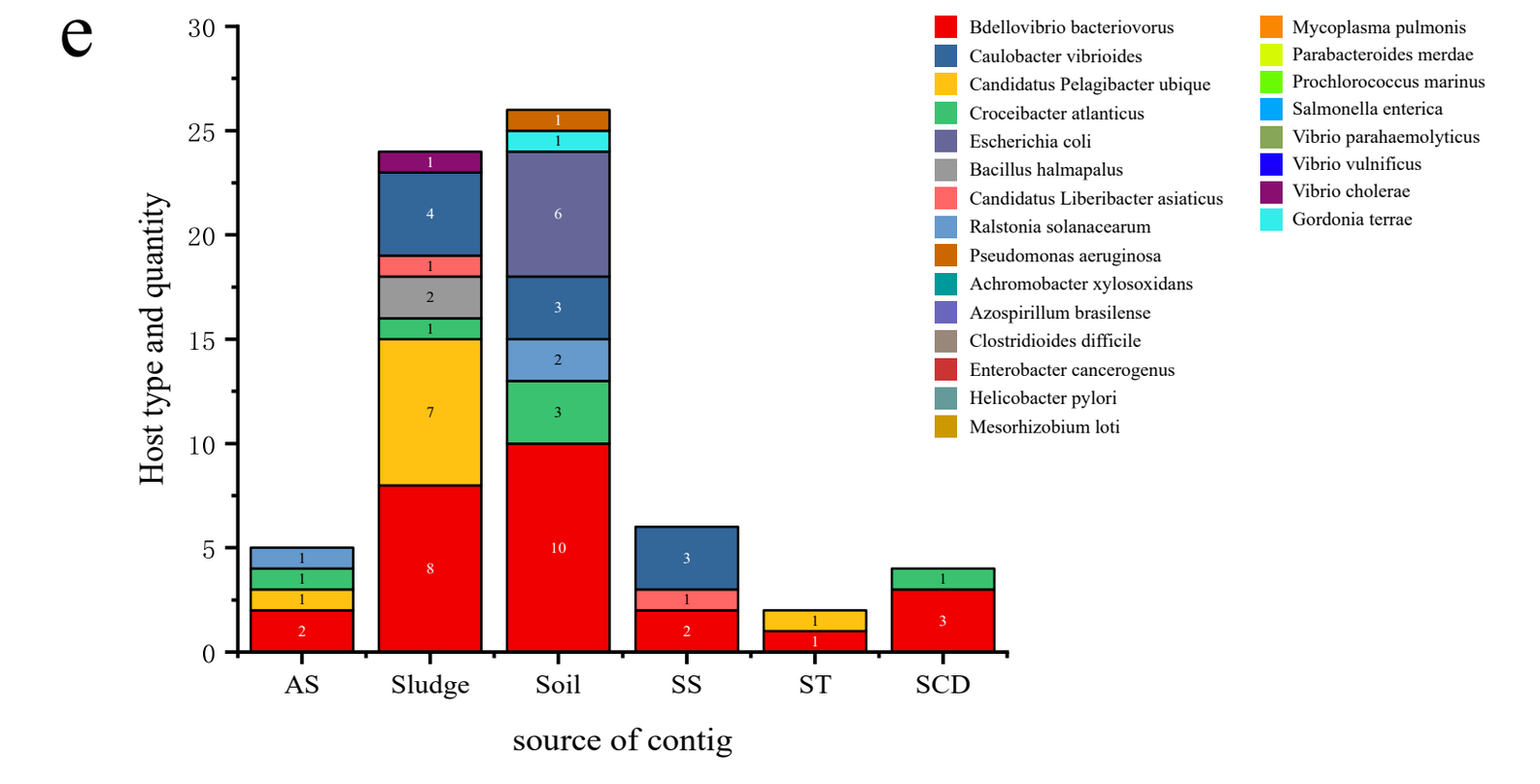
60. Tikhe CV, Husseneder C. 2017. Metavirome Sequencing of the Termite Gut Reveals the Presence of an Unexplored Bacteriophage Community. Front Microbiol 8:2548.

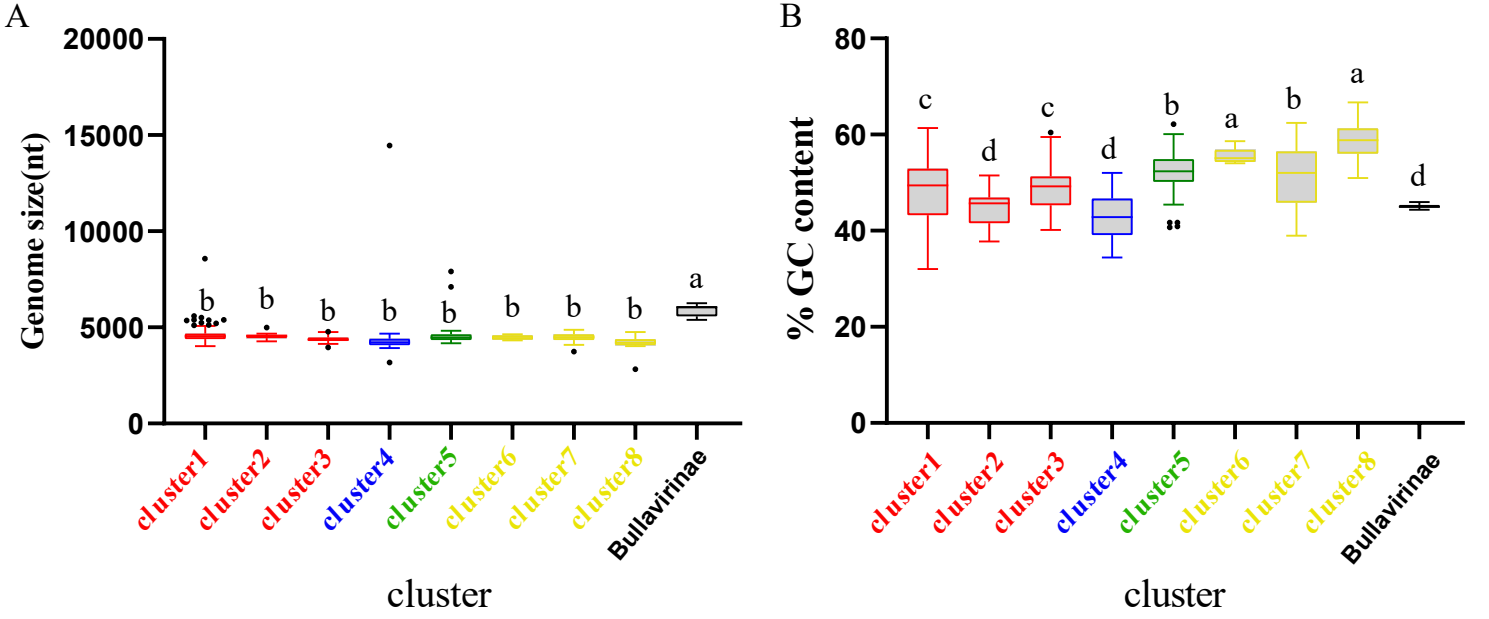
61. Rosario K, Dayaram A, Marinov M, Ware J, Kraberger S, Stainton D, Breitbart M, Varsani A. 2012. Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). J Gen Virol 93:2668-2681.
62. Chapagain S. 2019. Metaviromics: A metagenomics approach to understanding viruses doi:10.13140/RG.2.2.29763.76321.
63. Jiang J-Z, Fang Y-F, Wei H-Y, Zhu P, Liu M, Yuan W-G, Yang L-L, Guo Y-X, Jin T, Shi M, Yao T, Lu J, Ye L-T, Shi S-K, Wang M, Duan M, Zhang D-C. 2023. A remarkably diverse and well-organized virus community in a filter-feeding oyster. Microbiome 11:2.
64. Zhu P, Liu G-F, Liu C, Yang L-L, Liu M, Xie K-M, Shi S-K, Shi M, Jiang J-Z. 2022. Novel RNA viruses in oysters revealed by virome. iMeta 1:e65.
65. Angly FE, Willner D, Prieto-Davó A, Edwards RA, Schmieder R, Vega-Thurber R, Antonopoulos DA, Barott K, Cottrell MT, Desnues C, Dinsdale EA, Furlan M, Haynes M, Henn MR, Hu Y, Kirchman DL, McDole T, McPherson JD, Meyer F, Miller RM, Mundt E, Naviaux RK, Rodriguez-Mueller B, Stevens R, Wegley L, Zhang L, Zhu B, Rohwer F. 2009. The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes. PLoS Comput Biol 5:e1000593.
66. López-Bueno A, Tamames J, Velázquez D, Moya A, Quesada A, Alcamí A. 2009. High diversity of the viral community from an Antarctic lake. Science 326:858-61.
67. Hendrix RW, Lawrence JG, Hatfull GF, Casjens S. 2000. The origins and ongoing evolution of viruses. Trends Microbiol 8:504-8.
68. Hendrix RW, Smith MCM, Burns RN, Ford ME, Hatfull GF. 1999. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. Proceedings of the National Academy of Sciences 96:2192-2197.
69. Doore SM, Fane BA. 2016. The microviridae: Diversity, assembly, and experimental evolution. Virology 491:45-55.
70. Brentlinger Karie L, Hafenstein S, Novak Christopher R, Fane Bentley A, Borgon R, McKenna R, Agbandje-McKenna M. 2002. Microviridae, a Family Divided: Isolation, Characterization, and Genome Sequence of ϕMH2K, a Bacteriophage of the Obligate Intracellular Parasitic Bacterium Bdellovibrio bacteriovorus. Journal of Bacteriology 184:1089-1094.
71. Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, Bushman FD. 2013. Rapid evolution of the human gut virome. Proceedings of the National Academy of Sciences 110:12450-12455.
72. Brentlinger KL, Hafenstein S, Novak CR, Fane BA, Borgon R, McKenna R, Agbandje-McKenna M. 2002. Microviridae, a family divided: isolation, characterization, and genome sequence of phiMH2K, a bacteriophage of the obligate intracellular parasitic bacterium Bdellovibrio bacteriovorus. J Bacteriol 184:1089-94.
73. Xu W, Xuan G, Lin H, Wang J. 2022. Complete genome analysis of the newly isolated Vibrio phage vB_VpP_WS1 of the family Microviridae. Archives of Virology 167:1311-1316.
74. Guo R, Zheng K, Luo L, Liu Y, Shao H, Guo C, He H, Wang H, Sung Yeong Y, Mok Wen J, Wong Li L, Zhang Y-Z, Liang Y, McMinn A, Wang M. 2022. Characterization and Genomic Analysis of ssDNA Vibriophage vB_VpaM_PG19 within Microviridae, Representing a Novel Viral Genus. Microbiology Spectrum 10:e00585-22.
75. Li M, Lin H, Wang L, Wang J. 2021. Complete genome sequence of the extreme-pH-resistant Salmonella bacteriophage αα of the family Microviridae. Archives of Virology 166:325-329.
76. Lu H, Xiong W, Li Z, Yan P, Liu R, Liu X. 2022. Isolation and characterization of SGF3, a novel

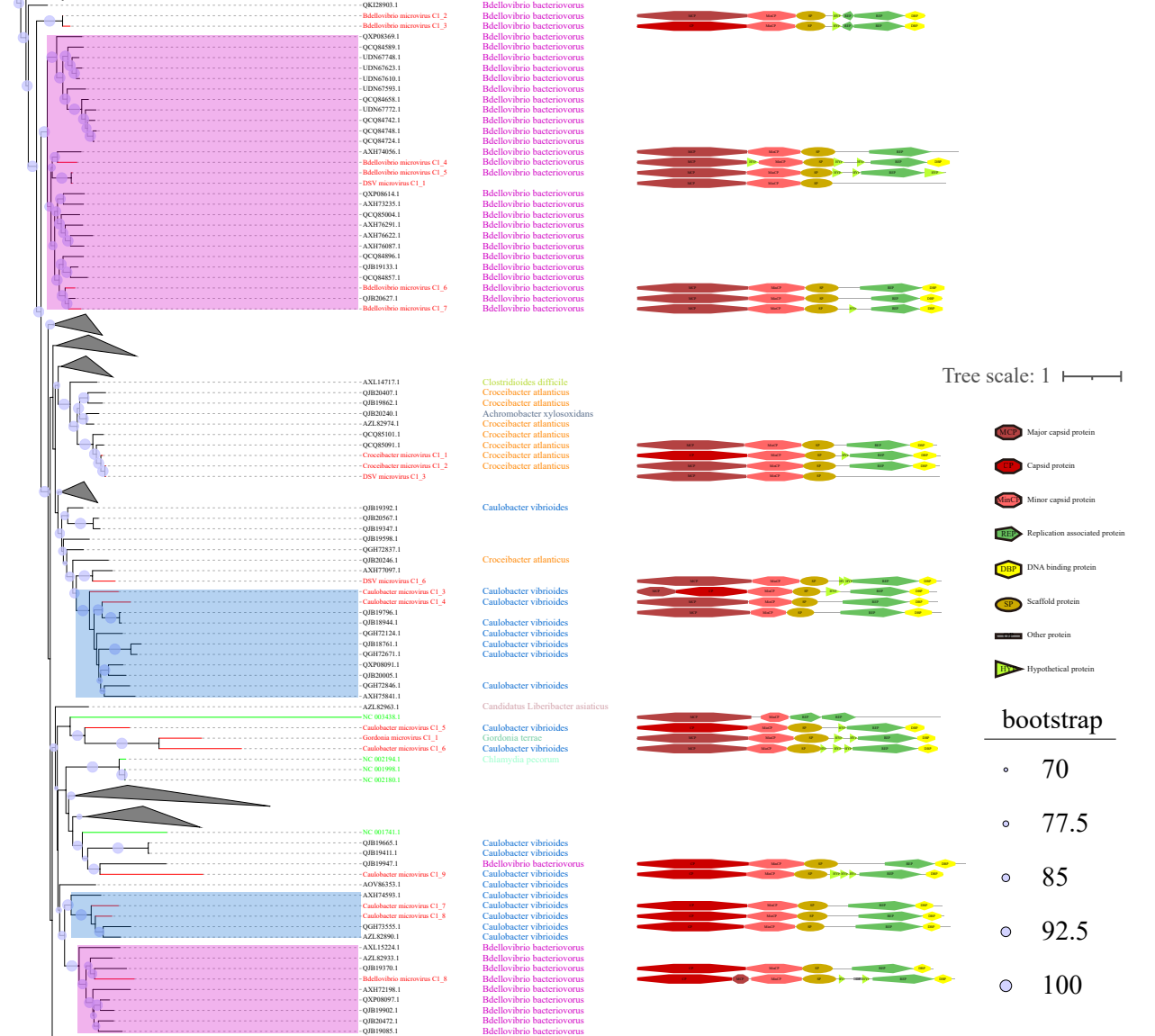
841 Microviridae phage infecting Shigella flexneri. Molecular Genetics and Genomics
842 297:935-945.
843 77. Champoiseau P, Jones J, Allen C. 2009. Ralstonia solanacearum Race 3 Biovar 2 Causes
844 Tropical Losses and Temperate Anxieties. Plant Health Progress 10.
845 78. Wicker E, Grassart L, Coranson-Beaudu R, Mian D, Guilbaud C, Fegan M, Prior P. 2007.
846 Ralstonia solanacearum Strains from Martinique (French West Indies) Exhibiting a New
847 Pathogenic Potential. Applied and Environmental Microbiology 73:6790-6801.
848 79. Wang C, Fang F, Li Y, Zhang L, Wu J, Li T, Zheng Y, Xu Q, Fan S, Chen J, Deng X, Zheng Z. 2022.
849 Biological Features and In Planta Transcriptomic Analyses of a Microviridae Phage (CLasMV1)
850 in "Candidatus Liberibacter asiaticus". Int J Mol Sci 23.

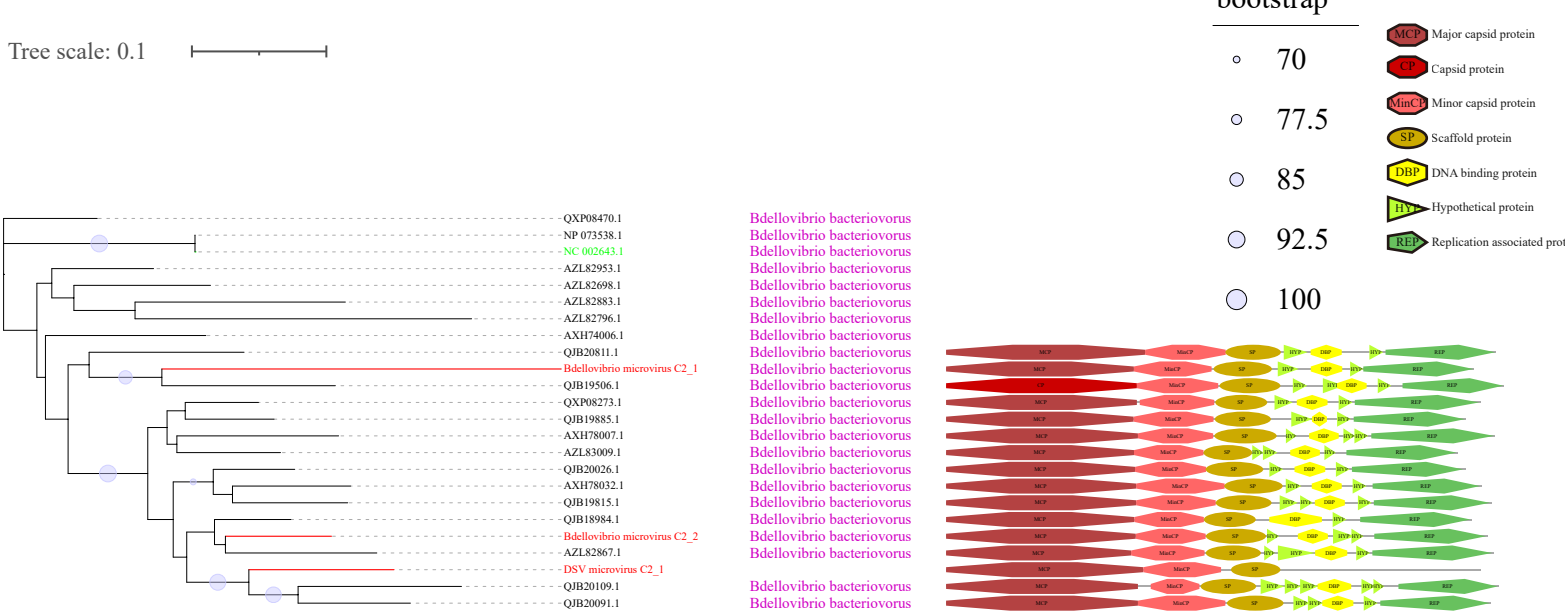

Family_Red
cluster_1
cluster_2
cluster_3
Family_Blue
cluster_4
Family_Green
cluster_5
cluster_7
cluster_8
Family_Yellow
cluster_6
Bullavirinae
NCBI nr (79.55%)
DSV (16.98%)
ICTV (3.47%)

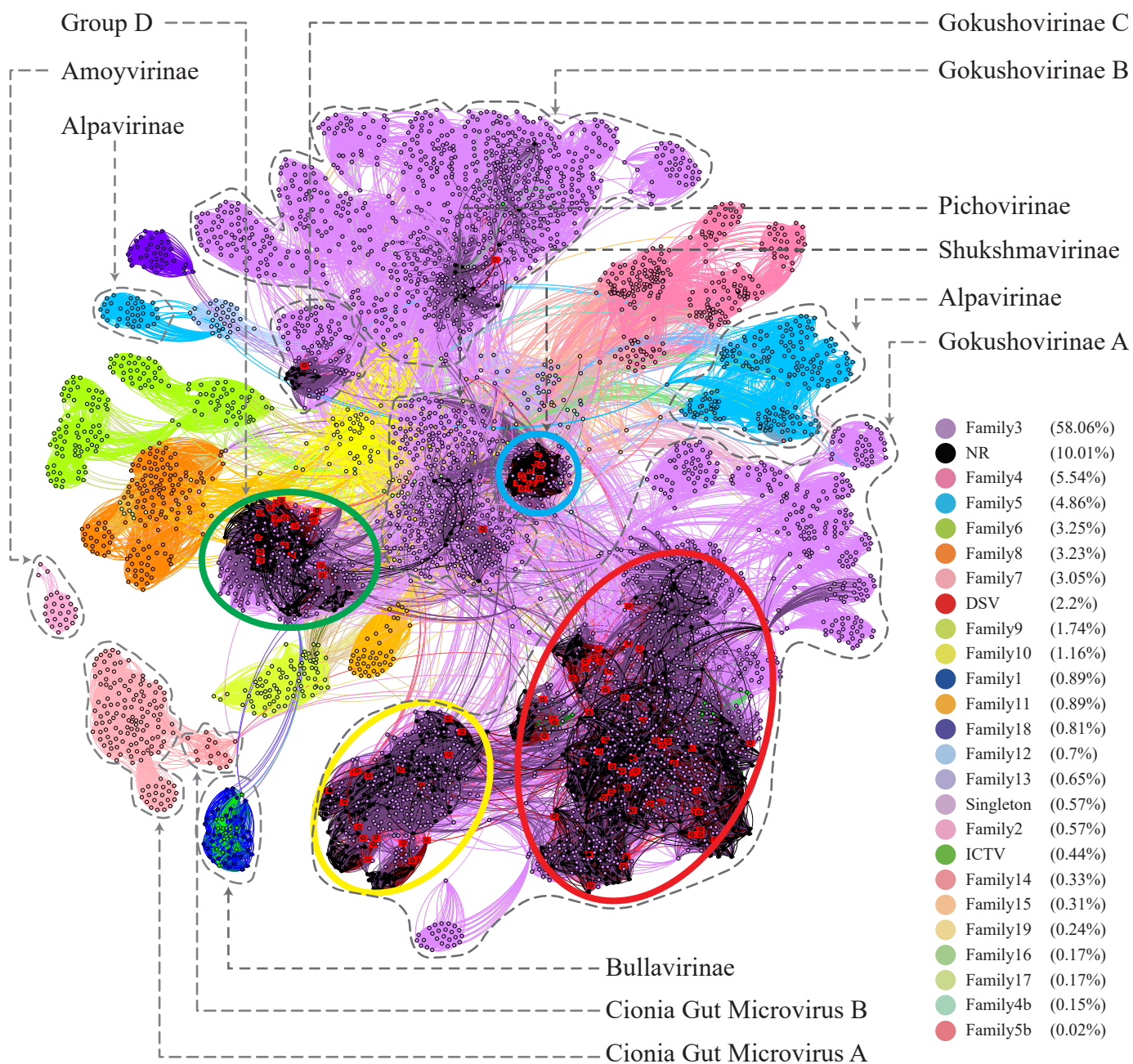


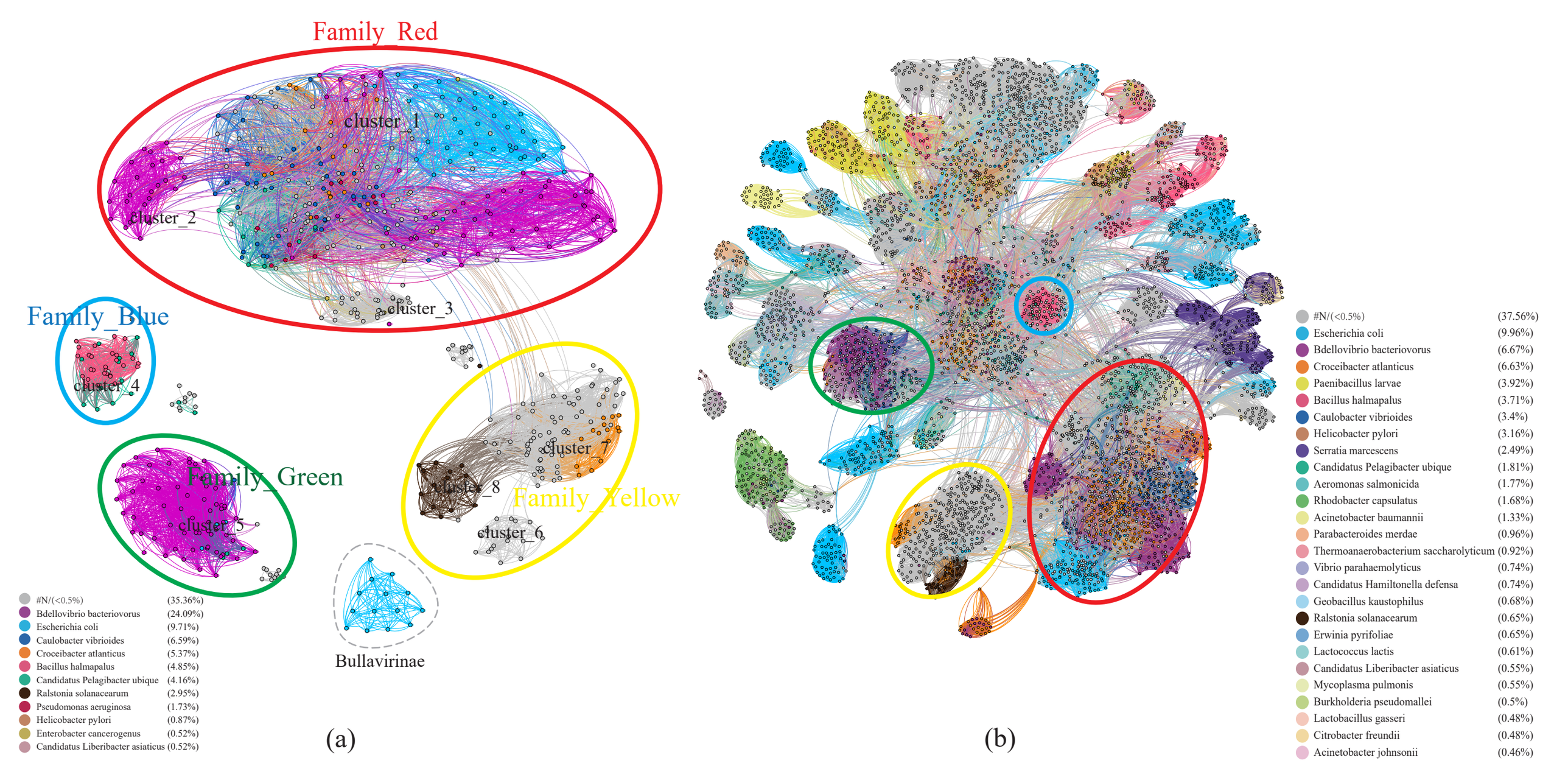



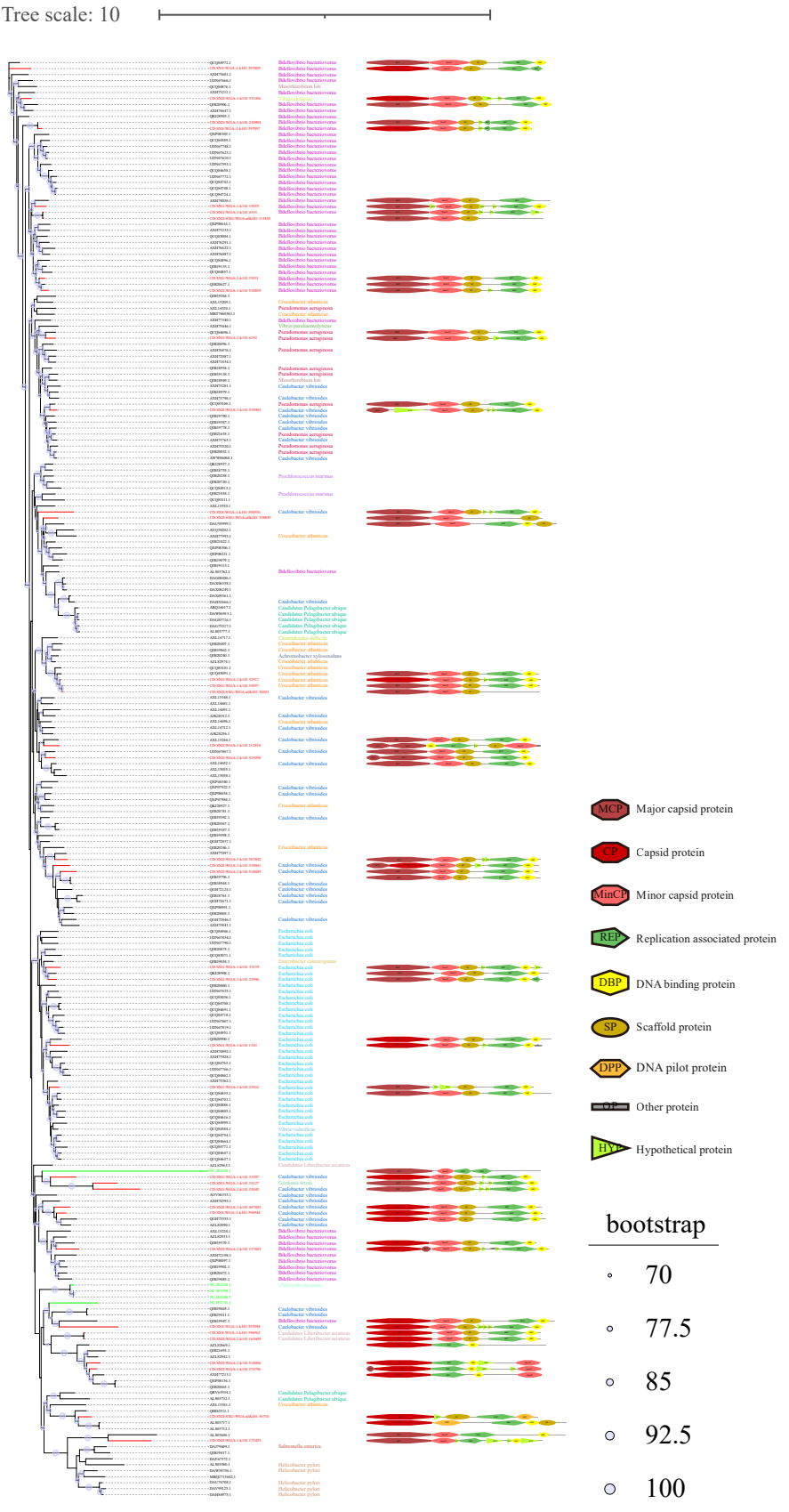

A
Genome size(nt)
20000
15000
10000
5000
0
b
b
b
b
b
b
b
b
a
cluster1
cluster2
cluster3
cluster4
cluster5
cluster6
cluster7
cluster8
Bullavirinae
cluster
B
% GC content
80
60
40
20
0
c
d
c
d
b
a
b
a
d
cluster1
cluster2
cluster3
cluster4
cluster5
cluster6
cluster7
cluster8
Bullavirinae
cluster

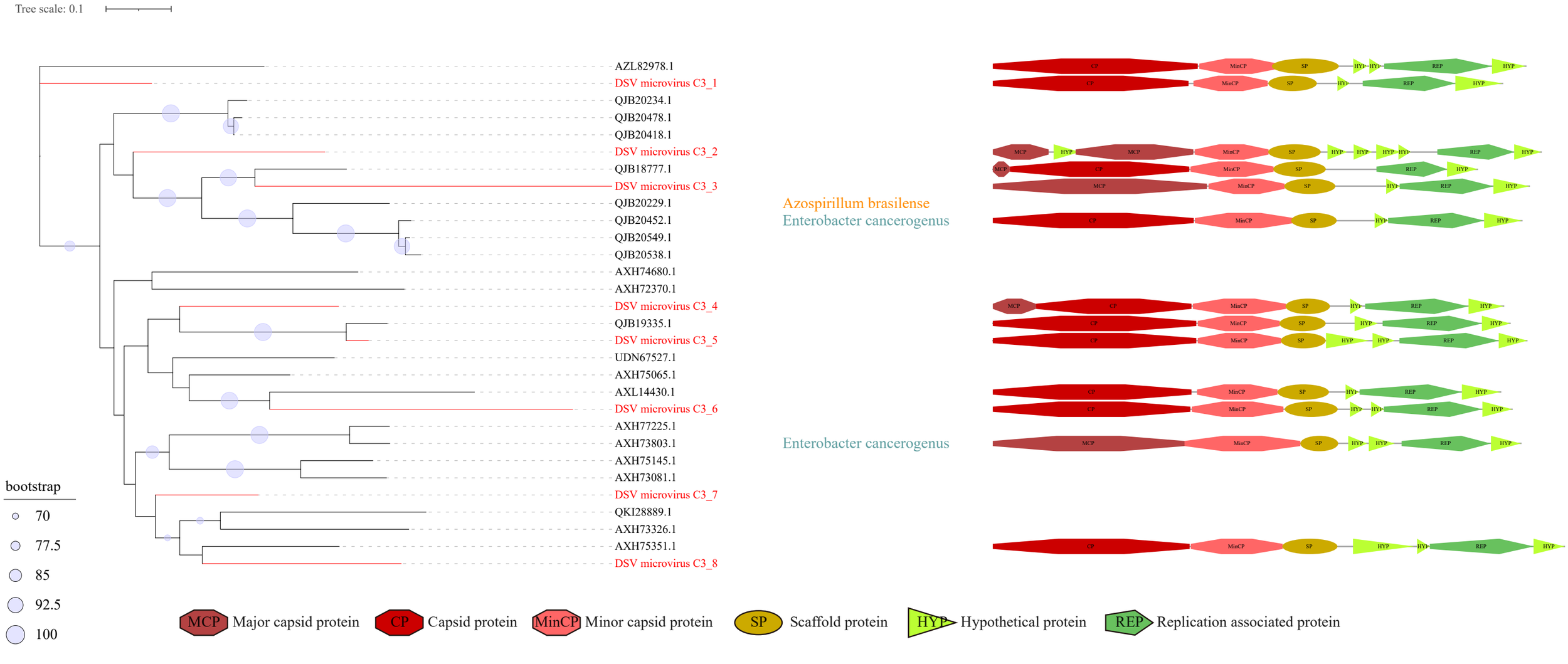

QKI28903.1
Bdellovibrio microvirus C1_2
Bdellovibrio microvirus C1_3
QXP08369.1
QCQ84589.1
UDN67748.1
UDN67623.1
UDN67610.1
UDN67593.1
QCQ84658.1
UDN67772.1
QCQ84742.1
QCQ84748.1
QCQ84724.1
AXH74056.1
Bdellovibrio microvirus C1_4
Bdellovibrio microvirus C1_5
DSV microvirus C1_1
QXP08614.1
AXH73235.1
QCQ85004.1
AXH76291.1
AXH76622.1
AXH76087.1
QCQ84896.1
QJB19133.1
QCQ84857.1
Bdellovibrio microvirus C1_6
QJB20627.1
Bdellovibrio microvirus C1_7
Bdellovibrio bacteriovorus
AXL14717.1
QJB20407.1
QJB19862.1
QJB20240.1
AZL82974.1
QCQ85101.1
QCQ85091.1
Crocebacter microvirus C1_1
Crocebacter microvirus C1_2
DSV microvirus C1_3
Clostridioides difficile
Crocebacter atlanticus
Crocebacter atlanticus
Achromobacter xylosoxidans
Crocebacter atlanticus
Crocebacter atlanticus
Crocebacter atlanticus
Crocebacter atlanticus
Crocebacter atlanticus
QJB19392.1
QJB20567.1
QJB19347.1
QJB19598.1
QGH72837.1
QJB20246.1
AXH77097.1
DSV microvirus C1_6
Caulobacter microvirus C1_3
Caulobacter microvirus C1_4
QJB19796.1
QJB18944.1
QGH72124.1
QJB18761.1
QGH72671.1
QXP08091.1
QJB20005.1
QGH72846.1
AXH75841.1
Caulobacter vibrioides
Crocebacter atlanticus
Caulobacter vibrioides
Caulobacter vibrioides
Caulobacter vibrioides
Caulobacter vibrioides
Caulobacter vibrioides
Caulobacter vibrioides
Caulobacter vibrioides
Caulobacter vibrioides
AZL82963.1
NC 003438.1
Caulobacter microvirus C1_5
Gordonia microvirus C1_1
Caulobacter microvirus C1_6
NC 002194.1
NC 001998.1
NC 002180.1
Candidatus Liberibacter asiaticus
Caulobacter vibrioides
Gordonia terrae
Caulobacter vibrioides
Chlamydia pecorum
NC 001741.1
QJB19665.1
QJB19411.1
QJB19947.1
Caulobacter microvirus C1_9
AOV86353.1
AXH74593.1
Caulobacter microvirus C1_7
Caulobacter microvirus C1_8
QGH73555.1
AZL82890.1
AXL15224.1
AZL82933.1
QJB19370.1
Bdellovibrio microvirus C1_8
AXH72198.1
QXP08097.1
QJB19902.1
QJB20472.1
QJB19085.1
Caulobacter vibrioides
Caulobacter vibrioides
Bdellovibrio bacteriovorus
Caulobacter vibrioides
Caulobacter vibrioides
Caulobacter vibrioides
Caulobacter vibrioides
Caulobacter vibrioides
Caulobacter vibrioides
Bdellovibrio bacteriovorus
Bdellovibrio bacteriovorus
Bdellovibrio bacteriovorus
Bdellovibrio bacteriovorus
Bdellovibrio bacteriovorus
Bdellovibrio bacteriovorus
Bdellovibrio bacteriovorus
Tree scale: 1
MCP
Major capsid protein
CP
Capsid protein
MiCP
Minor capsid protein
REP
Replication associated protein
DBP
DNA binding protein
SP
Scaffold protein
Other protein
HP
Hypothetical protein
bootstrap
70
77.5
85
92.5
100

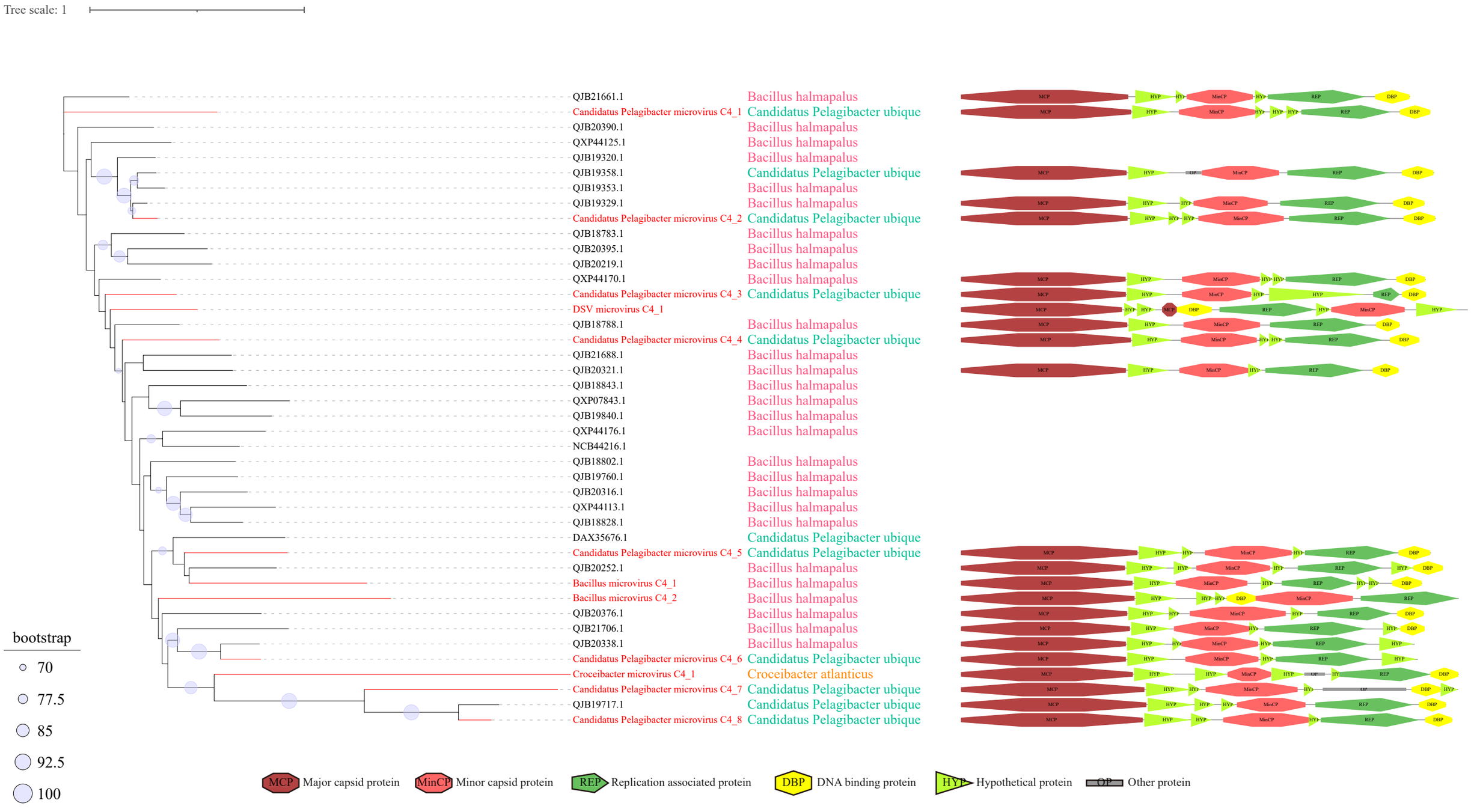

Tree scale: 0.1
bootstrap
70
77.5
85
92.5
100
MCP Major capsid protein
CP Capsid protein
MinCP Minor capsid protein
SP Scaffold protein
DBP DNA binding protein
HYP Hypothetical protein
REP Replication associated protein
QXP08470.1
NP 073538.1
NC 002643.1
AZL82953.1
AZL82698.1
AZL82883.1
AZL82796.1
AXH74006.1
QJB20811.1
Bdellovibrio microvirus C2_1
QJB19506.1
QXP08273.1
QJB19885.1
AXH78007.1
AZL83009.1
QJB20026.1
AXH78032.1
QJB19815.1
QJB18984.1
Bdellovibrio microvirus C2_2
AZL82867.1
DSV microvirus C2_1
QJB20109.1
QJB20091.1
Bdellovibrio bacteriovorus

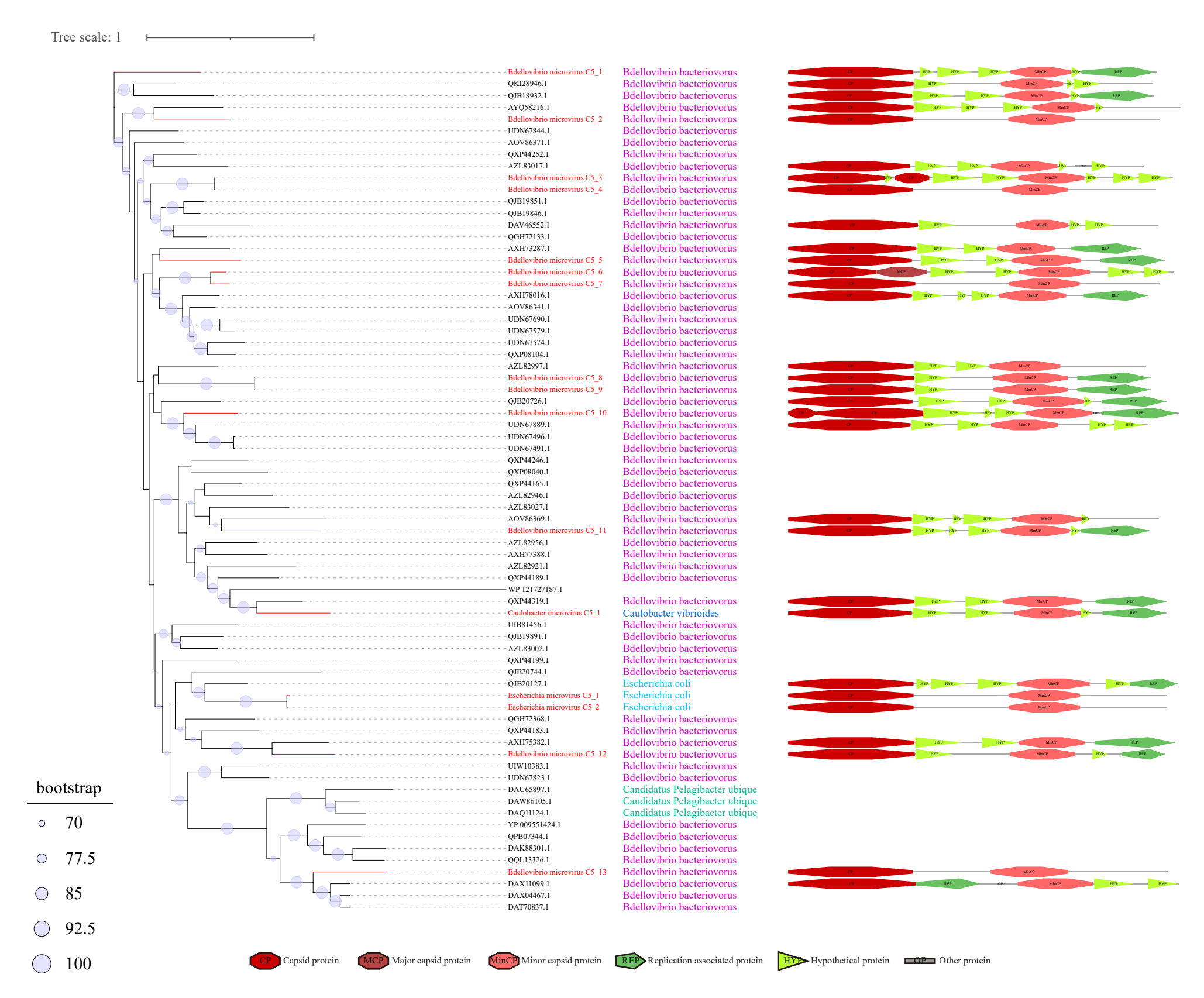

Group D
Amoyvirinae
Alpavirinae
Gokushovirinae C
Gokushovirinae B
Pichovirinae
Shukshmavirinae
Alpavirinae
Gokushovirinae A
Bullavirinae
Cionia Gut Microvirus B
Cionia Gut Microvirus A
Family3 (58.06%)
NR (10.01%)
Family4 (5.54%)
Family5 (4.86%)
Family6 (3.25%)
Family8 (3.23%)
Family7 (3.05%)
DSV (2.2%)
Family9 (1.74%)
Family10 (1.16%)
Family1 (0.89%)
Family11 (0.89%)
Family18 (0.81%)
Family12 (0.7%)
Family13 (0.65%)
Singleton (0.57%)
Family2 (0.57%)
ICTV (0.44%)
Family14 (0.33%)
Family15 (0.31%)
Family19 (0.24%)
Family16 (0.17%)
Family17 (0.17%)
Family4b (0.15%)
Family5b (0.02%)

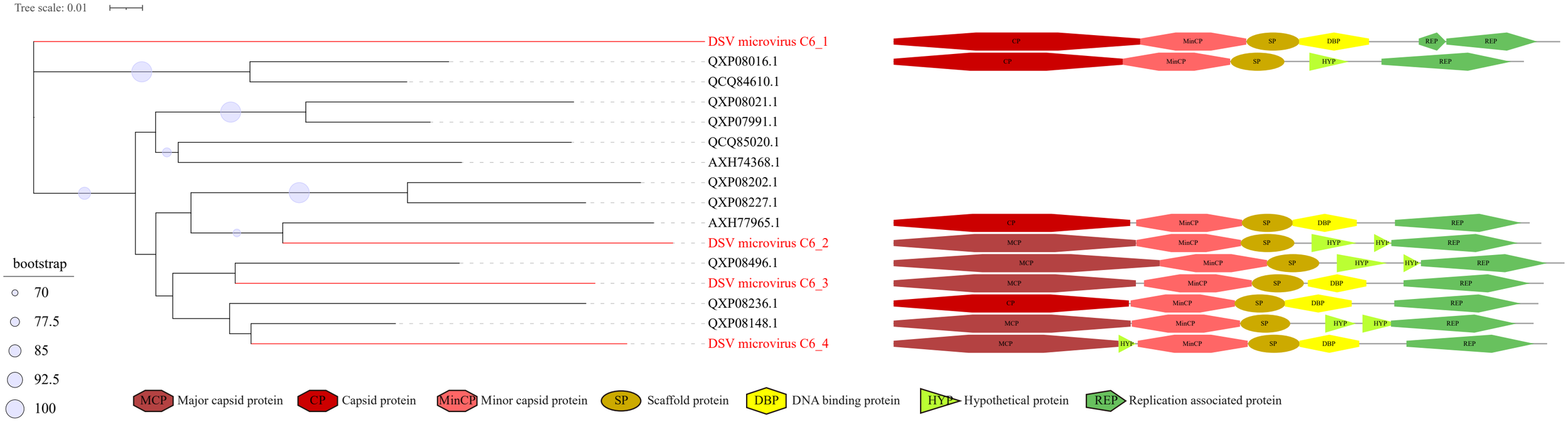

Family_Red
cluster_1
cluster_2
cluster_3
Family_Blue
cluster_4
Family_Green
cluster_5
Family_Yellow
cluster_6
cluster_7
cluster_8
Bullavirinae
(a)
(b)
#N/(<0.5%) (35.36%)
Bdellovibrio bacteriovorus (24.09%)
Escherichia coli (9.71%)
Caulobacter vibrioides (6.59%)
Croceibacter atlanticus (5.37%)
Bacillus halmapalus (4.85%)
Candidatus Pelagibacter ubique (4.16%)
Ralstonia solanacearum (2.95%)
Pseudomonas aeruginosa (1.73%)
Helicobacter pylori (0.87%)
Enterobacter cancerogenus (0.52%)
Candidatus Liberibacter asiaticus (0.52%)
#N/(<0.5%) (37.56%)
Escherichia coli (9.96%)
Bdellovibrio bacteriovorus (6.67%)
Croceibacter atlanticus (6.63%)
Paenibacillus larvae (3.92%)
Bacillus halmapalus (3.71%)
Caulobacter vibrioides (3.4%)
Helicobacter pylori (3.16%)
Serratia marcescens (2.49%)
Candidatus Pelagibacter ubique (1.81%)
Aeromonas salmonicida (1.77%)
Rhodobacter capsulatus (1.68%)
Acinetobacter baumannii (1.33%)
Parabacteroides merdae (0.96%)
Thermoanaerobacterium saccharolyticum (0.92%)
Vibrio parahaemolyticus (0.74%)
Candidatus Hamiltonella defensa (0.74%)
Geobacillus kaustophilus (0.68%)
Ralstonia solanacearum (0.65%)
Erwinia pyrifoliae (0.65%)
Lactococcus lactis (0.61%)
Candidatus Liberibacter asiaticus (0.55%)
Mycoplasma pulmonis (0.55%)
Burkholderia pseudomallei (0.5%)
Lactobacillus gasseri (0.48%)
Citrobacter freundii (0.48%)
Acinetobacter johnsonii (0.46%)

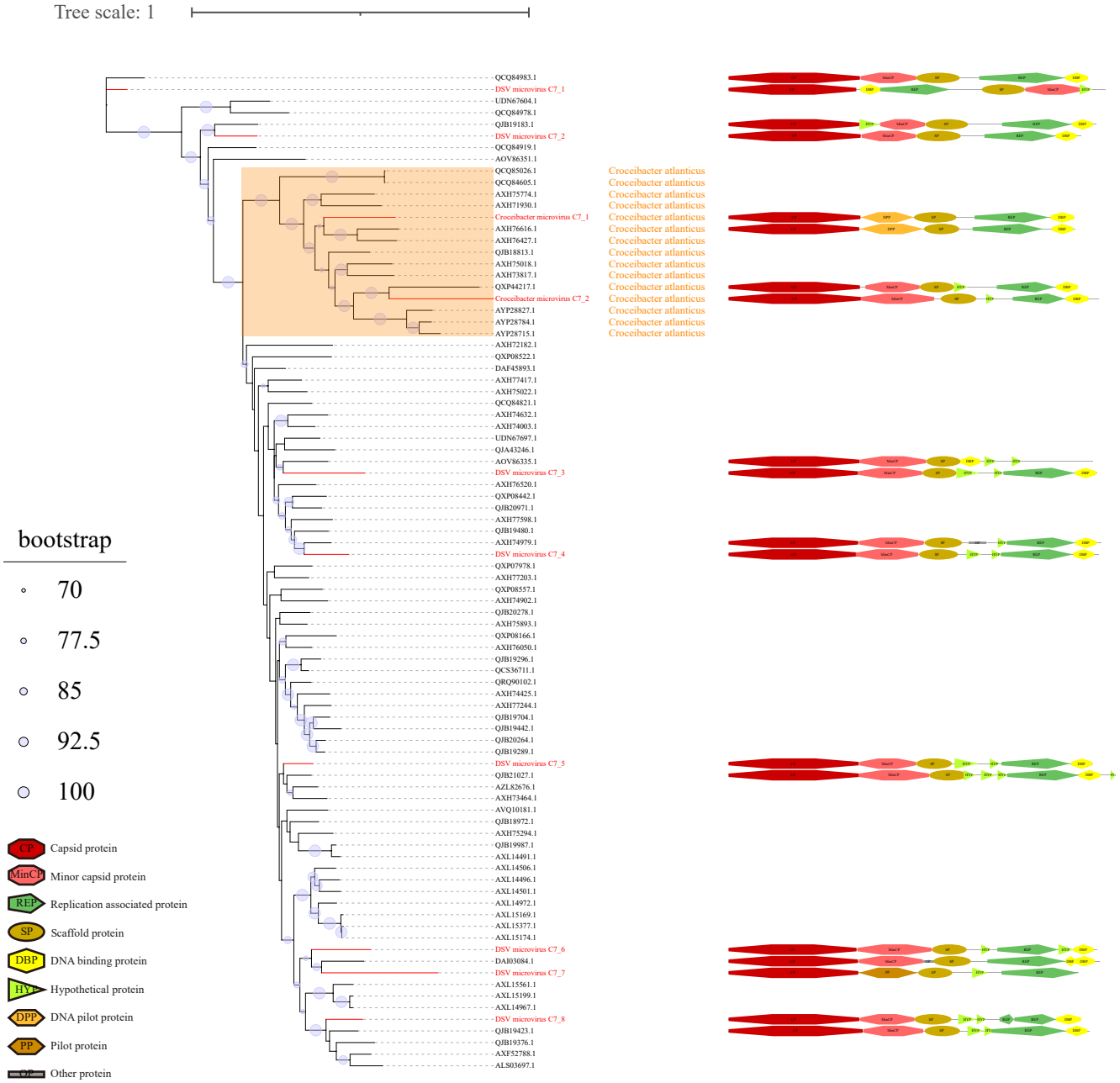

Tree scale: 10
MCP
Major capsid protein
CP
Capsid protein
MinCP
Minor capsid protein
REP
Replication associated protein
DBP
DNA binding protein
SP
Scaffold protein
DPP
DNA pilot protein
Other protein
HYP
Hypothetical protein
bootstrap
70
77.5
85
92.5
100

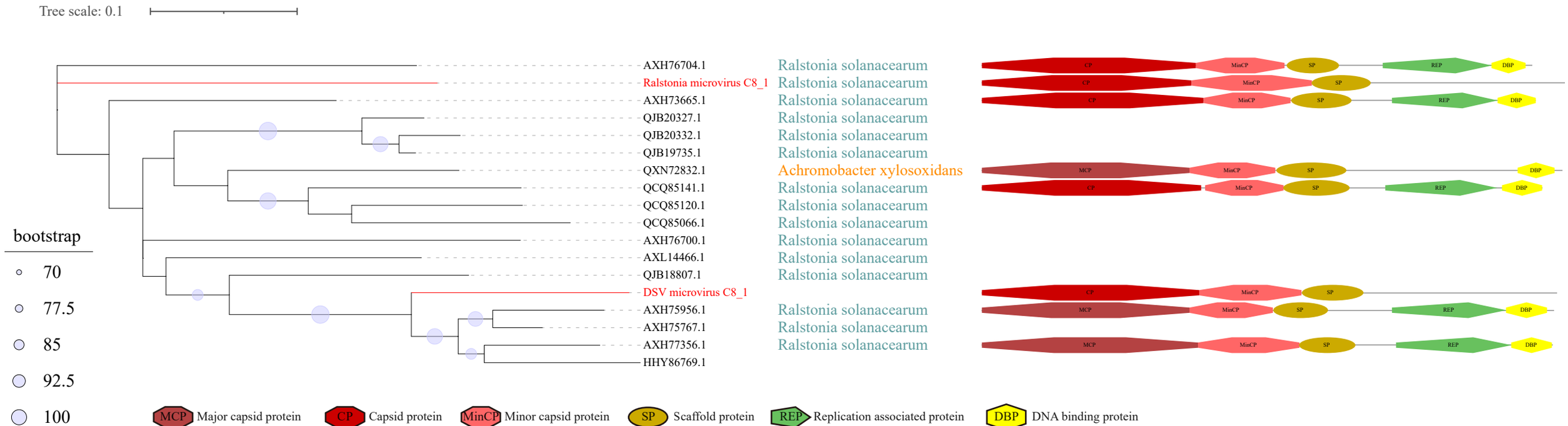

Tree scale: 0.1
AZL82978.1
DSV microvirus C3_1
QJB20234.1
QJB20478.1
QJB20418.1
DSV microvirus C3_2
QJB18777.1
DSV microvirus C3_3
QJB20229.1
QJB20452.1
QJB20549.1
QJB20538.1
AXH74680.1
AXH72370.1
DSV microvirus C3_4
QJB19335.1
DSV microvirus C3_5
UDN67527.1
AXH75065.1
AXL14430.1
DSV microvirus C3_6
AXH77225.1
AXH73803.1
AXH75145.1
AXH73081.1
DSV microvirus C3_7
QKI28889.1
AXH73326.1
AXH75351.1
DSV microvirus C3_8
bootstrap
70
77.5
85
92.5
100
Azospirillum brasilense
Enterobacter cancerogenus
Enterobacter cancerogenus
MCP
CP
MinCP
SP
HYP
REP
MCP Major capsid protein
CP Capsid protein
MinCP Minor capsid protein
SP Scaffold protein
HYP Hypothetical protein
REP Replication associated protein


Tree scale: 1
QJB21661.1 Bacillus halmapalus
Candidatus Pelagibacter microvirus C4_1 Candidatus Pelagibacter ubique
QJB20390.1 Bacillus halmapalus
QXP44125.1 Bacillus halmapalus
QJB19320.1 Bacillus halmapalus
QJB19358.1 Candidatus Pelagibacter ubique
QJB19353.1 Bacillus halmapalus
QJB19329.1 Bacillus halmapalus
Candidatus Pelagibacter microvirus C4_2 Candidatus Pelagibacter ubique
QJB18783.1 Bacillus halmapalus
QJB20395.1 Bacillus halmapalus
QJB20219.1 Bacillus halmapalus
QXP44170.1 Bacillus halmapalus
Candidatus Pelagibacter microvirus C4_3 Candidatus Pelagibacter ubique
DSV microvirus C4_1
QJB18788.1 Bacillus halmapalus
Candidatus Pelagibacter microvirus C4_4 Candidatus Pelagibacter ubique
QJB21688.1 Bacillus halmapalus
QJB20321.1 Bacillus halmapalus
QJB18843.1 Bacillus halmapalus
QXP07843.1 Bacillus halmapalus
QJB19840.1 Bacillus halmapalus
QXP44176.1 Bacillus halmapalus
NCB44216.1
QJB18802.1 Bacillus halmapalus
QJB19760.1 Bacillus halmapalus
QJB20316.1 Bacillus halmapalus
QXP44113.1 Bacillus halmapalus
QJB18828.1 Bacillus halmapalus
DAX35676.1 Candidatus Pelagibacter ubique
Candidatus Pelagibacter microvirus C4_5 Candidatus Pelagibacter ubique
QJB20252.1 Bacillus halmapalus
Bacillus microvirus C4_1 Bacillus halmapalus
Bacillus microvirus C4_2 Bacillus halmapalus
QJB20376.1 Bacillus halmapalus
QJB21706.1 Bacillus halmapalus
QJB20338.1 Bacillus halmapalus
Candidatus Pelagibacter microvirus C4_6 Candidatus Pelagibacter ubique
Croceibacter microvirus C4_1 Croceibacter atlanticus
Candidatus Pelagibacter microvirus C4_7 Candidatus Pelagibacter ubique
QJB19717.1 Candidatus Pelagibacter ubique
Candidatus Pelagibacter microvirus C4_8 Candidatus Pelagibacter ubique
bootstrap
70
77.5
85
92.5
100
MCP Major capsid protein
MinCP Minor capsid protein
REP Replication associated protein
DBP DNA binding protein
HYP Hypothetical protein
OP Other protein


Tree scale: 1
Bdellovibrio microvirus C5_1
QKI28946.1
QJB18932.1
AYQ58216.1
Bdellovibrio microvirus C5_2
UDN67844.1
AOV86371.1
QXP44252.1
AZL83017.1
Bdellovibrio microvirus C5_3
Bdellovibrio microvirus C5_4
QJB19851.1
QJB19846.1
DAV46552.1
QGH72133.1
AXH73287.1
Bdellovibrio microvirus C5_5
Bdellovibrio microvirus C5_6
Bdellovibrio microvirus C5_7
AXH78016.1
AOV86341.1
UDN67690.1
UDN67579.1
UDN67574.1
QXP08104.1
AZL82997.1
Bdellovibrio microvirus C5_8
Bdellovibrio microvirus C5_9
QJB20726.1
Bdellovibrio microvirus C5_10
UDN67889.1
UDN67496.1
UDN67491.1
QXP44246.1
QXP08040.1
QXP44165.1
AZL82946.1
AZL83027.1
AOV86369.1
Bdellovibrio microvirus C5_11
AZL82956.1
AXH77388.1
AZL82921.1
QXP44189.1
WP 121727187.1
QXP44319.1
Caulobacter microvirus C5_1
UIB81456.1
QJB19891.1
AZL83002.1
QXP44199.1
QJB20744.1
QJB20127.1
Escherichia microvirus C5_1
Escherichia microvirus C5_2
QGH72368.1
QXP44183.1
AXH75382.1
Bdellovibrio microvirus C5_12
UIW10383.1
UDN67823.1
DAU65897.1
DAW86105.1
DAQ11124.1
YP 009551424.1
QPB07344.1
DAK88301.1
QQL13326.1
Bdellovibrio microvirus C5_13
DAX11099.1
DAX04467.1
DAT70837.1
Bdellovibrio bacteriovorus
Caulobacter vibrioides
Escherichia coli
Candidatus Pelagibacter ubique
bootstrap
70
77.5
85
92.5
100
CP Capsid protein
MCP Major capsid protein
MinCP Minor capsid protein
REP Replication associated protein
HYP Hypothetical protein
OP Other protein


Tree scale: 0.01
DSV microvirus C6_1
QXP08016.1
QCQ84610.1
QXP08021.1
QXP07991.1
QCQ85020.1
AXH74368.1
QXP08202.1
QXP08227.1
AXH77965.1
DSV microvirus C6_2
QXP08496.1
DSV microvirus C6_3
QXP08236.1
QXP08148.1
DSV microvirus C6_4
bootstrap
70
77.5
85
92.5
100
CP
MinCP
SP
DBP
REP
CP
MinCP
SP
HYP
REP
CP
MinCP
SP
DBP
REP
MCP
MinCP
SP
HYP
HYP
REP
MCP
MinCP
SP
HYP
HYP
REP
MCP
MinCP
SP
DBP
REP
CP
MinCP
SP
DBP
REP
MCP
MinCP
SP
HYP
HYP
REP
MCP
HYP
MinCP
SP
DBP
REP
MCP Major capsid protein
CP Capsid protein
MinCP Minor capsid protein
SP Scaffold protein
DBP DNA binding protein
HYP Hypothetical protein
REP Replication associated protein


Tree scale: 1
QCQ84983.1
DSV microvirus C7_1
UDN67604.1
QCQ84978.1
QJB19183.1
DSV microvirus C7_2
QCQ84919.1
AOV86351.1
QCQ85026.1
QCQ84605.1
AXH75774.1
AXH71930.1
Croceibacter microvirus C7_1
AXH76616.1
AXH76427.1
QJB18813.1
AXH75018.1
AXH73817.1
QXP84217.1
Croceibacter microvirus C7_2
AYP28827.1
AYP28784.1
AYP28715.1
AXH72182.1
QXP08522.1
DAF45893.1
AXH77417.1
AXH75022.1
QCQ84821.1
AXH74632.1
AXH74003.1
UDN67697.1
QJA43246.1
AOV86335.1
DSV microvirus C7_3
AXH76520.1
QXP08442.1
QJB20971.1
AXH77598.1
QJB19480.1
AXH74979.1
DSV microvirus C7_4
QXP07978.1
AXH77203.1
QXP08557.1
AXH74902.1
QJB20278.1
AXH75893.1
QXP08166.1
AXH76050.1
QJB19296.1
QCS36711.1
QRQ90102.1
AXH74425.1
AXH77244.1
QJB19704.1
QJB19442.1
QJB20264.1
QJB19289.1
DSV microvirus C7_5
QJB21027.1
AZL82676.1
AXH73464.1
AVQ10181.1
QJB18972.1
AXH75294.1
QJB19987.1
AXL14491.1
AXL14506.1
AXL14496.1
AXL14501.1
AXL14972.1
AXL15169.1
AXL15377.1
AXL15174.1
DSV microvirus C7_6
DAI03084.1
DSV microvirus C7_7
AXL15561.1
AXL15199.1
AXL14967.1
DSV microvirus C7_8
QJB19423.1
QJB19376.1
AXF52788.1
ALS03697.1
Croceibacter atlanticus
Croceibacter atlanticus
Croceibacter atlanticus
Croceibacter atlanticus
Croceibacter atlanticus
Croceibacter atlanticus
Croceibacter atlanticus
Croceibacter atlanticus
Croceibacter atlanticus
Croceibacter atlanticus
Croceibacter atlanticus
Croceibacter atlanticus
Croceibacter atlanticus
Croceibacter atlanticus
bootstrap
70
77.5
85
92.5
100
CP Capsid protein
MinCP Minor capsid protein
REP Replication associated protein
SP Scaffold protein
DBP DNA binding protein
HYP Hypothetical protein
DPP DNA pilot protein
PP Pilot protein
OP Other protein


Tree scale: 0.1
AXH76704.1
Ralstonia microvirus C8_1
AXH73665.1
QJB20327.1
QJB20332.1
QJB19735.1
QXN72832.1
QCQ85141.1
QCQ85120.1
QCQ85066.1
AXH76700.1
AXL14466.1
QJB18807.1
DSV microvirus C8_1
AXH75956.1
AXH75767.1
AXH77356.1
HHY86769.1
Ralstonia solanacearum
Achromobacter xylosoxidans
bootstrap
70
77.5
85
92.5
100
MCP Major capsid protein
CP Capsid protein
MinCP Minor capsid protein
SP Scaffold protein
REP Replication associated protein
DBP DNA binding protein