

# An Insight into Phage Diversity at Environmental Habitats using Comparative Metagenomics Approach

Krupa Parmar<sup>1</sup> · Nishant Dafale<sup>1</sup> · Rajesh Pal<sup>1</sup> · Hitesh Tikariha<sup>1</sup> · Hemant Purohit<sup>1</sup>

Received: 18 May 2017 / Accepted: 12 September 2017 / Published online: 19 September 2017  
© Springer Science+Business Media, LLC 2017

**Abstract** Bacteriophages play significant role in driving microbial diversity; however, little is known about the diversity of phages in different ecosystems. A dynamic predator–prey mechanism called “kill the winner” suggests the elimination of most active bacterial populations through phages. Thus, interaction between phage and host has an effect on the composition of microbial communities in ecosystems. In this study, secondary phage metagenome data from aquatic habitats: wastewater treatment plant (WWTP), fresh, marine, and hot water spring habitat were analyzed using MG-RAST and STAMP tools to explore the diversity of the viruses. Differential relative abundance of phage families—*Siphoviridae* (34%) and *Myoviridae* (26%) in WWTP, *Myoviridae* (30%) and *Podoviridae* (23%) in fresh water, and *Myoviridae* (41%) and *Podoviridae* (8%) in marine—was found to be a discriminating factor among four habitats while *Rudoviridae* (9%), *Globuloviridae* (8%), and *Lipothrixviridae* (1%) were exclusively observed in hot water spring. Subsequently, at genera level, *Bpp-1-like virus*, *Chlorovirus*, and *T4-like virus* were found abundant in WWTP, fresh, and marine habitat, respectively. PCA analysis revealed completely disparate composition of phage in hot water spring from other three ecosystems. Similar analysis of relative abundance of functional features corroborated observations

from taxa analysis. Functional features corresponding to *phage packaging machinery*, *replication*, *integration and excision*, and *gene transfer* discriminated among four habitats. The comparative metagenomics approach exhibited genetically distinct phage communities among four habitats. Results revealed that selective distribution of phage communities would help in understanding the role of phages in food chains, nutrient cycling, and microbial ecology. Study of specific phages would also help in controlling environmental pathogens including MDR bacterial populations using phage therapy approach by selective mining and isolation of phages against specific pathogens persisting in a given environment.

## Abbreviations

WWTP Waste water treatment plant  
MDR Multi-drug resistant

## Introduction

Bacteriophages are the infectious virus that replicate within a particular bacterial host and one of the most common and diverse beings on earth. The life cycles of a bacteriophage entirely depend on the host bacteria and are found to be ubiquitous in nature with respect to the availability of host bacterium for infection. Bacteriophages are classified based on the size, structure, genetic material, and host range varying from simple to more elaborate and complex. The most common habitats where bacteriophages flourish are soil, oceans, deserts, wastewater, sewage, the gut of animals, and activated sludge [12, 35, 49]. Alterations have been observed between diverse habitats in terms of occurrence of microbial communities. The chemical,

**Electronic supplementary material** The online version of this article (doi:10.1007/s00284-017-1357-0) contains supplementary material, which is available to authorized users.

✉ Nishant Dafale  
nishant.dafale@gmail.com; na\_dafale@neeri.res.in

<sup>1</sup> Environmental Biotechnology & Genomics Division, CSIR-National Environmental Engineering Research Institute, Nagpur 440020, India

physical, and biological conditions prevailing in a particular habitat is thought to be one of the factors which shape the microbial community. Extremes of temperatures in hot springs favor the growth of hyperthermophiles, copiotroph flourish in wastewater, halophiles are observed in oceans, etc., according to the environment provided by particular habitat. A dynamic predator–prey mechanism called “kill the winner” [43] proposes that phages preferentially infect the most active host (bacteria) in any given environment. Thus, the interaction between phage and host chiefly maintains the composition of microbial communities [36, 39]. Bacteriophages play a major role in host mortality, carbon cycling [9], nutrient cycling [41], horizontal gene transfer leading to microbial diversity [45], and structuring food webs by being a nutrient source for protists [31, 40].

Marine water is one of the dense natural sources for phages, where  $10^7$  particles/ml have been found in seawater [19]. Majority of phages in aquatic environments are double-stranded DNA phages belonging to the order *Caudovirales* [45]. *Caudovirales* are classified into *Podoviridae* with short tails, *Siphoviridae* with long non-contractile tails, and *Myoviridae* with long contractile tails. In the different stages of wastewater treatment plants (WWTPs), phages have been observed in  $10^8$ – $10^{10}$  ml<sup>-1</sup> concentration [4, 47] 10–1000 times higher than in natural aquatic environments, suggesting WWTPs as an important reservoir and source of phages [42]. In hot water springs, bacteriophages infecting thermophilic archaea are found in abundance, suggesting extreme thermal environment harbors thermophiles. The species composition and abundance of a potential host in freshwater lake vary with lake trophic status, depth, watershed, and size [6, 21] which results in a variation of phages found in the freshwater habitat. Human gut consists of a diverse microbial community resulting in a diverse phage community as well. Through interactions with microbes, phages also influence the global flow of genes affecting host phenotype, adaptation to environment and evolution [36]. Metagenomic investigations of phages collected from various ecosystems have provided insight into local and global phage diversity [7, 23], genotypic distribution [7, 44], functional potential [15, 46], and replication strategy [24, 46].

Currently, the latest “omics” approaches will assist to analyze the bacteriophages directly from the sample which provides information about the total culturable and non-culturable phages present in a particular environment. Pre-processing of samples and concentration of phages [11] is required for the metagenomics analysis. Thus, the phages or particular genes that are highly abundant and specific to a particular niche can be determined. Metagenomic tools can also be used in a comparative approach to explore the differences in clades of various niches, study chronological

changes, and relate the results of investigational data. Metagenomics elucidates estimates of phage community structures and of the richness, evenness, and abundance of each phage genotype [19]. With the rapid rise in viral genomic and metagenomic study, a new database was developed recently coined as pVOG (prokaryotic Virus Orthologous Group) which can further expand the in-depth virus study [17]. Also server like METAVIR can be a great tool for taxonomic study of viruses along with their comparative study [37]. Microbial activities shape the biogeochemistry of the planet so it becomes important to understand the metabolic processes of phages as well. Most of the functional diversity of phages is maintained in all of the communities but the relative occurrences of metabolism vary and the differences between metagenomes predict the biogeochemical conditions of each environment [11]. Phages are an important factor for the regulation of microbial community composition and affect the cycling of carbon and other nutrients [41]. To identify the characteristics and potential specificities of phage communities in different habitats, the phage communities of hot water spring, freshwater, marine water, and wastewater treatment plants (from the previously published data) were compared in this study. The mining and characterization of bacteriophages from diverse habitats will assist in estimating the taxonomical and functional role of the phages. A habitat-specific mining of phages can help in discovering new phages to be used for formulation of phage biocontrol. Moreover, comparative study of phage abundance and distribution can be used to understand the impact of phages on biogeochemical cycle and recycling of nutrients by controlling bacterial hosts.

## Materials and Methods

### Metagenome Sample Selection

To study the diversity of the phages, four different habitats were selected based on disparate environmental conditions: (1) freshwater, (2) marine, (3) hot water spring, and (4) WWTP. The corresponding metagenome data were selected from the MG-RAST public database. All the above-mentioned metagenome data were derived from water sample of the mentioned sources, where virus were concentrated by filtration and centrifugation technique. Table 1 summarizes the features of the metagenomes selected in this study, illustrating MG-RAST ID, habitat, metagenome size, sequencing method, and virus purification method. The specific sample collection sites for habitat (1) freshwater were the Crim Dell Mouth (37.267°N, 76.721°W), the Pogonia Mouth (37.268°N, 76.727°W), and Matoaka open water (37.264°N, 76.722°W) from Lake

**Table 1** Source of metagenomic data from MG-RAST

Sr.No.	Accession number	Habitat	Size (bp)	Sequencing method	Virus purification method	Source
1	4523574.3	Fresh water	11,576,612	WGS	CsCl density gradient	[18]
2	4523575.3	Fresh water	18,445,791	WGS	CsCl density gradient	[18]
3	4523576.3	Fresh water	18,447,146	WGS	CsCl density gradient	[18]
4	4441095.3	Hot water spring	8,270,842	Other	By filtration	[38]
5	4441096.3	Hot water spring	22,557,192	Other	By filtration	[38]
6	4441623.3	Marine	42,535,763	WGS	CsCl density gradient	[14]
7	4441628.3	Marine	26,809,683	WGS	CsCl density gradient	[14]
8	4441629.3	Marine	26,809,683	WGS	CsCl density gradient	[14]
9	4450223.3	WWTP	108,954,161	WGS	CsCl density gradient	[42]
10	4450219.3	WWTP	70,775,098	WGS	CsCl density gradient	[42]
11	4450221.3	WWTP	75,816,433	WGS	CsCl density gradient	[42]
12	4450224.3	WWTP	119,334,582	WGS	CsCl density gradient	[42]

Matoaka, a temperate, eutrophic lake in southeastern Virginia, USA. (2) The samples for marine habitat were collected from the Gulf of Mexico—Texas coast (28.258, 87.6733), the British Columbia coastal waters (49.705, 124.351667), and from the Arctic Ocean (73.16167, 159.44667). (3) The samples for hot water spring were from Bear Paw (an unofficial name for LRNN374) (44°3321.994N, 110°505.232W) and Octopus (44°32Z.701N, 110°4752.402W) hot springs of Yellowstone National Park (YNP). Bear Paw hot spring (74 °C) is in the river group of the lower geyser basin of YNP, while Octopus (93 °C) is about 5 km away in the White Creek area. (4) The samples for WWTP habitat were collected from different stages of a WWTP in tropical climate in Singapore. The domestic wastewater mainly comprised water discharged from house-holds, commercial units, and manhole, and possibly some legally met effluent discharged from pre-treatment plants in industrial units, collected through a sewer system, and transported to the WWTP. The samples were collected from influent, activated sludge, effluent, and anaerobic digester.

### Bioinformatics Analysis

MG-RAST (version 3.3.6) [25] (<https://metagenomics.anl.gov/>) was used to analyze the metagenome datasets. To nullify the variation in data size, likely to be an issue in this study, multiple metagenomes from different sites of same habitats were selected and grouped to represent a single habitat. A hierarchical classification option was used as the data type and SEED-based subsystems [29] as the annotation source (with parameters: identity cut-off—60%, e-value cut-off— $1e-5$ , and alignment length cut-off—15 for amino acids) to generate a feature abundance profile for

each habitat. Similarly, a taxonomic abundance profile was generated using parameters described above, against the M5NR (novel non-redundant database used for taxonomic annotation and classification) annotation source and best classification was chosen according to the data type. Subsequently, a pie chart was prepared to show the distribution of microbial communities in each habitat.

### Statistical Analysis

Statistical analysis of Metagenomic Profiles (STAMP) (version v2.0.3) package [30] was implemented to analyze statistically significant differential abundance of phage features among the four habitats. A functional feature abundance table corresponding to the four habitats was generated via MG-RAST and imported to the STAMP tool for statistical analysis. A comparative heat-map using a normalized taxonomy abundance profile at the family taxa was prepared to compare the four niches with respect to their phage diversity. Similarly, a heat-map was generated in STAMP for the four habitats, considering functional level in the hierarchy of the MG-RAST functional profile by applying the ANOVA test with Hochberg's FDR multiple test correction [27]. The extended error bar plots were generated via two sample analysis by applying Benjamini–Hochberg's FDR multiple test correction. In order to consider features corresponding to phages, the subsystem–phages, prophages, transposable elements, plasmids—was compared using the same analysis pipeline. The results were curated to exhibit only features with a  $q$  value  $< 0.05$  and  $> 1\%$  difference in proportions of a given feature between the two groups.

## Results

### Classification of Microbial Community in Different Habitats

A comparison was made to study the abundance of microbial communities in four disparate habitats—marine, fresh water, hot water spring, and wastewater treatment plant. Figure 1 illustrates: in freshwater, 78% microbial species were found to be of bacteria while only 7% bacteriophages were observed. In marine water, 67% bacteria and 11% bacteriophages; whereas in hot water spring, 63% bacteria and 15% bacteriophages; and in WWTP, 83% bacteria with 4% bacteriophages were observed. The abundance of bacteria is very high in every habitat as compared with that of other microbial communities.

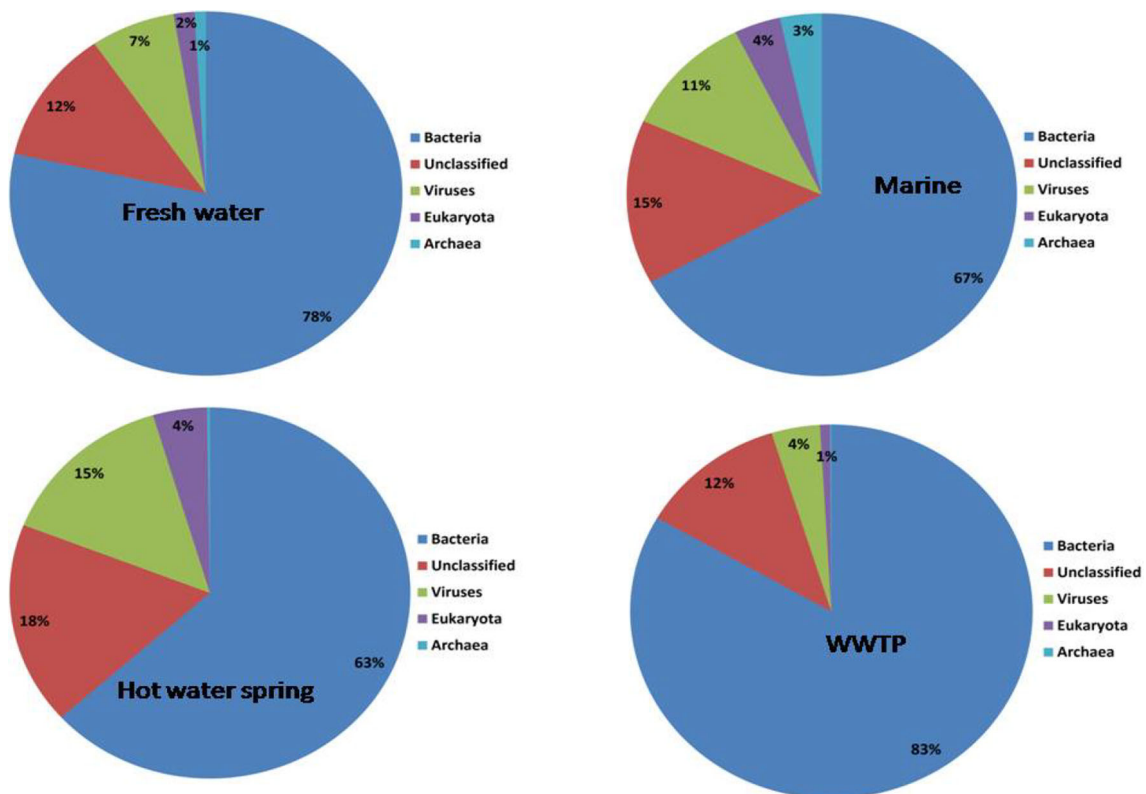
### Comparative Analyses of Phage Diversity in Selected Habitats

Principal component analysis (PCA) (emphasizes variation and brings out strong patterns in a dataset) (Fig. 2) illustrates habitat-based clustering of metagenomes suggesting a right selection of data for each habitat. PCA plot

generated from taxonomic and functional features corroborated each other. The result is in accordance with the fact that each habitat is unique with respect to its microbial communities and corresponding functional characteristics. A clear cluster of phage communities of WWTP, freshwater, and marine habitat was observed; however, freshwater and WWTP clusters were found to be much closer to each other. The clusters of marine and hot water spring were comparatively far from other habitats suggesting the phage communities were very specific because of extreme environment. The phage communities of hot water spring resulted in a very assorted cluster outlying from other habitats because of high temperature and diverse metabolic rates of surviving phage hosts.

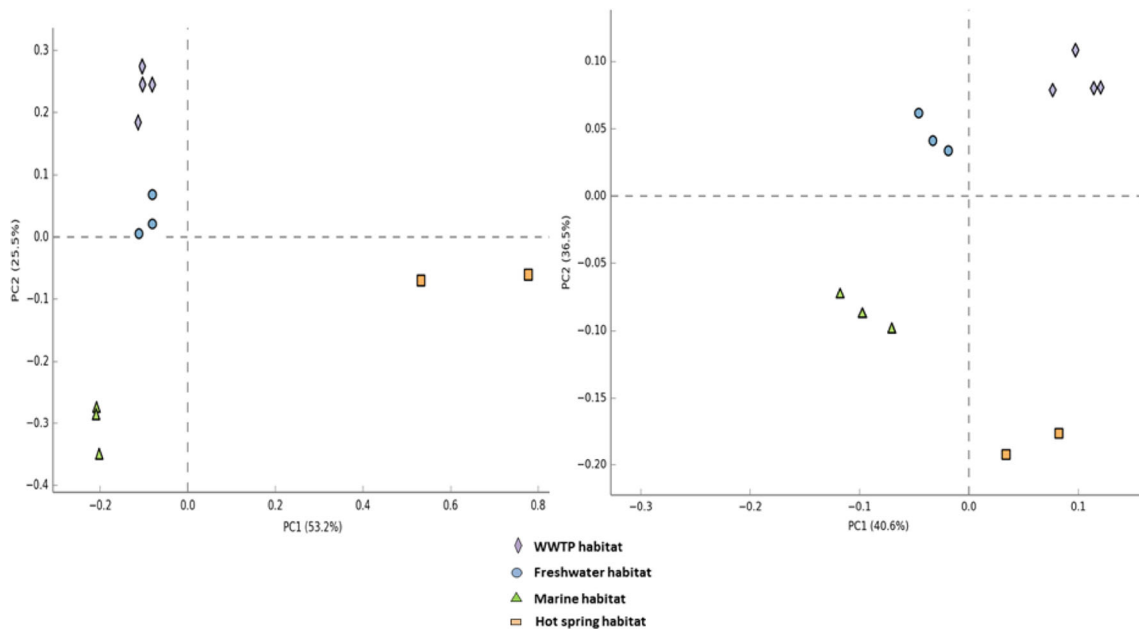
### Abundance of Diverse Phage Families in Different Habitats

Figure 3 shows another PCA plot indicating a correlation of four habitats with each other in terms of their phage diversity. It also accounts the phage families being dominant at each habitat, in the form of pie chart. As compared to Fig. 2 showing individual metagenome data, PCA plot in Fig. 3 was generated by using group analysis of MG-



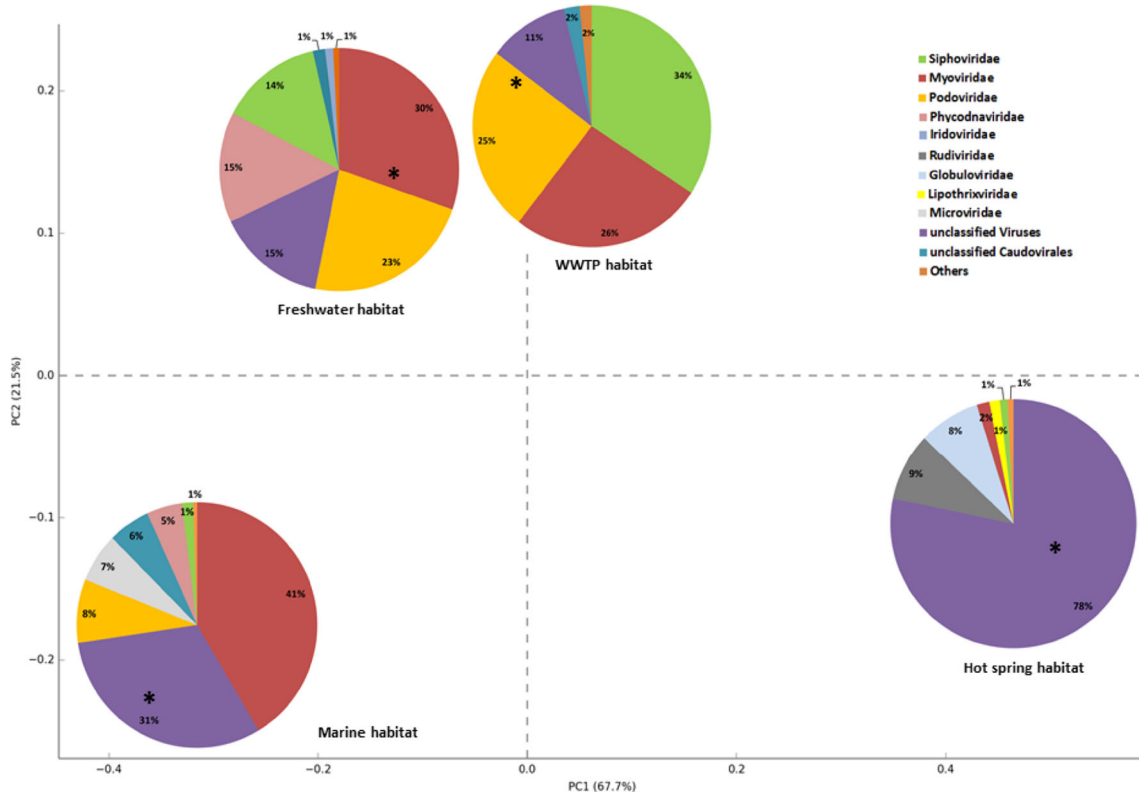
**Fig. 1** Comparative microbial diversity at different habitats. Pie chart was generated using microbial abundance data. Each chart represents the percentage abundance of microbial group in a specific

habitat. Abundance values are generated from normalized and statistical analysis carried on data retrieved from MG-RAST



**Fig. 2** Principal Component Analysis showing clustering of all the metagenome data based on **a** phage diversity at family level and **b** functional features. Generated using ANOVA statistical test Benjamini–Hochberg FDR multiple test correction. X- and Y-axis are Principal X component and Principal Y component, respectively.

The first figure represents the distribution of viral metagenome on the basis of their taxonomical data and second figure represents distribution and clustering of viral metagenome according to their annotated functional protein



**Fig. 3** Principal Component Analysis illustrating clustering of grouped metagenome data and pie chart showing an abundance of phage families at different habitats. Pie charts were generated using microbial abundance data at family level. X- and Y-axis are Principal

X component and Principal Y component, respectively. The unclassified viruses in hot spring mostly belong to hyperthermophilic archaeal viruses as identified through METAVIR server



RAST server which normalizes the data. Normalization in MG-RAST is performed by R-based algorithm which transforms/distributes the data so that all the data have same mean and standard deviation and they become more comparable as intersample variability is removed [25]. The abundance of phage families present in particular habitats revealed double-stranded DNA phages were most dominant. Phage families *Siphoviridae*, *Myoviridae*, *Podoviridae* from the order *Caudovirales* were the most abundant dsDNA phages in every habitat. In WWTP, *Siphoviridae* were observed the most abundant (34%) as compared with its abundance in other habitats. In fresh water, *Myoviridae* (30%), *Phycodnaviridae* (15%) were abundant with *Iridoviridae* (1%) exclusively reported. In marine habitat, *Myoviridae* (41%) was abundant with *Microviridae* (7%) exclusively. In hot water spring, 78% unclassified phages with 9% *Rudiviruses*, 8% *Globuloviridae*, and 1% *Lipothrixviridae* were observed exclusively. The maximum amount of unclassified viruses in hot water spring suggests a much exploration is needed in the untouched extreme habitat. As the data were derived from MG-RAST, annotation was carried out using RefSeq, COG, IMG database, etc., and there is a need for separate annotation of reads for this unclassified phages. Thus, the METAVIR server has been used to explore the unclassified viruses which were found to be hyperthermophilic archaeal viruses which are still least studied category of viruses. Apart from this, a small percentage of *Myoviridae* and *Siphoviridae* are also present.

### Comparative Phylogenetic Analysis of Phage Diversity Among Four Habitats

The similarity among the phage genera of four different habitats was studied (Fig. S1). When compared with other habitats, the genera *Bpp-1-like viruses*, *N4-like viruses*, *P22-like viruses*, *phiKZ-like viruses*, *AHJD-like viruses* were found more dominant in WWTP. *Bpp-1-like* and *N4-like viruses* belong to *Podoviridae* which infects *E. coli* in the sewage; *P22-like viruses* belong to *Podoviridae* infecting *Salmonella typhimurium*; *phiKZ-like viruses* belong to *Myoviridae* infecting *Pseudomonas*; *AHJD-like viruses* belong to *Podoviridae* infecting *Staphylococcus*. Freshwater habitat had *Chlorovirus* and *T7-like viruses* in dominance; *Chlorovirus* belongs to *Phycodnaviridae* infecting algae in freshwater with type species *Paramecium bursaria Chlorella phage 1*. *T7-like viruses* infect *E. coli* and some enteric bacteria [26]. Hot water spring resulted in *Rudivirus*, *Globulovirus*, and *Betalipothrixvirus* in abundance. *Rudivirus* are dsDNA phages that infect hyperthermophilic archaea of the kingdom *Crenarchaeota* [34, 50] type species which include *Sulfolobus islandicus rod-shaped phage*. Archaea serves as a natural host for

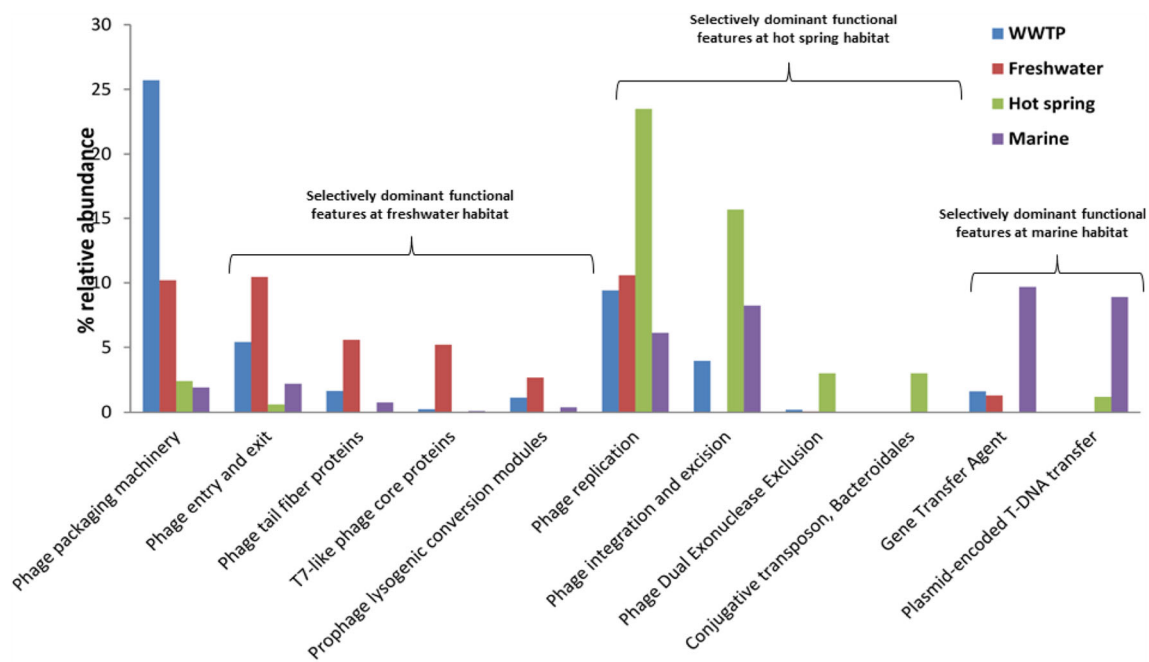
*Betalipothrixvirus*; it has dsDNA. In marine habitat, *T4-like virus* and *Bdellomicrovirus* were found in abundance; Gram-negative bacteria serve as hosts for *T4-like viruses*; and *Bdellomicrovirus* belongs to *Microviridae* containing ssDNA and infecting *Bdellovibrio* bacteria.

### Heat-Map Analyses of Relative Abundance of Phage Families and Functional Features

With respect to the metabolic functions, the MG-RAST assigned the sequences into 27 different functional categories. In the heat-map analyses (Fig. S2), *Phage packaging machinery* has been found most abundant in WWTP indicating a lytic life cycle suggesting a high phage dispersal rate. *Phage replication and integration and excision* are found as the most abundant function of hot water spring, while phage packaging is very less. *Gene transfer* is found to be the most abundant function in marine habitat, while *Phage entry and exit* and *gene transfer* play a major role in monitoring the diversity of freshwater. *Phage lysogeny conversion modules* include gene clusters that are thought to offer bacteria a selective advantage in their environments (or virulence niches) and are thus kept by prophages for the selective life cycle. *Phage tail fiber* function abundance indicates dominant tailed-phage diversity within freshwater habitat. The relative abundance of phage functional genes in four different habitats was studied (Fig. 4). *Phage packaging machinery* was found to be relatively most abundant in WWTP. In freshwater habitat, *phage entry and exit*, *phage tail fiber proteins*, *T7-like phage core proteins*, *prophage lysogenic conversion modules* were dominant. In hot water spring, *phage replication*, *phage integration*, and *excision*, *phage dual exonuclease exclusion*, *conjugative transposon*, *bacteroidales* were abundant. In marine habitat, only two functional features were found to be dominant: *Gene transfer agent* and *plasmid-encoded T-DNA transfer*. Thus, in PCA clear clusters were observed away from each other signifying a habitat-specific functionality among the phages.

### Discussion

As observed in the results, the abundance of microbial communities (especially bacteria) other than bacteriophages has been observed high as (i) the database of phage genomes has a much lesser number and size than other microbial genomes, (ii) the prophages are unidentified within microbial genomes, (iii) the horizontal gene transfer between phages and host bacteria is very high, and (iv) functional genes observed in hosts are also found in phages [28, 47]. The lack of conserved sequences like 16S rRNA



### Phage functional features

**Fig. 4** Relative abundance of phage functional features at different habitats. Graph shows the relative abundance of various proteins in all the four habitats. Phage replication proteins are the most abundant

in virus is attributable to fact that they are less annotated as compared with bacteria in a metagenomic sample. Also as virus genes are linked with bacteria due to exchange of genetic information between them, the database annotates most protein to be of bacterial origin. But the recent shift in study of virus gene pool present in the environment has lead to the development of many software, database, and algorithms which can lead to rapid and accurate identification of virus in a metagenomic sample [5].

Data retrieved from MG-RAST revealed that the viruses are rich in freshwater and wastewater treatment plant and least in marine habitat. Moreover when talking about the diversity, the more varied and fluctuating habitat, i.e., WWTP harbors more diverse kinds of bacteria, followed by the diverse kinds of viruses found in freshwater which also has an intake of varied chemicals. Hot spring lacks the viral diversity attributed to the fact that they have very specific genera of bacteria and archaea to survive and cope with harsh conditions.

The hot water spring temperature is found to be over the limit of eukaryotic life (near to 60 °C), limiting the microbial life to thermophilic bacteria, archaea, and their phages [22] which makes it to stand apart in the PCA analysis. WWTP family cluster was found to be very diverse as copiotroph flourishes abundantly in WWTP as it is rich in organic carbon and inorganic nutrients from waste and so are the phages corresponding to the same. The

ones followed by phage packaging machinery. Of the significant one, Prophage lysogenic conversion modules are the least abundant virus protein family in all the habitat

phage community of marine habitat was found to be different from other habitats as the marine environment provides a much-varied habitat like high salinity, higher temperatures of the sea floor, and diverse nutrient source from the food web placing it different from other habitats [8]. Disparate environmental conditions in each of these habitats provide a unique condition for the growth and development of diversity among the phage host. This diversity among the hosts results in a high diversity and specificity among the phages prevailing in each habitat. The phages also govern diversity by horizontal gene transfer between the hosts. Thus, the interaction of phages infecting the hosts chiefly shapes the composition of microbial communities [36, 39].

Phage hosts commonly prevalent in WWTP are imparted from the household domestic waste [48]. The phages in the WWTP are thought to be responsible for infecting bacterial hosts that remove phosphorous thus inhibiting the nutrient cycle [3]. Similarly, in a study of 191 known sequences in activated sludge, 95% were homologous to bacteriophage within the families *Myoviridae* (40.3%), *Siphoviridae* (31.9%), *Podoviridae* (25.6%), and unclassified (2.2%) phages [32]. The species composition and abundance of potential phage host in freshwater lake vary with lake trophic status, depth, watershed, and size [6] which results in a variation of phages found in the freshwater habitat. With both the freshwater and WWTP

swinging in richness and diversity, Siphoviridae were highly abundant in WWTP and Myoviridae in freshwater. As both the habitats are comparatively rich in nutrients and face lot of chemical stress from the environment, the more challenging condition arises in WWTP and so it possess diverse bacterial and archaeal community and the viral community. Further the high abundance of order Caudovirales in WWTP is due to the presence of pathogenic Enterobacteria phage as the WWTP water was derived from municipal area [4]. This can also be the reason for the presence of Caudovirales in freshwater where water channels from the human living area. Phage-mediated horizontal gene transfer has been observed in freshwater which contributes to genetic diversity [20]. Likewise, in a study of phages in East Lake, China, the genetic structure of the phage community revealed a high genetic diversity covering 23 phage families including members of *Myoviridae*, *Podoviridae*, *Siphoviridae*, *Phycodnaviridae*, and *Microviridae*, which infect bacteria or algae and the highest phage genetic diversity occurred in samples collected in August, followed by December and June, and the least diversity was in March [16].

The load of marine phages is dependent on their hosts, therefore any change in the abundance, metabolic rate, and generation time of the host affects phage abundance [13]. Review on marine viruses highlights that dominant *Cyanophages* belong to *Podovirus* followed by *Myovirus* and *Siphovirus* [10]. Our study along with the previous one suggests that *Microviruses* (icosahedral ssDNA phages) were particularly prevalent in marine habitats and were absent in other habitats [38]. They influence biogeochemical cycles, offer to regulate microbial biodiversity, recycle carbon through marine food webs, and are essential in controlling bacterial population explosions. Marine phages play an important part in shifting nutrients from living forms into dissolved organic matter and detritus, therefore, phages must influence the carbon, nitrogen, and phosphorus cycles, but the exact influences are currently not understood [13, 33].

Hyperthermophilic bacterial hosts are exclusively found in the extreme hot environment and thus the phages. Similarly, in a study of the phages of the extreme thermal environment in the Yellowstone National Park, the phages were found to have nearly identical morphologies to *Fuselloviruses*, *Rudiviruses*, and *Lipothrixviruses* isolated from Japan and Iceland [33]. Some hyperthermophilic phages contain supplementary metabolic genes that alter the host enzymatic capabilities [2]. Thermophilic archaea, *Thermoproteus archaea* and archaea *acidianus*, found in abundance in the boiling environment of hot water springs serve as a host for the *Lipothrixviridae*. Double-stranded DNA phages *Rudiviridae* which are specially found in hot spring infect hyperthermophilic archaea from

*Crenarchaeota* [34, 50]. Also high abundance of hyperthermophilic archaeal viruses suggests that there are still large groups of these dsDNA viruses whose host archaea and their relation are still not studied and understood.

Functional feature of a lytic life cycle in WWTP suggested a high number of phage hosts infected, as a result, abundant progeny phages would come into existence. On the other hand, high bacterial biomass gets mixed with the environment imparting its impact on the nutrient cycle. *Phage replication and integration and excision* indicated lysogeny form of a life cycle. This serves as a survival strategy against extremes of high temperatures of hot water spring and surviving by persisting with the host chromosome. Lysogeny also indicates a high gene transfer rate among the habitat which confers genetic diversity among the community. *Gene transfer* in marine habitat helps in the horizontal gene transfer governing microbial diversity which is evident from the high abundance of gene transfer agent and phage integration and excision protein [1]. The gene transfer among phages and hosts also plays an important role in the biogeochemical cycles as the enzymes required for nutrient recycle are also transferred. *Phage lysogeny conversion modules* in fresh water offer bacteria the selective advantage in their environments (or virulence niches) and are thus kept by prophages to make them more 'welcome' in their bacterial hosts' genomes. The environmental condition of each habitat provides a unique niche for the growth of specific microbial community which in turn serves as a host for the host-specific phages. The high abundance of phage replication and integration protein in hot spring habitat suggests that the virus dwells only inside thermophilic bacteria and archaea so that they can cope with harsh environment conditions.

The lysis of host by phages transfers the cell mass, nitrogen, and phosphorous into a pool of dissolved organic matter which influences nutrient cycling, nutrient flow within food webs [13, 41, 45] and impacts population size, biodiversity, and horizontal gene transfer [13, 41]. The environmental conditions and the host activities also influence the life cycle of phages suggesting it accept the lysogeny or the lytic phase. As the phages infect the most active host by lysis, "kill the winner" hypothesis, the dead debris of the hosts which are high in nutrients are in turn amalgamated with the environment helping in the recycling of nutrients. By lysogeny lifecycle, a high gene transfer results in a high microbial diversity. Every habitat in the study was extremely different from each other which provides a diverse niche for the growth of organisms and corresponding phages.

The study concludes that phage diversity varies with the habitats and there are specific phages for specific bacterial diversity at a given niche/habitat which can be seen in case of extreme conditions of hot water spring as it get clustered



separately in PCA analysis. Thus this study highlights that the uniqueness of phages increases with the more extreme environmental condition. It also brings forward the fact that phages replication and transfer mode also depend upon the habitat in which they flourish.

**Acknowledgements** The authors thank the Director, CSIR- NEERI for providing constant support and infrastructure for the research work, and funding from MLP project is gratefully acknowledged. Plagiarism has been checked using iThenticate software at CSIR-NEERI, Nagpur, and KRC number is \2016JUN\EGD\1.

**Author Contributions** HP and ND conceived and planned the experiments, also participated in the analyses and interpretation of the metagenome sequencing data; KP, HT, and RP analyzed the secondary phage metagenome data from different habitats using MG-RAST and STAMP tool to explore diversity. All the authors were involved in drafting the manuscript and approved the final manuscript.

#### Compliance with Ethical Standards

**Conflict of interest** The authors declare that they have no competing interests.

**Ethical Approval** This research did not involve human participants or animals.

**Informed Consent** This research did not involve animal and human participants and therefore is not applicable.

## References

- Adriaenssens EM, Van Zyl L, De Maayer P, Rubagotti E, Rybicki E, Tuffin M, Cowan DA (2015) Metagenomic analysis of the viral community in Namib Desert hypoliths. *Environ Microbiol* 17(2):480–495. doi:[10.1111/1462-2920.12528](https://doi.org/10.1111/1462-2920.12528)
- Anantharaman K, Duhaime MB, Breier JA, Wendt KA, Toner BM, Dick GJ (2014) Sulfur oxidation genes in diverse deep-sea viruses. *Science* 344(6185):757–760. doi:[10.1126/science.1252229](https://doi.org/10.1126/science.1252229)
- Barr JJ, Slater FR, Fukushima T, Bond PL (2010) Evidence for bacteriophage activity causing community and performance changes in a phosphorus-removal activated sludge. *FEMS Microbiol Ecol* 74(3):631–642. doi:[10.1111/j.1574-6941.2010.00967.x](https://doi.org/10.1111/j.1574-6941.2010.00967.x)
- Bibby K, Peccia J (2013) Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environ Sci Technol* 47(4):1945–1951. doi:[10.1021/es305181x](https://doi.org/10.1021/es305181x)
- Bolduc B, Youens-Clark K, Roux S, Hurwitz BL, Sullivan MB (2016) iVirus: facilitating new insights in viral ecology with software and community data sets imbedded in a cyberinfrastructure. *ISME J*. doi:[10.1038/ismej.2016.89](https://doi.org/10.1038/ismej.2016.89)
- Boucher D, Jardillier L, Debroas D (2006) Succession of bacterial community composition over two consecutive years in two aquatic systems: a natural lake and a lake-reservoir. *FEMS Microbiol Ecol* 55(1):79–97
- Brussaard CP, Baudoux AC, Rodríguez-Valera F (2016) In the marine microbiome. Springer, New York, pp 155–183
- Chow CET, Suttle CA (2015) Biogeography of viruses in the sea. *Annu Rev Virol* 2:41–66. doi:[10.1146/annurev-virology-031413-085540](https://doi.org/10.1146/annurev-virology-031413-085540)
- Cobián Güemes AG, Youle M, Cantú VA, Felts B, Nulton J, Rohwer F (2016) Viruses as winners in the game of life. *Annu Rev Virol* 3:197–214. doi:[10.1146/annurev-virology-100114-054952](https://doi.org/10.1146/annurev-virology-100114-054952)
- Corinaldesi C, Dell'Anno A, Danovaro R (2007) Viral infection plays a key role in extracellular DNA dynamics in marine anoxic systems. *Limnol Oceanogr* 52(2):508
- Dafale N, Lakhe S, Yadav K, Purohit H, Chakrabarti T (2008) Concentration and recovery of coliphages from water with bituminous coal. *Water Environ Res* 80(3):282–288. doi:[10.2175/106143007X220923](https://doi.org/10.2175/106143007X220923)
- Dafale NA, Hathi ZJ, Bit S, Purohit HJ (2015) Bacteriophage diversity in different habitats and their role in pathogen control in microbial factories. Springer, New York. doi:[10.1007/978-81-322-2595-9\\_17](https://doi.org/10.1007/978-81-322-2595-9_17)
- Danovaro R, Corinaldesi C, Dell'Anno A, Fuhrman JA, Middelburg JJ, Noble RT, Suttle CA (2011) Marine viruses and global climate change. *FEMS Microbiol Rev* 35(6):993–1034. doi:[10.1111/j.1574-6976.2010.00258.x](https://doi.org/10.1111/j.1574-6976.2010.00258.x)
- Dinsdale E, Edwards R, Hall D, Angly F, Breitbart M, Brulc J, Furlan M, Desnues C, Haynes M, Li L et al (2008) Functional metagenomic profiling of nine biomes. *Nature* 452:629–633. doi:[10.1038/nature06810](https://doi.org/10.1038/nature06810)
- Fancello L, Trape S, Robert C, Boyer M, Popgeorgiev N, Raoult D, Desnues C (2013) Viruses in the desert: a metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. *ISME J* 7(2):359–369. doi:[10.1038/ismej.2012.101](https://doi.org/10.1038/ismej.2012.101)
- Ge X, Wu Y, Wang M, Wang J, Wu L, Yang X et al (2013) Viral metagenomics analysis of planktonic viruses in East Lake, Wuhan, China. *Virol Sin* 28(5):280–290. doi:[10.1007/s12250-013-3365-y](https://doi.org/10.1007/s12250-013-3365-y)
- Grazziotin AL, Koonin EV, Kristensen DM (2016) Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. *Nucleic Acids Res*. doi:[10.1093/nar/gkw975](https://doi.org/10.1093/nar/gkw975)
- Green J, Rahman F, Saxton M, Willianson K (2015) Metagenomic assessment of viral diversity in Lake Matoaka, a temperate, eutrophic freshwater lake in south eastern Virginia, USA. *Aquat Microb Ecol* 75(2):117–128. doi:[10.3354/ame01752](https://doi.org/10.3354/ame01752)
- Hurwitz BL, Sullivan MB (2013) The Pacific Ocean Virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS ONE* 8(2):e57355. doi:[10.1371/journal.pone.0057355](https://doi.org/10.1371/journal.pone.0057355)
- Kenzaka T, Tani K, Nasu M (2010) High-frequency phage-mediated gene transfer in freshwater environments determined at single-cell level. *ISME J* 4(5):648–659. doi:[10.1038/ismej.2009.145](https://doi.org/10.1038/ismej.2009.145)
- Livermore JA, Emrich SJ, Tan J, Jones SE (2014) Freshwater bacterial lifestyles inferred from comparative genomics. *Environ Microbiol* 16(3):746–758. doi:[10.1111/1462-2920.12199](https://doi.org/10.1111/1462-2920.12199)
- López-López O, Cerdán ME, González-Siso MI (2013) Hot spring metagenomics. *Life* 3(2):308–320. doi:[10.3390/life3020308](https://doi.org/10.3390/life3020308)
- Martínez JM, Swan BK, Wilson WH (2014) Marine viruses, a genetic reservoir revealed by targeted viromics. *ISME J* 8(5):1079–1088. doi:[10.1038/ismej.2013.214](https://doi.org/10.1038/ismej.2013.214)
- McDaniel LD, Rosario K, Breitbart M, Paul JH (2014) Comparative metagenomics: natural populations of induced prophages demonstrate highly unique, lower diversity viral sequences. *Environ Microbiol* 16(2):570–585. doi:[10.1111/1462-2920.12184](https://doi.org/10.1111/1462-2920.12184)
- Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M et al (2008) The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of

- metagenomes. *BMC Bioinform* 9(1):386. doi:[10.1186/1471-2105-9-386](https://doi.org/10.1186/1471-2105-9-386)
26. Molineux IJ (2006) The T7 group. In: Calendar R (ed) *The bacteriophages*. Oxford University Press, Oxford, pp 277–301
27. Noble WS (2009) How does multiple testing correction work? *Nat Biotechnol* 27(12):1135–1137. doi:[10.1038/nbt1209-1135](https://doi.org/10.1038/nbt1209-1135)
28. Nobrega FL, Costa AR, Kluskens LD, Azeredo J (2015) Revisiting phage therapy: new applications for old resources. *Trends Microbiol* 23(4):185–191. doi:[10.1016/j.tim.2015.01.006](https://doi.org/10.1016/j.tim.2015.01.006)
29. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T et al (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 42(D1):D206–D214. doi:[10.1093/nar/gkt1226](https://doi.org/10.1093/nar/gkt1226)
30. Parks DH, Tyson GW, Hugenholtz P, Beiko RG (2014) STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30:3123–3124. doi:[10.1093/bioinformatics/btu494](https://doi.org/10.1093/bioinformatics/btu494)
31. Parmar KM, Hathi ZJ, Dafale NA (2016) Control of multidrug-resistant gene flow in the environment through bacteriophage intervention. *Appl Biochem Biotechnol*. doi:[10.1007/s12010-016-2265-7](https://doi.org/10.1007/s12010-016-2265-7)
32. Parsley LC, Consuegra EJ, Thomas SJ, Bhavsar J, Land AM, Bhuiyan NN et al (2010) Census of the viral metagenome within an activated sludge microbial assemblage. *Appl Environ Microbiol* 76(8):2673–2677. doi:[10.1128/AEM.02520-09](https://doi.org/10.1128/AEM.02520-09)
33. Periasamy D, Sundaram A (2013) A novel approach for pathogen reduction in wastewater treatment. *J Environ Health Sci Eng* 11(1):12. doi:[10.1186/2052-336X-11-12](https://doi.org/10.1186/2052-336X-11-12)
34. Prangishvili D, Stedman K, Zillig W (2001) Viruses of the extremely thermophilic archaeon *Sulfolobus*. *Trends Microbiol* 9(1):39–43. doi:[10.1016/S0966-842X\(00\)01910-7](https://doi.org/10.1016/S0966-842X(00)01910-7)
35. Prestel E, Salamitou S, DuBow MS (2008) An examination of the bacteriophages and bacteria of the Namib desert. *J Microbiol* 46(4):364. doi:[10.1007/s12275-008-0007-4](https://doi.org/10.1007/s12275-008-0007-4)
36. Rohwer F, Thurber RV (2009) Viruses manipulate the marine environment. *Nature* 459(7244):207. doi:[10.1038/nature08060](https://doi.org/10.1038/nature08060)
37. Roux S, Tournayre J, Mahul A, Debroas D, Enault F (2014) Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. *BMC Bioinform* 15(1):76. doi:[10.1186/1471-2105-15-76](https://doi.org/10.1186/1471-2105-15-76)
38. Schoenfeld T, Patterson M, Richardson PM, Wommack KE, Young M, Mead D (2008) Assembly of viral metagenomes from yellowstone hot springs. *Appl Environ Microbiol* 74(13):4164–4174. doi:[10.1128/AEM.02598-07](https://doi.org/10.1128/AEM.02598-07)
39. Shapiro OH, Kushmaro A, Brenner A (2010) Bacteriophage predation regulates microbial abundance and diversity in a full-scale bioreactor treating industrial wastewater. *ISME J* 4(3):327–336. doi:[10.1038/ismej.2009.118](https://doi.org/10.1038/ismej.2009.118)
40. Sime-Ngando T, Colombet J (2009) Virus and prophages in aquatic ecosystems. *Can J Microbiol* 55(2):95–109. doi:[10.1139/w08-099](https://doi.org/10.1139/w08-099)
41. Suttle CA (2007) Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol* 5(10):801–812. doi:[10.1038/nrmicro1750](https://doi.org/10.1038/nrmicro1750)
42. Tamaki H, Zhang R, Angly FE, Nakamura S, Hong PY, Yasunaga T et al (2012) Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate. *Environ Microbiol* 14(2):441–452. doi:[10.1111/j.1462-2920.2011.02630.x](https://doi.org/10.1111/j.1462-2920.2011.02630.x)
43. Thingstad TF, Lignell R (1997) Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat Microbial Ecol* 13(1):19–27. doi:[10.3354/ame013019](https://doi.org/10.3354/ame013019)
44. Tucker KP, Parsons R, Symonds EM, Breitbart M (2011) Diversity and distribution of single-stranded DNA phages in the North Atlantic Ocean. *ISME J* 5(5):822–830. doi:[10.1038/ismej.2010.188](https://doi.org/10.1038/ismej.2010.188)
45. Weinbauer MG (2004) Ecology of prokaryotic viruses. *FEMS Microbiol Rev* 28(2):127–181. doi:[10.1016/j.femsre.2003.08.001](https://doi.org/10.1016/j.femsre.2003.08.001)
46. Williamson SJ, Rusch DB, Yooshep S, Halpern AL, Heidelberg KB, Glass JI et al (2008) The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS ONE* 3(1):e1456. doi:[10.1371/journal.pone.0001456](https://doi.org/10.1371/journal.pone.0001456)
47. Wu Q, Liu WT (2009) Determination of virus abundance, diversity and distribution in a municipal wastewater treatment plant. *Water Res* 43(4):1101–1109. doi:[10.1016/j.watres.2008.11.039](https://doi.org/10.1016/j.watres.2008.11.039)
48. Ye L, Zhang T (2013) Bacterial communities in different sections of a municipal wastewater treatment plant revealed by 16S rDNA 454 pyrosequencing. *Appl Microbiol Biotechnol* 97(6):2681–2690. doi:[10.1007/s00253-012-4082-4](https://doi.org/10.1007/s00253-012-4082-4)
49. Zablocki O, van Zyl L, Adriaenssens EM, Rubagotti E, Tuffin M, Cary SC, Cowan D (2014) High-level diversity of tailed phages, eukaryote-associated viruses, and viroplasm-like elements in the metaviromes of antarctic soils. *Appl Environ Microbiol* 80(22):6888–6897. doi:[10.1128/AEM.01525-14](https://doi.org/10.1128/AEM.01525-14)
50. Zillig W, Prangishvili D, Schleper C, Elferink M, Holz I, Albers S et al (1996) Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic Archaea. *FEMS Microbiol Rev* 18(2–3):225–236. doi:[10.1111/j.1574-6976.1996.tb00239.x](https://doi.org/10.1111/j.1574-6976.1996.tb00239.x)